Genome-wide Association Study of Cannabis Dependence Severity, Novel Risk Variants, and Shared Genetic Risks

Richard Sherva, PhD; Qian Wang, MS; Henry Kranzler, MD; Hongyu Zhao, PhD; Ryan Koesterer, MS; Aryeh Herman, PsyD; Lindsay A. Farrer, PhD; Joel Gelernter, MD

IMPORTANCE Cannabis dependence (CAD) is a serious problem worldwide and is of growing importance in the United States because cannabis is increasingly available legally. Although genetic factors contribute substantially to CAD risk, at present no well-established specific genetic risk factors for CAD have been elucidated.

OBJECTIVE To report findings for DSM-IV CAD criteria from association analyses performed in large cohorts of African American and European American participants from 3 studies of substance use disorder genetics.

DESIGN, SETTING, AND PARTICIPANTS This genome-wide association study for DSM-IV CAD criterion count was performed in 3 independent substance dependence cohorts (the Yale-Penn Study, Study of Addiction: Genetics and Environment [SAGE], and International Consortium on the Genetics of Heroin Dependence [ICGHD]). A referral sample and volunteers recruited in the community and from substance abuse treatment centers included 6000 African American and 8754 European American participants, including some from small families. Participants from the Yale-Penn Study were recruited from 2000 to 2013. Data were collected for the SAGE trial from 1990 to 2007 and for the ICGHD from 2004 to 2009. Data were analyzed from January 2, 2013, to November 9, 2015.

MAIN OUTCOMES AND MEASURES Criterion count for DSM-IV CAD.

RESULTS Among the 14 754 participants, 7879 were male, 6875 were female, and the mean (SD) age was 39.2 (10.2) years. Three independent regions with genome-wide significant single-nucleotide polymorphism associations were identified, considering the largest possible sample. These included rs143244591 (β = 0.54, P = 4.32 × 10⁻¹⁰ for the meta-analysis) in novel antisense transcript RP11-206M11.7; rs146091982 (β = 0.54, P = 1.33 × 10⁻⁹ for the meta-analysis) in the solute carrier family 35 member G1 gene (SLC35G1); and rs77378271 (β = 0.29, P = 2.13 × 10⁻⁶ for the meta-analysis) in the CUB and Sushi multiple domains 1 gene (CSMD1). Also noted was evidence of genome-level pleiotropy between CAD and major depressive disorder and for an association with single-nucleotide polymorphisms in genes associated with schizophrenia risk. Several of the genes identified have functions related to neuronal calcium homeostasis or central nervous system development.

CONCLUSIONS AND RELEVANCE These results are the first, to our knowledge, to identify specific CAD risk alleles and potential genetic factors contributing to the comorbidity of CAD with major depression and schizophrenia.
fter nicotine, cannabis is the most widely abused drug worldwide. In the United States, the accelerated decriminalization of cannabis is based on the erroneous perception that it is relatively harmless. In fact, cannabis use produces craving, dependence, and drug-seeking behavior, as with the use of other substances. Despite these risks, the prevalence of cannabis use and cannabis use disorders has dramatically increased since 2001, and the political momentum to increase availability has continued. Use of cannabis early in life is associated with an increased risk for schizophrenia (SCZ), and sets of SCZ-associated risk alleles predict cannabis use. Cannabis use is also a risk factor for depressive symptoms, and a twin study showed cannabis dependence (CAD) to be associated with an elevated risk for major depressive disorder (MDD). Substance use and other psychiatric illnesses may share common genetic risk factors; or reverse causation, self-medication, or confounding by other factors may explain their co-occurrence.

Despite knowledge of the neurobiology of the endocannabinoid system and its response to tetrahydrocannabinol, little is known about specific genetic factors influencing susceptibility to CAD or cannabis abuse. A twin study showed that several aspects of cannabis use are heritable, including an early opportunity to use ($h^2 = 72\%$), early onset of use ($h^2 = 80\%$), lifetime use of cannabis 11 or more times ($h^2 = 76\%$), and cannabis abuse or dependence ($h^2 = 21\%-72\%$), where $h^2$ is heritability. Possible evidence of linkage of cannabis use on chromosome 16 and linkage and association encompassing the neuregulin 1 gene ($NRG1$ [OMIM 142445]; known as a possible SCZ risk gene) on chromosome 8 have been found. Despite several genome-wide association studies (GWAS) on cannabis-related traits, no genome-wide significant (GWS) associations were observed for initiation of use or CAD. Herein we report on findings for DSM-IV CAD criteria from association analyses performed in large cohorts of African American and European American participants from 3 studies of substance use disorder genetics who underwent genotyping with genome-wide microarrays. The primary cohort has been used in previous studies to identify genes associated with opioid (OD), cocaine (CD), alcohol (AD), and nicotine (ND) dependence and posttraumatic stress disorder.

**Methods**

**Participants and Diagnostic Procedures**

The samples included 6000 African American and 8754 European American participants (race was assigned based on genetic data; eMethods in the Supplement) from the following 3 studies: (1) the Yale-Penn Study cohort of small nuclear families and unrelated individuals (2020 individuals in 850 families and 6951 unrelated individuals), collected to study the genetics of substance dependence (AD, ND, and CD); (2) the GWAS data set from the Study of Addiction: Genetics and Environment (SAGE), collected to study the genetics of AD, ND, and CD (183 individuals in 89 families and 3707 unrelated individuals); and (3) the GWAS International Consortium on the Genetics of Heroin Dependence (ICGHD), a collaboration formed to identify genes associated with heroin dependence risk (66 individuals in 33 families and 1827 unrelated individuals). The SAGE and ICGHD data sets are publicly available via application. The present study received institutional review board approval from all participating institutions, and written informed consent was obtained from all study participants.

Participants from the Yale-Penn Study were recruited from 2000 to 2013. These participants were administered the Semi-Structured Assessment for Drug Dependence and Alcoholism to derive DSM-IV diagnoses of lifetime CAD and other major psychiatric traits. Data were collected for the SAGE trial from 1990 to 2007, and participants underwent phenotyping with the Semi-Structured Assessment for the Genetics of Alcoholism. Data were collected for the ICGHD from 2004 to 2009, and participants completed a comprehensive psychiatric diagnostic interview based on the Semi-Structured Assessment of the Genetics of Alcoholism-Australia. The method of phenotyping was similar across the 3 samples. Additional information about recruitment, genotyping, imputation, and quality control for the study cohorts is provided in eMethods in the Supplement.

**Statistical Analysis**

Data were analyzed from January 2, 2013, to November 9, 2015. Association analyses were performed using a count of DSM-IV CAD criteria (0-7) as the outcome variable and the imputed minor allele dosage (adjusted for sex, age, and the first 3 ancestry principal components) as a predictor variable. This ordinal trait model has greater power to detect genetic associations than a univariate model based on disease status because of greater information content and improved specificity of the dependence measure. Association tests were performed using linear association models embedded in generalized estimating equations to correct for correlations among related individuals. Analyses were performed separately within each data set and population group, and the results were combined by meta-analysis using the inverse variance method implemented in the program METAL. Genomic inflation factors ($\lambda$) were calculated within each subpopulation, and $P$ values were corrected accordingly. We performed a second correction for the $\lambda$ factor calculated after the meta-analysis.

For the primary analysis, individuals were included regardless of cannabis exposure. As secondary analyses, individuals who reported never having used cannabis were excluded, and the primary model was repeated adjusting for the background.
criterion counts for AD, CD, and OD. Participants from 2 genotyping batches in the Yale-Penn cohort (Yale-Penn 1 and Yale-Penn 2) were combined with the SAGE sample to form a discovery data set. A sample consisting of the ICGHD data and additional samples from the Yale-Penn cohort who did not undergo genotyping at the time of the discovery analyses (Yale-Penn 3) were used to replicate the top associations.

**Cross-Disorder Analysis**

We attempted to uncover shared genetic variation between CAD and 5 psychiatric disorders, including SCZ, MDD, bipolar affective disorder, attention-deficit/hyperactivity disorder, and autism spectrum disorder using the GWAS analysis reported herein and publicly available GWAS results from the Psychiatric Genomics Consortium (http://www.med.unc.edu/pgc/). To explore cross-disorder genetic relationships, we used stratified quintile-quintile (QQ) plots to evaluate the relative enrichment of single-nucleotide polymorphisms (SNPs) associated with both disorders. The QQ plots, which contrast the observed distribution of \( P \) values with the expected distribution under the null hypothesis (uniform in GWAS), were used to assess \( P \) value inflation in the GWAS results. Grouping associated SNPs for one disorder and comparing (across groups) the QQ plots of another disorder, however, could also reveal the enrichment of GWAS signals between disorders, which made them suitable for cross-disorder enrichment screening.

We also applied a statistical framework for pleiotropy analysis, Genetic Analysis Incorporating Pleiotropy and Annotation (GPA). The GPA was built as a mixture model with parameters estimated using an efficient expectation-maximization algorithm, where associated SNPs were modeled with a \( \beta[a, 1] \) distribution and unassociated SNPs with a uniform [0, 1] distribution. A likelihood ratio test assessed the significance of pleiotropy between disorders. The GPA also detected the SNPs that were pleiotropic by calculating the posterior probability of association with both disorders.

**Results**

Participant demographic characteristics and the correlation between the criterion counts for CAD and other substance use disorder traits are shown in Table 1. The DSM-IV CAD criterion counts were significantly \((P < .05)\) correlated with the criterion counts for AD, CD, OD, and ND. The correlations varied by sample and population and ranged from \( r^2 = 0.15 \) for OD to \( r^2 = 0.61 \) for CD criteria. The CAD criterion counts were significantlyheritable in European American (19%-25%; \( P = .006 \)) but not African American (10%-11%; \( P = .08 \)) participants. eFigure 1 in the Supplement shows a histogram of the CAD criterion count in African American and European American participants in each cohort; 3 or more criteria indicate a diagnosis of CAD. The criterion count distribution is very similar in African American and European American participants. In the Yale-Penn sample, where comorbid psychiatric diagnoses were available, CAD was significantly associated with MDD in African American participants (odds ratio, 1.07; \( P = .006 \)) but not SCZ, bipolar affective disorder, attention-deficit/hyperactivity disorder, or autism spectrum disorder. Cannabis dependence was not associated with any of these disorders in European American participants.

**GWAS Results**

Manhattan and QQ plots for the meta-analysis discovery GWAS results for African American and European American Yale-Penn 1 and 2 and SAGE cohorts are displayed in eFigures 2 and 3 in the Supplement. We found little evidence of \( P \) value inflation. Table 2 shows associations in the discovery sample with \( P = 1.0 \times 10^{-5} \) in African American or European American participants or the combined meta-analysis, trimmed for linkage disequilibrium. eTable 1 in the Supplement shows the same results, together with additional information about each SNP, including the results within each discovery sample subgroup, after excluding individuals with no cannabis exposure, and after adjusting for comorbid substance use disorders. We identified GWS associations with reliably imputed SNPs in 3 distinct regions (Table 2), specific to African American participants and the combined meta-analysis, trimmed for linkage disequilibrium. First, rs186825689 \((P = 1.86 \times 10^{-8} \text{ for the African American meta-analysis})\) is located 12.4 kb upstream from the gene encoding S100 calcium binding protein (S100B) with contributions from both informative African American samples. Second, rs143244591 \((P = 2.18 \times 10^{-6} \text{ for the African American meta-analysis})\) maps to a novel antisense transcript RPII-206M11.7 (Havana gene: OTTHUMG00000159583) located in the gene of the same name on chromosome 3 with at least nominally significant evidence in each of the 3 African American samples. Third, rs77378271 \((P = 2.76 \times 10^{-6} \text{ for the European American meta-analysis})\) is an intronic SNP in the CUB

---

**Table 1. Demographic Characteristics of the Study Cohorts**

<table>
<thead>
<tr>
<th>Samplea</th>
<th>Age, Mean (SD)</th>
<th>Female Sex, No./Total No.</th>
<th>DSM-IV CAD Criterion Count, Mean (SD)</th>
<th>CAD Diagnosis, No./Total No. of Participants</th>
<th>Correlation With DSM-IV CAD Criterion Count, ( r^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yale-Penn African American</td>
<td>41.3 (9.7)</td>
<td>2209/4750</td>
<td>1.7 (2.2)</td>
<td>1296/4750</td>
<td>0.40</td>
</tr>
<tr>
<td>Yale-Penn European American</td>
<td>38.6 (11.5)</td>
<td>1712/4221</td>
<td>2.0 (2.3)</td>
<td>1388/4221</td>
<td>0.27</td>
</tr>
<tr>
<td>SAGE African American</td>
<td>39.9 (7.3)</td>
<td>638/1250</td>
<td>1.4 (2.2)</td>
<td>276/1250</td>
<td>0.43</td>
</tr>
<tr>
<td>SAGE European American</td>
<td>38.4 (9.7)</td>
<td>1478/2640</td>
<td>1.0 (1.9)</td>
<td>434/2640</td>
<td>0.51</td>
</tr>
<tr>
<td>ICGHD</td>
<td>36.2 (9.1)</td>
<td>883/1893</td>
<td>3.2 (2.5)</td>
<td>1062/1893</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Abbreviations: AD, alcohol dependence; CAD, cannabis dependence; CD, cocaine dependence; ICGHD, International Consortium on the Genetics of Heroin Dependence; ND, nicotine dependence; OD, opioid dependence; SAGE, Study of Addiction: Genetics and Environment.

* Samples are described in the Participants and Diagnostic Procedure subsection of the Methods section.
and Sushi multiple domains 1 gene (CSMD1 [OMIM 608397]) with evidence of association in 3 of the 6 samples. We also identified consistent, non-GWS evidence of association in the combined sample of European American and African American participants with a large block of SNPs in and around the phosphatidylinositol 4-kinase type β gene (PI4K2B [OMIM 612101]), with consistent effect direction in every European American and African American population tested (minimum $P = 1.74 \times 10^{-7}$ for the meta-analysis). This signal was GWS when individuals without cannabis exposure were excluded (minimum $P = 2.98 \times 10^{-8}$ for the meta-analysis).

Replication Results

The SNPs in Table 2 were tested for CAD association in the 2 replication samples (ICGHD and Yale-Penn 3). Table 3 shows the replication cohort-specific results for these SNPs, with the meta-analysis results from the discovery phase and the discovery + replication phase. The smallest P value in the ICGHD cohort among the 13 SNPs that could be reliably imputed and analyzed (this cohort was European Australian) was at rs74823926 ($P = .064$) in an intergenic region on chromosome 1. Several associations, however, were replicated in the Yale-Penn 3 sample (Table 3). The P values for 2 of the 3 GWS SNPs improved after meta-analysis with the replication cohorts (rs143244591 in RP11-206M11.7, from $1.38 \times 10^{-8}$ to $4.32 \times 10^{-10}$; rs77378271 in CSMD1, from $2.84 \times 10^{-8}$ to $2.13 \times 10^{-9}$), as did the P value for another SNP, rs146091982 in the solute carrier family 35 member G1 (SLC35G1 [Ensembl ENSG00000176273]) (from $3.13 \times 10^{-7}$ to $2.33 \times 10^{-7}$). The signal in PI4K2B also improved ($P = 5.57 \times 10^{-8}$ for the full meta-analysis). However, rs186825689 near S100B was no longer GWS ($P = 8.27 \times 10^{-4}$) in the full meta-analysis. The Figure shows Manhattan plots for the regions encompassing RP11-206M11.7 (Figure, A), SLC35G1 (Figure, B), CSMD1 (Figure, C), and PI4K2B (Figure, D) in the discovery sample and after meta-analysis with the replication samples.

Cross-Disorder Analysis Results

The QQ plots of 5 Psychiatric Genomics Consortium traits (SCZ, bipolar affective disorder, autism spectrum disorder, etc.,) and Sushi multiple domains 1 gene (CSMD1 [OMIM 608397]) with evidence of association in 3 of the 6 samples. We also identified consistent, non-GWS evidence of association in the combined sample of European American and African American participants with a large block of SNPs in and around the phosphatidylinositol 4-kinase type β gene (PI4K2B [OMIM 612101]), with consistent effect direction in every European American and African American population tested (minimum $P = 1.74 \times 10^{-7}$ for the meta-analysis). This signal was GWS when individuals without cannabis exposure were excluded (minimum $P = 2.98 \times 10^{-8}$ for the meta-analysis).

Replication Results

The SNPs in Table 2 were tested for CAD association in the 2 replication samples (ICGHD and Yale-Penn 3). Table 3 shows the replication cohort-specific results for these SNPs, with the meta-analysis results from the discovery phase and the discovery + replication phase. The smallest P value in the ICGHD cohort among the 13 SNPs that could be reliably imputed and analyzed (this cohort was European Australian) was at rs74823926 ($P = .064$) in an intergenic region on chromosome 1. Several associations, however, were replicated in the Yale-Penn 3 sample (Table 3). The P values for 2 of the 3 GWS SNPs improved after meta-analysis with the replication cohorts (rs143244591 in RP11-206M11.7, from $1.38 \times 10^{-8}$ to $4.32 \times 10^{-10}$; rs77378271 in CSMD1, from $2.84 \times 10^{-8}$ to $2.13 \times 10^{-9}$), as did the P value for another SNP, rs146091982 in the solute carrier family 35 member G1 (SLC35G1 [Ensembl ENSG00000176273]) (from $3.13 \times 10^{-7}$ to $2.33 \times 10^{-7}$). The signal in PI4K2B also improved ($P = 5.57 \times 10^{-8}$ for the full meta-analysis). However, rs186825689 near S100B was no longer GWS ($P = 8.27 \times 10^{-4}$) in the full meta-analysis. The Figure shows Manhattan plots for the regions encompassing RP11-206M11.7 (Figure, A), SLC35G1 (Figure, B), CSMD1 (Figure, C), and PI4K2B (Figure, D) in the discovery sample and after meta-analysis with the replication samples.

Cross-Disorder Analysis Results

The QQ plots of 5 Psychiatric Genomics Consortium traits (SCZ, bipolar affective disorder, autism spectrum disorder, etc.,) and Sushi multiple domains 1 gene (CSMD1 [OMIM 608397]) with evidence of association in 3 of the 6 samples. We also identified consistent, non-GWS evidence of association in the combined sample of European American and African American participants with a large block of SNPs in and around the phosphatidylinositol 4-kinase type β gene (PI4K2B [OMIM 612101]), with consistent effect direction in every European American and African American population tested (minimum $P = 1.74 \times 10^{-7}$ for the meta-analysis). This signal was GWS when individuals without cannabis exposure were excluded (minimum $P = 2.98 \times 10^{-8}$ for the meta-analysis).

Replication Results

The SNPs in Table 2 were tested for CAD association in the 2 replication samples (ICGHD and Yale-Penn 3). Table 3 shows the replication cohort-specific results for these SNPs, with the meta-analysis results from the discovery phase and the discovery + replication phase. The smallest P value in the ICGHD cohort among the 13 SNPs that could be reliably imputed and analyzed (this cohort was European Australian) was at rs74823926 ($P = .064$) in an intergenic region on chromosome 1. Several associations, however, were replicated in the Yale-Penn 3 sample (Table 3). The P values for 2 of the 3 GWS SNPs improved after meta-analysis with the replication cohorts (rs143244591 in RP11-206M11.7, from $1.38 \times 10^{-8}$ to $4.32 \times 10^{-10}$; rs77378271 in CSMD1, from $2.84 \times 10^{-8}$ to $2.13 \times 10^{-9}$), as did the P value for another SNP, rs146091982 in the solute carrier family 35 member G1 (SLC35G1 [Ensembl ENSG00000176273]) (from $3.13 \times 10^{-7}$ to $2.33 \times 10^{-7}$). The signal in PI4K2B also improved ($P = 5.57 \times 10^{-8}$ for the full meta-analysis). However, rs186825689 near S100B was no longer GWS ($P = 8.27 \times 10^{-4}$) in the full meta-analysis. The Figure shows Manhattan plots for the regions encompassing RP11-206M11.7 (Figure, A), SLC35G1 (Figure, B), CSMD1 (Figure, C), and PI4K2B (Figure, D) in the discovery sample and after meta-analysis with the replication samples.

Cross-Disorder Analysis Results

The QQ plots of 5 Psychiatric Genomics Consortium traits (SCZ, bipolar affective disorder, autism spectrum disorder, etc.,)
Abbreviations: ICGHD, International Consortium on the Geneticsof Heroin Opiate (can population yields significant evidence of CAD-MDD pleiotropy (P < .01) in the Supplement) in European American participants, but enrichment of the MDD GWAS signal in the CAD GWAS (eFigure 4 in the Supplement) in European American participants, but no clear enrichment for the other 4 psychiatric disorders in either population group (eFigure 5 in the Supplement).

We used GPA to test the significance of pleiotropy between CAD and the same 5 psychiatric disorders (eMethods in the Supplement). For each disease pair, we estimated the percentage of SNPs shared by 2 diseases and tested the significance of pleiotropy (eTable 2 in the Supplement). The European American population yielded significant evidence of CAD-MDD pleiotropy (P = 2.39 × 10⁻³); genome wide, 1.7% of all imputed SNPs were estimated to be associated with both CAD and MDD. Of these, rs10954732 in P450 oxidoreductase (POR [OMIM 124015]) had the largest posterior probability (although not significant) of association with both traits (P = 2.59 × 10⁻⁶ for CAD; P = .02 for MDD; posterior probability, 0.70).

**Discussion**

We report herein the first GWS results for CAD to our knowledge. The sample includes a large proportion (83%-36%, depending on race and cohort) of individuals with CAD from 2 ancestral populations in 3 independent cohorts. We identified 3 regions with GWS SNPs imputed to the 1000 Genomes American cohort, SAGE European American cohort, Yale-Penn 2 African American cohort, Yale-Penn 2 European American cohort, Yale-Penn 3 African American cohort, Yale-Penn 3 European American cohort, and the ICGHD European American cohort. * Indicates effect allele (listed in Table 2) is associated with an increase in cannabis dependence (CAD) criterion count; −, effect allele is associated with a decrease in CAD criterion count; and x, a valid effect estimate could not be obtained. See the Statistical Analysis subsection of the Methods section for an explanation of Yale-Penn 1, 2, and 3.

\[\text{\textbullet}}\] Indicates genome-wide significant P values.

\[\text{\textbullet}}\] Indicates P < .05 in the replication sample.
reference panel that implicate several biological processes and provide insight into the biology of CAD, including evidence of an inflammatory component in the disorder, which may also mediate risk for SCZ and MDD. The smallest P value observed ($P = 4.32 \times 10^{-10}$) was at rs143244591 in RP11-206M11.7. Little is known about this antisense transcript or which, if any, genes it regulates. Minor alleles were protective. The next most significant locus was SLC35G1 (rs146091982, $P = 1.33 \times 10^{-9}$), a potential member of the drug/metabolite transporter superfamily (EamA, previously DUF6). Ubiquitously expressed, SLC35G1 binds stromal interaction molecule 1, a calcium sensor that communicates the calcium load within the endoplasmic reticulum to store-operated channels in the plasma membrane when calcium stores in the endoplasmic reticulum are depleted. The third GWS locus, CSMD1 (rs77378271; $P = 2.13 \times 10^{-8}$), is highly expressed in the growth cones of developing central nervous system neurons, where it likely acts as a regulator of complement activation and inflammation. Different SNPs in CSMD1 have been associated with SCZ at the GWS level. Thus, CSMD1 is the second
gene to be implicated in both disorders (after NRGI\textsuperscript{16}) and may explain at least part of their shared genetic susceptibility.

Two other established SCZ risk genes, RIMS1 (OMIM 606629) (minimum SNP, $P = 1.59 \times 10^{-5}$) and MEF2C (minimum SNP, $P = 5.22 \times 10^{-5}$), showed suggestive association with CAD. MEF2C is highly expressed in developing mammalian neurons and is thought to mediate calcium-dependent survival of neurons that have made the appropriate synaptic connections.\textsuperscript{43} From a biological perspective, RIMS1 acts as a scaffold protein that regulates synaptic vesicle exocytosis, affecting cannabinoid receptor 1 (CRI)-mediated long-term suppression of γ-aminobutyric acid release, ultimately mediating presynaptic forms of long-term plasticity.\textsuperscript{44} Minor alleles at rs142305709 in RIMS1 were associated with fewer CAD criteria in African American participants. We observed at least a nominally significant signal in both Yale-Penn African American analysis subsets and a nonsignificant trend in SAGE African American participants. It is conceivable that some of the differences between the SAGE and the Yale-Penn African-American cohort are owing to the lack of replication in the African-American group. The region containing PI4K2B, which became GWAS associated (rs143244591 on chromosome 3) has little supportive evidence for association from other SNPs in the region, possibly owing to low linkage disequilibrium. However, despite stringent imputation quality thresholds for the region, possibly owing to low linkage disequilibrium.

Moreover, there was no association in the replication sample, this signal may represent an imputation artifact. Second, although none of the GWAS SNPs identified in the full GWAS analysis are rare, they could be described as infrequent, with minor allele frequencies in a range sometimes associated with false-positive results (4%-6%). Also, of the GWAS regions, only CSMD1 showed evidence of association in European American and African American participants. We observed at least a nominally significant signal in both Yale-Penn African American analysis subsets and a nonsignificant trend in SAGE African American participants.

Limitations of the GWAS findings should be noted. One of the top SNPs associated with CAD (AD, CD, and OD) in each of the 3 European American subgroups. Two of the 3 regions identified remained GWS (eTable 1 in the Supplement).

**Effect of Exposure Status and Comorbidity**

Because the inclusion of genetically at-risk individuals who never initiated cannabis use might have influenced our results, we repeated the primary analyses in the discovery cohort after removing unexposed individuals. Two of the 3 regions identified remained GWS (eTable 1 in the Supplement).

The P-value for rs143244591 on chromosome 3 improved slightly ($P = 1.13 \times 10^{-6}$, meta-analysis exposed) and was associated at $P < 0.02$ in each of the African American subgroups. The signal at rs77378271 in CSMD1 was almost identical ($P = 2.95 \times 10^{-8}$, meta-analysis exposed) and showed association at $P < 5.07 \times 10^{-3}$ in 2 of the 3 European American subgroups and at $P = 4.46 \times 10^{-4}$ in 1 of the African American subgroups. In addition, the block of SNPs in and around PI4K2B became GWS with a consistent effect direction (minor alleles being protective) in every European American and African American population tested and became GWS (minimum $P = 2.98 \times 10^{-8}$, meta-analysis exposed, at rs147170184). The evidence for pleiotropy between CAD and MDD was attenuated substantially ($P = 6.0$) after excluding unexposed participants. That the removal of unexposed individuals from the analysis had a relatively minor effect on the primary findings and actually improved the strength of some suggests that any loss in power owing to the smaller sample was offset by an increase in phenotypic precision. In the pleiotropy analysis, which relies on genome-level association results and is not limited to the most significantly associated SNPs, the power loss apparently outweighed any increase in precision. The significance of each of the top SNPs was modestly attenuated after adjusting for the DSM-IV criterion counts for AD, CD, and OD (eTable 1 in the Supplement).

**Ion Homeostasis and Addiction**

The previously published GWAS of OD\textsuperscript{19} and CD\textsuperscript{20} in a subset of this sample each identified risk genes and pathways involved in the regulation of neuronal calcium and potassium, and the pathway involving synaptic long-term potentiation was also identified for OD. Also, a cross-disorder analysis identified calcium signaling in neurons as a pathway mediating 5 psychiatric diseases, including SCZ and MDD.\textsuperscript{34} The GWS association in SLC35G1 and GWS (in the discovery sample only) associations in and around STO08 suggest ion homeostasis may play a role in CAD risk.

**Shared Risk for CAD and Other Psychiatric Disorders**

Many previous studies\textsuperscript{7,8,45,46} have focused on the relationship between CAD and SCZ, whereas the correlation between CAD and MDD has received much less attention. Although depressive disorders are highly comorbid with CAD in clinical settings,\textsuperscript{47} to our knowledge no previous genomics study has explored CAD-MDD pleiotropy. We found some evidence for genetic correlation between the risks for CAD and MDD. The existence of shared genetic factors for CAD-MDD is supported by the overlap in SNPs nominally associated with both traits, although we found no significant evidence of pleiotropy at any single SNP. We also found limited support for the possibility that such a relationship exists for CAD and SCZ based on relatively strong signals for both traits with variants in CSMD1 (although not the same variants). Nongenetic explanations such as...
patients with SCZ or MDD mediating the symptoms of these disorders with cannabis use might also explain the comorbidity. These analyses are exploratory, and follow-up studies to validate and extend these findings are necessary.

Conclusions

This study provided the first GWAS evidence to our knowledge for SNPs associated with CAD via GWAS in 3 distinct genomic locations. These findings will lead our understanding of genetic vulnerability to CAD in new directions that can inform our understanding of the biology of CAD. We obtained entirely novel evidence of genetic overlap between CAD and MDD and conclude that CSMD1 may be a candidate gene that affects the risk for CAD and SCZ, a topic of considerable research interest. These results also suggest that common pathways (nervous system development, inflammation, and ion homeostasis) mediate the risk for multiple psychiatric disorders and dependence on multiple substances, including cannabis.

ARTICLE INFORMATION

Submitted for Publication: November 20, 2015; final revision received January 8, 2015; accepted January 9, 2015.


Author Affiliations: Section of Biomedical Genetics, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts (Sherva, Koesterer, Farrer); Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut (Wang, Zhao); Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia (Kranzler); Mental Illness Research Education and Clinical Center, Veterans Affairs (VA) Stars and Stripes Healthcare Network, Philadelphia VA Medical Center, Philadelphia, Pennsylvania (Kranzler); Department of Genetics, Yale School of Medicine, West Haven, Connecticut (Zhao); Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut (Zhao); VA Cooperative Studies Program Coordinating Center, West Haven, Connecticut (Zhao, Gelernter); Department of Psychiatry, Yale School of Medicine, New Haven, Connecticut (Herman, Gelernter); Department of Neurology, Boston University School of Medicine, Boston, Massachusetts (Farrer); Department of Ophthalmology, Boston University School of Medicine, Boston, Massachusetts (Farrer); Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts (Farrer); Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts (Farrer); Department of Psychiatry, VA Connecticut Healthcare Center, Yale University School of Medicine, West Haven, Connecticut (Gelernter); Department of Neurobiology, Yale School of Medicine, New Haven, Connecticut (Gelernter).

Author Contributions: Dr Sherva had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Kranzler, Koesterer, Farrer, Gelernter. Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Sherva, Wang. Critical revision of the manuscript for important intellectual content: Sherva, Kranzler, Zhao, Koesterer, Herman, Farrer, Gelernter. Statistical analysis: Sherva, Wang, Zhao, Koesterer, Farrer. Obtained funding: Kranzler, Gelernter. Administrative, technical, or material support: Kranzler, Farrer, Gelernter. Study supervision: Kranzler, Zhao, Farrer, Gelernter.

Conflict of Interest Disclosures: Dr Kranzler reports being a consultant or an advisory board member for Alkermes, Indivior, Lundbeck, and Otsuka (unrelated to the present study) and being a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which is supported by AbbVie, Ethypharm, Lilly, Lundbeck, and Pfizer. No other disclosures were reported.

Funding/Support: This study was supported by grants R01 DA028909, R01 DA12690, R01 DA12849, R01 DA18432, R01 AA13330, and R01 AA07535 from the National Institutes of Health (NIH), the Veterans Affairs Connecticut Healthcare Center, and the Philadelphia Veterans Affairs Mental Illness Research, Education and Clinical Center. Funding support for the Study of Addiction: Genetics and Environment (SAGE) was provided by grant U01 HG004422 from the NIH Genes, Environment and Health Initiative. SAGE is one of the genome-wide association studies funded as part of the Gene Environment Association Studies under the Genes, Environment and Health Initiative. The portion of the genotyping that was performed at the Johns Hopkins University Center for Inherited Disease Research was supported by grant U01 HG004438 from the NIH Genes, Environment and Health Initiative. HHSN268200782096C (High Throughput Genotyping for Studying the Genetic Contributions to Human Disease) from the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, NIH. Support for collection of datasets and samples was provided by grant U01 AA008401 (Collaborative Study on the Genetics of Alcoholism), grant R01 CA128932 (Collaborative Genetic Study of Nicotine Dependence), and grant R01 DA034233 (Family Study of Cocaine Dependence) from the NIH. The International Consortium on the Genetics of Heroin Dependence (principal investigator, Elliot Nelson, MD) was funded by HHSN268200782096C (NIH contract: High Throughput Genotyping for Studying the Genetic Contributions to Human Disease) and HHSN26820110100I (NIH contract: High Throughput Genotyping for Studying the Genetic Contributions to Human Disease).

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Information: The publicly available datasets used for analysis were obtained from dbGAP (http://www.ncbi.nlm.nih.gov/gap) through accession numbers phs000092.v1 and phs000277.v1.

Additional Contributions: Work in recruitment and assessment was performed at Yale University School of Medicine and the APT Foundation by James Poling, PhD; at McLean Hospital by Roger Weiss, MD; at the Medical University of South Carolina by Kathleen Brady, MD, PhD, and Raymond Antonic, MD; and at the University of Pennsylvania by David Oslin, MD. Genotyping services for a part of our genome-wide association study were provided by the Center for Inherited Disease Research and Yale University (Center for Genome Analysis), which is fully funded by Federal contract N01-HG-65403 from the NIH to The Johns Hopkins University. Ann Marie Lacabelle, BS, Catherine Aldi, BS, and Christa Robinson, BS, provided technical assistance. The Semi-structured Assessment for Drug Dependence and Alcoholism interviewers, led by Yari Nuhez and Michelle Silvinsky, devoted substantial time and effort to phenotype the study sample. John Farrell, PhD, Section of Biomedical Genetics, Boston University School of Medicine, provided database management assistance. None of these individuals were compensated for this contribution to this study. Assistance with phenotype harmonization, genotype cleaning, and general study coordination was provided by the Gene Environment Association Studies Coordinating Center (supported by grant U01 HG004446 from the NIH). Assistance with data cleaning was provided by the National Center for Biotechnology Information.

REFERENCES


