HHS Public Access

Author manuscript

Behav Genet. Author manuscript; available in PMC 2021 May 01.

Published in final edited form as:

Behav Genet. 2020 May ; 50(3): 175–183. doi:10.1007/s10519-020-09999-3.

A Family-Based Genome Wide Association Study of Externalizing Behaviors

Peter B. Barr[1], **Jessica E. Salvatore**[1,2], **Leah Wetherill**[3], **Andrey Anokhin**[4], **Grace Chan**[5], **Howard J. Edenberg**[6], **Samuel Kuperman**[7], **Jacquelyn Meyers**[8], **John Nurnberger**[3], **Bernice Porjesz**[8], **Mark Shuckit**[9], **Danielle M. Dick**[1,10]

[1]Department of Psychology, Virginia Commonwealth University, Richmond, VA

[2]Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, School of Medicine, Virginia Commonwealth University, Richmond, VA

[3]Department of Medical and Molecular Genetics, School of Medicine, Indiana University, Indianapolis, IN

[4]Department of Psychiatry, School of Medicine, Washington University in St. Louis, St Louis, MO

[5]Department of Psychiatry, School of Medicine, University of Connecticut, Farmington, CT

[6]Department of Biochemistry and Molecular Biology, School of Medicine, Indiana University, Indianapolis, IN

[7]Department of Psychiatry, Carver College of Medicine, University of Iowa, Iowa City, IA

[8]Department of Psychiatry & Behavioral Sciences, SUNY Down State Medical Center, SUNY Downstate, Brooklyn, NY

[9]Department of Psychiatry, School of Medicine, University of California, San Diego, San Diego, CA

[10]Department of Human and Molecular Genetics, School of Medicine, Virginia Commonwealth University, Richmond, VA

Abstract

Shared genetic factors contribute to the high degree of comorbidity among externalizing problems (e.g. substance use and antisocial behavior). We leverage this common genetic etiology to identify

Corresponding Author: Peter Barr, PhD, Department of Psychology, Virginia Commonwealth University, 8 North Harrison St., Richmond, VA. 23284, (804) 828-1266, pbarr2@vcu.edu.

Conflict of Interest: The authors have no conflicts of interest to report.

Compliance with Ethical Standards

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

genetic influences externalizing problems in participants from the Collaborative Study on the Genetics of Alcoholism (European ancestry = 7,568; African ancestry = 3,274). We performed a family-based genome-wide association study (GWAS) on externalizing scores derived from criterion counts of five DSM disorders (alcohol dependence, alcohol abuse, illicit drug dependence, illicit drug abuse, and either antisocial personality disorder or conduct disorder). We meta analyzed these results with a similar measure of externalizing in an independent sample, Spit for Science (combined sample N = 15,112). We did not discover any robust genome-wide significant signals. Polygenic scores derived from the ancestry-specific GWAS summary statistics predicted externalizing problems in an independent European ancestry sample, but not in those of African ancestry. However, these PRS were no longer significant after adjusting for multiple testing. Larger samples with deep phenotyping are necessary for the discovery of SNPs related to externalizing problems.

Substance use disorders and antisocial behaviors are highly comorbid (Kessler et al., 2005), and load onto a single, highly heritable ($h^2 \sim 80\%$), latent factor generally referred to as the externalizing spectrum (Krueger et al., 2002). These disorders are associated with substantial costs to society (National Drug Intelligence Center, 2011; Sacks et al., 2015; World Health Organization, 2014), and to affected individuals and their families (Breslau et al., 2008; McLeod et al., 2012). Gene identification efforts for many of these disorders on the externalizing spectrum are still in their infancy, and to date few replicable associations have emerged from genome-wide association studies (GWAS) of externalizing spectrum phenotypes including alcohol consumption (Kranzler et al., 2019; Liu et al., 2019), alcohol use disorders (Kranzler et al., 2019; Walters et al., 2018), antisocial behaviors (Tielbeek et al., 2017), and cannabis use (Pasman et al., 2018), among others. For example, few genetic variants beyond those related to alcohol metabolism are consistently linked to alcohol misuse and dependence (Bierut et al., 2011; Gelernter et al., 2014; Liu et al., 2019; Walters et al., 2018).

Evidence from multiple twin and family studies indicates that there is a substantial degree of genetic overlap among externalizing disorders including alcohol dependence and abuse, illicit drug dependence and abuse, antisocial personality disorder, and conduct disorder (Hicks et al., 2011; Kendler et al., 2003; Krueger et al., 2002). Thus our understanding of the molecular genetic basis of externalizing spectrum disorders may benefit from an alternative GWAS approach that reflects the shared genetic architecture among these disorders (Dick et al., 2008; McGue et al., 2013). In the present study, we took this approach and conducted a GWAS of an externalizing disorder score in the Collaborative Study on the Genetics of Alcoholism, which includes individuals of European and African ancestry. We recognize that our sample size is underpowered, so in addition to SNP level replication we examined whether polygenic scores derived from summary statistics predicted externalizing problems in an independent sample.

Methods

Samples

Collaborative Study on the Genetics of Alcoholism (COGA)—COGA participants came from alcohol dependent probands identified through alcohol treatment programs at six U.S. sites and were invited to participate if they had a sufficiently large family (usually sibships > 3 with parents available) with two or more members in the COGA catchment area (Begleiter et al., 1995; Edenberg, 2002). The Institutional Review Boards at all sites approved this study and written consent was obtained from all participants. A portion of the original COGA participants were assessed twice, and a subset of the adolescent and young adult offspring in COGA families have been assessed multiple times since 2004 (Bucholz et al., 2017). Participants completed the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA), a valid and reliable polydiagnostic interview (Bucholz et al., 1994). Those 18 years of age or older received the adult SSAGA and those who were under the age of 18 received an adolescent version of the SSAGA designed for those between ages 12–17 (Bucholz et al., 1994; Hesselbrock et al., 1999). We used data from the interview where participants endorsed the maximum total number of DSM-IV alcohol dependence criteria (i.e., the phenotype on which study probands were initially recruited) in order to capture their maximal phenotypic expression of externalizing behaviors. We included data from all COGA participants to create the discovery phenotype (n = 16,167). Genetic analyses were limited to those with genome-wide data (n = 11,562). A full description of genotyping and quality control is available in the supplementary material (Section 1).

Spit for Science (S4S)

Spit for Science (S4S) is an ongoing study of college students enrolled at a large, urban university in the Mid-Atlantic region (Dick et al., 2014). Spit for Science is a cohort sequential, longitudinal study of incoming students, with the goal of understanding genetic and environmental influences on health and well-being in college students. Participants enrolled in the study in the fall of their freshman year, and follow-up assessments were completed in each subsequent spring semester. Individuals who did not participate in the first wave of data collection had the opportunity to join the study the following spring of their freshman year. Those who participated during their first year and remained enrolled at the university were eligible to complete yearly follow-up assessments each spring. Participants who completed the assessments were eligible to provide a DNA sample. Of those who completed the initial survey, 98% provided a DNA sample. A full description of genotyping, quality control, ancestry assignment, and the derivation of ancestry-specific principal components for this sample can be found in the Supplemental Information (Section 1) and in Peterson et al. (2017).

Measures

Externalizing Score—We created a composite measure of externalizing behaviors using the first principal component extracted from a principal components analysis. In COGA, we used DSM-IV clinical criterion counts (American Psychiatric Association, 1994) for five externalizing disorders: (1) alcohol dependence, (2) alcohol abuse, (3) illicit drug dependence (cocaine, marijuana, sedatives, stimulants, opioids, and other drugs), (4) illicit

drug abuse, and (5) antisocial behavior measured as either DSM-IV antisocial personality disorder or DSM-III-R conduct disorder criteria (American Psychiatric Association, 1987) depending on age. For those younger than 18, we used conduct disorder (CD) criteria as our index of antisocial behavior, as ASPD is only assessed in those 18 years and older. We omitted other disorders often included on the externalizing spectrum, including nicotine dependence and attention-deficit hyperactivity disorder (ADHD) because these have weaker genetic overlap with the above phenotypes (Young et al., 2002). In S4S, we used both clinical and non-clinical phenotypes meant to match the phenotypes used in COGA. Indicators included: (1) alcohol dependence, (2) alcohol abuse, (3) high school antisocial behavior, (4) college antisocial behavior, and (5) a count of illicit drug use. The first principal component accounted for approximately 68% of the common variance among the externalizing variables in COGA and 42% in S4S. A full description of the externalizing score construction can be found in the supplemental material (Section 2).

Family-based GWAS

We utilized a generalized estimating equation (GEE) framework to control for relatedness via a kinship matrix (Chen & Yang, 2010) in the EA and AA family samples separately. Autosomal SNPs with minor allele frequency (MAF) ≥ 3% (estimated using founders in the EA and AA families separately) were included in the discovery analysis to test for association using an additive model of SNP effects. Due to the skewed distribution of externalizing scores, we transformed scores for the analysis (square root, left anchored at 1). Sex, age of last observation, GWAS array, and the first ten principal components computed from GWAS data were included as covariates. Because we focused primarily on the results from the combined-ancestry meta analysis, we used established thresholds for genome wide significance ($p < 5 \times 10^{-8}$). Finally, we used LD score regression (Bulik-Sullivan et al., 2015) on ancestry-specific summary statistics to 1) determine the degree to which inflation in the association test statistics was due to population stratification rather than polygenic signal; and 2) estimate the narrow sense, SNP-based heritability (h^2_{snp}).

GWAS meta-analysis

We performed meta-analysis across the EA and AA GWAS samples in COGA (referred to as within COGA meta analysis) in METAL using an inverse-variance weighting fixed effects approach (Willer et al., 2010). We focus on the meta analysis results, as there was not sufficient sample size to focus on ancestry specific results. Genomic control was applied in METAL to all GWAS results. We utilized the publicly available online tool FUMA (Functional Mapping and Annotation of Genome-Wide Association Studies) (Watanabe et al., 2017) to evaluate *cis*-acting chromatin interactions and expression quantitative trait loci (eQTLs). Gene-based test results are provided for genes with p-values less than the adjusted alpha, based on a correction of 18,012 protein-coding genes ($0.05/18012=2.78\times10^{-6}$). Genetic background for linkage disequilibrium (LD) was based on the observed LD patterns in each ancestry sample. For meta-analysis results, we used LD from the full 1000 genomes reference panel (The 1000 Genomes Project Consortium, 2015).

Validation Methods

In order to validate the results from the current GWAS, we employed a series of follow up analyses. First, we meta-analyzed GWAS results across COGA and S4S (referred to as cross-sample meta analysis) in METAL using an sample-size weighted approach (Willer et al., 2010) to examine whether the inclusion of additional samples improved estimates for top SNPs in the within COGA meta-analysis. Second, we examined whether genome-wide significant or suggestive SNPs identified in the within COGA meta analysis were associated with externalizing in S4S. Because we were underpowered to detect individual SNP effects, we also employed a "holistic replication" strategy used by others in gene-identification (Karlsson Linnér et al., 2019; Okbay et al., 2016) that assessed the number of lead SNPs with concordant signs, the number of lead SNPs with $p \leq 0.05$, and the proportion of lead SNPs with $p \leq 0.05$ given all lead SNPs in the validation cohort.

Finally, we used summary statistics to create genome-wide polygenic scores (PGS) in the S4S sample. As our discovery sample included individuals of EA and AA ancestry, we selected the same ancestral populations in S4S for our validation effort (62% Female; European ancestry N = 2,761; African ancestry N = 1,175). We conducted all analyses separately by ancestry using ancestry specific GWAS results for constructing PGS. In order to maximize available sample size, we ran 10 additional GWAS in S4S using a leave-one-out (LOO) strategy in which 10% of each sample was omitted from the GWAS. PGS were then constructed from the meta-analyzed results of the ancestry specific GWAS in COGA and the S4S sample in which each hold out was not included, similar to previous GWAS (Otowa et al., 2016). For example, the PGS for the first 10% of EA removed in S4S were constructed from GWAS weights of the EA results in COGA meta analyzed with GWAS weights of the remaining 90% of S4S respondents using a sample-size based meta analysis (Willer et al., 2010)

We used the *clump* and *score* procedures separately by ancestry in PLINK to sum each individual's total number of minor alleles from the score SNPs, with each SNP weighted by the negative log of the association *p* value and sign of the association (beta) statistic from the ancestry-specific results. Clumping was done with respect to the linkage disequilibrium (LD) pattern in the appropriate 1000 Genome Phase 3 ancestry sample (EUR and AFR) using a 500kb physical distance and an LD threshold of $r^2 >= 0.25$. Thus, the polygenic scores were constructed from SNPs that captured independent genetic association signals from the discovery GWAS. We calculated a series of scores that included SNPs meeting increasingly stringent *p*-value thresholds ($p < .0001$, $p < .001$, $p < .01$, $p < .05$, $p < .10$, $p < .20$, $p < .30$, $p < .40$, and $p < .50$). All models included sex and the first 10 within-ancestry principal components as covariates.

Results

Externalizing behaviors and score descriptive statistics

Descriptive statistics for each of the externalizing behaviors and the composite externalizing score within each ancestry group are shown in Table I. Males reported higher externalizing behaviors compared to females in both EA (Male mean [SD] = 0.3 [1.05]; female mean [SD]

= −0.24 [0.84]; p < 0.0001) and AA (Male mean [SD] = 0.35 [1.06]; female mean [SD] = −0.22 [0.86]; p < 0.0001) samples. Although EA and AA did not differ in their mean externalizing scores, EA respondents had significantly higher levels of alcohol dependence, alcohol abuse, and illicit drug abuse symptoms compared to AA respondents. AA respondents had significantly higher levels of illicit drug dependence and ASPD/CD symptoms compared to EA respondents.

Within COGA GWAS Meta Analysis

Figures 1 shows the Manhattan plot for the association results in the meta-analysis (lambda = 1.004) for the 15,537,641 SNPs (genotyped and imputed) for the EA (lambda = 1.034) and AA GWAS (lambda = 1.026). We identified 3 independent, genome-wide significant (GWS) SNPs in the meta analysis. The top SNP ($p = 3.91\times10^{-9}$) rs2376620, was on chromosome 6 in the *CDKN1A* (cyclin dependent kinase inhibitor 1A) gene. Inspection of the ancestry-specific results showed this effect was driven primarily by the association in EA families ($p = 2.43\times10^{-8}$, MAF = 0.148). Although this SNP was not significantly associated in the AA families (p = 0.009, MAF = 0.242), the association was in the same direction, which improved evidence of association in the meta-analysis. The next SNP identified in the meta-analysis was rs2433198 ($p = 1.78\times10^{-8}$) in the region *GCOM1/MYZAP* (GRINL1A complex locus /myocardial zonula adherens protein) on chromosome 15, with support from both the EA families ($p = 8.75\times10^{-6}$, MAF = 0.450) and the AA families ($p = 2.86\times10^{-4}$, MAF = 0.467). The final GWS SNP rs12928255 ($p = 1.93\times10^{-8}$) was in the *PKD1L2* (polycystin 1 like 2) gene on chromosome 16, driven by associations in both EA families ($p = 1.26\times10^{-7}$, MAF = 0.381) and AA families (p = 0.024, MAF = 0.218). All independent SNPs with $p < 1.0\times10^{-7}$ from the meta analysis are shown in Table II. Estimates of SNP-based heritability (h^2_{snp}) from the summary statistics using LD score regression (Bulik-Sullivan et al., 2015) did not differ significantly from zero in either the EA or AA results.

Gene-based analyses partially confirmed findings from the GWAS. Figure 2 presents the gene-based Manhattan plot of the meta-analyzed results. Gene-based meta-analysis revealed significant association in the *GCOM1/MYZAP* region on chromosome 15 ($p = 8.09 \times 10^{-7}$). Additionally, all three genome-wide findings are in genes with chromatin and eQTL interactions with other nearby genes (see supplemental information for detailed plots).

Cross Sample Meta Analysis

Table III presents the top independent SNPs from the cross sample meta analysis, which included all of the COGA and S4S samples (N = 15,112). We present two sets of results, the cross sample meta analysis results for the top 3 SNPs identified in the within COGA meta analysis (to test if initial SNPs replicated); and the actual top SNPs identified in the cross sample meta analysis (to test if the increased sample resulted in additional signals). For each of the top genome-wide significant SNPs identified in the within COGA meta analysis (rs2376620, rs2433198, rs12928255), the effect was reduced in the cross sample meta analysis and these SNPs were no longer genome-wide significant. Of the top SNPs in the cross-sample meta analysis, none were significant at the genome-wide threshold. Overall, the larger cross sample meta-analysis did not support replication of SNPs identified in the primary within COGA meta analysis.

In order to further explore whether the SNPs identified in the within COGA meta analysis would replicate in S4S, we looked at the association statistics for the S4S specific GWAS. None of the individuals SNPs replicated. Only 3 of the seven lead SNPs had concordant signs across each sample. We would expect 50% of the SNPs to have the same sign by chance, and 5% of lead SNPs to have a $p < 0.05$, so this also failed replication.

Polygenic Validation

Results for the PGS analysis can be found in Table IV. Among the Spit for Science participants of European ancestry, estimates for the effect of PGS on externalizing were consistently in the expected direction, whereby higher polygenic scores were associated with higher levels of externalizing behavior. As *p* values for inclusion in the PGS became less stringent, polygenic scores became more strongly associated with the outcome, in terms of ΔR^2. Overall the polygenic score that included SNPs meeting $p < .10$ from the COGA GWAS/S4S LOO GWAS meta-analysis was most strongly associated with externalizing in S4S, though the variance the scores explained was small (approximately 0.20%) and the p-value was not significant after correcting for a false discovery rate of 5% (Benjamini & Hochberg, 1995).

Among individuals of African ancestry, we did not find any evidence of significant associations in the validation sample.

Discussion

Our goal was to capitalize on the genetic architecture of externalizing spectrum disorders to identify molecular genetic influences that these disorders share. To do this, we conducted a family-based genome-wide association study of an externalizing score in a predominantly clinically ascertained sample. We identified 3 independent genome-wide significant loci from the meta-analysis of the European and African ancestry groups in COGA. However we did not find evidence for replication of these loci and we did not discover any robust genome-wide significant signals or polygenic associations. We also note that supplementary analyses using an alternative linear mixed modeling approach, where all populations are analyzed together with principal components and a genetic relatedness matrix (GRM) to adjust for population stratification (Sul et al., 2018; Wojcik et al., 2019), produced highly similar results (supplementary information, Section 3).

The most complete analysis, a meta analysis that included both COGA and S4S, did not produce any genome-wide significant results. Additionally, none of the signals from the within COGA meta analysis replicated. Only one SNP from the within COGA meta analysis was located in a gene (*PKD1L2* on chromosome 16) with prior evidence of association with externalizing problems (Anney et al., 2008). It is possible the failure of this SNP to replicate in the combined sample meta analysis was a lack of statistical power. It is also possible that some or all of the associations from the within COGA meta analysis were false positives.

Differences in sample characteristics between COGA and Add Health could have introduced additional noise into our estimates. While we tried to make the phenotypes across sample match as closely as possible, they were not identical. There were important differences

between samples, as COGA was enriched for families with a high density of alcohol dependence, while S4S is college based. These differences can influence the performance of polygenic scores (Savage et al., 2018). There was also greater variation in the age of COGA participants, whereas young adults in the S4S sample may have yet to reach the highest risk period for the onset of these externalizing problems.

We found initial evidence of polygenic replication in the European ancestry sample of S4S, but not the African ancestry sample. However, the PGS were no longer significant after correcting for multiple testing. The lack of replication is most likely due to lack of power in the discovery GWAS. Other issues, including population differences (e.g. allele frequencies or LD structure), or gene-environment interactions could also influence the poor performance of PGS. Because the LD blocks in African ancestry groups are much shorter, tagging SNPs correlated with a causal SNP becomes less likely (Campbell & Tishkoff, 2008). Even larger samples likely will be necessary for gene discovery in African ancestry populations (Dick et al., 2017). Additionally, individuals of African ancestry tend to identify as African-Americans (Banda et al., 2015), and AA individuals experience more adverse environments on average compared to EA individuals (Williams & Mohammed, 2009). This is important in view of findings from twin studies that genetic influences on externalizing behaviors are reduced and shared environmental amplified in adverse environmental conditions (Tuvblad et al., 2006).

These analyses have several important limitations. We recognize that both our GWAS and additional meta analyses were severely underpowered. Sample sizes in the hundreds of thousands or more are necessary to identify robust genetic signals. Second, our measure of externalizing was mostly limited to criteria for psychiatric disorders. There are other heritable dimensions of personality, like sensation seeking and impulsivity, related to these behaviors not captured in the current phenotype (Krueger et al., 2002).

Despite these limitations, the current analyses aimed at using refined phenotypes to help us to better understand genetic pathways of risk for externalizing disorders. Because individual phenotypes may serve as indicators of some broader underlying risk, efforts focused on understanding the genetic comorbidity underlying multiple disorders with a shared genetic architecture are likely to be fruitful. Future efforts utilizing new methods, such as Genomic SEM (Grotzinger et al., 2019), that can leverage summary statistics from multiple well-powered GWAS of traits on the externalizing spectrum may be able to identify genetic variants through examining the overlapping genetic architecture among these traits.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The Collaborative Study on the Genetics of Alcoholism (COGA), Principal Investigators B. Porjesz, V. Hesselbrock, H. Edenberg, L. Bierut, includes eleven different centers: University of Connecticut (V. Hesselbrock); Indiana University (H.J. Edenberg, J. Nurnberger Jr., T. Foroud); University of Iowa (S. Kuperman, J. Kramer); SUNY Downstate (B. Porjesz); Washington University in St. Louis (L. Bierut, J. Rice, K. Bucholz, A. Agrawal); University of California at San Diego (M. Schuckit); Rutgers University (J. Tischfield, A. Brooks); Department of

Biomedical and Health Informatics, The Children's Hospital of Philadelphia; Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA (L. Almasy), Virginia Commonwealth University (D. Dick), Icahn School of Medicine at Mount Sinai (A. Goate), and Howard University (R. Taylor). Other COGA collaborators include: L. Bauer (University of Connecticut); J. McClintick, L. Wetherill, X. Xuei, Y. Liu, D. Lai, S. O'Connor, M. Plawecki, S. Lourens (Indiana University); G. Chan (University of Iowa; University of Connecticut); J. Meyers, D. Chorlian, C. Kamarajan, A. Pandey, J. Zhang (SUNY Downstate); J.-C. Wang, M. Kapoor, S. Bertelsen (Icahn School of Medicine at Mount Sinai); A. Anokhin, V. McCutcheon, S. Saccone (Washington University); J. Salvatore, F. Aliev, J. Su, S. I-Chun Kuo, B. Cho (Virginia Commonwealth University); and Mark Kos (University of Texas Rio Grande Valley). A. Parsian and H. Chin are the NIAAA Staff Collaborators.

We continue to be inspired by our memories of Henri Begleiter and Theodore Reich, founding PI and Co-PI of COGA, and also owe a debt of gratitude to other past organizers of COGA, including Ting-Kai Li, P. Michael Conneally, Raymond Crowe, and Wendy Reich, for their critical contributions. This national collaborative study is supported by NIH Grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA). Research reported in this publication was supported by the National Institute on Alcohol Abuse and Alcoholism of the National Institutes of Health under award numbers K02AA018755 and K01AA024152. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Funding: This study was funded by the National Institutes of Health through the National Institute on Alcohol Abuse and Alcoholism (U10AA008401, K02AA018755, K01AA024152) and the National Institute on Drug Abuse (U10AA008401).

References

American Psychiatric Association. (1987). Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition. Washington, D. C.

American Psychiatric Association. (1994). Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) Washington, D.C.

Anney RJL, Lasky-Su J, Ó'Dúshláine C, Kenny E, Neale BM, Mulligan A, … Gill M (2008). Conduct disorder and ADHD: Evaluation of conduct problems as a categorical and quantitative trait in the international multicentre ADHD genetics study. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 147B(8), 1369–1378. doi:10.1002/ajmg.b.30871 [PubMed: 18951430]

Banda Y, Kvale MN, Hoffmann TJ, Hesselson SE, Ranatunga D, Tang H, … Risch N (2015). Characterizing Race/Ethnicity and Genetic Ancestry for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. Genetics, 200(4), 1285. [PubMed: 26092716]

Begleiter H, Reich T, Hesselbrock V, Porjesz B, Li TK, Schuckit MA, … Rice JP (1995). The Collaborative Study on the Genetics of Alcoholism. Alcohol Health & Research World, 19, 228–236.

Benjamini Y, & Hochberg Y (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological), 57(1), 289–300.

Bierut LJ, Goate AM, Breslau N, Johnson EO, Bertelsen S, Fox L, … Edenberg HJ (2011). ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. Mol Psychiatry, 17, 445. doi:10.1038/mp.2011.12410.1038/mp.2011.124https://www.nature.com/articles/mp2011124#supplementary-informationhttps://www.nature.com/articles/mp2011124#supplementary-information [PubMed: 21968928]

Breslau J, Lane M, Sampson N, & Kessler RC (2008). Mental disorders and subsequent educational attainment in a US national sample. Journal of psychiatric research, 42(9), 708–716. doi:10.1016/j.jpsychires.2008.01.016 [PubMed: 18331741]

Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI, … Schuckit MA (1994). A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. Journal of Studies on Alcohol, 55(2), 149–158. doi:10.15288/jsa.1994.55.149 [PubMed: 8189735]

Bucholz KK, McCutcheon VV, Agrawal A, Dick DM, Hesselbrock VM, Kramer J, Porjesz B (2017). Comparison of Parent, Peer, Psychiatric, and Cannabis Use Influences Across Stages of Offspring

Alcohol Involvement: Evidence from the COGA Prospective Study. Alcoholism: Clinical and Experimental Research, 41(2), 359–368. doi:10.1111/acer.13293

Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics, C., … Neale BM (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet, 47(3), 291–295. doi:10.1038/ng.3211 [PubMed: 25642630]

Campbell MC, & Tishkoff SA (2008). African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu Rev Genomics Hum Genet, 9, 403–433. doi:10.1146/annurev.genom.9.081307.164258 [PubMed: 18593304]

Chen M-H, & Yang Q (2010). GWAF: an R package for genome-wide association analyses with family data. Bioinformatics, 26(4), 580–581. doi:10.1093/bioinformatics/btp710 [PubMed: 20040588]

Dick DM, Aliev F, Wang JC, & et al. (2008). Using dimensional models of externalizing psychopathology to aid in gene identification. Archives of general psychiatry, 65(3), 310–318. [PubMed: 18316677]

Dick DM, Barr P, Guy M, Nasim A, & Scott D (2017). Review: Genetic research on alcohol use outcomes in African American populations: A review of the literature, associated challenges, and implications. The American Journal on Addictions, 26(5), 486–493. doi:10.1111/ajad.12495 [PubMed: 28240821]

Dick DM, Nasim A, Edwards AC, Salvatore JE, Cho SB, Adkins A, … Kendler KS (2014). Spit for Science: launching a longitudinal study of genetic and environmental influences on substance use and emotional health at a large US university. Frontiers in Genetics, 5, 47. doi:10.3389/fgene.2014.00047 [PubMed: 24639683]

Edenberg HJ (2002). The collaborative study on the genetics of alcoholism: an update. Alcohol Res Health, 26(3), 214–218. [PubMed: 12875050]

Gelernter J, Kranzler HR, Sherva R, Almasy L, Koesterer R, Smith AH, … Farrer LA (2014). Genome-wide association study of alcohol dependence:significant findings in African- and European-Americans including novel risk loci. Mol Psychiatry, 19(1), 41–49. doi:10.1038/mp.2013.145 [PubMed: 24166409]

Grotzinger AD, Rhemtulla M, de Vlaming R, Ritchie SJ, Mallard TT, Hill WD, … Tucker-Drob EM (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. Nature Human Behaviour. doi:10.1038/s41562-019-0566-x

Hesselbrock M, Easton C, Bucholz KK, Schuckit M, & Hesselbrock V (1999). A validity study of the SSAGA--a comparison with the SCAN. Addiction, 94(9), 1361–1370. doi:10.1046/j.1360-0443.1999.94913618.x [PubMed: 10615721]

Hicks BM, Schalet BD, Malone SM, Iacono WG, & McGue M (2011). Psychometric and Genetic Architecture of Substance Use Disorder and Behavioral Disinhibition Measures for Gene Association Studies. Behavior genetics, 41(4), 459–475. doi:10.1007/s10519-010-9417-2 [PubMed: 21153693]

Karlsson Linnér R, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, … Social Science Genetic Association, C. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nature Genetics, 51(2), 245–257. doi:10.1038/s41588-018-0309-3 [PubMed: 30643258]

Kendler KS, Prescott CA, Myers J, & Neale MC (2003). The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. Archives of general psychiatry, 60(9), 929–937. doi:10.1001/archpsyc.60.9.929 [PubMed: 12963675]

Kessler RC, Chiu WT, Demler O, & Walters EE (2005). Prevalence, Severity, and Comorbidity of Twelve-month DSM-IV Disorders in the National Comorbidity Survey Replication (NCS-R). Archives of general psychiatry, 62(6), 617–627. doi:10.1001/archpsyc.62.6.617 [PubMed: 15939839]

Kranzler HR, Zhou H, Kember RL, Vickers Smith R, Justice AC, Damrauer S, … Gelernter J (2019). Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. Nat Commun, 10(1), 1499. doi:10.1038/s41467-019-09480-8 [PubMed: 30940813]

Krueger RF, Hicks BM, Patrick CJ, Carlson SR, Iacono WG, & McGue M (2002).Etiologic connections among substance dependence, antisocial behavior and personality: Modeling the externalizing spectrum. Journal of abnormal psychology, 111(3), 411–424. doi:10.1037//0021-843x.111.3.411 [PubMed: 12150417]

Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, … Psychiatry HA-I (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nature Genetics. doi:10.1038/s41588-018-0307-5

McGue M, Zhang Y, Miller MB, Basu S, Vrieze S, Hicks B, … Iacono WG (2013). A Genome-Wide Association Study of Behavioral Disinhibition. Behavior genetics, 43(5), 363–373. doi:10.1007/s10519-013-9606-x [PubMed: 23942779]

McLeod JD, Uemura R, & Rohrman S (2012). Adolescent mental health, behavior problems, and academic achievement. J Health Soc Behav, 53(4), 482–497.doi:10.1177/0022146512462888 [PubMed: 23197485]

National Drug Intelligence Center. (2011). National drug threat assessment. Retrieved from Washington, DC:

Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, … Benjamin DJ (2016). Genome-wide association study identifies 74 loci associated with educational attainment. Nature, 533, 539–542. doi:10.1038/nature17671 [PubMed: 27225129]

Otowa T, Hek K, Lee M, Byrne EM, Mirza SS, Nivard MG, … Hettema JM(2016). Meta-analysis of genome-wide association studies of anxiety disorders. Mol Psychiatry, 21, 1485. doi:10.1038/mp.2016.11 [PubMed: 26857599]

Pasman JA, Verweij KJH, Gerring Z, Stringer S, Sanchez-Roige S, Treur JL, … International Cannabis, C. (2018). GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. Nature Neuroscience, 21(9), 1161–1170. doi:10.1038/s41593-018-0206-1 [PubMed: 30150663]

Peterson RE, Edwards AC, Bacanu SA, Dick DM, Kendler KS, & Webb BT (2017). The utility of empirically assigning ancestry groups in cross-population genetic studies of addiction. Am J Addict, 26(5), 494–501. doi:10.1111/ajad.12586 [PubMed: 28714599]

Sacks JJ, Gonzales KR, Bouchery EE, Tomedi LE, & Brewer RD (2015). 2010 National and State Costs of Excessive Alcohol Consumption. American Journal of Preventitive Medicine, 49(5), e73–e79. doi:10.1016/j.amepre.2015.05.031

Savage JE, Salvatore JE, Aliev F, Edwards AC, Hickman M, Kendler KS, … Dick DM (2018). Polygenic Risk Score Prediction of Alcohol Dependence Symptoms Across Population-Based and Clinically Ascertained Samples. Alcoholism: Clinical and Experimental Research, 42(3), 520–530. doi:10.1111/acer.13589

Sul JH, Martin LS, & Eskin E (2018). Population structure in genetic studies: Confounding factors and mixed models. PLoS Genet, 14(12), e1007309. doi:10.1371/journal.pgen.1007309 [PubMed: 30589851]

The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. Nature, 526, 68. doi:10.1038/nature1539310.1038/nature15393https://www.nature.com/articles/nature15393#supplementary-informationhttps://www.nature.com/articles/nature15393#supplementary-information [PubMed: 26432245]

Tielbeek JJ, Johansson A, Polderman TC, & et al. (2017). Genome-wide association studies of a broad spectrum of antisocial behavior. JAMA Psychiatry, 74(12), 1242–1250. doi:10.1001/jamapsychiatry.2017.3069 [PubMed: 28979981]

Tuvblad C, Grann M, & Lichtenstein P (2006). Heritability for adolescent antisocial behavior differs with socioeconomic status: gene–environment interaction. Journal of Child Psychology and Psychiatry, 47(7), 734–743. doi:10.1111/j.1469-7610.2005.01552.x [PubMed: 16790008]

Walters RK, Polimanti R, Johnson EC, McClintick JN, Adams MJ, Adkins AE, andMe Research T (2018). Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. Nature Neuroscience, 21(12), 1656–1669. doi:10.1038/s41593-018-0275-1 [PubMed: 30482948]

Watanabe K, Taskesen E, van Bochoven A, & Posthuma D (2017). Functional mapping and annotation of genetic associations with FUMA. Nat Commun, 8(1), 1826. doi:10.1038/s41467-017-01261-5 [PubMed: 29184056]

Willer CJ, Li Y, & Abecasis GR (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics, 26(17), 2190–2191. doi:10.1093/bioinformatics/btq340 [PubMed: 20616382]

Williams DR, & Mohammed SA (2009). Discrimination and racial disparities in health: evidence and needed research. J Behav Med, 32(1), 20–47. doi:10.1007/s10865-008-9185-0 [PubMed: 19030981]

Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, … Carlson CS (2019). Genetic analyses of diverse populations improves discovery for complex traits. Nature, 570(7762), 514–518. doi:10.1038/s41586-019-1310-4 [PubMed: 31217584]

World Health Organization. (2014). Global status report on alcohol and health, 2014. Retrieved from

Young SE, Stallings MC, Corley RP, Krauter KS, & Hewitt JK (2002). Genetic and environmental influences on behavioral disinhibition. American Journal of Medical Genetics, 96(5), 684–695. doi:10.1002/1096-8628(20001009)96:5<684::AID-AJMG16>3.0.CO;2-G

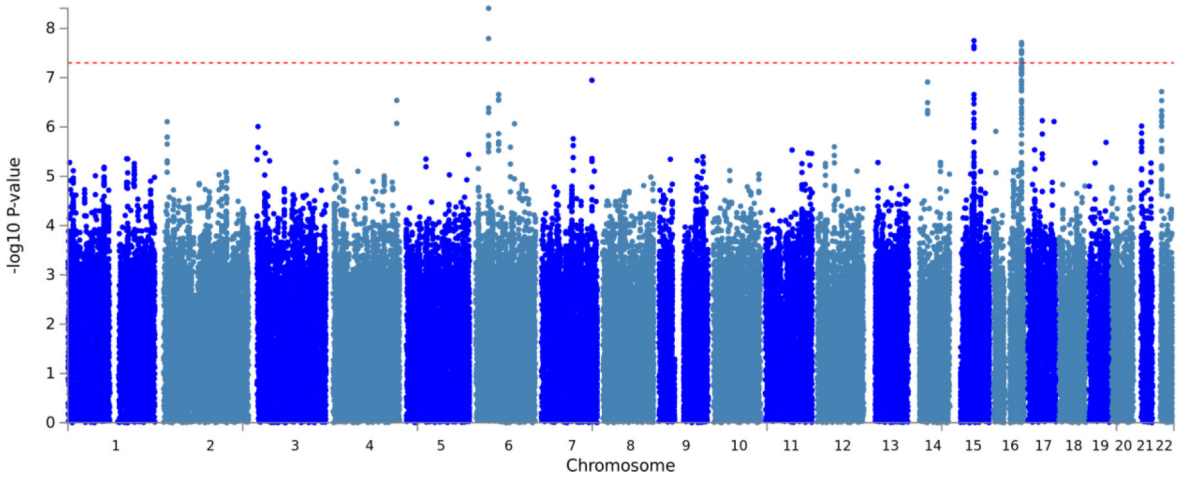

Figure 1.

Manhattan plots for inverse-variance weighted meta analysis of ancestry specific GWAS results. The y axis represents – log(P values) for association of variants with externalizing problem scores in each ancestry group (European Ancestry N = 7568; African Ancestry N = 3274). The dotted horizontal line represents the threshold for genome-wide significance ($p < 5 \times 10^{-8}$).

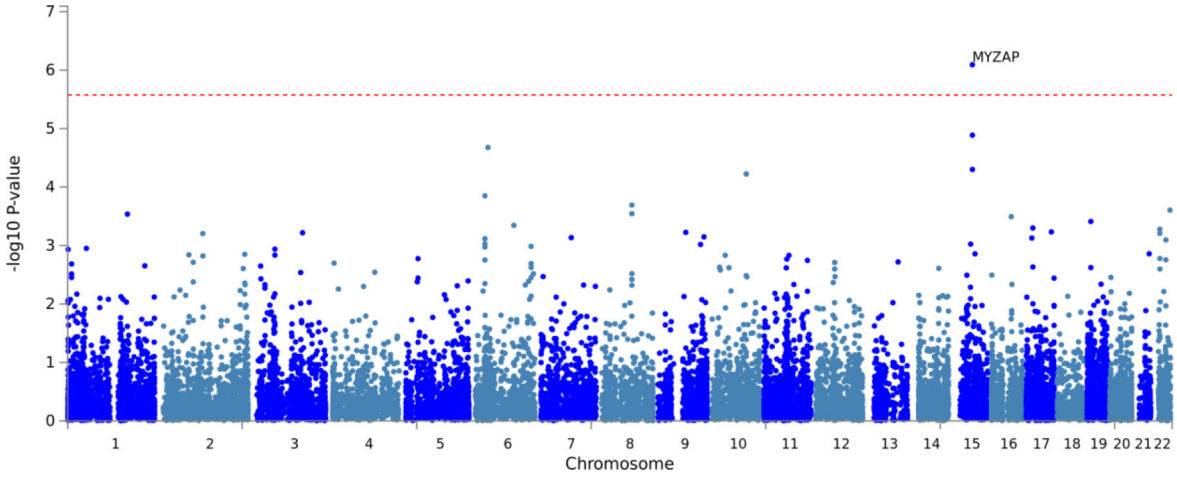

Figure 2.

Manhattan plot for meta-analytic results from gene-based analyses (inverse variance fixed effects meta-analysis). The y axis represents – log(P values) for association of variants with externalizing problem scores. The dotted horizontal line represents the threshold for genome-wide significance ($p < 2.78 \times 10^{-6}$).

Table I:

Descriptive Statistics of the COGA Discovery Sample

	EA		AA			
	N/Mean	%/SD	N/Mean	%/SD	Range	
Female	4000	53.5%	1,788	52.6%	-	
Age	35.1	15.4	31.4	12.7	12.0 – 91.0	*
Externalizing Score	0.0	1.0	0.1	1.0	−0.9 – 3.9	
DSM-IV Symptom Counts:						
Alcohol Dependence	2.1	2.4	1.9	2.7	0 – 7	*
Alcohol Abuse	1.0	1.3	0.7	1.2	0 – 4	*
Illicit Drug Dependence	1.2	2.2	1.8	2.5	0 – 7	*
Illicit Drug Abuse	0.5	0.9	0.4	0.8	0 – 4	*
Antisocial Personality Disorder/Conduct Disorder	1.7	1.9	2.2	2.1	0 – 7	*

* Differences across EA and AA p <.05 (Chi-square/ T-test)

EA = European ancestry, AA = African ancestry

Table II:

Top SNPs from within COGA meta analysis (N = 10,842)

RSID	CHR	Dir.	Meta B	Meta SE	Meta P	AA B	AA SE	AA P	AA MAF	EA B	EA SE	EA P	EA MAF
rs2376620	6	--	−0.050	0.008	3.91E-09	0.035	0.013	9.11E-03	0.242	0.059	0.011	2.43E-08	0.148
rs2433198	15	--	−0.035	0.006	1.78E-08	0.040	0.011	2.86E-04	0.467	0.033	0.007	8.75E-06	0.450
rs12928255	16	++	0.039	0.007	1.93E-08	−0.031	0.014	2.44E-02	0.218	−0.042	0.008	1.26E-07	0.381
rs10149887	14	++	0.074	0.014	1.23E-07	0.076	0.017	7.80E-06	0.114	0.068	0.023	2.90E-03	0.037
rs8135590	22	++	0.038	0.007	1.92E-07	0.029	0.015	5.63E-02	0.182	0.040	0.008	6.50E-07	0.358
rs72884541	6	-?	−0.155	0.030	2.19E-07	0.155	0.030	1.53E-07	0.037	-	-	-	-
rs1901817	16	++	0.036	0.007	3.21E-07	0.039	0.013	3.72E-03	0.240	0.035	0.008	1.63E-05	0.313

EA = European ancestry; AA = African ancestry; Meta = METAL inverse variance fixed effects results

CHR = Chromosome; Dir. = direction of association effect in AA (left) and EA (right); B = beta; SE = standard error; P = p value

Table III:

Results from the Cross Sample Meta Analysis (N = 15,112)

RSID	CHR	Dir.	Z-score	P	MAF
Within COGA Meta Analysis SNPs					
rs2376620	6	--++	−4.592	4.39E-06	0.646
rs2433198	15	--+-	−4.594	4.34E-06	0.542
rs12928255	16	++--	4.332	1.48e-05	0.574
Cross Sample Meta Analysis SNPs					
rs2932207	15	----	−5.449	5.08E-08	0.642
rs146787096	8	----	−5.374	7.69E-08	0.710
rs113552766	15	-?-?	−5.136	2.80E-07	0.301

Within COGA Meta Analysis SNPs present the results of the top SNPs identified from the initial within COGA meta analysis in the Cross Sample meta analysis. Cross Sample Meta Analysis SNPs represent the top SNPs identified from the combined sample meta analysis. Meta analysis was run using sample-size weighted meta analysis in METAL.

CHR = Chromosome; Dir. = direction of association effect in COGA EA, COGA AA, S4S EA, and S4S AA; P = p value; MAF = minor allele frequency

Table IV:

Polygenic Score Prediction in Validation Sample

PGS P-Value Threshold	European Ancestry (N = 2,761)				African Ancestry (N = 1,175)			
	Beta	**SE**	**P**	**R^2 (%)**	**Beta**	**SE**	**P**	**R^2 (%)**
p < .0001	−0.010	0.012	0.391	0.026%	0.012	0.017	0.479	0.042%
p < .001	0.007	0.012	0.560	0.012%	0.027	0.017	0.125	0.197%
p < .01	0.008	0.012	0.505	0.016%	0.011	0.018	0.549	0.030%
p < .05	0.018	0.012	0.134	0.081%	0.014	0.018	0.419	0.055%
p < .10	**0.029**	**0.012**	**0.018**	**0.203%**	0.010	0.018	0.581	0.025%
p < .20	0.015	0.012	0.214	0.055%	0.016	0.018	0.377	0.065%
p < .30	0.012	0.012	0.314	0.036%	0.015	0.018	0.401	0.059%
p < .40	0.014	0.012	0.254	0.047%	0.009	0.018	0.622	0.020%
p < .50	0.013	0.012	0.281	0.042%	0.010	0.018	0.583	0.025%

Bolded estimates indicates p < .05. All models include age, sex, and within-ancestry principal components as covariates.