Genome-Wide Association Study of Generalized Anxiety Symptoms in the Hispanic Community Health Study/Study of Latinos

Erin C. Dunn,1,2,3* Tamar Sofer,4 Linda C. Gallo,5 Stephanie M. Gogarten,4 Kathleen F. Kerr,4 Chia-Yen Chen,1,2,3,6 Murray B. Stein,7 Robert J. Ursano,9 Xiuxing Guo,9 Yucheng Jia,9 Qibin Qi,10 Jerome I. Rotter,9,11 Maria Argos,12 Jianwen Cai,13 Frank J. Penedo,14 Krista Perreira,15 Sylvia Wassertheil-Smoller,10 and Jordan W. Smoller1,2,3

1Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts
2Department of Psychiatry, Harvard Medical School, Boston, Massachusetts
3Stanley Center for Psychiatric Research, The Broad Institute of Harvard and MIT, Cambridge, Massachusetts
4Department of Biostatistics, University of Washington, Seattle, Washington
5Department of Psychology, San Diego State University, La Jolla, California
6Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts
7Department of Psychiatry, University of California San Diego, San Diego, California
8Department of Psychiatry, Uniformed Services University of the Health Sciences, Bethesda, Maryland
9Department of Pediatrics, Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute, Harbor-UCLA Medical Center, Torrance, California
10Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York
11Department of Medicine, Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute, Harbor-UCLA Medical Center, Torrance, California
12School of Public Health, University of Illinois at Chicago, Chicago, Illinois
13Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina
14Department of Medical Social Sciences, Northwestern University, Chicago, Illinois
15College and Arts and Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

Manuscript Received: 9 December 2015; Manuscript Accepted: 14 March 2016

Although generalized anxiety disorder (GAD) is heritable and aggregates in families, no genomic loci associated with GAD have been reported. We aimed to discover potential loci by conducting a genome-wide analysis of GAD symptoms in a large, population-based sample of Hispanic/Latino adults. Data came from 12,282 participants (aged 18–74) in the Hispanic Community Health Study/Study of Latinos. Using a shortened Spielberger Trait Anxiety measure, we analyzed the following: (i) a GAD symptoms score restricted to the three items tapping diagnostic features of GAD as defined by DSM-V; and (ii) a total trait anxiety score based on summing responses to all ten items. We first calculated the heritability due to common variants (h²_SNPH) and then conducted a genome-wide association study (GWAS) of GAD symptoms. Replication was attempted in three independent Hispanic cohorts (Multi-Ethnic Study of Atherosclerosis, Women’s Health Initiative, Army STARRS). The GAD symptoms score showed evidence of modest heritability (7.2%; P = 0.03), while the total trait anxiety score did not (4.97%; P = 0.20). One genotyped SNP (rs78602344) intronic to thrombospondin 2 (THBS2) was associated with the GAD symptoms score (P = 0.02).
nominally associated \( (P = 5.28 \times 10^{-5}) \) in the primary analysis adjusting for psychiatric medication use and significantly associated with the GAD symptoms score in the analysis excluding medication users \( (P = 4.18 \times 10^{-8}) \). However, meta-analysis of the replication samples did not support this association. Although we identified a genome-wide significant locus in this sample, we were unable to replicate this finding. Evidence for heritability was also only detected for GAD symptoms, and not the trait anxiety measure, suggesting differential genetic influences within the domain of trait anxiety. © 2016 Wiley Periodicals, Inc.

**Key words:** genetic association study; anxiety; Hispanics/Latinos

**INTRODUCTION**

Generalized anxiety disorder (GAD) is a mental disorder characterized by persistent uncontrollable worry and symptoms of arousal (e.g., restlessness, insomnia, muscle tension, irritability) [Hoge et al., 2012; American Psychiatric Association, 2013; Stein and Sareen, 2015]. GAD is common in the United States and worldwide [Grant et al., 2005; Kessler et al., 2005a,b; Wittchen and Jacobi, 2005; Wittchen et al., 2011]. Retrospective epidemiological studies suggest the past year prevalence of GAD is 3.1% and lifetime prevalence is 5.7% [Kessler et al., 2005a,b]. Even higher estimates have been observed from prospective studies (14.2% lifetime; 4.2% past year) [Moffitt et al., 2010]. Though GAD is about half as common in Hispanics/Latinos compared to Whites [Grant et al., 2005; Asnaani et al., 2010], Hispanics/Latinos represent one of the fastest growing populations in the United States [Passel et al., 2011; Brown, 2014, June 26], making the population burden of GAD in the United States, therefore, quite large. GAD is also a highly comorbid disorder, with about 90% of people with GAD experiencing at least one other DSM-IV Axis 1 or Axis 2 disorder [Grant et al., 2005]. Given its prevalence and profound social and economic costs [Hoffman et al., 2008; Newman et al., 2013], it is of strong interest to identify factors associated with the development of GAD.

Exploration of the role of genetic factors in the etiology of GAD is warranted as GAD appears attributable, in part, to genetic variation [Shimada-Sugimoto et al., 2015]. Family studies have found first degree relatives of people with GAD have six times the odds of having GAD compared to first degree relatives of those without GAD [Hettema et al., 2001]. Twin studies also suggest GAD is moderately heritable, with 32% of the variation in the population risk of GAD being attributable to genetic variation [Hettema et al., 2001]. Despite evidence of family aggregation, there have not yet been any published genome-wide association studies (GWAS) of GAD or GAD symptoms. Given the recent pursuit of GWAS for other anxiety disorders, notably post-traumatic stress disorder [Guffanti et al., 2013; Logue et al., 2013; Xie et al., 2013] and panic disorder [Otowa et al., 2009 2010; Erhardt et al., 2011], as well as efforts to examine domains related to GAD, including anxiety sensitivity [Davies et al., 2015], or composite indicators of anxiety disorder [Otowa et al., 2014], we sought to identify genomic loci linked to GAD by conducting a genome-wide analysis of GAD symptoms. We used a dimensional measure of trait anxiety symptoms chosen to match DSM-5 criteria for GAD. Use of a dimensional measure enables an examination of the full range of quantitative variation, rather than extremes in this quantitative distribution (e.g., cases versus controls) and may be a statistically more powerful approach to identify variants associated with GAD [Plomin et al., 2009].

In this report, we present results from the first GWAS of GAD symptoms, where we found a genome-wide significant association between a SNP intronic to thrombospindin 2 \((THBS2)\) and GAD symptoms in a large, diverse, and population-based sample of Hispanic/Latino adults. This finding did not replicate in a meta-analysis of three independent samples of Hispanic/Latino adults. We also present results from a SNP-chip heritability analysis, where we found evidence of modest heritability in GAD symptoms (7.2%), but no statistically significant heritability for a broader measure of trait anxiety symptoms.

**MATERIALS AND METHODS**

**Overview**

The Hispanic Community Health Study/Study of Latinos (HCHS/SOL) is a community-based prospective cohort study following 16,415 self-identified Hispanic/Latino adults (aged 18–74 at screening) and 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], participants were recruited via a stratified two-stage area probability sample of households across four cities in the United States (Chicago, IL; Miami, FL; Bronx, NY; San Diego, CA). The majority of the sample self-identified with the following background groups: Central American \((n = 1,730)\), Cuban \((n = 2,348)\), Dominican \((n = 1,460)\), Mexican \((n = 6,471)\), Puerto Rican \((n = 2,728)\), and South American \((n = 1,068)\). 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 12,254 respondents who consented to provide blood for the purpose of genotyping and had complete outcome and relevant covariates information (to be described
later), as well as non-missing records of antianxiety and antidepressants medication use.

**Phenotype Definition**

Anxiety symptoms were assessed at baseline using a 10-item Spielberger State-Trait Anxiety Inventory (STAI-T) administered in the participant’s preferred language (Spanish or English) [Bromberger and Matthews, 1996; Bergua et al., 2015]. This a short form version of the 20 item STAI-T [Spielberger, 1989], which is a valid and commonly used measure of trait anxiety symptoms in population-based studies (see e.g.: [De Moor et al., 2006; Caravati-Jouvenceaux et al., 2011]) that has been shown to correlate highly with other anxiety measures [Spielberger and Reheiser, 2009]. The abbreviated 10-item STAI-T short form has shown excellent internal consistency reliability in the full HCHS/SOL sample (α = 0.93) and for both the English (α = 0.92) and Spanish (α = 0.94) versions of the instrument [Wassertheil-Smoller et al., 2014]. It has been shown in other studies to correlate highly with the full version (α = 0.96) [Bromberger and Matthews, 1996]. For each item, participants were asked to indicate how they generally feel (0 = almost never; 1 = sometimes; 2 = often; 3 = almost always). Using the STAI short form, we created a GAD symptoms score by summing the three items (i.e., feeling nervous or restless; worrying over things that don’t matter; getting in a state of tension or turmoil as you think about recent concerns and interests) that are diagnostic criteria for GAD as defined by the DSM-5 [American Psychiatric Association, 2013]. The GAD symptoms score demonstrated moderate internal consistency reliability (α = 0.70) in the full HCHS/SOL sample. For comparison, we also examined a total trait anxiety score based on summing responses to all 10 items (i.e., the three GAD symptom score items noted above plus the following seven items: I feel satisfied with myself; I lack self-confidence; I feel secure; I feel inadequate; I am a steady person; I wish I could be as happy as others seem to be; I feel like a failure). Both phenotypes were coded so that higher scores indicated higher levels of anxiety.

To account for the possibility that current use of antidepressant or anxiolytic medications might affect anxiety scores, we applied an imputation algorithm to increase the scores of medication users. This algorithm was used in a previous GWAS of depressive symptoms [Hek et al., 2013] and was similar to an algorithm used to adjust blood pressure for persons on antihypertensive medications [Levy et al., 2000]. Antidepressant or anxiolytic medication use was determined by pill bottles brought by the participant to the baseline interview. Antidepressants were included, as this class of drugs are commonly prescribed to treat generalized anxiety symptoms [Kapczinski et al., 2003; Milea et al., 2010]. This algorithm assumed that: (i) the anxiety score of a respondent taking these psychotropic medications is lower (i.e., indicating fewer symptoms) than would be expected if the respondent were not taking these medications (thus, we assume that the medications are effective in reducing symptoms); (ii) respondents with high anxiety scores, on average, respond less to these medications than respondents with lower anxiety scores. The algorithm therefore, replaced the anxiety score of respondents on medications (n = 1,068) with the mean anxiety score of all respondents taking these medications that had the same or a higher anxiety score. For example, a medication user with an observed anxiety score of 10 would have a revised score of 21.07 (derived by taking the average anxiety score of medication users with an anxiety score value of 10 or greater). Anxiety scores for medication users were increased by 6.2 points on average above the raw score (raw scores ranged from 0 to 30).

**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 Americanpopulations, 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) according to 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 (note: participants could have been genetically related due to being drawn from the same household or different households living in the same community). Twelve thousand eight-hundred three 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 1 reference panel [1000 Genomes Project et al., 2012] and IMPUTE2 software [Howie et al., 2009, 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).

Principal components (PCs) and kinship coefficients were computed in an iterative manner to estimate both population structure and relatedness between study individuals such that the PCs were not affected by relatedness, and kinship estimates are not affected by ancestry. The process began with estimating relatedness using KING-robust [Manichaikul et al., 2010], followed by iterative estimation of PCs and kinship coefficients using PC-AiR [Conomos et al., 2015] and PC-Relate (https://www.bioconductor.org/packages/release/bioc/html/GENESIS.html), and is described comprehensively elsewhere [Conomos, 2014]. Consequently, 19 individuals who were identified to have primarily east Asian ancestry were excluded from analysis. For association analysis, the kinship
matrix was based on an independent set of SNPs selected with LD pruning.

Statistical Analyses

All analyses 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 reflect 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 PCs of ancestry. SNP annotation was performed using ANNOVAR [Wang et al., 2010] (http://annovar.openbioinformatics.org/en/latest/).

Heritability Analysis

We estimated “SNP-chip heritability,” or the narrow-sense heritability due to the additive effect of common variants (genotyped and imputed), 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 [Conomos et al., 2016; Schick et al., 2016]. For this analysis, the kinship matrix was calculated based on PC-relate using all autosomal SNPs, and the model was fit on a set of 10,414 unrelated individuals by removing participants so that the unrelated set did not have first-, second-, or third-degree relatives [Yang et al., 2010]. We conducted this analysis examining the GAD symptoms score as well as the total trait anxiety score to evaluate and compare SNP-chip heritability estimates across these phenotypes.

GWAS Analysis

We performed a GWAS using the linear mixed-effect model approach. All SNPs were modeled additively and the standard $5 \times 10^{-8}$ 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 0.5, we prioritized genotyped over imputed SNPs, we preferred imputed SNPs with higher quality score (info), lower $P$-values, and 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].

Secondary Analysis

As a secondary analysis, we repeated our analyses in the subset of non-medication users ($n = 11,456$; 91.5% of the sample) and using an untransformed score that did not consider medication use (i.e., the raw phenotype score).

Replication

We attempted replication of these results using data from three independent cohorts. Additional details about these cohorts are presented in Supplemental Materials. Briefly, the Women’s Health Initiative (The Women’s Health Initiative Study Group, 1998, WHI; www.whi.org [Wassertheil-Smoller et al., 2004]) provided data on Hispanic/Latina women ($n = 3,352$; mean age 60.0; SD = 6.57), where anxiety symptoms were measured using a single item (i.e., have you been a very nervous person in the past four weeks). The Multi-Ethnic Study of Atherosclerosis (MESA; http://www.mesa-nhlbi.org) [Bild et al., 2002] provided data from Hispanic/Latino adults ($n = 1,449$; mean age 61.38; SD = 10.30) where anxiety symptoms were measured using a scale identical to the HCHS/SOL. Finally, the Army Study To Assess Risk and Resilience in Service members (Army STARRS; http://www.armystarrs.org) [Ursano et al., 2014] provided data from Hispanic/Latino adults ($n = 3,394$; mean age 25.98; SD = 5.00), where anxiety symptoms were captured using a five-item scale designed to match DSM-IV criteria for GAD.

We meta-analyzed GWAS results across the three independent samples. As we were interested in testing whether the direction of effect was the same in the replication (as the discovery), one sided $P$-values were used [Heller et al., 2015]. Inverse variance weighted fixed-effect meta-analysis was conducted using METAL (http://www.sph.umich.edu/csg/abecasis/metal/)[Willer et al., 2010]).

<table>
<thead>
<tr>
<th>TABLE I. Results of Genome-Wide Complex Trait Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original scores</strong></td>
</tr>
<tr>
<td>GAD symptoms score</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Total trait anxiety score</td>
</tr>
<tr>
<td><strong>Accounting for medication use</strong></td>
</tr>
<tr>
<td>GAD symptoms score</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Total trait anxiety score</td>
</tr>
<tr>
<td><strong>Medication users removed</strong></td>
</tr>
<tr>
<td>GAD symptoms score</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Total trait anxiety score</td>
</tr>
</tbody>
</table>

$V(G)/V_p \times 100 = SNV$ heritability estimate ($h^2_{SNV} \times 100$). All phenotypes were treated as continuous. All models adjusted for sex, age, education (five-levels), principal components, study center, and sampling weights, and included random effects for the design variables kinship, household, and block unit, the three study design variables. $P$-values were calculated using the likelihood ratio test. The total trait anxiety score was derived by summing responses to all 10 items [see Supplemental Materials for listing of all items]. The GAD symptoms score was based on summing three items (i.e., feeling nervous or restless; worrying over things that don’t matter; getting in a state of tension or turmoil as you think about recent concerns and interests) that are diagnostic criteria for GAD as defined by the DSM-V.
A total of 12,282 Hispanic/Latino respondents were in the analysis. As expected, the GAD symptom score (skew = 0.63; kurtosis = 2.48) and total trait anxiety score (skew = 0.87; kurtosis = 3.21) were skewed towards lower values. No transformations of the outcome were performed as linear regression is robust to minor violations of normality [van Belle, 2002].

**FIG. 1.** Quantile–quantile (QQ) plots and Manhattan plots for GAD symptoms score 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 the GAD symptoms score derived from the linear regression model. The dotted line shows the genome-wide significance level ($5 \times 10^{-8}$). The displayed $P$-value corresponds to SNPs with effective $N > 30$. [Color figure can be seen in the online version of this article, available at http://wileyonlinelibrary.com/journal/ajmgb].
### Table II. Genome-Wide Association Study (GWAS) Results for the Top Loci \((P < 0.01 / 10^6)\) with the GAD Symptoms Score Imputed for Medication Use

<table>
<thead>
<tr>
<th>SNP</th>
<th>CHR</th>
<th>Position</th>
<th>AlleleA</th>
<th>AlleleB</th>
<th>MAF</th>
<th>Minor allele</th>
<th>Geno. n</th>
<th>Beta</th>
<th>SE</th>
<th>P-value</th>
<th>Location</th>
<th>Closest gene (&lt;20 kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10749727</td>
<td>1</td>
<td>62505653</td>
<td>A</td>
<td>G</td>
<td>0.005</td>
<td>A</td>
<td>12,282</td>
<td>1.97</td>
<td>0.36</td>
<td>5.28E-08</td>
<td>1.41E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs78602344</td>
<td>6</td>
<td>169626581</td>
<td>T</td>
<td>C</td>
<td>0.113</td>
<td>B</td>
<td>12,282</td>
<td>0.26</td>
<td>0.05</td>
<td>1.41E-07</td>
<td>4.04E-07</td>
<td>DRAF2</td>
</tr>
<tr>
<td>rs10994985</td>
<td>10</td>
<td>63713219</td>
<td>T</td>
<td>C</td>
<td>0.046</td>
<td>B</td>
<td>12,282</td>
<td>0.38</td>
<td>0.08</td>
<td>3.48E-07</td>
<td>3.70E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs3753389</td>
<td>1</td>
<td>160807153</td>
<td>T</td>
<td>C</td>
<td>0.378</td>
<td>A</td>
<td>12,282</td>
<td>0.16</td>
<td>0.03</td>
<td>3.70E-07</td>
<td>4.04E-07</td>
<td>CD244</td>
</tr>
<tr>
<td>rs12353562</td>
<td>X</td>
<td>96354053</td>
<td>G</td>
<td>A</td>
<td>0.099</td>
<td>B</td>
<td>12,282</td>
<td>0.23</td>
<td>0.05</td>
<td>4.04E-07</td>
<td>2.04E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs144417828</td>
<td>4</td>
<td>3740444</td>
<td>G</td>
<td>A</td>
<td>0.010</td>
<td>B</td>
<td>12,282</td>
<td>0.84</td>
<td>0.17</td>
<td>4.04E-07</td>
<td>2.04E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs148349076</td>
<td>16</td>
<td>54271147</td>
<td>C</td>
<td>T</td>
<td>0.004</td>
<td>B</td>
<td>12,282</td>
<td>1.29</td>
<td>0.26</td>
<td>6.78E-07</td>
<td>4.04E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs7094998</td>
<td>10</td>
<td>61306319</td>
<td>A</td>
<td>G</td>
<td>0.210</td>
<td>B</td>
<td>12,282</td>
<td>0.19</td>
<td>0.04</td>
<td>7.13E-07</td>
<td>4.04E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs17729883</td>
<td>8</td>
<td>9256631</td>
<td>T</td>
<td>C</td>
<td>0.220</td>
<td>B</td>
<td>12,282</td>
<td>0.19</td>
<td>0.04</td>
<td>7.29E-07</td>
<td>4.04E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs11803917</td>
<td>1</td>
<td>15611187</td>
<td>C</td>
<td>T</td>
<td>0.071</td>
<td>B</td>
<td>12,282</td>
<td>0.32</td>
<td>0.06</td>
<td>7.41E-07</td>
<td>4.04E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs75192612</td>
<td>7</td>
<td>88641306</td>
<td>A</td>
<td>G</td>
<td>0.004</td>
<td>A</td>
<td>12,282</td>
<td>0.16</td>
<td>0.03</td>
<td>7.49E-07</td>
<td>4.04E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs3110650</td>
<td>17</td>
<td>36075905</td>
<td>T</td>
<td>C</td>
<td>0.002</td>
<td>A</td>
<td>12,282</td>
<td>1.86</td>
<td>0.38</td>
<td>8.94E-07</td>
<td>4.04E-07</td>
<td>THBS2</td>
</tr>
</tbody>
</table>

**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 III. Genome-Wide Association Study (GWAS) Results for the Top Loci \((P < 0.01 / 10^6)\) with the GAD Symptoms Score, After Excluding Medication Users

<table>
<thead>
<tr>
<th>SNP</th>
<th>CHR</th>
<th>Position</th>
<th>AlleleA</th>
<th>AlleleB</th>
<th>MAF</th>
<th>Minor allele</th>
<th>Geno. n</th>
<th>Beta</th>
<th>SE</th>
<th>P-value</th>
<th>Location</th>
<th>Closest gene (&lt;20 kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs78602344</td>
<td>6</td>
<td>169626581</td>
<td>T</td>
<td>C</td>
<td>0.115</td>
<td>B</td>
<td>11,456</td>
<td>0.27</td>
<td>0.05</td>
<td>4.18E-08</td>
<td>1.81E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs141474992</td>
<td>4</td>
<td>75958947</td>
<td>A</td>
<td>C</td>
<td>0.006</td>
<td>B</td>
<td>11,456</td>
<td>1.04</td>
<td>0.20</td>
<td>1.83E-07</td>
<td>4.04E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs79812562</td>
<td>14</td>
<td>64593580</td>
<td>C</td>
<td>A</td>
<td>0.008</td>
<td>B</td>
<td>11,456</td>
<td>1.02</td>
<td>0.20</td>
<td>1.95E-07</td>
<td>4.04E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs146964092</td>
<td>6</td>
<td>39754875</td>
<td>A</td>
<td>G</td>
<td>0.001</td>
<td>B</td>
<td>11,456</td>
<td>2.16</td>
<td>0.42</td>
<td>2.34E-07</td>
<td>4.04E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs7350124</td>
<td>8</td>
<td>10625045</td>
<td>T</td>
<td>C</td>
<td>0.182</td>
<td>B</td>
<td>11,455</td>
<td>0.21</td>
<td>0.04</td>
<td>2.61E-07</td>
<td>4.04E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs186222942</td>
<td>20</td>
<td>24548719</td>
<td>G</td>
<td>A</td>
<td>0.004</td>
<td>B</td>
<td>11,456</td>
<td>0.37</td>
<td>0.09</td>
<td>3.16E-07</td>
<td>4.04E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs11776020</td>
<td>8</td>
<td>8809696</td>
<td>A</td>
<td>G</td>
<td>0.455</td>
<td>B</td>
<td>11,455</td>
<td>0.17</td>
<td>0.03</td>
<td>3.46E-07</td>
<td>4.04E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs115013535</td>
<td>6</td>
<td>55628681</td>
<td>C</td>
<td>A</td>
<td>0.004</td>
<td>B</td>
<td>11,456</td>
<td>1.30</td>
<td>0.26</td>
<td>4.16E-07</td>
<td>4.04E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs186294317</td>
<td>11</td>
<td>20267060</td>
<td>G</td>
<td>A</td>
<td>0.002</td>
<td>B</td>
<td>11,456</td>
<td>1.88</td>
<td>0.37</td>
<td>5.09E-07</td>
<td>4.04E-07</td>
<td>THBS2</td>
</tr>
<tr>
<td>rs11756502</td>
<td>6</td>
<td>169633185</td>
<td>C</td>
<td>T</td>
<td>0.315</td>
<td>B</td>
<td>11,456</td>
<td>0.16</td>
<td>0.03</td>
<td>5.09E-07</td>
<td>4.04E-07</td>
<td>THBS2</td>
</tr>
</tbody>
</table>
As shown in Table I, the GAD symptom score showed evidence of modest heritability ($h^2_{SNP} = 7.2\%$; $P = 0.03$), while the total trait anxiety score did not ($h^2_{SNP} = 4.97\%$; $P = 0.20$). Building from these results, we conducted a GWAS only on the GAD symptom score.

**Discovery Sample: GWAS**

The Manhattan and QQ plots are shown in Figure 1. As shown in the QQ plots, there was no evidence of inflation in either the GWAS of the full sample or the analysis that excluded medication users ($\lambda = 1.02$). No SNPs achieved genome-wide significance in the full sample, which included imputed scores for medication users (Table II). However, one genotyped SNP (rs78602344), located on chromosome six at position 169626581, emerged from both analyses. This SNP was the second most significant result in the full sample ($P = 1.41 \times 10^{-7}$) and the most significant result ($P = 4.18 \times 10^{-8}$) in the analysis excluding medication users (Table III). The SNP is intronic to thrombospondin 2 ($THBS2$), a gene that mediates cell-to-cell and cell-to-matrix interactions. Several other SNPs in the region also showed support for association (Fig. 2).

A second SNP with a low $P$-value in both analyses was rs17729883 (full sample $P = 7.29 \times 10^{-7}$; excluding medication users $P = 5.09 \times 10^{-7}$) located on chromosome eight. This genotyped SNP was located in an intron of an uncharacterized gene (LOC106379231; Supplemental Fig. S1).

All GWAS results at $P < 1 \times 10^{-5}$ are shown in the Supplemental Materials for the GAD symptom score for the full sample (Supplemental Table SI), excluding medication users (Supplemental Table SII), and for the original, non-transformed score (Supplemental Table SIII).

To determine which 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 according to our selection criteria detailed above. Replication power estimates were based on the projected samples sizes of each replication dataset (WHI = 3,000; MESA = 1,500; Army STARRS = 3,000) 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,” effectively attenuating the observed effect size. A prior study showed that attenuated effect size estimates tend to be closer than uncorrected estimates to effects seen in independent replication studies [Zhong and Prentice, 2010].

Our power analysis suggested that one SNP [rs78602344] would have excellent power in a meta analysis of the three replication cohorts after the winner’s curse bias correction (estimated power $= 0.96$; all other SNPs had weak power ($\leq 0.70$). We therefore carried forward this single SNP for replication.

**Replication Samples: GWAS Results**

In the replication phase, one SNP [rs78602344] was evaluated in three independent samples. This SNP was not significantly associated with the GAD symptom score in a meta analysis of the replication sites (Table IV).

**DISCUSSION**

The current study involved three major innovations in efforts to identify the genetic basis of generalized anxiety. First, to our knowledge, this was the first GWAS of GAD symptoms. Prior genetic association studies of GAD have focused on candidate gene polymorphisms, most of which have showed inconsistent results [Smoller, 2016]. Among GWAS, extant studies have focused on other anxiety disorders, including post-traumatic stress disorder [Guffanti et al., 2013; Logue et al., 2013; Xie et al., 2013] and panic disorder [Otowa et al., 2009, 2010; Erhardt et al., 2011], or have examined more global symptoms of trait anxiety in children [Trzaskowski et al., 2013] or composite indicators of anxiety disorder in adults [Otowa et al., 2014], but have not yet examined general symptoms of anxiety in adults. Second, our study was also the first to provide SNP-chip heritability estimates of GAD symptoms. Such analyses are important to provide upper- and lower-bound estimates of the additive genetic contribution to GAD. Finally, we conducted these genetic association analyses in Hispanics/Latinos, a large and growing US population group. Previous studies have largely focused on individuals of European ancestry.
### TABLE IV. Replication Results of rs78602344 for GAD Symptoms

#### A. Adjusting for medication use

<table>
<thead>
<tr>
<th>SNP</th>
<th>CHR</th>
<th>Position</th>
<th>AlleleA</th>
<th>AlleleB</th>
<th>MAF</th>
<th>Minor allele</th>
<th>Geno.</th>
<th>n</th>
<th>Beta</th>
<th>SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Army NSS1</td>
<td>6</td>
<td>169626581</td>
<td>T</td>
<td>C</td>
<td>0.108</td>
<td>C</td>
<td>G</td>
<td>12,282</td>
<td>−0.26</td>
<td>0.05</td>
<td>1.41E-07</td>
</tr>
<tr>
<td>Army NSS2</td>
<td>6</td>
<td>169626581</td>
<td>T</td>
<td>C</td>
<td>0.111</td>
<td>C</td>
<td>G</td>
<td>453</td>
<td>0.13</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>Army PPDS</td>
<td>6</td>
<td>169626581</td>
<td>T</td>
<td>C</td>
<td>0.108</td>
<td>C</td>
<td>G</td>
<td>1,533</td>
<td>0.17</td>
<td>0.30</td>
<td>0.57</td>
</tr>
<tr>
<td>MESA</td>
<td>6</td>
<td>169626581</td>
<td>T</td>
<td>C</td>
<td>0.133</td>
<td>C</td>
<td>G</td>
<td>1,441</td>
<td>0.02</td>
<td>0.06</td>
<td>0.73</td>
</tr>
<tr>
<td>WHI</td>
<td></td>
<td></td>
<td>C</td>
<td>T</td>
<td>0.206</td>
<td>T</td>
<td>G</td>
<td>2,950</td>
<td>−</td>
<td>0.04</td>
<td>0.54</td>
</tr>
<tr>
<td>Meta analysis</td>
<td></td>
<td></td>
<td>T</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td>7,785</td>
<td>0.03</td>
<td>0.03</td>
<td>0.43</td>
</tr>
</tbody>
</table>

CHR, chromosome. In the geno. (genotyping) column, G, genotyped and I, imputed. All imputed SNPs had info scores (indicating imputation quality) >0.83. AlleleA is the tested allele. The Army STARRs dataset comprised three different cohorts: the New Soldiers Study 1, the New Soldiers Study 2, and the Post-Deployment Study. The SNP identified in the discovery analysis and carried forward to the replication (rs78602344) was neither genotyped nor imputed in WHI. We, therefore, used the best proxy SNP (rs9505953) in closest LD ($r^2 = 0.15$).

#### B. Excluding medication users

<table>
<thead>
<tr>
<th>SNP</th>
<th>CHR</th>
<th>Position</th>
<th>AlleleA</th>
<th>AlleleB</th>
<th>MAF</th>
<th>Minor allele</th>
<th>Geno.</th>
<th>n</th>
<th>Beta</th>
<th>SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11,456</td>
<td>−0.27</td>
<td>0.05</td>
<td>4.18E-08</td>
</tr>
<tr>
<td>Army NSS1</td>
<td>6</td>
<td>169626581</td>
<td>T</td>
<td>C</td>
<td>0.108</td>
<td>C</td>
<td>G</td>
<td>1,372</td>
<td>0.11</td>
<td>0.34</td>
<td>0.74</td>
</tr>
<tr>
<td>Army NSS2</td>
<td>6</td>
<td>169626581</td>
<td>T</td>
<td>C</td>
<td>0.111</td>
<td>C</td>
<td>G</td>
<td>431</td>
<td>1.00</td>
<td>0.66</td>
<td>0.13</td>
</tr>
<tr>
<td>Army PPDS</td>
<td>6</td>
<td>169626581</td>
<td>T</td>
<td>C</td>
<td>0.108</td>
<td>C</td>
<td>G</td>
<td>1,430</td>
<td>−0.04</td>
<td>0.27</td>
<td>0.88</td>
</tr>
<tr>
<td>MESA</td>
<td>6</td>
<td>169626581</td>
<td>T</td>
<td>C</td>
<td>0.133</td>
<td>C</td>
<td>G</td>
<td>1,369</td>
<td>0.01</td>
<td>0.07</td>
<td>0.92</td>
</tr>
<tr>
<td>WHI</td>
<td></td>
<td></td>
<td>C</td>
<td>T</td>
<td>0.205</td>
<td>T</td>
<td>G</td>
<td>2,513</td>
<td>−0.01</td>
<td>0.04</td>
<td>0.77</td>
</tr>
<tr>
<td>Meta analysis</td>
<td></td>
<td></td>
<td>T</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td>7,115</td>
<td>0.01</td>
<td>0.03</td>
<td>0.71</td>
</tr>
</tbody>
</table>

CHR, chromosome. In the geno. (genotyping) column, G, genotyped and I, imputed. All imputed SNPs had info scores (indicating imputation quality) >0.83. AlleleA is the tested allele. The Army STARRs dataset comprised three different cohorts: the New Soldiers Study 1, the New Soldiers Study 2, and the Post-Deployment Study. The SNP identified in the discovery analysis and carried forward to the replication (rs78602344) was neither genotyped nor imputed in WHI. We, therefore, used the best proxy SNP (rs9505953) in closest LD ($r^2 = 0.15$).
Two findings emerged from the current study. First, results from the SNP-chip heritability analysis suggested that about 7.2% of the variance in GAD symptoms was explained by common genetic variants. This SNP heritability estimate is lower than those found for phobic anxiety ($h^2_{SNP} = 21\%$; $P = 0.01$) [Walter et al., 2013] and anxiety sensitivity ($h^2_{SNP} = 45\%$; 95%CI = 32%, 56%) [Davies et al., 2015] in adults, and also lower relative to estimates for a composite measure of anxiety traits in children, which was derived by summing measures of negative affect, negative cognition, fear, and social anxiety ($h^2_{SNP} = 16\%$; $P = 0.07$) [Trzaskowski et al., 2013]. The lower heritability estimates observed in this study relative to other studies conducted in adults may be due to the use of symptom scale, rather than a diagnostic measure of GAD. Interestingly, we also found that the total trait anxiety score, derived by summing all items on the scale (rather than just the three corresponding to GAD symptoms) carried no significant heritable signal. This result suggests that not all symptoms on existing anxiety scales may be equally influenced by additive genetic variation. Future studies using dimensional measures of anxiety symptoms may benefit from conducting similar analyses to determine whether an existing scale should be used in its entirety.

Second, we identified one genotyped SNP (rs78602344) located on chromosome six that was common to analyses accounting for psychiatric medication use or excluding medication users. Although not genome-wide significant in the former analysis, this SNP was genome-wide significant after excluding medication users ($P = 4.18 \times 10^{-8}$). This SNP is intronic to thrombospondin 2 (THBS2), a gene that mediates cell-to-cell and cell-to-matrix interactions. Several other SNPs in the region also showed support for association. However, this association was not supported in a meta-analysis of the three independent Hispanic/Latino replication samples ($n > 7,000$). We suspect that GWAS of GAD symptoms will likely share a similar trajectory as depressive symptoms, where increasing larger sample sizes and refinement of the phenotype will lead to the identification of associated loci [CONVERGE Consortium, 2015; Dunn et al., 2015].

We note several limitations of the current study. First, the outcomes were based on a brief inventory of trait anxiety symptoms. Although the widespread use of this anxiety measure in population-based studies allowed us to carry out the current analyses, future studies of diagnostic measures of GAD as well as more robust measures of GAD symptoms (from more detailed and specific measures or repeated phenotyping) are needed. Second, the replication samples were smaller and both more demographically and phenotypically heterogeneous than the HCHS/SOL discovery sample. Unfortunately, replication efforts are currently hampered by a lack of available data on anxiety symptoms in racial/ethnic minority populations. Third and relatedly, only one SNP was carried forward to the replication phase. This single SNP was the only one with high replication power. Moreover, greater insights are needed regarding the most optimal strategy to account for medication use in genetic association studies of quantitative traits. Future studies are needed to examine the suitability of different techniques and the extent to which different adjustment methods lead to different results (e.g., whether they substantially reduce variance if a substantial portion of the sample is assigned the same score; whether empirical data, such as medication efficacy, can be used to inform the adjustment strategy).

In conclusion, although the GWAS revealed a genome-wide significant locus in the discovery sample, we were unable to replicate this in independent samples. These findings underscore the need for even larger studies of GAD symptoms.

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, 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 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, U1L-TR-001079, U1L-TR-000040, 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 HHSN268271201100004C. 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 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 by the National Institute of Mental Health of the National Institutes of Health under Award Number K01MH102403 (Dunn) and K24MH094614 (Smoller), by a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation (Dunn), and a Tepper Family MGH Research Scholar Award (Smoller).

The authors thank Sandy Li and Jenna Kiely 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 Lavange et al. [2010] and is also available on the study website http://www.cscc.unc.edu/hchs/.
REFERENCES


SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher’s web-site.