

UCLA

UCLA Previously Published Works

Title

Genome-wide association study of depressive symptoms in the Hispanic Community Health Study/Study of Latinos

Permalink

<https://escholarship.org/uc/item/9r90n125>

Authors

Dunn, Erin C
Sofer, Tamar
Wang, Min-Jung
[et al.](#)

Publication Date

2018-04-01

DOI

10.1016/j.jpsychires.2017.12.010

Peer reviewed



HHS Public Access

Author manuscript

J Psychiatr Res. Author manuscript; available in PMC 2018 October 17.

Published in final edited form as:

J Psychiatr Res. 2018 April ; 99: 167–176. doi:10.1016/j.jpsychires.2017.12.010.

Genome-wide association study of depressive symptoms in the Hispanic Community Health Study/Study of Latinos

Erin C. Dunn^{a,b,c,*}, Tamar Sofer^{d,e}, Min-Jung Wang^a, Thomas W. Soare^{a,b,c}, Linda C. Gallo^f, Stephanie M. Gogarten^g, Kathleen F. Kerr^g, Chia-Yen Chen^{a,b,c,h}, Murray B. Steinⁱ, Robert J. Ursano^j, Xiuqing Guo^k, Yucheng Jia^k, Jie Yao^k, Jerome I. Rotter^k, Maria Argos^l, Jianwen Cai^m, Krista Ferreiraⁿ, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium¹, Sylvia Wassertheil-Smoller^o, and Jordan W. Smoller^{a,b,c}

^aPsychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, United States

^bDepartment of Psychiatry, Harvard Medical School, Boston, MA, United States

^cStanley Center for Psychiatric Research, The Broad Institute of Harvard and MIT, Cambridge, MA, United States

^dDivision of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, United States

^eDepartment of Medicine, Harvard Medical School, Boston, MA, United States

^fDepartment of Psychology, San Diego State University, San Diego, CA, United States

^gDepartment of Biostatistics, University of Washington, Seattle, WA, United States

^hAnalytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, United States

ⁱDepartment of Psychiatry, University of California San Diego, La Jolla, CA, United States

^jDepartment of Psychiatry, Uniformed Services University of the Health Sciences, Bethesda, MD, United States

^kInstitute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute and Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA, United States

^lSchool of Public Health, University of Illinois at Chicago, Chicago, IL, United States

*Corresponding author. Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, 185 Cambridge Street, Simches Research Building 6th Floor, Boston, MA 02114, United States. edunn2@mgh.harvard.edu (E.C. Dunn).

¹Members are listed in the supplementary material.

Data availability statement

Genotype data and GWAS summary results of all discovery results can be requested via dbGaP study accession phs000880; phenotype data can be requested via dbGaP study accession phs000810.

Disclosure

Drs. Dunn and Sofer take responsibility for the integrity of the data and accuracy of the analyses. All authors have reviewed and approved the final manuscript. None of the authors had any financial or other conflicts of interest.

Conflicts of interest

None.

^mDepartment of Biostatistics, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

ⁿCollege of Arts and Sciences, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

^oDepartment of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, United States

Abstract

Although genome-wide association studies (GWAS) have identified several variants linked to depression, few GWAS of non-European populations have been performed. We conducted a genome-wide analysis of depression in a large, population-based sample of Hispanics/Latinos. Data came from 12,310 adults in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). Past-week depressive symptoms were assessed using the 10-item Center for Epidemiological Studies of Depression Scale. Three phenotypes were examined: a total depression score, a total score modified to account for psychiatric medication use, and a score excluding anti-depressant medication users. We estimated heritability due to common variants (h_{SNP}^2), and performed a GWAS of the three phenotypes. Replication was attempted in three independent Hispanic/Latino cohorts. We also performed sex-stratified analyses, analyzed a binary trait indicating probable depression, and conducted three trans-ethnic analyses. The three phenotypes exhibited significant heritability ($h_{SNP}^2 = 6.3 - 6.9\%$; $p = .002$) in the total sample. No SNPs were genome-wide significant in analyses of the three phenotypes or the binary indicator of probable depression. In sex-stratified analyses, seven genome-wide significant SNPs (one in females; six in males) were identified, though none were supported through replication. Four out of 24 loci identified in prior GWAS were nominally associated in HCHS/SOL. There was no evidence of overlap in genetic risk factors across ancestry groups, though this may have been due to low power. We conducted the largest GWAS of depression-related phenotypes in Hispanic/Latino adults. Results underscore the genetic complexity of depressive symptoms as a phenotype in this population and suggest the need for much larger samples.

Keywords

Genetic association study; Depression; Depressive symptoms; Hispanics/Latinos

1. Introduction

After many years of effort, genome-wide association studies (GWAS) have started to make progress in identifying variants linked to depression (Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium et al., 2013), a common and burdensome disorder (Kessler et al., 2005; Mathers et al., 2008) known to be driven, in part, by genetic susceptibility (Sullivan et al., 2000). Prior to 2016, only two genetic variants associated with depression were reported using GWAS. These two loci were detected in a sample of Chinese women with re-current major depressive disorder (MDD) (CONVERGE Consortium, 2015). However, during 2016, another 22 variants were identified within samples of European

American adults: (1) four new loci were detected through a meta-analysis and proxy-phenotype analysis across three large cohorts (Okbay et al., 2016); (2) an additional 17 loci across 15 regions were identified using “crowd-sourced” data collected by consumer genomics platforms (Hyde et al., 2016); and (3) one new locus was identified through efforts to examine a broad depression phenotype comprising lifetime major depressive disorder and depressive symptoms (Direk et al., 2016).

Nevertheless, few genetic association studies of racial/ethnic minority populations have been performed, reflecting a general under-representation of non-European populations in psychiatric genetics research (Dalvie et al., 2015). Genetic studies of Hispanics/Latinos are especially relevant, as Hispanics/Latinos are one of the fastest growing populations in the US, increasing in size by 43% from 2000 to 2010 (Brown, 2014; Passel et al., 2011) and expected to double by 2050 to represent more than 25% of the US population (Projections of the Resident Population by Race, Hispanic Origin, and Nativity, 2025 and 2050 and 2003). Though epidemiological studies have found MDD is less common in Hispanics/Latinos compared to non-Hispanic whites (e.g., lifetime prevalence of 15.2% in adult Hispanics compared to 22.1% in non-Hispanic whites) (Alegria et al., 2008), recent data suggest the prevalence of depression may be similar (Avenevoli et al., 2015) or upwards of 7% higher among young Hispanics compared to their White counterparts (High School Youth Risk Behavior Survey data, 2015). Further, twin studies suggest that depression, as measured by brief diagnostic tools designed to capture DSM-IV diagnoses (Sheehan et al., 1998), is at least as heritable (around 40%) among some Hispanics, including in Mexican Americans (Olvera et al., 2011) compared to European Americans (Sullivan et al., 2000).

To address this gap, the current study used data from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), a large population-based sample, to: (1) estimate the common variant contribution (h_{SNP}^2) to depressive symptoms; (2) conduct the largest GWAS of Hispanics/Latinos to date; and (3) perform trans-ethnic generalization analyses to understand the potential genetic overlap of depression across ancestry groups. Only two prior GWAS of depressive symptoms have been performed in Hispanics/Latinos. One study of Hispanic women ($n = 3158$) did not find any SNPs that achieved genome-wide significance (Dunn et al., 2016), whereas the other study ($n = 1443$) found one SNP (rs1127233), intronic to the *MUC13* gene, reaching genome-wide significance ($p = 3.85 \times 10^{-8}$), though this result was not replicated (Ware et al., 2015). Although we suspected at the outset that these analyses were likely underpowered, given that much larger samples were previously needed to identify replicable loci, our hope was to highlight the need for genetic association studies in more diverse populations, provide empirical data regarding the genetic architecture of depressive symptoms in Hispanics/Latinos, and make available summary statistics that could be useful for future analyses and to build more ancestrally diverse research consortia (see Data Availability Statement following the Discussion).

2. Materials and methods

2.1. Overview

Data came from the HCHS/SOL, a community-based prospective cohort study following 16,415 self-identified Hispanic/Latino adults (aged 18–74 at screening). HCHS/SOL was designed to examine the distribution and determinants of chronic health conditions, including diabetes, pulmonary disease, and cardiovascular disease. As described elsewhere (Lavange et al., 2010; Sorlie et al., 2010), participants were recruited via a stratified two-stage area probability sample of households across four cities in the US (Chicago, IL; Miami, FL; Bronx, NY; San Diego, CA). The majority of the sample self-identified with the following background groups: Central American ($n = 1730$), Cuban ($n = 2348$), Dominican ($n = 1460$), Mexican ($n = 6471$), Puerto Rican ($n = 2728$), and South American ($n = 1068$). Baseline examinations were conducted between 2008 and 2011. Institutional Review Boards at each field center approved the study and all participants provided written informed consent. In the current study, we analyzed data from 12,310 respondents who consented to provide blood for genotyping and had complete depressive symptoms data and relevant covariates information, as well as non-missing records of antidepressant medication use.

2.2. Phenotype definition

Depressive symptoms were assessed during the first visit using the Andresen version of the 10-item Center for Epidemiological Studies of Depression Scale (CES-D-10) (Andresen et al., 1994). The CES-D-10 captured core symptoms of depression in the past week, including depressed mood and behavioral symptoms (e.g., felt depressed; felt lonely; could not get going). The CES-D-10 has good predictive accuracy ($k = .97$) relative to the full 20-item CES-D (Andresen et al., 1994). It was also shown in the current study population to have excellent internal consistency reliability (Cronbach's alphas = .80–.86), test-retest reliability ($r_s = .41$ –.70), and moderate to high correlations with other measures of depressive or mental health symptoms, including the Spielberger State-Trait Anxiety Inventory ($r = .72$, $p < .001$), the Patient Health Questionnaire-9 ($r = .80$, $p < .001$), and the Short Form-12's Mental Component Summary ($r = -.65$, $p < .001$) (Gonzalez et al., 2017). While the inter-rater reliability of CES-D-10 has not previously been assessed, the full 20-item CES-D has demonstrated good-to-excellent inter-rater reliability ($r_s = .67$ –.76) (Shinar et al., 1986; van de Rest, van der Zwaluw, Beekman, de Groot and Geleijnse, 2010). Total depressive symptoms scores were calculated by summing responses to each item; item responses ranged from 0 = rarely or none of the time (< 1 day) to 3 = all of the time (5–7 days). Higher total scores (which varied between 0 and 30) indicated higher levels of depressive symptoms.

To account for the possibility that current use of antidepressant medications might affect depressive symptoms scores, we applied a previously described (Dunn et al., 2016; Hek et al., 2013) algorithm to adjust the scores of medication users (see Supplemental Materials).

2.3. SNP genotyping, quality control and imputation

Blood samples from consenting respondents were sent to Illumina Microarray Services for genotyping on the Illumina SOL HCHS Custom 15041502 B3 array. This array comprised

the Illumina Omni 2.5M array (HumanOmni2.5-8v1-1) and additional custom content (e.g., ancestry-informative markers, variants characteristic of Amerindian populations, known GWAS hits, and other candidate gene markers) selected for HCHS/SOL.

Quality assurance/quality control (QA/QC) was performed by Illumina, LA Biomed, and the HCHS/SOL Genetic Analysis Center (GAC) per established methods (Laurie et al., 2010) to generate recommended SNP and sample-level quality filters. In brief, samples were checked for annotated versus genetic sex, gross chromosomal anomalies (Laurie et al., 2012), call rates, batch effects, duplicate sample discordance, Mendelian errors, population structure, and relatedness (genetic relatedness occurred when participants were drawn from the same household or from different households living in the same community). 12,803 unique study samples passed these criteria. SNPs that passed the Illumina/LA Biomed assay failure indicator were further checked for Hardy-Weinberg equilibrium, MAF, duplicate probe discordance, and missing call rate. A total of 2,232,944 SNPs passed both quality and informativeness filters (unduplicated on the array and polymorphic).

Genome-wide imputation was carried out on all 12,803 samples together using the 1000 Genomes Project phase 3 reference panel (1000 Genomes Project Consortium, 2012) and IMPUTE2 software (Howie et al., 2009; Howie et al., 2011). Genotypes were first pre-phased with SHAPEIT2 (v2.r644) and then imputed with IMPUTE2 (v2.3.0). Only variants with at least two copies of the minor allele present in any of the four 1000 Genomes continental panels were imputed, yielding a total of 25,568,744 imputed variants. Overall imputation quality was assessed both by looking at the distribution of imputed quality metrics by different MAF levels and by examining results from the IMPUTE2 internal masking experiments (as some genotyped variants were “masked”, meaning removed from the imputation basis).

2.4. Statistical analyses

In these analyses, we used a linear mixed-effect model, which accounted for the correlations between individuals due to genetic relatedness (kinship), shared household, and the complex sampling design (Conomos et al., 2016; Schick et al., 2016). The variance components were estimated using restricted maximum likelihood (REML). Fixed effects included the covariates: log(sampling weight), which reflects the differences in sampling probabilities of study individuals and is included to prevent potential selection bias; field center; age; sex; education (1 = no high school diploma or GED – referent; 2 = at most a high school diploma or GED, 3 = greater than high school or GED; 4 = bachelors degree, 5 = masters, professional, or doctorate degree); and the top five principal components (PCs) of ancestry. SNP annotation was performed using ANNOVAR (Wang et al., 2010) (<http://annovar.openbioinformatics.org/en/latest/>).

2.5. Heritability analysis

We estimated “SNP-chip heritability”, or the narrow-sense heritability due to the additive effect of genotyped common variants, by first fitting a “null” linear mixed model that included all covariates, PCs, and random effects (but did not include genotypes), and then calculating the proportion of variance attributable to relatedness out of all phenotypic

variance. For this analysis, the model was fit on all individuals to maximize statistical power and allow for comparability with the GWAS results. We calculated the kinship matrix using all autosomal SNPs, and ran the analysis on subsets of genetically unrelated individuals (about 80% of the analytic sample size) by randomly removing participants so that none of the pairwise estimated kinship coefficients were larger than $2^{-11/2}$, which excluded all first-, second-, or third-degree relatives (Yang et al., 2010). Results are presented as SNP-heritability estimates (h^2_{SNP}) and 95% confidence intervals (CIs), calculated using Haseman-Elston regression (Sofer, 2017).

2.6. GWAS analysis

We performed a GWAS using the linear mixed-effect model approach on three phenotypes: (1) the original, untransformed depressive symptoms score; (2) the depressive symptoms score adjusting for medication users via the procedure described under “phenotype definition”; and (3) the original, untransformed depressive symptoms score only in the subset of participants who were non-anti-depressant medication users ($n = 11,486$; 93.3% of the sample). All SNPs were modeled additively and the standard 5×10^{-8} was used as the threshold for genome-wide statistical significance. In addition, we report the set of SNPs with p -value $< 1 \times 10^{-6}$ according to the following selection criteria: out of SNPs that were less than 500,000 base pairs apart, and their correlation was higher than .5 ($R^2 > .5$), we prioritized (1) genotyped over imputed SNPs, (2) imputed SNPs with higher quality scores (info), and (3) imputed SNPs with lower p -values; for SNPs with similar p -values and imputation quality score (or genotyped), we prioritized SNPs with higher MAF. Quantile-quantile (QQ) and Manhattan plots were generated using the R package GWASTools (Gogarten et al., 2012). Regional association plots were generated using Locus Zoom (Pruim et al., 2010).

2.7. Replication

We performed a replication test of our top results using data from three independent cohorts. Additional details about these cohorts are presented in supplementary material. Briefly, the Women’s Health Initiative (1998; Wassertheil-Smoller et al., 2004) (WHI; www.whi.org) provided data on Hispanic/Latina women ($n = 3138$; mean age 60.0; $sd = 6.57$), where depressive symptoms were measured using a 6-item version of the CES-D (Wassertheil-Smoller et al., 2004) (items were: felt depressed; sleep was restless; enjoyed life; had crying spells; felt sad; felt people disliked you). The Multi-Ethnic Study of Atherosclerosis (Bild et al., 2002) (MESA; <http://www.mesa-nhlbi.org>) provided data from Hispanic/Latino adults ($n = 1449$; mean age 61.39; $sd = 10.30$) where depressive symptoms were measured using a 20-item version of the CES-D (Andresen et al., 1994). Finally, the Army Study To Assess Risk and Resilience in Service members (Ursano et al., 2014) (Army STARRS; <http://www.armystarrs.org>) provided data from Hispanic/Latino adults ($n = 3361$; mean age = 22.77; $sd = 5.01$), where depressive symptoms were captured using a four-item Composite International Diagnostic Interview Screening Scale (CIDI-SC) for major depressive episode (MDE), which included the items: feeling sad or depressed; down about how things are going; little or no pleasure in things; feeling down on yourself or worthless in the past 30 days (Kessler et al., 2013).

We meta-analyzed association results across these three independent samples. Inverse variance weighted fixed-effect meta-analysis was conducted using METAL (<http://www.sph.umich.edu/csg/abecasis/metal/>; (Willer et al., 2010). We used one sided p-values in the replication phase (Sofer et al., 2017).

2.8. Secondary analyses

We performed two secondary analyses. First, we conducted sex-stratified analyses, in light of evidence showing a two-fold elevated risk of depression in women compared to men (Kessler et al., 2005) and perhaps higher heritability of depression (Kendler et al., 2001). Second, to facilitate comparisons to other studies using case-control designs, we analyzed a binary trait indicating probable depression, using a cut-off threshold of CES-D-10 score ≥ 10 to define cases of high depressive symptoms ($n = 3979$). This cut-point has been shown to have 96% sensitivity and 100% specificity against the CES-D-20 to screen for depressive symptoms in adults (Andresen et al., 1994), and has been used in previous studies based on the same study population (Wassertheil-Smoller et al., 2014). In this analysis, we included anti-depressant medication users as cases. To minimize false negatives, controls were defined as participants with CES-D-10 scores 6 or below ($n = 6499$).

2.9. Trans-ethnic analyses

We performed three sets of trans-ethnic analyses to determine the extent to which there was genetic overlap in risk for depressive symptoms across ancestry groups and to place our study results in the context of prior work. First, to evaluate the extent of sharing of SNP-level associations across populations, we examined the p-values in HCHS/SOL of 25 loci previously identified as genome-wide significant in prior GWAS studies of Chinese, European, and Hispanic/Latino ancestry populations (CONVERGE Consortium, 2015; Direk et al., 2016; Hyde et al., 2016; Okbay et al., 2016; Ware et al., 2015) and compared the direction of effect observed between our study and the original study. One-sided binomial tests were performed to determine whether the observed same-direction allelic effects with the primary studies were greater than expected by chance.

Second, to more deeply probe whether prior SNP-effects were population-specific, we performed a specific generalization analysis (Sofer, 2017) to evaluate the strength of association between the larger set of loci identified within the GWAS studies described above. The generalization testing algorithm we used (Sofer, 2017) accounts for multiple testing by controlling for the directional False Discovery Rate (FDR) at the .05 level (Sofer, 2017); it does so by selecting SNPs reported in prior GWAS (also called “discovery study”) by first calculating their Benjamini and Hochberg (BH)-adjusted (Benjamini and Hochberg, 1995) p-values and then calculating a data-adaptive p-value threshold based on these adjusted values, so that all SNPs with adjusted p-values lower than the threshold can potentially be generalized under the set FDR level (.05). Once SNPs are selected, the algorithm assigns these SNPs with an r-value, based on their p-value in both the discovery study and the HCHS/SOL GWAS of the original CES-D score. An r-value $< .05$ indicates that the SNP is associated in the same direction in both the discovery GWAS and HCHS/SOL GWAS. We studied the generalization of genetic associations from European

ancestry populations to Hispanics/Latinos, using the three largest GWAS of depression to date (Direk et al., 2016; Okbay et al., 2016; Wray et al., 2017) as the discovery datasets.

Third, to evaluate the extent to which GWAS findings in European samples were, in aggregate, significantly associated with depressive symptoms in Hispanics/Latinos, we constructed unweighted genetic risk scores (GRS) (Moon et al., 2017; Qi et al., 2017) using summary statistics from three published GWAS of European populations (Direk et al., 2016; Wray et al., 2017; Okbay et al., 2016) and then evaluated the strength of association between these GRS in HCHS/SOL. Eight sets of GRS were developed using SNPs with p-value thresholds in the discovery GWAS of $p_T < .05$, $p_T < .01$, $p_T < .001$, $p_T < 10^{-4}$, $p_T < 10^{-5}$, $p_T < 10^{-6}$, $p_T < 10^{-7}$, and $p_T < 10^{-8}$; within each p-value threshold, SNPs were pruned to identify independent SNPs within a 500kb window. Therefore, all HCHS/SOL participants had 24 GRS values (eight p-value thresholds within 3 GWAS discovery samples), which were generated by summing the number of trait-increasing alleles, determined by the effect sign in the discovery GWAS. We then used the models from the primary association analysis (i.e. mixed models, adjusted to the same fixed and random effects) to evaluate the significance of the association between each GRS and the original, untransformed CES-D scores in the full HCHS/SOL sample. When at least one of the GRSs based on the discovery study showed significant association in HCHS/SOL with the original depressive symptoms score, we then repeated the analysis with all depression measures, allowing us to evaluate whether one GRS measure was more consistently generalizable than others.

3. Results

There were 12,310 Hispanic/Latino adults in the analysis; sample demographic characteristics can be found in the supplementary material.

3.1. Discovery sample: SNP heritability

Depressive symptoms were significantly and moderately heritable for each of the three phenotypes in the total sample – *original score*: $h_{\text{SNP}}^2 = 6.4\%$, 95% CI=(1.6, 11.1)%, $p = .002$; *the score accounting for medication use*: $h_{\text{SNP}}^2 = 6.3\%$, 95% CI=(1.6, 11.1)%, $p = .002$; and *excluding medication users*: $h_{\text{SNP}}^2 = 6.9\%$, 95% CI=(2.3, 12)%, $p = .002$ (Supplemental Table 1). However, depressive symptom scores did not show evidence of significant heritability among unrelated individuals for any phenotype: *original score*: $h_{\text{SNP}}^2 = 4\%$, 95% CI=(0, 10.1)%, $p = .82$; *the score accounting for medication use*: $h_{\text{SNP}}^2 = 2.9\%$, 95% CI=(0, 8.9)%, $p = .153$; or *excluding medication users*: $h_{\text{SNP}}^2 = 3.8\%$, 95% CI=(0, 10.3), $p = .104$ (Supplemental Table 1).

3.2. Discovery sample: GWAS

Fig. 1 displays the Manhattan and QQ plots for the three phenotypes. The QQ plots show no evidence of genomic inflation in the GWAS of the original score ($\lambda = 1.01$), the analysis accounting for medication use ($\lambda = 1.01$), or excluding medication users ($\lambda = 1.00$).

No SNP reached genome-wide significance in either the analysis based on scores accounting for medication users (Table 1) or that excluded medication users (Table 2). The most significant SNP in the analysis accounting for medication use was an imputed rare variant located on chromosome 19 position 46976251 ($p = 1.89 \times 10^{-7}$).

Among SNPs with the lowest p-values, several were common to both analyses, suggesting robustness of their association with the investigated depression phenotypes, but perhaps not surprising since a minority of participants were medication users. For instance, rs116270819 ($p = 6.69 \times 10^{-8}$) was the top locus in the analysis excluding medication users and also emerged with a low p-value (2.07×10^{-7}) in the analysis adjusting for medication use (Table 1).

To further interpret these results and select the SNPs to carry forward for replication, we estimated replication power for all SNPs with p-values $< 1 \times 10^{-6}$ in at least one of the two analyses. Replication power estimates were based on the projected samples sizes of each replication dataset (WHI = 3000; MESA = 1500; Army STARRS = 3000) and using MAF, outcome standard deviation, and estimated effect sizes from the discovery sample. Our power calculations incorporated a method (Zhong and Prentice, 2008) to reduce bias due to “winner’s curse,” by attenuating the observed effect size. Prior studies have previously shown that attenuated effect size estimates tend to be closer than uncorrected estimates to effects seen in independent replication studies (Zhong and Prentice, 2010).

The power estimate suggested that two SNPs (rs2004237 and rs34208798) would have good to excellent power in a meta-analysis of the three replication cohorts after correcting for the winner’s curse bias (estimated power = 90% and 62%, respectively) and thus we carried forward these SNPs for replication. One of these SNPs, rs2004237, was the second most significant result ($p = 7.54 \times 10^{-8}$) in the analysis excluding medication users. This imputed SNP is located on chromosome 19 and is intronic to fibrillin 3 (*FBN3*), a gene that encodes fibrillins, a constituent of the extracellular matrix. Several other SNPs in the region, some of which were genotyped, also showed support for this association (Fig. 2a). The second SNP carried forward was also imputed (rs34208798) and had a low p-value in the analysis excluding medication users ($p = 1.87 \times 10^{-7}$). This SNP is intronic to the TRNA-YW Synthesizing Protein 1 Homolog gene (*TYWI*), which encodes for Wybutosin (yW) and is involved in the modification of tRNAs (Landgraf et al., 2016). SNPs in this region also showed support for this association (Fig. 2b).

All GWAS results at $p < 1 \times 10^{-5}$ are shown in the supplementary material for the depressive symptom score for the full sample adjusting for medication use (Supplemental Table 2), excluding medication users (Supplemental Table 3), and for the original, non-transformed score (Supplemental Table 4).

3.3. Replication samples: GWAS

In the replication stage, the two selected SNPs (rs2004237 and rs34208798) were evaluated in three independent samples. Neither SNP was significantly associated with the depressive symptom score in a meta-analysis of the replication data, either accounting for medication use or excluding medication users (Supplemental Table 5 and Supplemental Table 6).

3.4. Secondary analyses

In the sex-stratified analyses, we found that among females, one SNP (rs72662446) reached genome-wide significance ($p = 6.36 \times 10^{-9}$) in the analysis accounting for medication use (Supplemental Table 7), and was also the most significant result ($p = 7.48 \times 10^{-8}$) in the analysis excluding medication users (Supplemental Table 8). In males, two SNPs (rs111365740, $p = 3.70 \times 10^{-8}$; and rs9709324, $p = 4.20 \times 10^{-8}$) achieved genomewide significance in the analysis accounting for medication use (Supplemental Table 9). In the analysis excluding medication users (Supplemental Table 10), four SNPs reached genome-wide significance (rs113403132, $p = 7.74 \times 10^{-9}$; rs6993028, $p = 1.27 \times 10^{-8}$; rs144850488, $p = 2.69 \times 10^{-8}$; rs17800303, $p = 3.59 \times 10^{-8}$); the top hit from this analysis also had a low p-value in the analysis accounting for medication use (rs113403132, $p = 1.33 \times 10^{-7}$) (Supplemental Table 9). Five of these SNPs were selected for replication (1 in females and 4 in males); none showed significant evidence of association (Supplemental Tables 11–15).

We also conducted a GWAS with depressive symptoms treated as a binary outcome; none of the SNPs reached genome-wide significance (Supplemental Fig. 1; Supplemental Table 16).

3.5. Trans-ethnic analyses

3.5.1. Examination of published GWAS loci—Among the published genome-wide significant loci for depression, four out of 25 loci were nominally associated in this study (Table 3), with two showing the same direction of effect as the primary study. Two of these loci (rs4481363, rs4543289) were significant after correcting for multiple testing ($0.05/25 = p < .002$), with one showing a similar direction of effect as the primary study. These associations were strongest in the analysis that excluded medication users. Overall, 13 out of 25 (52%) loci showed consistent directions of effect across the three depression phenotypes as compared to the prior GWAS studies. There was no evidence of excess same-direction effects (one-sided binomial test $p = .50$ for all analyses).

3.5.2. Generalization analysis—We performed formal generalization analyses to study whether previously reported associations generalize to Hispanics/Latinos, using each of the three discovery GWAS. Based on the adaptive selection rule for selecting SNPs from the discovery GWAS to follow-up in the HCHS/SOL, 3 SNPs, 3594 SNPs, and 23,957 SNPs were selected for the generalization testing in the CHARGE, Okbay et al., and the combined PGC MDD2 and 23andme cohorts, respectively. The generalization testing algorithm calculated an FDR-controlling r-value for each of these SNPs, where that the FDR control is applied to the tests of the generalization null hypotheses. None of these SNPs generalized (all r-values were > 0.05).

3.5.3. Examination of genetic risk scores—At the Bonferroni-corrected p-value threshold of .0006 (i.e., .05/80 tests), the GRS constructed from SNPs with p-values < 0.001 in the combined PGC MDD2 and 23andme cohorts (Wray et al., 2017) were significantly associated with all four depression phenotypes (Supplemental Table 17) - *original depressive symptom score*: $\beta = 0.012$, 95% CI=(0.006,0.017), $p = .00004$; *adjusting for medication use*: $\beta = 0.013$, 95% CI=(0.007,0.019), $p = .00004$; *excluding medication users*: $\beta = 0.011$, 95% CI=(0.005,0.017), $p = .0001$; binary depression: OR = 1.001, 95% CI=(1.000, 1.001), $p = .$

0004. The $p < 10^{-4}$ GRS was also significantly associated with depression in the analysis excluding medication users ($p = .0004$). No other significant associations were detected for any phenotype with other GRS cut-points in the PGC-MDD2/23andme analysis or for any GRS generated from either the CHARGE or Okbay et al. datasets (Supplemental Table 17).

4. Discussion

The current study involved several advances in efforts to identify the genetic basis of depression. First, our study was the largest GWAS of depressive symptoms in Hispanics/Latinos. Only two prior GWAS of depressive symptoms in Hispanics/Latinos have been performed, the largest of which contained one-third the number of participants included here (Dunn et al., 2016). Second, we conducted trans-ethnic analyses by evaluating whether previously published individual loci and aggregate polygenic variation derived from GWAS studies of Chinese and European ancestry populations were linked to depression in Hispanics/Latinos. Finally, our primary analyses were based on a dimensional measure of depressive symptoms, which enabled us to examine the full range of quantitative variation, in addition to extremes in this quantitative distribution (e.g., cases versus controls); prior studies suggest such an approach may be statistically more powerful for identifying variants associated with depression (Plomin et al., 2009).

In our primary analysis, no genome-wide significant loci were found, though we did detect seven genome-wide significant associations in sex-stratified analyses. However, neither these loci nor any others with suggestive evidence were replicated across three independent cohorts. Our heritability estimates were comparable in magnitude to those observed in at least one prior study (Dunn et al., 2016), but were not statistically significant in all instances. The lack of genome-wide significant loci is likely due to low statistical power, emphasizing the need for larger and thus better powered studies.

There was also little evidence of overlap in genetic vulnerability to depression across ancestry groups. In our SNP-based analysis, four out of 25 loci previously associated with depression were nominally associated in this HCHS/SOL sample. These four loci were first observed in two studies of European populations, one of which examined depressive symptoms (Okbay et al., 2016) and the other probable major depressive disorder (Hyde et al., 2016). Two of these loci showed the same direction of effect as the primary study, suggesting that at least some genetic contributors to depression are common across ancestral populations. However, there was generally only modest support for shared genetic effects between Hispanics/Latinos and European Americans as shown in the analyses examining more SNPs or aggregate polygenic variation as measured through a genetic risk score. Prior trans-ethnic analyses of MDD have also suggested possible population differences in genetic liability to depression across racial/ethnic groups (Bigdeli et al., 2017). The lack of observed overlap in genetic risk factors for depression across ancestry groups could be due to low power in our sample.

Several limitations are noted. First, our outcome was based on a measure of depressive symptoms during the past week, rather than a diagnostic interview or depressive symptoms captured over a longer period. Thus, we could only study recent symptoms of depression and

did not have information about how long the depressive symptoms lasted. However, as we describe previously, the CES-D has been widely used in epidemiological studies, has demonstrated excellent reliability (Gonzalez et al., 2017; Shinar et al., 1986; van de Rest et al., 2010) and validity (Radloff, 1977), and appears to be a robust measure of depressive symptoms in Hispanic/Latino populations (Gonzalez et al., 2017). Future genetic studies of lifetime depressive disorder in Hispanic/Latino samples are warranted. Second, the replication samples were smaller and more demographically heterogeneous than HCHS/SOL, which along with other factors (e.g., variation in depressive symptom measures; gene-environment interplay; different LD patterns with a causal SNP) could have led to null results. Unfortunately, replication efforts are currently impeded by the lack of available data in racial/ethnic minority populations. Efforts now underway to carry out genome-wide analyses in more diverse samples hold promise to address such challenges.

In summary, results from this study suggest that larger samples and perhaps more refined phenotypes are needed to identify genetic variants associated with depression in Hispanics/Latinos. This will enable the important work of extending genomic studies to populations of diverse ancestry.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The Hispanic Community Health Study/Study of Latinos (HCHS/SOL) was carried out as a collaborative study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (N01-HC65233), University of Miami (N01-HC65234), Albert Einstein College of Medicine (N01-HC65235), Northwestern University (N01-HC65236), and San Diego State University (N01-HC65237). The following Institutes/Centers/Offices contribute to the HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological Disorders and Stroke, NIH Institution-Office of Dietary Supplements. The Genetic Analysis Center at the University of Washington was supported by NHLBI contract HHSN268201300005C. MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-001079, UL1-TR-000040, UL1-TR-001420, UL1-TR-001881, and DK063491. The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. The Army Study of Risk and Resilience in Servicemembers (Army STARRS) was sponsored by the Department of the Army and funded under cooperative agreement number U01MH087981 with the U.S. Department of Health and Human Services, National Institutes of Health, and National Institute of Mental Health (NIH/NIMH). The Psychiatric Genomics Consortium was funded by NIH grants U01 MH109528 and U01 MH1095320.

The contents are solely the responsibility of the authors and do not necessarily represent the views of the Department of Health and Human Services, the National Institutes of Health, the Veterans Administration, Department of the Army, or the Department of Defense.

The current study is supported the National Institute of Mental Health of the National Institutes of Health under Award Number K01MH102403 (Dunn) and by a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation (Dunn). Dr. Smoller is a Tepper Family MGH Research Scholar. Dr. Sofer was supported by the National Heart, Lung, and Blood Institute under Award Number 1R35HL135818.

The authors thank Stephanie Gomez and Emily Moya for their assistance in conducting the literature search for this paper.

The authors also thank the staff and participants of HCHS/SOL for their important contributions. A complete list of staff and investigators has been provided by Sorlie P, et al. in *Annals of Epidemiology* (2010), issue 20 (8), pages 629-641 and is also available on the study website <http://www.csc.unc.edu/hchs/>.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.jpsychires.2017.12.010>.

References

- Alegria M, Canino G, Shrout PE, Woo M, Duan N, Vila D, et al. Prevalence of mental illness in immigrant and non-immigrant U.S. Latino groups. *Am J Psychiatr*. 2008; 165(3):359–369. DOI: 10.1176/appi.ajp.2007.07040704 [PubMed: 18245178]
- Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *Am J Prev Med*. 1994; 10(2):77–84. [PubMed: 8037935]
- Avenevoli S, Swendsen J, He JP, Burstein M, Merikangas KR. Major depression in the national comorbidity survey-adolescent supplement: prevalence, correlates, and treatment. *J Am Acad Child Adolesc Psychiatr*. 2015; 54(1):37–44. . e32. DOI: 10.1016/j.jaac.2014.10.010
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*. 1995; 57:289–300.
- Bigdeli T, Ripke S, Bacanu S, Abdellaoui A, Andlauer T, Beekman A, et al. Genetic effects influencing risk for major depressive disorder in China and Europe. *Mol Psychiatr*. 2017; 7:e1074.
- Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol*. 2002; 156(9):871–881. [PubMed: 12397006]
- Brown A. US Hispanic and Asian populations growing, but for different reasons. 2014. Jun 26, Retrieved from <http://www.pewresearch.org/fact-tank/2014/06/26/u-s-hispanic-and-asian-populations-growing-but-for-different-reasons/>
- Conomos MP, Laurie CA, Stilp AM, Gogarten SM, McHugh CP, Nelson SC, et al. Genetic diversity and association studies in US hispanic/latino populations: applications in the hispanic community health study/study of Latinos. *Am J Hum Genet*. 2016; 98(1):165–184. DOI: 10.1016/j.ajhg.2015.12.001 [PubMed: 26748518]
- CONVERGE Consortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*. 2015; 523(7562):588–591. DOI: 10.1038/nature14659 [PubMed: 26176920]
- Dalvie S, Koen N, Duncan L, Abbo C, Akena D, Atwoli L, et al. Large scale genetic research on neuropsychiatric disorders in african populations is needed. *EBioMedicine*. 2015; 2(10):1259–1261. DOI: 10.1016/j.ebiom.2015.10.002 [PubMed: 26629498]
- Design of the Women’s Health Initiative clinical trial and observational study. The Women’s Health Initiative Study Group. *Contr Clin Trials*. 19:61–109. <http://www.ncbi.nlm.nih.gov/pubmed/9492970>.
- Direk N, Williams S, Smith JA, Ripke S, Air T, Amare AT, et al. An analysis of two genome-wide association meta-analyses identifies a new locus for broad depression phenotype. *Biol Psychiatr*. 2016.
- Dunn EC, Wiste A, Radmanesh F, Almli LM, Gogarten SM, Sofer T, et al. Genome-wide association study (GWAS) and genome-wide by environment interaction study (GWEIS) of depressive symptoms in african american and hispanic/latina women. *Depress Anxiety*. 2016; 33(4):265–280. DOI: 10.1002/da.22484 [PubMed: 27038408]

- Genomes Project, C. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491(7422):56–65. DOI: 10.1038/nature11632 [PubMed: 23128226]
- Gogarten SM, Bhangale T, Conomos MP, Laurie CA, McHugh CP, Painter I, et al. GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. *Bioinformatics*. 2012; 28(24):3329–3331. DOI: 10.1093/bioinformatics/bts610 [PubMed: 23052040]
- Gonzalez P, Nunez A, Merz E, Brintz C, Weitzman O, Navas EL, et al. Measurement properties of the center for epidemiologic studies depression scale (CES-D 10): findings from HCHS/SOL. *Psychol Assess*. 2017; 29(4):372–381. DOI: 10.1037/pas0000330 [PubMed: 27295022]
- Hek K, Demirkan A, Lahti J, Terracciano A, Teumer A, Cornelis MC, et al. A genome-wide association study of depressive symptoms. *Biol Psychiatr*. 2013; 73(7):667–678. DOI: 10.1016/j.biopsych.2012.09.033
- High School Youth Risk Behavior Survey data. 2015. Retrieved from <https://nccd.cdc.gov/youthonline/App/Default.aspx>
- Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009; 5(6):e1000529.doi: 10.1371/journal.pgen.1000529 [PubMed: 19543373]
- Howie BN, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3: Genes, Genomes Genomics*. 2011; 1:457–470.
- Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat Genet*. 2016; 48(9):1031–1036. DOI: 10.1038/ng.3623 [PubMed: 27479909]
- Kendler KS, Gardner CO, Neale MC, Prescott CA. Genetic risk factors for major depression in men and women: similar or different heritabilities and same or partly distinct genes? *Psychol Med*. 2001; 31(4):605–616. [PubMed: 11352363]
- Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Arch Gen Psychiatr*. 2005; 62:593–602. [PubMed: 15939837]
- Kessler RC, Calabrese JR, Farley PA, Gruber MJ, Jewell MA, Katon W, et al. Composite International Diagnostic Interview screening scales for DSM-IV anxiety and mood disorders. *Psychol Med*. 2013; 43(8):1625–1637. DOI: 10.1017/S0033291712002334 [PubMed: 23075829]
- Landgraf BJ, McCarthy EL, Booker SJ. Radical S-Adenosylmethionine enzymes in human health and disease. *Annu Rev Biochem*. 2016; 85:485–514. DOI: 10.1146/annurev-biochem-060713-035504 [PubMed: 27145839]
- Laurie CC, Doheny KF, Mirel DB, Pugh EW, Bierut LJ, Bhangale T, et al. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol*. 2010; 34(6): 591–602. DOI: 10.1002/gepi.20516 [PubMed: 20718045]
- Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet*. 2012; 44(6):642–650. DOI: 10.1038/ng.2271 [PubMed: 22561516]
- Lavange LM, Kalsbeek WD, Sorlie PD, Aviles-Santa LM, Kaplan RC, Barnhart J, et al. Sample design and cohort selection in the hispanic community health study/study of Latinos. *Ann Epidemiol*. 2010; 20(8):642–649. DOI: 10.1016/j.annepidem.2010.05.006 [PubMed: 20609344]
- Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium. Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, et al. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatr*. 2013; 18(4):497–511.
- Mathers C, Ma Fat D, Boerma JT. The Global Burden of Disease: 2004 Update. 2008 (Retrieved from Geneva, Switzerland).
- Moon JY, Wang T, Sofer T, North KE, Isasi CR, Cai J., et al. Objectively measured physical activity, sedentary behavior and genetic predisposition to obesity in US hispanics/latinos: results from the hispanic community health study/study of Latinos (HCHS/SOL). *Diabetes*. 2017.
- Okbay A, Baselmans BM, De Neve JE, Turley P, Nivard MG, Fontana MA, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through

- genome-wide analyses. *Nat Genet.* 2016; 48(6):624–633. DOI: 10.1038/ng.3552 [PubMed: 27089181]
- Olvera RL, Bearden CE, Velligan DI, Almasy L, Carless MA, Curran JE, et al. Common genetic influences on depression, alcohol, and substance use disorders in Mexican-American families. *Am J Med Genet B Neuropsychiatr Genet.* 2011; 156B(5):561–568. DOI: 10.1002/ajmg.b.31196 [PubMed: 21557468]
- Passel JS, Cohn DV, Lopez MH. Census 2010: 50 Million Latinos, Hispanics Account for More than Half of Nation's Growth in Past Decade. 2011. Retrieved from www.pewhispanic.org
- Plomin R, Haworth CM, Davis OS. Common disorders are quantitative traits. *Nat Rev Genet.* 2009; 10:872–878. [PubMed: 19859063]
- Projections of the Resident Population by Race, Hispanic Origin, and Nativity: 2025 and 2050. Hyattsville, MD: US Census Bureau; 2003.
- Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* 2010; 26(18):2336–2337. [PubMed: 20634204]
- Qi Q, Stilp AM, Sofer T, Moon JY, Hidalgo B, Szpiro AA, et al. Genetics of type 2 diabetes in U.S. Hispanic/Latino individuals: results from the hispanic community health study/study of Latinos (HCHS/SOL). *Diabetes.* 2017; 66(5):1419–1425. DOI: 10.2337/db16-1150 [PubMed: 28254843]
- Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas.* 1977; 1:385–401.
- Schick UM, Jain D, Hodonsky CJ, Morrison JV, Davis JP, Brown L, et al. Genome-wide association study of platelet count identifies ancestry-specific loci in hispanic/latino Americans. *Am J Hum Genet.* 2016; 98(2):229–242. DOI: 10.1016/j.ajhg.2015.12.003 [PubMed: 26805783]
- Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatr.* 1998; 59(Suppl. 20):22–33. quiz 34–57.
- Shinar D, Gross CR, Price TR, Banko M, Bolduc PL, Robinson RG. Screening for depression in stroke patients: the reliability and validity of the center for epidemiologic studies depression scale. *Stroke.* 1986; 17(2):241–245. [PubMed: 3961834]
- Sofer T. Confidence intervals for heritability via Haseman-Elston regression. *Stat Appl Genet Mol Biol.* 2017; 16(4):259–273. DOI: 10.1515/sagmb-2016-0076 [PubMed: 28862991]
- Sofer T, Heller R, Bogomolov M, Avery CL, Graff M, North KE, Reiner AP, Thornton TA, Rice K, Benjamini Y, Laurie CC, Kerr KF. A powerful statistical framework for generalization testing in GWAS, with application to the HCHS/SOL. *Genet Epidemiol.* 2017; 41(3):251–258. [PubMed: 28090672]
- Sorlie PD, Aviles-Santa LM, Wassertheil-Smoller S, Kaplan RC, Daviglius ML, Giachello AL, et al. Design and implementation of the hispanic community health study/study of Latinos. *Ann Epidemiol.* 2010; 20(8):629–641. DOI: 10.1016/j.annepidem.2010.03.015 [PubMed: 20609343]
- Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta analysis. *Am J Psychiatr.* 2000; 157:1552–1562. [PubMed: 11007705]
- Ursano RJ, Colpe LJ, Heeringa SG, Kessler RC, Schoenbaum M, Stein MB, Army SC. The Army study to assess risk and resilience in servicemembers (Army STARRS). *Psychiatry.* 2014; 77(2): 107–119. DOI: 10.1521/psyc.2014.77.2.107 [PubMed: 24865195]
- van de Rest O, van der Zwaluw N, Beekman AT, de Groot LC, Geleijnse JM. The reliability of three depression rating scales in a general population of Dutch older persons. *Int J Geriatr Psychiatr.* 2010; 25(10):998–1005. DOI: 10.1002/gps.2449
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from next-generation sequencing data. *Nucleic Acids Res.* 2010; 3:e164.
- Ware EB, Mukherjee B, Sun YV, Diez-Roux AV, Kardina SL, Smith JA. Comparative genome-wide association studies of a depressive symptom phenotype in a repeated measures setting by race/ethnicity in the Multi-Ethnic Study of Atherosclerosis. *BMC Genet.* 2015; 16:118.doi: 10.1186/s12863-015-0274-0 [PubMed: 26459564]

- Wassertheil-Smoller S, Arredondo EM, Cai J, Castaneda SF, Choca JP, Gallo LC, et al. Depression, anxiety, antidepressant use, and cardiovascular disease among Hispanic men and women of different national backgrounds: results from the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol.* 2014; 24(11):822–830. DOI: 10.1016/j.annepidem.2014.09.003 [PubMed: 25439033]
- Wassertheil-Smoller S, Shumaker S, Ockene J, Talavera GA, Greenland P, Cochrane B, et al. Depression and cardiovascular sequela in postmenopausal women: the Women’s Health Initiative (WHI). *Arch Intern Med.* 2004; 164:289–298. [PubMed: 14769624]
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010; 26(17):2190–2191. [PubMed: 20616382]
- Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Sullivan PF. Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depressive disorder. *bioRxiv.* 2017.
- Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet.* 2010; 42(7):565–569. DOI: 10.1038/ng.608 [PubMed: 20562875]
- Zhong H, Prentice RL. Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. *Biostatistics.* 2008; 9(4):621–634. DOI: 10.1093/biostatistics/kxn001 [PubMed: 18310059]
- Zhong H, Prentice RL. Correcting “winner’s curse” in odds ratios from genomewide association findings for major complex human diseases. *Genet Epidemiol.* 2010; 34(1):78–91. DOI: 10.1002/gepi.20437 [PubMed: 19639606]

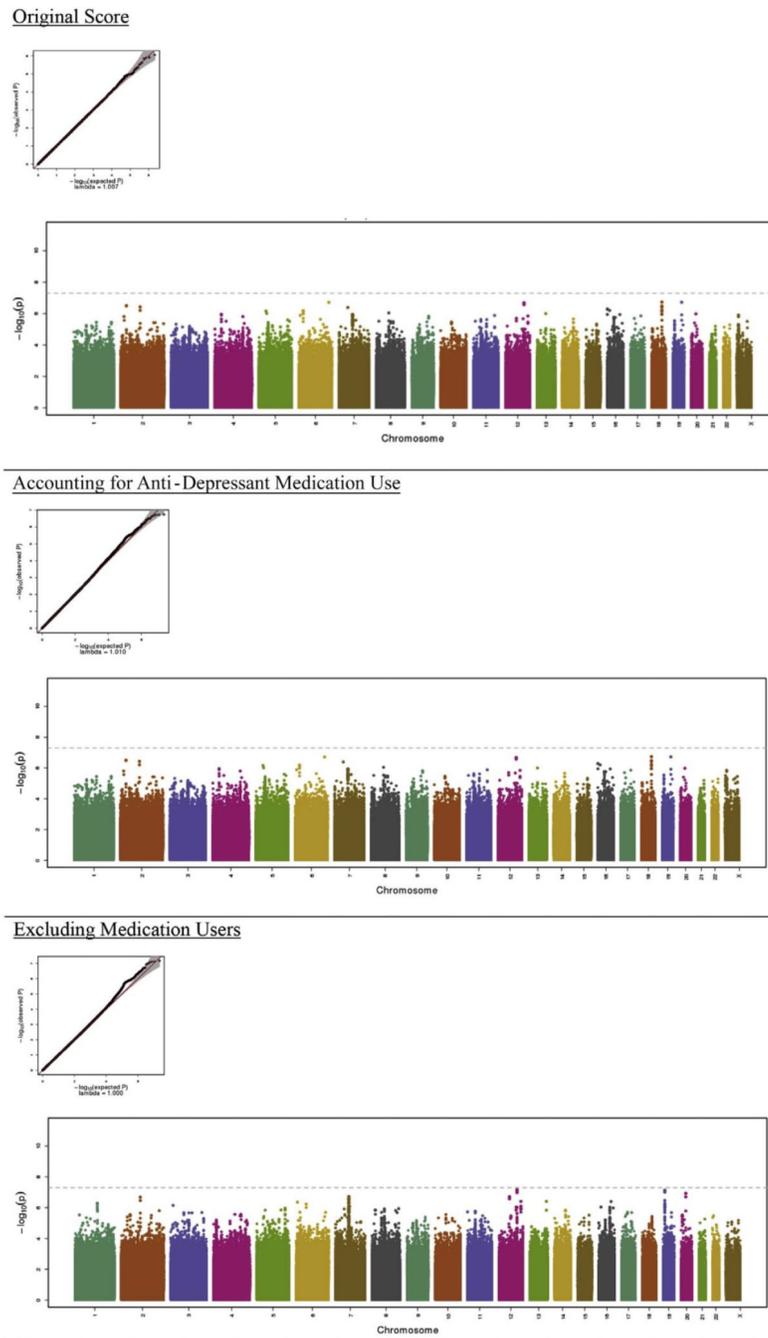


Fig. 1. Quantile-quantile (QQ) plots and Manhattan plots for depressive symptoms from the Hispanic Community Health Study/Study of Latinos. The quantile-quantile plots (“QQ-plots”), which present the observed by expected p-values on the $-\log_{10}$ scale, indicate conformity of the observed results to what would be expected under the null. In the Manhattan plots, the x-axis is the chromosomal position and the y-axis is the $-\log_{10}$ p-value for the association between each SNP and depressive symptoms derived from the linear

regression model. The dotted line shows the genome-wide significance level (5×10^{-8}). The displayed p-value corresponds to SNPs with effective $N > 30$.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

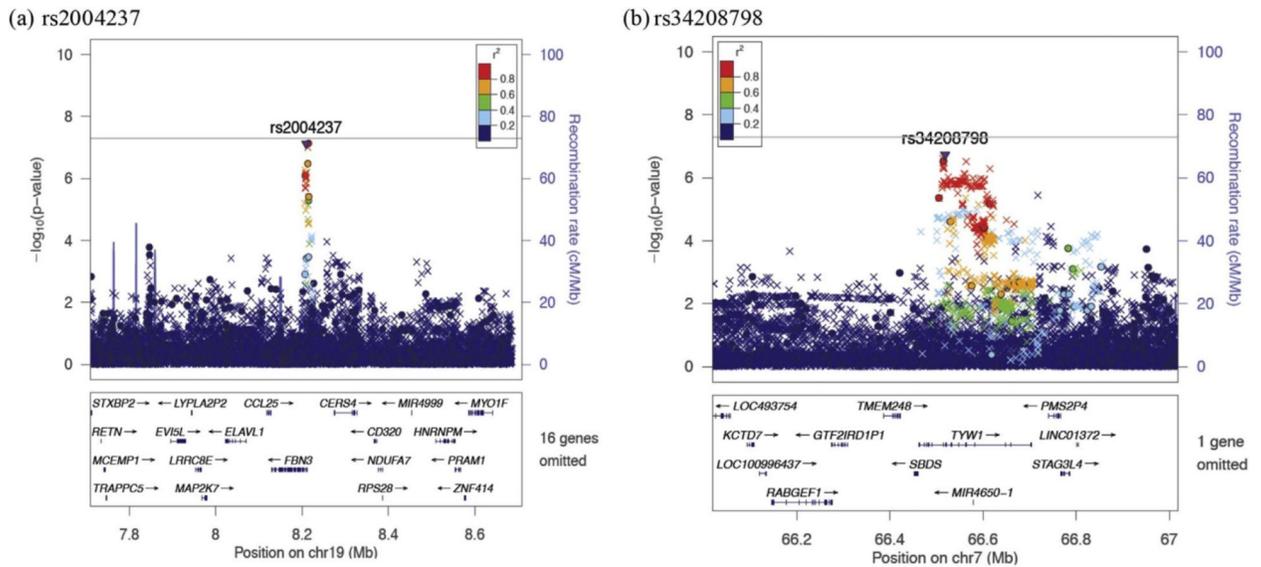


Fig. 2.

Regional association plot for top SNPs identified in the analysis excluding medication users. The regional association plot was generated using LocusZoom (<http://csg.sph.umich.edu/locuszoom/>) The left-side y-axis refers to the $-\log$ of the p-value corresponding to the test of association between each SNP (denoted as a colored dot, if genotyped, or X, if imputed) and depressive symptoms. SNPs are colored based on the level of linkage disequilibrium (LD) between each SNP and the index, genotyped, SNP (purple diamond). r^2 values are determined based on the HCHS/SOL data.

Table 1

Genome-wide association study results for the top loci ($p < 1 \times 10^{-6}$) with depressive symptoms scores for anti-depressant medication use.

SNP	CHR	position	alleleA	alleleB	MAF	minor allele	geno.	n	Beta	SE	pval	Location	Closest gene (< 20kb)
19:46976251:C:G	19	46976251	C	G	0.006	G	I	12310	-2.85	0.55	1.89E-07		
rs139628768	6	154761127	C	A	0.008	A	I	12310	-2.39	0.46	1.92E-07	Intron	CNKSR3
rs116270819	12	102245668	C	T	0.012	T	I	12310	-2.02	0.39	2.07E-07		
2:29851995:A:G	2	29851995	A	G	0.001	G	I	12310	-6.01	1.18	3.16E-07	Intron	ALK
rs187816263	18	58502295	G	C	0.004	C	I	12309	-3.46	0.68	3.26E-07		
rs13417200	2	107972669	T	C	0.006	C	I	12310	-2.73	0.54	3.71E-07	Intron	LOC105373536
7:39836135:C:T	7	39836135	C	T	0.004	T	I	12310	-3.57	0.7	4.05E-07	Intron	CICP22
rs374519284	16	295403	T	C	0.002	C	I	12310	-4.22	0.84	5.16E-07	Intron	ITFG3
rs144133307	18	58346558	A	G	0.004	A	I	12309	3.22	0.65	5.95E-07		
16:8248204:T:C	16	8248204	T	C	0.002	C	I	12310	-5.28	1.06	6.28E-07		
rs139550651	6	21825124	C	T	0.008	T	I	12310	2.32	0.47	6.52E-07	Intron	CASC15
5:35257239:T:C	5	35257239	T	C	0.004	C	I	12310	-3.72	0.75	6.89E-07		
rs72662446	8	59113186	C	T	0.032	T	G	12309	-1.14	0.23	9.04E-07		
rs56712382	5	39209149	G	GCA	0.279	G	I	12310	0.49	0.1	9.75E-07	Intron	FYB

CHR = chromosome. In the geno. (genotyping) column, G = genotyped and I = imputed. All imputed SNPs had info scores (indicating imputation quality) 0.70. AlleleA is the tested allele. Position is given in genome build GRCh37/hg19.

Table 2

Genome-wide association study results for the top loci ($p < 1 \times 10^{-5}$) with depressive symptoms, after excluding medication users.

SNP	CHR	position	alleleA	alleleB	MAF	minor allele	geno.	Beta	SE	pval	Location	Closest gene (< 20kb)
rs116270819	12	102245668	C	T	0.011	T	I	-2.02	0.38	6.69E-08		
rs2004237	19	8209194	T	C	0.313	C	I	0.45	0.08	7.54E-08	Intron	<i>FBN3</i>
rs62123284	19	8214984	A	G	0.312	G	G	0.45	0.08	7.68E-08		
rs77219363	20	24489072	C	CCAGGCTGATCT	0.034	C	I	1.25	0.24	1.17E-07	Intron	<i>SYNDIG1</i>
rs34208798	7	66518851	G	A	0.186	A	I	-0.52	0.10	1.87E-07	Intron	<i>TYW1</i>
rs189585011	12	59015294	T	A	0.005	A	I	-3.46	0.66	1.94E-07	Intron	<i>LOC100506869; LOC101927653</i>
rs13417200	2	107972669	T	C	0.006	C	I	-2.69	0.52	2.08E-07		
rs10228308	7	66515614	T	C	0.188	C	G	-0.51	0.10	3.13E-07	Intron	<i>TYW1</i>
rs6603149	19	8213231	A	G	0.382	G	G	0.40	0.08	3.41E-07	Intron	<i>FBN3</i>
rs902168	13	124492461	T	G	0.007	G	G	-2.33	0.46	3.88E-07	Intron	<i>ZNF664/ZNF664-FAM101A</i>
rs186703566	12	108834681	T	A	0.004	A	I	-3.21	0.63	3.88E-07		
rs7198462	16	77395589	T	C	0.010	C	I	-2.02	0.40	3.98E-07	Intron	<i>ADAMTS18</i>
rs12198471	6	6906545	C	T	0.196	T	G	0.50	0.10	4.39E-07		
rs112849100	1	154630337	T	C	0.002	C	I	-4.26	0.85	5.15E-07		
rs140262007	6	46983470	T	C	0.023	C	I	-1.33	0.27	5.84E-07	Intron	<i>ADGRF1</i>
rs10611601	7	70044230	TAA	T	0.434	TAA	I	-0.41	0.08	8.04E-07	Intron	<i>AUTS2</i>
16:8248204:T:C	16	8248204	T	C	0.002	C	I	-5.13	1.04	9.05E-07		
rs146173796	12	97310831	A	T	0.005	T	I	-3.07	0.62	9.08E-07	Intron	<i>NEDD1</i>
rs4985364	16	70806674	C	T	0.172	C	I	0.50	0.10	9.37E-07	Non-coding/intron	<i>VAC14/VAC14-AS1</i>
rs74495014	12	102270520	T	C	0.014	C	I	-1.76	0.36	9.54E-07	Upstream	<i>DRAM1</i>

These analyses were based on an analytic sample of $n = 11486$. CHR = chromosome. In the geno. (genotyping) column, G = genotyped and I = imputed. All imputed SNPs had info scores (indicating imputation quality) > 0.62 . AlleleA is the tested allele. Position is given in genome build GRCh37/hg19.

Table 3

Results for the loci identified in published GWAS of depression.

Pub.	Pop.	SNP	CHR	Position	Alleles		MAF	MA	Geno	Adjusted Depressive Symptoms Score				Excluding Medication Users				Binary Depression				
					a	b				n	Beta	SE	p-value	n	Beta	SE	p-value	n	Beta	SE	p-value	
Ware et al., 2015	Hispanics	rs1127233	3	124627024	T	G	0.23	G	I	12310	-0.085	0.094	.368	11486	0.017	0.090	.850	10478	-0.014	0.036	.701	- +--
Okbay et al., 2016	Europeans	rs7973260	12	118375486	A	G	0.16	A	G	12310	-0.120	0.108	.265	11486	-0.075	0.102	.466	10478	-0.030	0.041	.462	- +++
Okbay et al., 2016	Europeans	rs62100776	18	50754633	A	T	0.28	T	I	12309	-0.059	0.094	.526	11485	-0.039	0.089	.659	10477	-0.003	0.035	.944	+ +++
Okbay et al., 2016	Europeans	rs4346787*	5	164483794	G	A	0.37	G	I	12310	-0.238	0.086	.006	11486	-0.297	0.082	.0003	10478	-0.079	0.032	.015	+ +++
Okbay et al., 2016	Europeans	rs4481363*	5	164474719	C	A	0.39	C	I	12310	-0.238	0.085	.005	11486	-0.298	0.081	.0002	10478	-0.081	0.032	.011	- +++
Direk et al., 2016	Europeans	rs9825823	3	61082153	T	C	0.39	C	G	12310	0.084	0.083	.312	11486	0.141	0.080	.077	10478	0.001	0.032	.970	- ---
Hyde et al., 2016	Europeans	rs10514299	5	87663610	c	T	0.15	T	G	12310	0.041	0.114	.718	11486	0.065	0.109	.552	10478	0.036	0.043	.402	- ---
Hyde et al., 2016	Europeans	rs1518395	2	58208074	A	G	0.36	A	I	12310	-0.031	0.084	.713	11486	-0.117	0.080	.143	10478	0.002	0.032	.955	- +--
Hyde et al., 2016	Europeans	rs2179744	22	41621714	G	A	0.43	A	G	12310	-0.070	0.084	.404	11486	-0.095	0.080	.232	10478	-0.014	0.032	.661	+ +++
Hyde et al., 2016	Europeans	rs11209948	1	72811904	G	T	0.29	G	I	12310	0.039	0.089	.665	11486	0.028	0.085	.741	10478	0.001	0.034	.988	- ---
Hyde et al., 2016	Europeans	rs454214	5	88003403	C	T	0.41	T	G	12310	-0.016	0.083	.848	11486	-0.073	0.078	.349	10478	0.018	0.031	.568	+ +--
Hyde et al., 2016	Europeans	rs301806	1	8482078	C	T	0.38	C	G	12309	0.072	0.084	.393	11485	0.066	0.080	.409	10478	0.048	0.032	.132	+ ---
Hyde et al., 2016	Europeans	rs1475120	6	105389953	G	A	0.41	G	G	12310	-0.124	0.084	.138	11486	-0.017	0.079	.828	10478	-0.060	0.032	.057	+ +++
Hyde et al., 2016	Europeans	rs10786831	10	106614571	A	G	0.39	A	G	12310	0.099	0.084	.237	11486	0.083	0.080	.298	10478	0.016	0.032	.616	+ ---
Hyde et al., 2016	Europeans	rs12552	13	53625781	A	G	0.36	A	G	12310	0.068	0.084	.419	11486	0.071	0.080	.375	10478	0.027	0.032	.402	+ ---
Hyde et al., 2016	Europeans	rs6476606	9	37005561	A	G	0.44	A	G	12310	0.180	0.083	.030	11486	0.229	0.079	.004	10478	0.071	0.031	.024	+ ---
Hyde et al., 2016	Europeans	rs8025231	15	37648402	A	C	0.50	C	I	12310	-0.076	0.081	.352	11486	-0.104	0.077	.181	10478	-0.037	0.031	.231	- +++
Hyde et al., 2016	Europeans	rs12065553	1	80793118	A	G	0.22	G	G	12310	-0.070	0.098	.475	11486	-0.054	0.093	.566	10478	-0.008	0.037	.821	- +++
Hyde et al., 2016	Europeans	rs1656369	3	158280085	T	A	0.28	A	G	12310	0.009	0.089	.920	11486	-0.014	0.084	.870	10478	0.002	0.034	.946	--+++
Hyde et al., 2016	Europeans	rs4543289*	5	164484948	T	G	0.36	T	I	12310	-0.237	0.086	.006	11486	-0.314	0.082	.0001	10478	-0.075	0.032	.020	+ +++
Hyde et al., 2016	Europeans	rs2125716	12	84941429	A	G	0.25	A	G	12310	-0.036	0.093	.697	11486	-0.104	0.089	.238	10478	-0.018	0.036	.621	+ +++
Hyde et al., 2016	Europeans	rs2422321	1	73293393	G	A	0.37	G	I	12310	0.065	0.083	.438	11486	0.039	0.079	.625	10478	0.003	0.032	.926	- ---
Hyde et al., 2016	Europeans	rs7044150	9	2982931	T	C	0.24	T	G	12310	-0.008	0.097	.933	11486	-0.026	0.093	.775	10478	-0.016	0.036	.659	+ +++
CONVERGE Consortium, 2015	Chinese	rs12415800	10	69624180	A	G	0.08	A	I	12310	-0.126	0.146	.391	11486	-0.107	0.138	.441	10478	-0.052	0.058	.370	- +++
CONVERGE Consortium, 2015	Chinese	rs35936514	10	126244970	C	T	0.17	T	I	12310	0.043	0.109	.692	11486	0.061	0.103	.551	10478	-0.002	0.042	.955	+ ---

CHR = chromosome. In the geno, (genotyping) column, G = genotyped and I = imputed. All imputed SNPs had info scores (indicating imputation quality) 0.70. Allele a is the tested allele. MAF = minor allele frequency. MA = minor allele. Position is given in genome build GRCh37/hg19. +/- = Consistency of direction: + protective; - negative; sign prior to separation (!) indicates direction of effect in prior study; signs after separation indicate effects in HCHS/SOL for the 3 phenotypes (left to right: adjusted score; excluding medication users; binary depression).

* Loci are in high linkage disequilibrium and may be tagging the same signal ($r^2 > .8$).