Addiction Biology

SSA SOCIETY FOR THE STUDY OF ADDICTION

ORIGINAL ARTICLE

doi:10.1111/adb.12578

Shared additive genetic variation for alcohol dependence among subjects of African and European ancestry

Leslie A. Brick[1,2], Matthew C. Keller[3], Valerie S. Knopik[1,2], John E. McGeary[1,2,4] & Rohan H.C. Palmer[5,1,2]

Division of Behavioral Genetics, Department of Psychiatry, Rhode Island Hospital, Providence, RI USA[1], Department of Psychiatry and Human Behavior, Alpert Medical School, Brown University, Providence, RI USA[2], Institute for Behavior Genetics, Department of Psychology and Neuroscience, University of Colorado at Boulder, Boulder, CO USA[3], Providence Veterans Affairs Medical Center, Providence, RI USA[4] and Behavior Genetics of Addiction Laboratory, Department of Psychology, Emory University, Atlanta, GA USA[5]

ABSTRACT

Alcohol dependence (AD) affects individuals from all racial/ethnic groups, and previous research suggests that there is considerable variation in AD risk between and among various ancestrally defined groups in the United States. Although the reasons for these differences are likely due in part to contributions of complex sociocultural factors, limited research has attempted to examine whether similar genetic variation plays a role across ancestral groups. Using a pooled sample of individuals of African and European ancestry (AA/EA) obtained through data shared within the Database for Genotypes and Phenotypes, we estimated the extent to which additive genetic similarity for AD between AA and EAs using common single nucleotide polymorphisms overlapped across the two populations. AD was represented as a factor score by using Diagnostic and Statistical Manual dependence criteria, and genetic data were imputed by using the 1000 Genomes Reference Panel. Analyses revealed a significant single nucleotide polymorphism-based heritability of 17 percent (SE = 5) in EAs and 24 percent (SE = 15) in AAs. Further, a significant genetic correlation of 0.77 (SE = 0.46) suggests that the allelic architecture influencing the AD factor for EAs and AAs is largely similar across the two populations. Analyses indicated that investigating the genetic underpinnings of alcohol dependence in different ethnic groups may serve to highlight core etiological factors common to both groups and unique etiological factors specific to each ethnic group.

Keywords African ancestry, alcohol dependence, heritability.

Correspondence to: Rohan H. C. Palmer, Behavior Genetics of Addiction Laboratory, Department of Psychology at Emory University, 36 Eagle Row, Room 379, Atlanta, GA 30322, USA. E-mail: rohan.palmer@emory.edu

INTRODUCTION

Alcohol dependence (AD) is a global problem that affects individuals from all racial/ethnic groups and all levels of socioeconomic standing. Previous studies suggest that there is considerable variation in patterns of drinking and alcohol use disorders across different US racial/ethnic groups (Caetano, Clark & Tam 1998; Chartier & Caetano 2010; Vaeth, Wang-Schweig & Caetano 2017). For example, analysis of college-aged students (i.e. ages 18–24) from the National Epidemiological Survey on Alcohol and Related Conditions has shown that higher rates of alcohol consumption are observed among individuals of European and Native American ancestry compared with individuals of African or Asian ancestry (Chen, Dufour & Yi 2005). Similarly, among adults, National Epidemiological Survey on Alcohol and Related Conditions data have shown both lifetime and past 12-month AD to be significantly lower for individuals of African, Asian or Hispanic ancestry relative to European ancestry (EA; Hasin *et al.* 2007). While there is evidence to suggest that the risk for AD may be greater, in part, for some individuals as a function of their economic standing and sociocultural factors (Caetano & Clark 1998a, 1998b, 1999; Swendsen *et al.* 2009), there have been a limited number of studies that have attempted to examine whether genetic variation might also play a role in observed differences among African Americans. A review of the published literature using a combination of search terms (Supporting Information)

in PubMed (in October 2016) revealed 14 genome-wide association studies (GWASs) that have examined genetic variants related to alcohol consumption and/or dependence with some combination of analyses in ancestrally mixed samples or ancestry-specific (i.e. identified by using samples; Bierut *et al.* 2010; Panhuysen *et al.* 2010; Johnson *et al.* 2011; Zuo *et al.* 2012; Zuo *et al.* 2012; Zuo *et al.* 2013; Zuo *et al.* 2013; Gelernter *et al.* 2014; Ulloa *et al.* 2014; Yang *et al.* 2014; Zuo *et al.* 2014; Xu *et al.* 2015; Zuo *et al.* 2015). Of these, none have empirically compared/contrasted the additive genetic effects in subjects of African ancestry (AA) and EA. While the definitions of ancestry are nuanced and complex, here we have focused our review on studies that considered genetic ancestry as best defined by the International HapMap (2003) or 1000 Genomes Project (1KG) reference panels (e.g. various reference panels comprised slightly different populations; thus, individuals from sample data may be characterized differently depending on the reference panel being used, which can lead to different 'definitions' of ancestry and potentially different results; Gibbs *et al.* 2003; Auton *et al.* 2015).

Indeed a demonstrable gap exists in the literature on alcohol genetics such that AAs are substantially underrepresented in twin studies, candidate gene studies and GWAS (Desalu *et al.* 2017; Dick *et al.* 2017). This insufficient attention in the alcohol literature on distinctions between AAs and EAs in etiology likely has profound implications on our ability to determine specific etiological influences as studies of distinct ethnic backgrounds afford a number of advantages such as differences in allele frequency and linkage disequilibrium (LD; Kristiansson, Naukkarinen & Peltonen 2008; Dick *et al.* 2017) that might facilitate the identification of causal variants as well as environmental differences (Minster *et al.* 2016; Chartier *et al.* 2017; Dick *et al.* 2017) that may provide key insights into GXE relationships. These scientific advantages underscore the importance of genetic research in under-represented populations in order to move the field forward for all ethnicities.

In general, GWASs that incorporate multiple ancestral populations are limited by small sample sizes, making the detection of single nucleotide polymorphism (SNP) effects in the smallest groups more difficult. Statistical power to detect a true significant effect varies as a function of sample size, disease prevalence, LD between the measured marker(s) and the causal locus and the minor allele frequencies (MAFs) of the marker (Purcell, Cherny & Sham 2003; Hong & Park 2012). Given differences in LD, MAF and sample sizes across ancestral populations (Manolio, Brooks & Collins 2008), candidate gene studies and GWAS use strict methods or covariates to limit confounding of association signals. At the same time, differences in genomic characteristics among ancestral groups highlight the strengths of conducting candidate gene and genome-wide types of association studies across multiple ancestral groups, to the extent that sufficient statistical power is achieved to conduct analyses within group and/or to examine loci across groups. This has largely been seen in studies of candidate biological systems, some of which have shown that increased power can be gained by studying other ethnic groups where certain alleles are more commonly observed in comparison with subjects of EA. For example, some of the most reliable effects on AD in GWAS center around variation in and around the chromosome 4 ADH cluster. Genes that play a role in the alcohol metabolizing system and associated genes on chromosome 4 (ADH1B, ADH1C and ADH4) and chromosome 12 (ALDH2) have been observed in individuals of Korean, Chinese, African and EA (Frank *et al.* 2012; Park *et al.* 2013; Quillen *et al.* 2014). Among the studies including AAs, Gelernter *et al.* (2014) conducted a GWAS by using a pooled sample of 16 087 individuals of EA and AA and was the only study to date to explore convergence of genome-wide significant findings across the subpopulations. Further, this study sought to identify novel risk loci for substance dependence phenotypes and is the first published GWAS of AA with significant results. The report by Gelernter *et al.* (2014) was the first to provide some indication of shared genetic effects across populations around a nominal GWAS finding but did not estimate the population-specific additive genetic effects for AD across EAs and AAs. The current study expands upon the Gelernter *et al.* (2014) paper by using pooled samples of EAs and AAs (respectively) to estimate the extent of shared additive genetic influence for AD between AA and EAs. The data for this project are the result of sharing agreements imposed by the National Institutes of Health and principal investigators that support collaborative work by submitting their data to the Database for Genotypes and Phenotypes (dbGAP). This is, to our knowledge, the first study of its kind to estimate these genome-wide effects by using molecular data.

MATERIALS AND METHODS

Sample

All study data was accessed as part of the National Human Genome Research Institute's Gene Environment Association Study Initiative (dbGaP). For all analyses, data from four dbGaP datasets were pooled, including The Study of Addiction: Genetics and Environment (SAGE; study accession phs000092.v1.p1), the Alcohol Dependence GWAS in European and African Americans (Yale Study; study accession phs000425.v1.p1), the Australian twin-family study of alcohol use disorder (OZ-ALC; study accession phs000181.v1.p1), and the GWAS of

Heroin Dependence (Heroin GWAS study; study accession phs000277.v1.p1). Table 1 describes the set of samples.

Assessments

Each study collected DSM-IV symptoms (coded as present or absent) for AD by using the Semi-Structured Assessment for the Genetics of Alcoholism (SAGE study), the adapted Semi-Structured Assessment for the Genetics of Alcoholism OZ (OZ-ALC study), or the Semi Structured Assessment for Drug Dependence and Alcoholism (Yale Study, Heroin GWAS) (Bucholz *et al.* 1994; Hesselbrock *et al.* 1999; Pierucci-Lagha *et al.* 2005). All responses were limited to individuals who had previously been exposed to alcohol (and possibly other drugs).

Genotyping, quality control and genetic imputation

The GWAS data were obtained through the National Center for Biotechnology Information's dbGAP, where more detailed protocols are available. For each sample set, quality control (QC) and imputation of autosomal SNPs were conducted separately by study and are explained in the succeeding texts. Genotyping in SAGE was conducted by using the Illumina Human 1 M BeadChip. Genotyping in the Yale Study was conducted on the Illumina HumanOmni1-Quad v1.0 microarray.

Table 1 Descriptions of samples that were aggregated to identify genetic factors related to alcohol dependence.

Study	*N*	*Description*
Study of Addiction: Genetics and Environment (SAGE)	4316	A multiethnic sample of unrelated individuals from three large, complementary data sets designed to study drug addiction: the Collaborative Study on the Genetics of Alcoholism (COGA), the Family Study of Cocaine Dependence (FSCD) and the Collaborative Genetic Study of Nicotine Dependence (COGEND)
Alcohol Dependence GWAS in European and African Americans (Yale Study)	2909	A case–control study focusing on AAs and EAs who meet DSM-IV criteria for AD. The sample was collected over the course of ongoing projects that focused on oversampling of alcohol-dependent AAs and also included measures on cocaine and opioid dependence. The sample was originally collected to identify sibling pairs suitable for linkage analysis.
Australian twin-family study of alcohol use disorder (OZ-ALC)	6701	A family study deriving from two general population volunteer cohorts of twins in Australia totaling over 11 000 families. Two cohorts of twins born between 1940 and 1961 (cohort 1) or 1964–1971 (cohort 2) were assessed by using a shared protocol to discover genes related to alcohol use. Data from these studies were compiled into a case–control family-based GWAS that focused on alcohol use and dependence.
Genome-Wide Association Study of Heroin Dependence (Heroin GWAS)	6410	A collaboration of investigators from the United States and Australia to identify genes associated with heroin dependence by using a case–control study. Data on participants from the Heroin Study *who were assessed for dependence on alcohol* consisted of the following from ongoing genetic studies of substance dependence conducted by investigators at Yale and collaborating institutions: 1. Cases (i.e. individuals who reported participation in pharmacotherapy maintenance treatment for opioid dependence at some point in their life) and assessed controls (i.e. individuals not dependent on heroin) from the Comorbidity and Trauma Study; 2. Cases (i.e. individuals dependent on heroin) from the Heroin Dependence in Western Australia; 3. Controls (i.e. individuals who did not meet criteria for illicit drug dependence but may have been dependent on alcohol or nicotine) from the OZ-ALC Study; 4. Assessed controls (i.e. individuals who did not meet criteria for substance use dependence) and cases (i.e. individuals dependent on opioids with heroin listed as the most used opioid).

Genotyping for the OZ-ALC study was conducted on Illumina HumanCNV370-Duov1 BeadChip. Finally, the participants from the Heroin GWAS were genotyped on three separate platforms: Illumina Human610 Quad v1, Illumina Human660W Quad v1 and HumanCNV370 Quad v3.0.

Genomic data across all study samples were imputed [within study sample (by ethnic group)] up to Phase III of the 1000 Genomes Project (1KG) in order to maximize similar genetic coverage across samples. Data management was conducted by using SNP & VARIATION SUITE v8. x (Golden Helix, Inc, Bozeman, MT, www.goldenhelix.com; SNP & VARIATION SUITE version 8.4.4), PLINK version 1.9 (Purcell *et al.* 2007) and R version 3.1.1. Genetic imputation was conducted by using MINIMAC (version 3) via the Michigan Imputation Server (https://imputationserver.sph.umich.edu/index.html#!pages/home). Ancestry determination of sample data and imputation of genotypes utilized data from Phase III of the 1000 Genomes Project (1KGP) reference sample (Auton *et al.* 2015).

A series of steps across three phases was conducted to: (1) prepare the data for imputation, (2) impute the data, and (3) prepare the data for analyses. A flowchart of this procedure is presented in Fig. 1.

Phase 1: Imputation preparation

Step 1: Identify major ancestral populations within sample data by using the 1KGP reference sample.

Subject ancestry was determined by using the Phase 3 reference panel from the 1KGP, which is composed of 2504 individuals across 26 populations and contains genotyping data for 84.4 million markers. The major ancestral groups captured in this data are East Asian, South Asian, African, European, and American. For ancestry determination, we restricted the number of markers in the 1KGP to include only the union of markers present in each of our sample data sets (2 240 710). We then removed any markers with a MAF less than 5 percent and a call rate (CR) less than 99 percent. Finally, we used a subset of the resulting set of 1KGP data based on LD ($r^2 < 0.5$), resulting in a final set of 423 738 markers to be used for ancestry determination.

Quality control was conducted in each of the study samples separately prior to being combined with the reference panel. Each study sample set was subset to include autosomal SNPs with MAF greater than 10 percent and a CR of 95 percent. Using allele information compiled from the marker map of the reference panel data, we compared the allele frequencies (across all populations) and strand orientation of our data to the reference panel. Markers that had MAF differences of greater than 20 percent when compared with the reference panel were removed. Markers whose stand orientation could not be resolved (e.g. flipped) were also removed.

After QC of the sample data and preparation of the reference panel data were complete, the study samples were combined with the reference data (separately) to determine ancestry within each study. Principal component analysis was conducted within each study to examine population stratification. Plots of genetic components were examined visually and statistically to determine ancestral groups. First, scatterplots comparing the first, second and third components, which largely distinguish

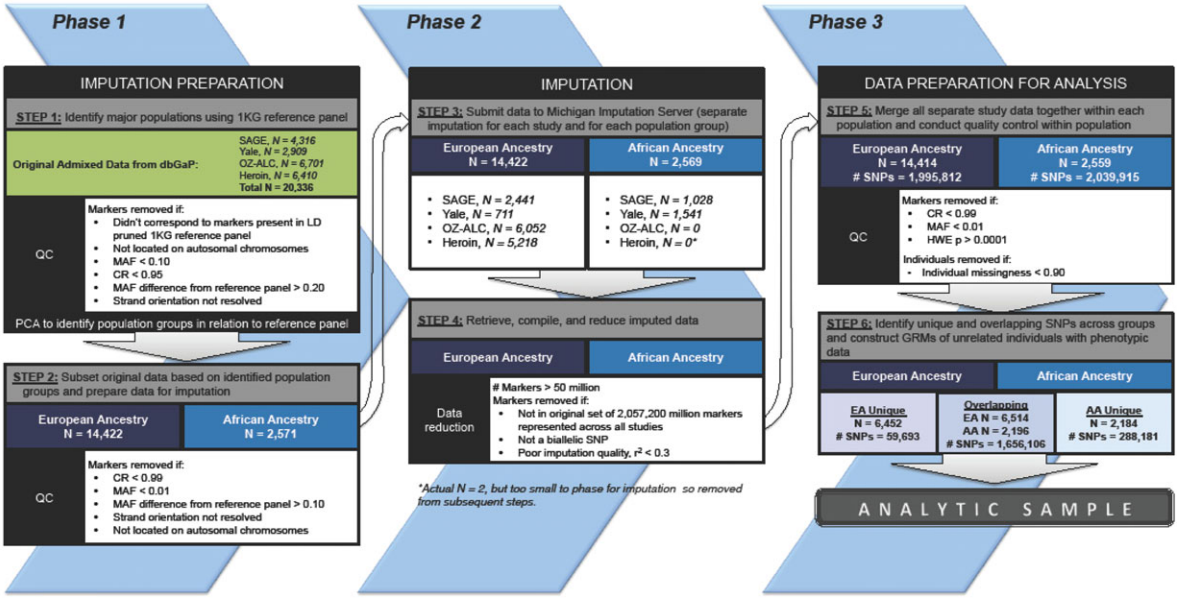


Figure 1 Flow chart of data preparation and imputation. (a) First component (PC1; *y*-axis) plotted against second component (PC2; *x*-axis). b. Third component (PC3; *y*-axis) plotted against second component (PC2; *x*-axis)

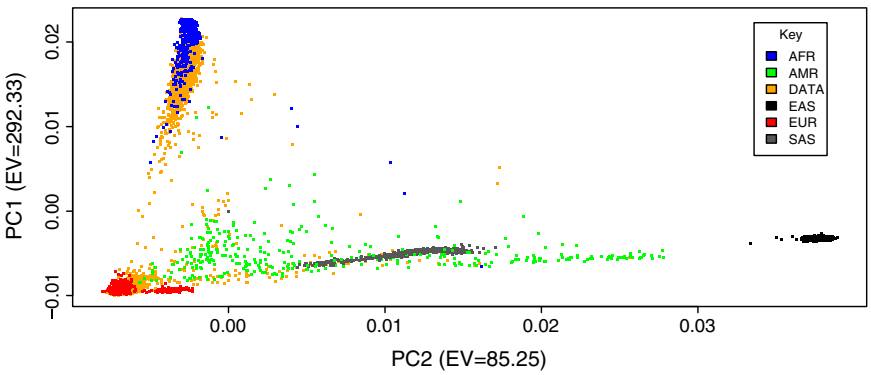
among African, East Asian, South Asian and European groups, were plotted to determine ancestry of the sample data compared with the reference data. For example, Fig. 2 presents scatterplots of principal components of the SAGE data with the 1KG reference data. In Fig. 2a, the first principal component is plotted against the second principal component. The first component in each of the data sets separated African and European ancestral groups, which represented the larges two subgroups in each of samples examined in this study. Subsequently, we calculated the mean and standard deviations of the first principal component in the reference panel and retained individuals in the sample data whose eigenvector value fell within two standard deviations (i.e. 98 percent of the 1KG ansestral distributions) of the African and European reference panel component means. As such, the current study clusters individuals into two groups, AA or EA, based on their proximity to established ancestral groups within the 1KG reference panel data. To reduce further population stratification, we conducted multidimensional outlier detection within the identified AA/EA groups in the sample data by using a multiplier value of 1.5. This procedure computed a distance score based on the median centroid vector calcutated by using the first three principal components. Any individuals determined to be outliers from the AA and EA samples were removed from the sample data. The resulting set of 2571 AA and 14 422 EA individuals in the sample data were selected for imputation out of the admixed sample of 20 336 available via dbGaP.

Step 2: Subset original data based on identifed population groups and prepare data for imputation.

After the AA and EA individuals were identified in Step 1, the original sample data were subset into the two respective groups identified by principal component analysis to be imputed separately. QC was conducted in each group, and markers with CR $<$ 95 percent or MAF $<$ 1 percent were removed. Using allele information compiled from the marker map of the reference panel data, we compared the allele frequencies (specific to expected allele frequencies based on the 1KG African or European populations) and strand orientation of our data to the reference panel. Markers that had MAF differences of greater than 10 percent when compared with the reference panel population were removed. Markers whose stand orientation could not be resolved (e.g. flipped) were also removed. Individuals who had greater than 95 percent missingness were also removed. In the Heroin GWAS study, too few individuals of AA (n = 2) were identified and data were not able to be imputed for this study. The final set of data for AA (n = 2569) and EA (n = 14 422) individuals within each study was separated into autosomal chromosome files for submission to the Michigan Imputation Server.

Phase 2: Imputation of genotypes in identified EAs and AAs (separately)

a. First component (PC1; y-axis) plotted against second component (PC2; x-axis).



b. Third component (PC3; y-axis) plotted against second component (PC2; x-axis).

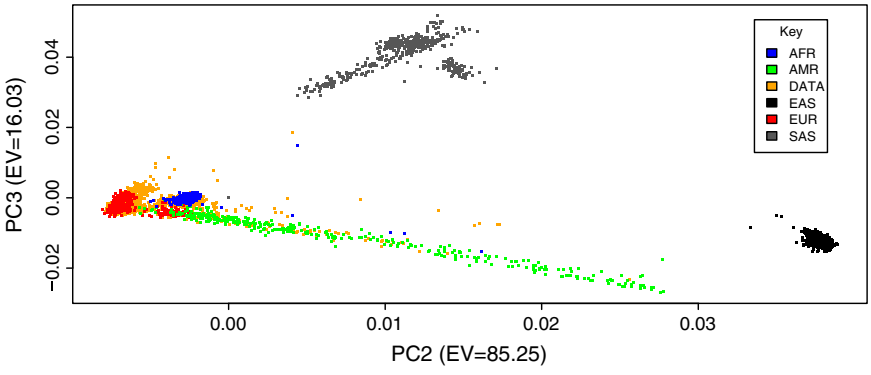


Figure 2 Scatterplots of genetic principal components of the SAGE sample data plotted with 1000 Genomes Reference Panel ancestral groups data. Note: EV, eigenvalue; 1000 Genomes Referance Panel ancestral groups included: AFR, African; AMR, Americas; DATA, SAGE sample data; EAS, East Asian; EUR, European; SAS, South Asian.

Step 3: Submit data for imputation.

Ancestral groups within each study were imputed separately on the Michigan Imputation Server (https://imputationserver.sph.umich.edu/index.html) by using Minimac3 with the 1KG Phase 3 reference panel and SHAPEIT phasing.

Step 4: Retrieve, compile and reduce imputed data.

After imputation was completed, each file totalled over 50 million markers. Files were subset based on the union of the aforementioned 2 057 200 million markers present across all the studies, and markers that did not represent biallelic SNPs were removed. This approach ensured that different markers present across various genotyping platforms were represented in the final set of markers, thereby limiting bias based on representation from a specific platform. Finally, markers with an imputation quality score $r^2 < 0.3$ were removed. A small number of individuals from the SAGE sample (n = 10 AAs, n = 7 EAs) had missingness patterns that precluded imputation and were removed, resulting in a sample size of 2559 AAs and 14 414 EAs with imputed data.

Phase 3: Data preparation for analysis

Step 5: Merge all study data together within each population and conduct QC.

Following imputation, all imputed study data for EAs across each study were merged. Likewise, data for AAs across each imputed study were merged. QC was conducted within each ancestry group separately to select individuals with missingness <10 percent and markers with CR > 99 percent, MAF > 1 percent, and that passed HWE test ($P < 0.0001$) (see Table S1 for summary of the number of markers across EAs/AAs that drop out at each step). After QC, data for 2559 AAs and 13 461 EAs were available for analyses.

Step 6: Identify unique and overlapping SNPs across ancestral groups and construct genetic-relatedness matrices of unrelated individuals.

Following QC, we identified a common set of SNPs across both populations (1 656 106 in EAs and AAs) and a set of SNPs that survived QCs in one group but not the other (N = 288 181 unique to AAs; N = 59 693 unique to EAs). Each set of SNPs was then used to contruct genetic relationship matrices (GRMs). The GRMs were computed by using the Genomewide Complex Trait Analysis software tool (version 1.25.3) and to maximally select one of any pair of individuals who were more related than second cousins (Yang *et al.* 2011). Subsetting the data for unrelated individuals was done to control for cryptic relatedness, which could artificially inflate SNP heritability estimates (see succeeding texts). The ancestry specific GRMs used in univariate genetic analyses were composed of 2257 unrelated AA individuals and 8722 unrelated EA individuals. Of the unrelated indivuduals in each population, separate GRMs were computed for overlapping and sample-specific SNPs. Thus, joint analyses for each ancestry group by using the GRM constructed from overlapping markers as well as the GRM constructed from sample-specific markers provided the total amount of variation in AD attributable to genetic variants. In addition, bivariate genetic models described in the succeeding texts used a combined GRM of 11 314 individuals who had overlapping SNPs to provide an estimate of genetic correlation between populations.

Derivation of phenotypes and sample characteristics

Data for the seven DSM-IV AD symptoms were pooled to determine and confirm the factor structure of the AD latent variable in EAs and AAs. A latent variable approach was used for several reasons: (1) the AD factor is a continuous phenotype, thereby boosting power to detect effects compared with a dichotomous trait; (2) the factor score represents the shared variance across the seven items and thus does not separate individuals based on the DSM-IV cutoff, thereby capitalizing on the entire spectrum of alcohol use issues; and (3) it is consistent with previous work that suggests that the underlying dimension of risk for AD comprises a single factor (Harford & Muthén 2001; Ray *et al.* 2008; Palmer *et al.* 2015). Data for participants in each study were subset to include only those participants who were unrelated and were genetically determined to be EA or AA; consequently, data for 6514 genetically determined EAs and 2196 genetically determined AAs who had phenotypic data were used for factor analysis.

The final sample of EA individuals were 53.22 percent male and ranged in age from 16 to 82 (mean age = 40.16, standard deviation = 10.42). Of these individuals, 34.35 percent came from the SAGE study, 8.26 percent came from the Yale study, 15.13 percent came from the OZ-ALC study, and 42.26 percent came from the Heroin GWAS study The final sample of AA individuals were 51.09 percent male and ranged in age from 16 to 79 (mean age = 40.48, standard deviation = 8.84). Of these AA individuals, 37.93 percent came from the SAGE study and 62.07 percent came from the Yale study. The OZ-ALC and Heroin GWAS studies did not contain enough individuals of AA to impute genetic data and thus did not contribute to the final sample of AAs used in the current study.

The factor structure of AD symptoms within each ancestral group was determined by randomly splitting each subpopulation in half to create exploratory and confirmatory subsets. Exploratory and confirmatory factor analyses (EFA/CFA) were conducted in MPLUS [version 7] (Muthén & Muthén 1998–2015) by using weighted least-squares mean variance estimation. Missing data were handled in MPLUS with full information maximum

likelihood estimation. The exploratory subsets consisted of 1098 AA participants and 3255 EA participants. The confirmatory subsets consisted of 1098 AA participants and 3260 EA participants. Scree plots, consistency with previous empirical research and examination of fit indices (e.g. root mean square error of approximation, comparable fit index and Tucker–Lewis index) were used to determine the factor structure for each 1KGP-defined ancestral group (Hu & Bentler 1999; Yu 2002). EFA and CFA models indicated that a single latent factor represented AD symptoms (see Fig. S1 for scree plot and Table S2 for model fit for EFA/CFA across ancestral groups). Measurement invariance by using a sequential approach to fit a single latent factor across the two ancestral groups supported configural invariance (Meredith & Teresi 2006; Dimitrov 2010). As such, the same factor structure, but non-equivalent error variances and item thresholds, was supported for each group (see Table S3 for results of invariance testing). Consequently, the latent factor will be estimated separately for AA and EA individuals to acknowledge these observed differences.

Based on consensus from EFA/CFA, separate factor scores (mean = 0, standard deviation = 1) from a one-factor solution were extracted for EAs and AAs to be used in genetic analyses. Specifically, these analyses yielded factor scores derived *within* each 1KGP-defined ancestral group that represent latent indicators of AD based on the seven DSM-IV symptoms specific to that ancestral group. The factor model, including unstandardized factor loadings, is presented in Fig. 3 along with fit indices for each CFA model in Table 2. To account for the effects of covariates on phenotypic variance, each factor score was residualized to account for variation due to age, sex and study of origin, and the residualized scores were used for all subsequent genetic analyses (Wurm & Fisicaro 2014).

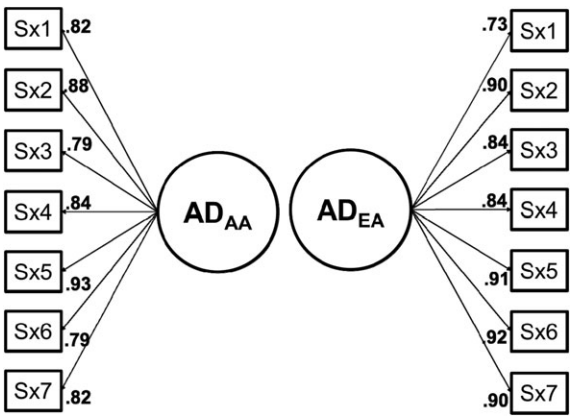


Figure 3 Confirmatory factor analysis of DSM-IV Alcohol Dependence among individuals of African or European ancestry. Note: AD, alcohol dependence; AA, African ancestry; EA, European ancestry; Sx1, tolerance; Sx2, withdrawal; Sx3, use longer than intended; Sx4, failure to quit; Sx5, great time spent using/recovering; Sx6, activities foregone; Sx7, continued use despite problems

Table 2 Model fit for alcohol dependence factors among African ancestry and European ancestry groups.

Model fit information	AD_{AA}	AD_{EA}
χ^2 (14)	259.800	302.649
P-value	<0.001	<0.001
CFI	0.986	0.995
RMSEA [90% CI]	0.089 [0.080, 0.099]	0.056 [0.051, 0.062]

Note: AD, alcohol dependence; AA, African ancestry; EA, European ancestry; CFI, confirmatory fit index; RMSEA, root mean square error of approximation.

Estimation of variance/covariance explained by the SNPs

Genomic-relatedness restricted maximum likelihood estimation (GREML) was used to decompose phenotypic variance in the EA and AA AD factors into additive effects of genotyped and imputed SNPs (Yang *et al.* 2013). In this approach, genetic similarity captured in each GRM is modeled as a random effect to account for variance in the residualized AD factor score for each ancestry group. Two separate variance components were included in each ancestry-specific linear model: one component composed of genetic variance due to markers that overlapped between ancestry groups and one that represented genetic variance due to sample-specific markers (i.e. markers that passed QC for one group but not the other). Total SNP heritability for each model represents the total variation across each component for EAs and AAs separately. In bivariate GREML models, the covariance between two groups can be described by a standard bivariate linear mixed model in which covariance is reflected by the covariance between the genetic and environmental/residual factors influencing each trait. Only genetic variance attributable to overlapping markers was used in the bivariate model. As such, with the current data, the additive genetic correlation ($r_{G\text{-}SNP}$) reflects shared genetic variance tagged by the genotyped SNPs. While the bivariate model does not provide a direct test of univariate h^2_{SNP} differences between the two populations, the model estimates $r_{G\text{-}SNP}$, which is interpreted as the extent to which the genetic variants influencing AD in EAs and AAs are correlated (ranging in value from −1 to 1 (Lee *et al.* 2012; de Candia *et al.* 2013). Consequently, analyses were designed to determine the SNP heritability (h^2_{SNP}) within each ancestral group as well as the genetic correlation across EAs and AAs (using the set of overlapping SNPs that survive QC across EAs' and AAs). Assuming the lack of direct evolutionary pressures related to alcohol use, we hypothesized

that SNP heritability estimates would be similar across EAs and AAs and that the genetic correlation would be high (e.g. >0.60). Given our observation of non-overlapping SNP sets following sample QC, we also explored the extent to which these SNPs might be an additional source of genetic variation for EAs and AAs.

RESULTS

Prevalence of alcohol dependence items across ancestral groups

Demographics of sample data by ancestral group are presented in Table 3. Prevalence rates and correlations between AD symptoms are presented in Table 4. For both AAs and EAs, 'using longer than intended' (58 percent endorsed in AAs and 57 percent endorsed in EAs) was the highest endorsed item and giving up activities (33 percent endorsed in AAs and 25 percent endorsed in EAs) was the lowest endorsed item. Phenotypic tetrachoric correlations among all items were generally high (all greater than 0.58).

Table 3 Demographics of sample data by ancestral group.

	EA		*AA*	
	N	%	*N*	%
Sex				
Male	3467	53.22%	1122	51.09%
Female	3048	46.78%	1074	48.91%
Study				
SAGE	2238	34.35%	833	37.93%
YALE	538	8.26%	1363	62.07%
OZ-ALC	986	15.13%	0	0.00%
Heroin	2753	42.26%	0	0.00%
	M	SD	*M*	SD
Age	40.16	10.42	40.48	8.84

Note: EA, European ancestry; AA, African ancestry; *N*, sample size; M, mean; SD, standard deviation.

Phenotypic variance attributable to AD among EAs and AAs

Total SNP-based heritability estimates of the AD factor were similar across EAs and AAs. See Table 5 for a summary of univariate results within EA and AA groups (also see Table S3 for a summary of results using AD diagnosis instead of the factor score). Partitioning of the total genetic variance for EAs by using multiple GRMs in a single linear model revealed a significant SNP heritability estimate of 0.17 (SE = 0.05, $P < 0.001$) for variation in AD, which was attributable to SNPs that *overlapped* with AAs, and a significant estimate of 0.10 (SE = 0.04, $P < 0.001$) for variation in AD, which was attributable to SNPs that were *sample-specific* to EAs (i.e. markers that passed QC for EAs but not for AAs).

Partitioning of the total genetic variance for AAs revealed a significant SNP heritability estimate of 0.24 (SE = 0.15, $P = 0.028$) for variation in AD, which was attributable to SNPs that *overlapped* with EAs, and a non-significant estimate of 0.07 (SE = 0.14, $P = 0.313$) for variation in AD, which was attributable to SNPs that were *sample-specific* to AAs (i.e. markers that passed QC for AAs but not for EAs).

Table 4 Item endorsement and tetrachoric correlations of DSM-IV Alcohol Dependence for each ancestral group.

Symptom	%	*N*	*Correlation*					
African ancestry			1	2	3	4	5	6
1. Tolerance	47%	1033						
2. Withdrawal	53%	1170	0.66					
3. Longer than intended	58%	1279	0.70	0.63				
4. Attempt to quit	56%	1238	0.76	0.68	0.69			
5. Time spent	57%	1251	0.69	0.89	0.68	0.74		
6. Giving up activities	33%	724	0.66	0.59	0.66	0.68	0.67	
7. Continued use	46%	998	0.69	0.66	0.66	0.71	0.70	0.72
European ancestry								
1. Tolerance	49%	2880						
2. Withdrawal	26%	1660	0.58					
3. Longer than intended	57%	3278	0.67	0.71				
4. Attempt to quit	40%	2608	0.65	0.75	0.74			
5. Time spent	30%	1656	0.60	0.86	0.71	0.72		
6. Giving up activities	25%	1421	0.62	0.82	0.73	0.75	0.86	
7. Continued use	41%	2292	0.65	0.78	0.78	0.75	0.76	0.85

Table 5 Univariate SNP heritability [h^2_{SNP} (SE)] of AD factor for EAs and AAs.

	h^2_{SNP} *gene set A*	h^2_{SNP} *gene set B*	*Total SNP heritability*
European ancestry	0.17 (0.05)***	0.10 (0.04)**	0.27 (0.05)[a]
African ancestry	0.24 (0.15)*	0.07 (0.14)	0.30 (0.15)[a]

Note: Table presenting the univariate SNP heritability of AD factors for EAs and AAs by using subsets of SNPs [gene set A comprises SNPs that survive within-ancestral-group quality control procedures (QC) across both populations; gene set B includes SNPs that differentially survive QC across ancestral groups; total SNP heritability reflects the genome-wide effects of gene sets A and B within each ancestral group]. *$P < 0.05$. **$P < 0.01$. ***$P < 0.001$. [a]Significance test not available for total heritability in model, as the likelihood ratio test is conducted only on the specific variance components within the model.

Further examination of the additive genetic effects of AD in EAs and AAs by chromosome for those SNPs that survive QC across EAs and AAs (see Fig. S2) indicated that longer chromosomes did not account for significantly more phenotypic variation (for EAs: $R^2 = 0.15$, $\beta = 6.96 \times 10^{-11}$, $t(20) = 1.87$, $P = 0.076$; for AAs: $R^2 = 0.01$, $\beta = 1.98 \times 10^{-11}$, $t(20) = 0.39$, $P = 0.704$).

Genetic correlation attributable to overlapping markers across EAs and AAs

Bivariate analyses revealed a significant genetic correlation between EAs and AAs ($r_{G\text{-}SNP} = 0.77$, SE = 0.46, $P = 0.030$) for SNPs that survived QC across both ancestral groups. Overall, this suggests that there is moderate evidence for convergence across EA and AAs for a subset of genome-wide SNPs that contribute to the additive genetic variance of AD.

DISCUSSION

The results from this study are, to our knowledge, the first to directly compare the SNP-based genetic liability for AD across individuals of African and EA. The inclusion of racial and ethnic minority groups in genetic research (when used appropriately and ethically) is essential to progress in understanding the role that genetic and sociocultural factors play in racial and ethnic health disparities. Large-scale GWASs have primarily concentrated on European populations, with very little representation of individuals of AA (Need & Goldstein 2009). Yet despite the tendency for genetically informed studies to focus on populations of EA, psychological and epidemiological research has found that compared with their European American counterparts, African Americans initiate drinking at an older age and report overall lower rates and levels of use and higher levels of abstinence (Quality, 2016; Zapolski *et al.* 2014). Further, African American drinkers report significantly higher rates of social consequences and AD symptoms compared with European Americans (Mulia *et al.* 2009).

Evidence from the current study supported a moderately shared genetic liability for the AD factor score across EA and AA groups, yet empirical research has identified social, cultural, health, environmental, historical, economic and numerous other demographic factors that contribute to observed disparities in AD risk and consequences among African Americans (Zapolski *et al.* 2014). It is likely that the intersectionality of multiple other risk factors (Mereish & Bradford 2014), such as sexual orientation and gender, as well as specific individual and environmental influences, may impart an impact on risk for substance use above and beyond the observed genetic effects (McGue, Elkins & Iacono 2000). Although it was beyond the score of the present study to explore sociocultural factors, the current results do provide a framework for beginning to explore these potential sources of variation in the context of genetic variation (i.e. gene–environment interaction).

Single nucleotide polymorphism-based heritability estimates found in this study are consistent with previous work using the GREML method to examine various parameterizations of the AD phenotype. Our findings indicate that 27 percent of the phenotypic variation of the AD factor in EAs and 30 percent of the phenotypic variation in AA were attributable to additive genetic effects when examining a set of the same genetic markers across the two populations. These estimates are similar to the 30 percent SNP-based heritability estimated by Palmer *et al.* (2015) by using an AD factor in a sample in EAs from only the SAGE subsample, within the margin of error. Recent studies that utilized AD diagnosis, rather than the factor score, have estimated a heritability of 21 percent in a Caucasian sample (Vrieze *et al.* 2013) and 22 percent in an African American sample (Yang *et al.* 2014). Kos *et al.* (2013) recently estimated that 38 percent and 35 percent of the variation in AD diagnosis risk are attributable to common SNPs in EAs and AAs, respectively; however, their study did not limit the SNPs used in the estimations of GRMs to be overlapping across populations, and thus, different markers for each ancestral groups could have contributed to the observed genetic variance in the heritability estimates. For example, the current data suggested that SNPs that differentially survive QC across our groups may contribute an additional 7 percent genetic variance to the AA AD factor

and an additional 10 percent genetic variance to the EA AD factor. Although the statistical significance of the AA value is precluded by the large standard error, power simulations conducted by Visscher *et al.* (2014) indicate that with a larger sample, the standard error will decrease rendering the effect significant. Finally, our study also partitioned additive genetic effects of overlapping SNPs in EAs/AAs that survived QC across each chromosome. While analyses revealed that longer chromosomes did not account for more genetic variance, underscoring the polygenic nature of AD, most chromosomes accounted for between 0 and 5 percent of the total variation, regardless of ancestry.

The significant genetic correlation found in this study suggests that the allelic architecture influencing the AD factor for EAs and AAs is largely shared across the two populations. In a single population, genetic correlations arise from pleiotropy or cosegregations of causal variants among genes influencing multiple traits. In the current analysis, the significant genetic correlation represents these genetic contributions influencing a single trait (AD) measured across two populations. Consistent with the conclusion reached by de Candia *et al.* (2013) and given that it is unlikely that different causal variants across ancestral populations would be in LD with the same SNP, the common causal variants tagged by the SNPs that survive QC in both 1KGP-defined ancestral groups likely predate the European-African divergence; SNPs that differentially fall out across ancestral groups may also predate the European-African divergence but are differentially selected in the current sample during QC (e.g. some may be dropped due to the violation of the HWE assumption). Future work should examine the AD factor in other populations to delineate whether these results apply broadly.

One important consideration for how this study informs future studies is the fact that genetic markers contributing to the AD factor may have different allele frequencies and different effect sizes across different ancestral populations. Likewise, differences in haplotype structure across ancestral populations will affect our ability to identify a common set of variants across ancestral groups. The current study focuses on the aggregate effect of common variants (i.e. MAF > 0.01 in each population). As such, this approach does not examine the effects of individual loci. The current study provides an overall estimate of how the aggregate effect of genome-wide SNPs differs between two homogenous groups (i.e. EAs and AAs; see methods section). It is possible that genetic influences on AD may vary as a function of differences in allele frequency across populations (e.g. variants that are more common in one population may contribute more to the overall effect) at the level of the individual loci. However, one recent study concluded that when variants common in both populations are examined, differences in allele frequencies have a minimal influence on genetic correlation based solely on effect sizes (Brown *et al.* 2016). The GREML approach used in this study treats SNPs as statistically random and therefore does not estimate individual effects. Future studies combining multiple populations to identify sets of individual SNPs that contribute to the SNP heritability of AD should consider allele frequency and haplotype structure differences. Another point of considerations is that although the present study focused on common biallelic variants that were present across both ancestral populations, little is known about how rare variation (e.g. copy number variants, multiallelic makers and exome variation) contributes to AD. Kos *et al.* (2013) show support for modest effects of rare and uncommon loci on the susceptibility for AD that were captured from GWAS signals and then aggregated.

The results from this study must be interpreted within the context of the samples and reference panels used in analyses. Differences in estimates may arise due to ascertainment differences across the studies, although we attempted to address this by residualizing our factor scores in part on the study from which the data were obtained. AD symptomology and, subsequently, the covariance among the symptoms may differ across studies due to study specific inclusion/exclusion criteria, individual presentation of symptoms or family history status. For example, individuals in the Heroin GWAS were recruited via a large case–control study focused on opioid dependence, while the SAGE sample consists of a case–control study made up of alcoholic probands recruited from treatment facilities. In addition, given that our ancestry determination relied heavily on the 1KG reference panel data, our findings may not generalize to other populations or individuals with substantial admixture. More research is needed to examine additive genetic effects in additional populations.

In summary, this study demonstrated that (1) approximately 59 percent (i.e., rG-SNP-squared) of the genetic variation for AD that is tagged by measured and retained genome-wide SNPs is shared across EAs and AAs and (2) additional sources of genetic variation may be captured by studying variants that differentially survive QC in one population but not the other. Overall, these observations underscore the reciprocal value of whole genome alcohol studies of ethnically divergent populations.

Acknowledgements

The authors gratefully acknowledge Mr. Bryce Christensen and Golden Helix (Bozeman, Montana) for their assistance with data preparation for genomic imputation analyses. Further, we acknowledge the services of the Michigan Imputation Server team for the use of their publicly available resource for genomic imputation.

Lastly, we acknowledge the many grant-funded research projects (described in the succeeding texts) that assembled all of the data used in the current study, as well as the NIH's database for Genotypes and Phenotypes, which provides a platform for data sharing.

Funding support for the SAGE study was supported by NIH and NHLBI grant # R01HL117004; study enrollment was supported by the Sandler Family Foundation, the American Asthma Foundation, the RWJF Amos Medical Faculty Development Program, and Harry Wm. and Diana V. Hind Distinguished Professor in Pharmaceutical Sciences II.

Funding support for the OZ-ALC GWAS was provided through the Center for Inherited Disease Research (CIDR) and the National Institute on Alcohol Abuse and Alcoholism (NIAAA). CIDR-OZ-ALC GWAS was funded as part of the NIAAA grant 5 R01 AA013320-04. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the CIDROZ-ALC GWAS. Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of OZ-ALC database was provided by MARC: Risk Mechanisms in Alcoholism and Comorbidity (MARC; P60 AA011998-11). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, and the NIH contract 'High throughput genotyping for studying the genetic contributions to human disease' (HHSN268200782096C).

Funding support for the Yale Study was provided through the Center for Inherited Disease Research (CIDR) and the Genetics of Alcohol Dependence in American Populations (CIDR-Gelernter Study). CIDR-Gelernter Study is a GWAS funded as part of the Genetics of Alcohol Dependence in American Populations. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the Genetics of Alcohol Dependence in American Populations. Assistance with data cleaning was provided by the National Center for Biotechnology Information. Cleared 1/4/10 Gelernter: Whole Genome Association (CIDR-Gelernter Study) Dataset 21 January 2010 version Page 5 of 7 Support for collection of datasets and samples were provided by the Genetics of Alcohol Dependence in American Populations (R01 AA011330). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, and the NIH contract 'High throughput genotyping for studying the genetic contributions to human disease' (HHSN268200782096C). The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gap through dbGaP accession number [phs000425].

Finally, funding support for the GWAS of Heroin Dependence was provided by R01DA17305.

Role of Funding Sources

This project was supported by a K01 grant (K01AA021113 awarded to Dr. Palmer) from the National Institute on Alcohol Abuse and Alcoholism (NIAAA), a T32 (T32MH019927, which supports the postdoctoral training of Dr. Brick), an R56 (MH108650) from the National Institute on Mental Health (NIMH; awarded to Dr. McGeary), an R01 (AA025593) from the National Institute on Alcohol Abuse and Alcoholism (NIAAA; awarded to Dr. McGeary), and an R01 (MH100141) from the National Institute on Mental Health (NIMH; awarded to Dr. Keller). The views expressed in this article do not necessarily reflect the position or policy of the Department of Veteran Affairs.

CONFLICT OF INTEREST

All of the listed authors declare that they have no conflicts of interests.

AUTHORS CONTRIBUTION

Drs. Palmer and Keller designed the current project. Drs. Palmer and Brick conducted the analyses under the supervision of Dr. Keller. Drs. Palmer and Brick conducted literature searches and cowrote the first draft of the manuscript, which was later reviewed and edited by Drs. Keller, Knopik and McGeary. All authors contributed to and have approved the final manuscript.

References

Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR (2015) A global reference for human genetic variation. Nature 526:68–74.

Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E, Fisher S, Fox L, Howells W, Bertelsen S, Hinrichs AL, Almasy L, Breslau N, Culverhouse RC, Dick DM, Edenberg HJ, Foroud T, Grucza RA, Hatsukami D, Hesselbrock V, Johnson EO, Kramer J, Krueger RF, Kuperman S, Lynskey M, Mann K, Neuman RJ, Nothen MM, Nurnberger JI Jr, Porjesz B, Ridinger M, Saccone NL, Saccone SF, Schuckit MA, Tischfield JA, Wang JC, Rietschel M, Goate AM, Rice JP (2010) A genome-wide association study of alcohol dependence. Proc Natl Acad Sci U S A 107:5082–5087.

Brown BC, Ye CJ, Price AL, Zaitlen N, Consortium AGENTD (2016) Transethnic genetic-correlation estimates from summary statistics. Am J Hum Genet 99:76–88.

Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI Jr, Reich T, Schmidt I, Schuckit MA (1994) A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. J Stud Alcohol 55:149–158.

Caetano R, Clark CL (1998a) Trends in alcohol-related problems among Whites, Blacks, and Hispanics: 1984–1995. Alcohol Clin Exp Res 22:534–538.

Caetano R, Clark CL (1998b) Trends in alcohol consumption patterns among Whites, Blacks and Hispanics: 1984 and 1995. J Stud Alcohol 59:659–668.

Caetano R, Clark CL (1999) Trends in situational norms and attitudes toward drinking among Whites, Blacks, and Hispanics: 1984–1995. Drug Alcohol Depend 54:45–56.

Caetano R, Clark CL, Tam T (1998) Alcohol consumption among racial/ethnic minorities: theory and research. Alcohol Health Res World 22:233–241.

Chartier K, Caetano R (2010) Ethnicity and health disparities in alcohol research. Alcohol Res Health 33:152–160.

Chartier KG, Karriker-Jaffe KJ, Cummings CR, Kendler KS (2017) Environmental influences on alcohol use: informing research on the joint effects of genes and the environment in diverse U.S. populations. Am J Addict.

Chen C, Dufour M, Yi H (2005) Alcohol consumption among young adults ages 18–24 in the United States: results from the 2001–2002 NESARC survey. Alcohol Res Health 28:269–280.

de Candia TR, Lee SH, Yang J, Browning BL, Gejman PV, Levinson DF, Mowry BJ, Hewitt JK, Goddard ME, O'Donovan MC, Purcell SM, Posthuma D, Visscher PM, Wray NR, Keller MC (2013) Additive genetic variation in schizophrenia risk is shared by populations of African and European descent. Am J Hum Genet 93:463–470.

Desalu JM, Zaso MJ, Kim J, Belote JM, Park A (2017) Interaction between ADH1B*3 and alcohol-facilitating social environments in alcohol behaviors among college students of african descent. Am J Addict 26:349–356.

Dick DM, Barr P, Guy M, Nasim A, Scott D (2017) (Invited review) genetic research on alcohol use outcomes in African American populations: a review of the literature, associated challenges, and implications. Am J Addict 26:486–493.

Dimitrov DM (2010) Testing for factorial invariance in the context of construct validation. Measurement and Evaluation in Counseling and Development 43:121–149.

Frank J, Cichon S, Treutlein J, Ridinger M, Mattheisen M, Hoffmann P, Herms S, Wodarz N, Soyka M, Zill P, Maier W, Mossner R, Gaebel W, Dahmen N, Scherbaum N, Schmal C, Steffens M, Lucae S, Ising M, Muller-Myhsok B, Nothen MM, Mann K, Kiefer F, Rietschel M (2012) Genome-wide significant association between alcohol dependence and a variant in the ADH gene cluster. Addict Biol 17:171–180.

Gelernter J, Kranzler HR, Sherva R, Almasy L, Koesterer R, Smith AH, Anton R, Preuss UW, Ridinger M, Rujescu D, Wodarz N, Zill P, Zhao H, Farrer LA (2014) Genome-wide association study of alcohol dependence:significant findings in African- and European-Americans including novel risk loci. Mol Psychiatry 19:41–49.

Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, Ch'ang LY, Huang W, Liu B, Shen Y, Tam PKH (2003) The international HapMap project. Nature 426:789–796.

Harford TC, Muthén BO (2001) The dimensionality of alcohol abuse and dependence: a multivariate analysis of DSM-IV symptom items in the National Longitudinal Survey of Youth. J Stud Alcohol 62:150–157.

Hasin DS, Stinson FS, Ogburn E, Grant BF (2007) Prevalence, correlates, disability, and comorbidity of DSM-IV alcohol abuse and dependence in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Arch Gen Psychiatry 64:830–842.

Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V (1999) A validity study of the SSAGA—a comparison with the SCAN. Addiction 94:1361–1370.

Hong EP, Park JW (2012) Sample size and statistical power calculation in genetic association studies. Genomics Inform 10:117–122.

Hu LT, Bentler PM (1999) Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model 6:1–55.

International HapMap (2003) The International HapMap Project. Nature 426:789–796.

Johnson C, Drgon T, Walther D, Uhl GR (2011) Genomic regions identified by overlapping clusters of nominally-positive SNPs from genome-wide studies of alcohol and illegal substance dependence. PLoS One 6: e19210.

Kos MZ, Yan J, Dick DM, Agrawal A, Bucholz KK, Rice JP, Johnson EO, Schuckit M, Kuperman S, Kramer J (2013) Common biological networks underlie genetic risk for alcoholism in African-and European-American populations. Genes Brain Behav 12:532–542.

Kristiansson K, Naukkarinen J, Peltonen L (2008) Isolated populations and complex disease gene identification. Genome Biol 9:109.

Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR (2012) Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics 28:2540–2542.

Manolio TA, Brooks LD, Collins FS (2008) A HapMap harvest of insights into the genetics of common disease. J Clin Invest 118:1590–1605.

McGue M, Elkins I, Iacono WG (2000) Genetic and environmental influences on adolescent substance use and abuse. Am J Med Genet 96:671–677.

Meredith W, Teresi JA (2006) An essay on measurement and factorial invariance. Med Care 44:S69–S77.

Mereish EH, Bradford JB (2014) Intersecting identities and substance use problems: sexual orientation, gender, race, and lifetime substance use problems. J Stud Alcohol Drugs 75:179–188.

Minster RL, Hawley NL, Su CT, Sun G, Kershaw EE, Cheng H, Buhule OD, Lin J, Reupena MS, Viali S, Tuitele J, Naseri T, Urban Z, Deka R, Weeks DE, McGarvey ST (2016) A thrifty variant in CREBRF strongly influences body mass index in Samoans. Nat Genet 48:1049–1054.

Mulia N, Ye Y, Greenfield TK, Zemore SE (2009) Disparities in alcohol-related problems among white, black, and Hispanic Americans. Alcohol Clin Exp Res 33:654–662.

Muthén LK, Muthén BO (1998-2015) Mplus User's Guide, Seventh Edition edn. Los Angeles, CA: Muthén & Muthén.

Need AC, Goldstein DB (2009) Next generation disparities in human genomics: concerns and remedies. Trends Genet 25:489–494.

Palmer RH, McGeary JE, Heath AC, Keller MC, Brick LA, Knopik VS (2015) Shared additive genetic influences on DSM-IV criteria for alcohol dependence in subjects of European ancestry. Addiction 110:1922–1931.

Panhuysen CI, Kranzler HR, Yu Y, Weiss RD, Brady K, Poling J, Farrer LA, Gelernter J (2010) Confirmation and generalization

of an alcohol-dependence locus on chromosome 10q. Neuropsychopharmacology 35:1325–1332.
Park BL, Kim JW, Cheong HS, Kim LH, Lee BC, Seo CH, Kang TC, Nam YW, Kim GB, Shin HD, Choi IG (2013) Extended genetic effects of ADH cluster genes on the risk of alcohol dependence: from GWAS to replication. Hum Genet 132:657–668.
Pierucci-Lagha A, Gelernter J, Feinn R, Cubells JF, Pearson D, Pollastri A, Farrer L, Kranzler HR (2005) Diagnostic reliability of the Semi-structured Assessment for Drug Dependence and Alcoholism (SSADDA). Drug Alcohol Depend 80:303–312.
Purcell S, Cherny SS, Sham PC (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics 19:149–150.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575.
Quillen EE, Chen XD, Almasy L, Yang F, He H, Li X, Wang XY, Liu TQ, Hao W, Deng HW, Kranzler HR, Gelernter J (2014) ALDH2 is associated to alcohol dependence and is the major genetic determinant of "daily maximum drinks" in a GWAS study of an isolated rural Chinese sample. Am J Med Genet B Neuropsychiatr Genet 165B:103–110.
Ray LA, Kahler CW, Young D, Chelminski I, Zimmerman M (2008) The factor structure and severity of DSM-IV alcohol abuse and dependence symptoms in psychiatric outpatients. J Stud Alcohol Drugs 69:496–499.
SNP & Variation Suite (Version 8.4.4) SNP & Variation Suite (Version 8.x) [Software]. Bozeman, MT: Golden Helix, Inc. Available from http://goldenhelix.com/products/citing.html.
Swendsen J, Conway KP, Degenhardt L, Dierker L, Glantz M, Jin R, Merikangas KR, Sampson N, Kessler RC (2009) Socio-demographic risk factors for alcohol and drug dependence: the 10-year follow-up of the national comorbidity survey. Addiction 104:1346–1355.
Ulloa AE, Chen J, Vergara VM, Calhoun V, Liu J (2014) Association between copy number variation losses and alcohol dependence across African American and European American ethnic groups. Alcohol Clin Exp Res 38:1266–1274.
Vaeth PA, Wang-Schweig M, Caetano R (2017) Drinking, alcohol use disorder, and treatment access and utilization among U.S. racial/ethnic groups. Alcohol Clin Exp Res 41:6–19.
Visscher PM, Hemani G, Vinkhuyzen AA, Chen GB, Lee SH, Wray NR, Goddard ME, Yang J (2014) Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. PLoS Genet 10: e1004269.
Vrieze SI, McGue M, Miller MB, Hicks BM, Iacono WG (2013) Three mutually informative ways to understand the genetic relationships among behavioral disinhibition, alcohol use, drug use, nicotine use/dependence, and their co-occurrence: twin biometry, GCTA, and genome-wide scoring. Behav Genet 43:97–107.
Wurm L, Fisicaro S (2014) What residualizing predictors in regression analyses does (and what it does not do). Journal of Memory and Language 72:37–48.
Xu K, Kranzler HR, Sherva R, Sartor CE, Almasy L, Koesterer R, Zhao H, Farrer LA, Gelernter J (2015) Genomewide association study for maximum number of alcoholic drinks in European Americans and African Americans. Alcohol Clin Exp Res 39:1137–1147.
Yang C, Li C, Kranzler HR, Farrer LA, Zhao H, Gelernter J (2014) Exploring the genetic architecture of alcohol dependence in African-Americans via analysis of a genomewide set of common variants. Hum Genet 133:617–624.
Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88:76–82.
Yang J, Lee SH, Goddard ME, Visscher PM (2013) Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. Methods Mol Biol 1019:215–236.
Yu C (2002) Evaluation of model fit indices for latent variable models with categorical and continuous outcomes (Vol. 30). Los Angeles: University of California, Los Angeles.
Zapolski TC, Pedersen SL, McCarthy DM, Smith GT (2014) Less drinking, yet more problems: understanding African American drinking and related problems. Psychol Bull 140:188–223.
Zuo L, Gelernter J, Zhang CK, Zhao H, Lu L, Kranzler HR, Malison RT, Li CS, Wang F, Zhang XY, Deng HW, Krystal JH, Zhang F, Luo X (2012) Genome-wide association study of alcohol dependence implicates KIAA0040 on chromosome 1q. Neuropsychopharmacology 37:557–566.
Zuo L, Wang K, Wang G, Pan X, Zhang X, Zhang H, Luo X (2014) Common PTP4A1-PHF3-EYS variants are specific for alcohol dependence. Am J Addict 23:411–414.
Zuo L, Wang K, Zhang XY, Krystal JH, Li CS, Zhang F, Zhang H, Luo X (2013) NKAIN1-SERINC2 is a functional, replicable and genome-wide significant risk gene region specific for alcohol dependence in subjects of European descent. Drug Alcohol Depend 129:254–264.
Zuo L, Zhang CK, Sayward FG, Cheung KH, Wang K, Krystal JH, Zhao H, Luo X (2015) Gene-based and pathway-based genome-wide association study of alcohol dependence. Shanghai Arch Psychiatry 27:111–118.
Zuo L, Zhang F, Zhang H, Zhang XY, Wang F, Li CS, Lu L, Hong J, Krystal J, Deng HW, Luo X (2012) Genome-wide search for replicable risk gene regions in alcohol and nicotine co-dependence. Am J Med Genet B Neuropsychiatr Genet 159B:437–444.
Zuo L, Zhang XY, Wang F, Li CS, Lu L, Ye L, Zhang H, Krystal JH, Deng HW, Luo X (2013) Genome-wide significant association signals in IPO11-HTR1A region specific for alcohol and nicotine codependence. Alcohol Clin Exp Res 37:730–739.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

Table S1. Number of markers that pass at each step of quality control of imputed data by ancestry group.
Table S2. Model fit and loadings for exploratory and confirmatory factor analytic models.
Table S3. Univariate SNP heritability [h^2_{SNP} (SE)] of AD diagnosis for EAs and AAs.
Figure S1. Scree plots from EFA of AD in European ancestry (EA) and African ancestry (AA) groups.
Figure S2. SNP-based heritability, h^2_{SNP}, by chromosome for individuals of European and African ancestry using only SNPs that survive QC across both ancestral groups.
Text S1. PubMed search and results.