Genome-wide association study identifies loci associated with liability to alcohol and drug dependence that is associated with variability in reward-related ventral striatum activity in African- and European-Americans.

Leah Wetherill[1], Dongbing Lai[1], Emma C. Johnson[2], Andrey Anokhin[2], Lance Bauer[3], Kathleen K. Bucholz[2], Danielle M. Dick[4], Ahmad R. Hariri[5], Victor Hesselbrock[3], Chella Kamarajan[6], John Kramer[7], Samuel Kuperman[7], Jacquelyn L. Meyers[6], John I. Nurnberger Jr.,[8] Marc Schuckit[9], Denise M. Scott[10], Robert E. Taylor[11], Jay Tischfield[12], Bernice Porjesz[6], Alison M. Goate[13], Howard J. Edenberg[1,14], Tatiana Foroud[1], Ryan Bogdan[15]**# & Arpana Agrawal[2]**#

1. Indiana University. Department of Medical and Molecular Genetics, Indiana University School of Medicine. Indianapolis, IN.
2. Washington University. Washington University School of Medicine, Department of Psychiatry. Saint Louis, MO. USA.
3. University of Connecticut. University of Connecticut School of Medicine, Department of Psychiatry. Farmington, CT.
4. Virginia Commonwealth University. Department of Psychology & College Behavioral and Emotional Health Institute, Virginia Commonwealth University. Richmond, VA.
5. Duke Institute for Brain Sciences, Dept. of Psychology, Duke University, Durham, NC.
6. SUNY. Henri Begleiter Neurodynamics Laboratory, Department of Psychiatry and Behavioral Sciences, SUNY Downstate Medical Center. Brooklyn, NY.
7. University of Iowa. University of Iowa Roy J and Lucille A Carver College of Medicine, Department of Psychiatry. Iowa City, IA.
8. Indiana University. Department of Psychiatry, Indiana University School of Medicine. Indianapolis, IN.
9. University of California San Diego. University of California San Diego, Department of Psychiatry. San Diego, CA.
10. Howard University, Departments of Pediatrics and Human Genetics, Washington, DC
11. Howard University, Department of Pharmacology, Washington, DC
12. Rutgers University, Dept. of Genetics, Piscataway, NJ.
13. Department of Neuroscience, Icahn School of Medicine at Mt. Sinai, New York, NY
14. Indiana University. Department of Biochemistry and Molecular Biology, Indiana University School of Medicine. Indianapolis, IN.
15. Washington University in Saint Louis, Department of Psychological and Brain Sciences, Saint Louis, MO, USA

**These authors made equal contributions.

#Corresponding author(s):

Arpana Agrawal
Washington University School of Medicine
Dept. of Psychiatry, 660 S. Euclid, CB 8134,
Saint Louis, MO 63110, USA
arpana@wustl.edu
1-314-286-1778

Ryan Bogdan
Washington University in St. Louis,
BRAINLab, Department of Psychological and Brain Sciences, and Division of Biology and Biomedical Sciences
St. Louis, Missouri 63110, USA
bogdan.ryan@gmail.com
1-314-935-7957

Running Title: GWAS of alcohol and drug dependence

Keywords: African American; alcohol dependence; drug dependence; European American; genetics; GWAS; heritability; neural reward; ventral striatum; fMRI

Submitted on: Feb 1, 2019

Abstract: 248; Introduction: 593; Discussion: 2,022

Abstract

Genetic influences on alcohol and drug dependence partially overlap, however specific loci underlying this overlap remain unclear. We conducted a genome-wide association study (GWAS) of a phenotype representing alcohol or illicit drug dependence (*ANYDEP*) among 7,291 European-Americans (EA; 2,927 cases) and 3,132 African-Americans (AA: 1,315 cases) participating in the family-based Collaborative Study on the Genetics of Alcoholism. *ANYDEP* was heritable (h^2 in EA=0.60, AA=0.37). The AA GWAS identified 3 regions with genome-wide significant (GWS; p<5E-08) single nucleotide polymorphisms (SNPs) on chromosomes 3 (rs34066662, rs58801820) and 13 (rs75168521, rs78886294), and an insertion-deletion on chromosome 5 (chr5:141988181). No polymorphisms reached GWS in the EA. One GWS region (chromosome 1: rs1890881) emerged from a trans-ancestral meta-analysis (EA+AA) of *ANYDEP*, and was attributable to alcohol dependence in both samples. Four genes (AA: *CRKL, DZIP3, SBK3*; EA: *P2RX6*) and 4 sets of genes were significantly enriched within biological pathways for hemostasis and signal transduction. GWS signals did not replicate in two independent samples but there was weak evidence for association between rs1890881 and alcohol intake in the UK Biobank. Among 118 AA and 481 EA individuals from the Duke Neurogenetics Study, rs75168521 and rs1890881 genotypes were associated with variability in reward-related ventral striatum activation. This study identified novel loci for substance dependence and provides preliminary evidence that these variants are also associated with individual differences in neural reward reactivity. Gene discovery efforts in non-European samples with distinct patterns of substance use may lead to the identification of novel ancestry-specific genetic markers of risk.

INTRODUCTION

Reducing the widespread prevalence[1-3] and devastating worldwide impact[4,5] of alcohol and illicit drug dependence is hindered by limited etiologic insight that impedes prevention and treatment advances. In the United States (US), 12.5% of the population meets criteria for a lifetime history of alcohol dependence[3] while 2.6% meet criteria for DSM-IV drug dependence during their lifetime[2]. Notably, individuals are often comorbid for multiple substance use disorders[2], and common latent genetic factors[6,7] explain a large proportion of the moderate to high heritability of dependence on individual substances (h^2=50-70%[8-10]). The common genetic architecture of dependence liability is also underscored by evidence from genome-wide association studies (GWAS) documenting genetic correlations between alcohol-related measures and cannabis and cigarette use[11,12]. Leveraging the common genetic architecture underlying general substance dependence liability to identify markers of dependence risk through GWAS would complement existing efforts targeting individual substances (e.g., [12-16]) to elucidate underlying etiologic risk factors for general and specific substance dependence liability.

It is estimated that an overwhelming proportion of participants in existing GWAS are of European ancestry.[17,18] Data generated from GWAS of individuals of European ancestry are less applicable to other ancestral groups and when applied to non-European cohorts may result in inaccurate estimations of risk that may further perpetuate racial health and healthcare disparities. Studies suggest that even when discovery samples of non-European individuals are small, including them in individual discovery analyses and trans-ancestral analyses can result in novel insights into the genetic architecture of the disorder and in polygenic prediction[12,19,20]. Differences in prevalence and patterns of substance dependence across ancestrally diverse groups in the United States[21] underscore the importance of conducting GWAS on these

phenotypes in these groups. In particular, the study of African-Americans, one the largest minorities represented in GWAS data in the US, provides an opportunity to address this notable disparity in genomic research.

Here, we conduct a GWAS of a phenotype representing alcohol or illicit drug (i.e., cannabis, cocaine, sedatives, stimulants and/or opioids) dependence (*ANYDEP*) among 7,291 European-Americans (EA; 2,927 cases) and 3,132 African-Americans (AA: 1,315 cases) participating in the family-based Collaborative Study on the Genetics of Alcoholism (COGA). COGA participants were recruited from extended families, most of which were ascertained for alcohol dependence. The *ANYDEP* phenotype is particularly well suited for this ascertained sample as drug dependence more commonly co-occurs with alcohol dependence than with dependence on any other substance[22,23]. We conducted ancestry-specific analyses in EAs and AAs followed by a trans-ancestral meta-analysis (EA+AA) to identify loci associated with *ANYDEP*, i.e., dependence on any one or a combination of alcohol, cannabis, cocaine, sedatives, stimulants and/or opioids. For genome-wide significant (GWS) associations, we performed secondary analyses evaluating associations with individual alcohol and drug dependence diagnoses, and to examine whether the exclusion of those cases who met criteria only for alcohol dependence altered the association. Replication was attempted in two small independent samples that contained EA and AA individuals and substance dependence phenotypes, the Study of Addiction: Genes and Environment (EA: 630 cases, 1,020 controls: AA 387 cases, 415 controls)[24] and the Yale-Penn AA study (AA: 1,525 cases, 485 controls)[25]. Further, any GWS associations with *ANYDEP* in the EA sample were tested for association with alcohol intake among 452,264 individuals from the UK Biobank[26] and cannabis use from a meta-analysis

conducted on 184,765 individuals[27]. Finally, given the proposed role of reward-related neural response in the etiology of addiction[28,29], we examined whether GWS loci were correlated with reward-related ventral striatum reactivity as measured with blood-oxygen-level dependent (BOLD) functional magnetic resonance imaging (fMRI) in the independent Duke Neurogenetics Study (EA n=481, AA n=118)[30,31].

MATERIALS AND METHODS

Sample: COGA is a large family-based study that recruited alcohol dependent probands from treatment facilities across seven sites in the United States[32,33]. Probands and their extended families were invited to participate. Additional individuals and their families were recruited from the same communities using a variety of resources (e.g., dental clinics). Institutional review boards at all sites approved the study, and all participants provided informed consent. All participants were administered a version of the Semi-Structured Assessment for the Genetics of Alcoholism interview (SSAGA; those aged <18 years were administered a child version, the C-SSAGA)[34,35]. Phenotypic data were available on 16,809 individuals. A substantial portion of the sample (n=12,146) has been genotyped. Because the number of individuals of other ethnicities was small, only EA (n=7,983) and AA (n=3,685) individuals were included in these analyses. As the study was ascertained for alcohol dependence, individuals who reported never drinking alcohol even once in their life were excluded from analyses (n=550). The majority of those individuals reported not ever using other drugs, with the exception of cannabis (n=63). The final analytic sample (n=7,291 EA and 3,132 AA) included those with both genotypic and phenotypic data.

Measures: ANYDEP was defined as a binary variable where cases met lifetime criteria for DSM-IV dependence[36] on alcohol, cannabis, cocaine, sedatives, stimulants and/or opioids (for prescription drugs, non-prescription use was specified) or any combination thereof. We did not include nicotine dependence as it was not assessed in earlier versions of the SSAGA, and was therefore missing for those who were only interviewed using older SSAGAs. Controls did not meet DSM-IV dependence criteria for alcohol or any drug listed above but were required to have consumed at least 1 drink of alcohol. Of the controls, 32.9% met lifetime criteria for DSM-IV alcohol or drug abuse (analyses excluding these individuals are described in the **Discussion**). For GWS SNPs, alcohol dependence and each individual DSM-IV drug dependence diagnosis was also examined against this uniform set of controls. As COGA was primarily ascertained for alcohol dependence, we created a variable for secondary analysis, *drug_noalc*, where individuals with alcohol dependence were excluded from the *ANYDEP* cases (and remained excluded from controls). Numbers of individuals for each phenotype are in **Table 1**.

Phenotype analysis:

The prevalence of alcohol and drug dependence were compared across ancestral groups using chi-square tests. The number of DSM-IV criteria endorsed by individuals in AA and EA families were compared (total number and for each drug) using an ordinary least squares regression that accounted for sex. Birth cohorts (1890-1929, 1930-1949, 1950-1969, ≥1970) were included in all COGA analyses as covariates to account for secular trends (see[37]) across this wide range of birth years. Over and above birth cohort, age was not a significant predictor of *ANYDEP* ($p > 0.3$). A confirmatory factor analysis of substance dependence diagnoses was fitted to the data separately for EAs and AAs, to determine phenotypic patterns of comorbidity

using MPLUSv8[38]. The comparative fit index (CFI), standardized root mean square residual (SRMR), and the root mean square error of approximation (RMSEA) were used to assess model fit. Heritability for *ANYDEP* and *drug_noalc* was estimated using SOLAR[39] in the EA and AA families separately, using familial relatedness (but not GWAS data) alone.

Genotyping, Quality Review, Ancestry and Imputation

Multiple genome-wide arrays were used to genotype the COGA sample[23,40-42] (see **Supplemental Text**). A subset of 47,000 common (minor allele frequency (MAF) > 0.1 in the combined sample), independent (defined as $R^2 < 0.5$) and high quality (missing rate < 2% and Hardy-Weinberg Equilibrium (HWE) p-values > 0.001) SNPs that were genotyped across all arrays were used to assess duplicate samples, confirm the reported pedigree structure and compute ancestral principal components (see **Supplemental Text** for details). After assignment of individuals in a family to a specific population, family-wise ancestry was designated according to the majority of individual family members (see Lai et al, accompanying paper). Only AA and EA families were included in subsequent analyses, due to low numbers of other groups. Only variants with non A/T or C/G alleles, missing rates < 5%, MAF > 3%, and HWE p values > 0.0001 were used for imputation. Genotypes were imputed to 1000 Genomes using the cosmopolitan reference panel (Phase 3, version 5, NCBI GRCh37; **Supplemental Text**) using SHAPEIT2[43] and Minimac3[44]. Imputed SNPs with $R^2 < 0.30$ were excluded, and genotype probabilities were converted to best-guess genotypes if ≥ 0.90. Because some individuals within a family were genotyped on different arrays, families were again evaluated for Mendelian inconsistencies using Pedcheck[45], and imputed SNPs were cleaned as described above. All

genotyped and imputed SNPs with missing rates <25%, MAF ≥1% and HWE p >1 x 10^{-6} were included in analyses.

Genome wide association studies and meta-analysis:

Association analysis was performed separately in AA and EA families using a generalized estimating equation (GEE) framework to account for family relatedness by considering each family as a cluster. The GEE employs a logistic regression model (i.e., binomial distribution) to account for relatedness in the R package GWAF[46]. Gender, birth cohort, GWAS array indicator, and the first four principal components (as in[23]) were included as covariates in the model. A trans-ancestral (EA+AA) GWAS was performed by meta-analyzing summary statistics from the EA and AA GWAS using inverse-variance weighting in METAL[47], with genomic control.

Annotation of results and gene-based analyses: Overall plotting (e.g., regional association) and annotation of individual loci was conducted in FUMA[48]. For gene and gene-set based analysis, MAGMA, as implemented in FUMA, was used. FUMA was utilized to conduct gene-set analyses that examined whether genes were enriched in curated classification systems, by molecular function, biological process or other criteria. Gene sets were defined for 4,728 curated gene sets (including canonical pathways) and 6,166 GO terms. Differential expression of prioritized genes was conducted using the GENE2FUNC option in FUMA, which examines whether genes of interest from the GWAS are overrepresented in differentially expressed gene sets in 53 specific tissue types from The Genotype-Tissue Expression (GTEx) database[49]. Although this database is comprised of primarily EA individuals, it is one of few publicly available databases available, and therefore was utilized for the AA results as well. To further

prioritize possible causal genes, we used S-PrediXcan[50] to impute genetically-regulated gene expression in twelve brain tissues and whole blood. The prediction models were trained on reference transcription data from GTEx (brain) and the Depression Genes and Network (DGN) (whole blood) (all available from the PredictDB Data Repository, http://predictdb.org, downloaded on 6/6/2018). Analyses were restricted to the EA data, as the prediction models used by the tools were built using only individuals of European ancestry. We used GTEx v7 to extract gene expression values. Finally, individual genome-wide significant (GWS) SNPs and genes were examined against SNP and gene-based summary statistics for 3,798 GWAS of 2,824 traits, available through http://atlas.ctglab.nl/ (accessed on 10/1/2018).

Replication:

Data from two dbGaP samples with individuals of EA and AA ancestry that were ascertained for alcohol and substance dependence were utilized for replication of GWS SNPs. These included the Study of Addiction: Genes and Environment (phs000092.v1.p1, SAGE: non-overlapping individuals numbered EA: 630 cases and 1,020 controls; AA: 387 cases and 415 controls)[24] and the Yale-Penn AA sample (phs000425.v1.p1) with 1,525 cases and 485 controls[25]. Any overlapping participants as well as the first and second degree of relatives ($\pi \geq 0.2$) of COGA members in SAGE or Yale-Penn were excluded from the replication samples. Cases and controls were defined as described above. Covariates included sex and the first 3 principal components. For SAGE, birth cohorts as defined in COGA were included as covariates while for Yale-Penn AA, age was used (as recommended in prior publications of this sample[51,52]). Effect sizes across COGA and replication samples were meta-analyzed in METAL [47]. For SNPs associated with *ANYDEP* in the EA families, we also examined summary statistics for association with alcohol

intake frequency in 452,264 individuals from the UK Biobank (http://geneatlas.roslin.ed.ac.uk/, accessed 11/26/2018)[26] and with cannabis use from the current largest GWAS of the phenotype [n=184,765[27]].

Neuroimaging analysis of GWS loci:

We examined whether GWS SNPs identified in our COGA GWAS of *ANYDEP* (i.e., rs34066662, rs75168521, rs1890881; the *indel* was not available) were associated with reward-related brain function in the Duke Neurogenetics Study (DNS), an independent neuroimaging sample containing non-Hispanic AA (n=118) and EA (n=481) undergraduate students aged 18-22 years[31] (see **Supplemental Text**). A number guessing paradigm was used to elicit ventral striatum (VS) reactivity associated with positive and negative feedback linked to monetary gains and losses while bold-oxygen-level dependence (BOLD) functional magnetic resonance imaging (fMRI) data were acquired[53]. Statistical Parametric Mapping version 8 (SPM8) software was used to extract parameter estimates for the contrast of Positive Feedback > Negative Feedback from maximal voxels within left and right VS regions of interest (ROIs). Imaging acquisition protocol, task, ROIs, and preprocessing details are described in the **Supplemental Text**. Extracted parameter estimates from VS activity in each hemisphere were regressed on genotype (coded as 0 vs 1 or more copies of the minor allele) of GWS loci while co-varying for sex, and 3 (AA) or 2 (EA) ancestral principal components using Full Information Maximum Likelihood in MPlus v7.3[38]. Confidence intervals on estimates were derived via bootstrapping (n=10,000). To adjust for multiple comparisons, we used a Bonferroni-corrected p-value threshold ($p<0.00625$), to account for our hypothesized 8 tests [i.e., rs34066662 and rs75168521 in both brain hemispheres among AAs (4 tests); rs1890881 in both brain hemispheres among AAs and EAs (4

tests)]. As rates of drug dependence, but not alcohol problems (see[31]), are low in the DNS sample, structural equation models linking genotype to substance dependence with reward-related response as a mediator were not fitted to these data.

RESULTS

Phenotypic analyses: Alcohol dependence was the largest contributor to *ANYDEP,* followed by cannabis and cocaine dependence (**Table 1)**. Alcohol and drug dependence were correlated with each other in both EAs and AAs (**Supplemental Table S1**). However, correlations between dependence on alcohol and individual drugs were higher (r = 0.55 – 0.82) in EA relative to AA (r = 0.33 – 0.77), especially for cannabis dependence (see **Supplemental Table S1**). A single factor solution fit the lifetime dependence diagnoses data adequately in both EAs and AAs (EA: Comparative Fit Index=0.989, Standardized Root Mean-square Residual=0.046, Root Mean Square Error of Approximation=0.042; AA: Comparative Fit Index=0.978, Standardized Root Mean-square Residual=0.09, Root Mean Square Error of Approximation: 0.047) and factor loadings were greater than 0.75, with the exception of cannabis dependence in AA (loading=0.35).

Other phenotypic differences across ancestral groups in this sample were apparent. For instance, while alcohol dependence was the most common contributing diagnosis in EAs (81%, vs. 67% in AAs), cannabis (EA: 42%; AA: 52%) and cocaine dependence (EA: 27%; AA: 44%) were more common in AAs. Across ancestral groups, *ANYDEP* cases endorsed a similar number of criteria across all substances (i.e., 7 criteria x 6 substances; mean = 10.8, SD = 7.3; beta=0.11, SE=0.24, p=0.66). Among *ANYDEP* cases, EAs endorsed significantly more alcohol dependence criteria

than AAs (EA: mean 4.44, SD 2.09; AA; mean 3.89 SD 2.21; beta=0.45, SE=0.06, p < 0.0001). Conversely, AAs endorsed a greater number of cannabis (AA: mean 2.76, SD 2.29; EA: mean 2.24, SD 2.27; beta=0.41, SE=0.07, p < 0.0001) and cocaine dependence (AA mean 2.82 SD 3.10; EA: mean 1.65, SD: 2.59, beta=1.22, SE=0.09, p < 0.0001) criteria than EAs, and these differences were significant even after accounting for sex and birth cohorts. The heritability of *ANYDEP* and *drug_noalc* were 0.60 (standard error (SE) = 0.043) and 0.59 (SE=0.085) respectively, in the EA families. Although the heritability of *ANYDEP* was lower in the AA families (0.37; SE=0.065), the heritability of *drug_noalc* was slightly higher (0.63; SE=0.106).

GWAS findings: No GWS loci emerged in the EA GWAS (**Figure 1a**; **Supplemental Figure S1A**). The lowest p-value (p = 8.6E-08; **Table 2**) was obtained for rs74611272, an intergenic SNP on chromosome 7. In contrast, three GWS regions were identified in the AA GWAS: on chromosome 3 (rs34066662: p = 1.77E-08 & rs58801820: p = 1.89E-08; **Figure 1b**; **Supplemental Figure S1B**), chromosome 13 (rs75168521: p = 3.31E-08 & rs78886294: p = 4.38E-08) and an insertion–deletion (*indel*) on chromosome 5 (5:141988181, mapped to rs527904740, p = 4.48E-08). As shown in **Table 2**, the effects of these variants were ancestry-specific. In addition, one locus on chromosome 1 was GWS in the trans-ancestral (EA+AA) meta-analysis (**Figure 1c**, **Supplemental Figure S1C)**, with the most significant SNP, rs1890881 (p = 3.77E-08; EA p = 8.95E-05; AA p = 1.94E-05) in an intron of *RABGAP1L* (RAB GTPase Activating Protein 1 Like).

Specificity of GWAS SNPs to alcohol or drug dependence: In the AA GWAS, all drugs contributed to the chromosome 3 GWS signal, albeit at nominal levels of significance, while all

drugs except opioids and sedatives contributed to the chromosome 5 and 13 signals (**Table 3**). Alcohol dependence was also associated with these loci (**Table 3**); however, when a smaller subset of individuals who met criteria for drug but not alcohol dependence was studied (i.e., *drug_noalc*), the loci on chromosome 3, 5 and 13 remained nominally associated (all p>8.61E-04; **Table 4**), suggesting that these signals were only partially attributable to shared genetic liability between alcohol and illicit drug dependence. In contrast, the trans-ancestral signal on chromosome 1 was due primarily to association with cocaine and alcohol dependence in both the EA and AA subsamples (**Table 4**). When individuals with alcohol dependence only were excluded from the study (*drug_noalc*), there was no association (p=0.26) in the EAs and the association in AAs decreased in significance to p=0.04.

Biological annotation: Regional association plots for the chromosome 3 and chromosome 13 GWS loci from the AA GWAS are presented in **Figure 2**. The two SNPs on chromosome 3, rs34066662 and rs58801820, are in complete linkage disequilibrium (LD; HapMap AFR sample: r2>0.996; D'=1). There was evidence that one or both of these SNPs were eQTLs for Nephrocystin 3 *(NPHP3)* in the sigmoid colon (GTEx v6: p = 4E-06) and the adrenal gland (GTEx v7: p = 7E-06; reference expression data primarily drawn from Europeans).

The regional association plot for the GWS *indel* on chromosome 5 (chr5:141988181, mapped to rs527904740) is presented in **Supplemental Figure S2A**. The next most significantly associated variant, rs74911483 (chr5:141990602, p=6.13E-8), is shown in **Supplemental Figure S2B**, and was in high LD with this variant (AFR: r2=0.87, D'=0.94[54]). Both variants were in the intron of the Fibroblast Growth Factor 1 (*FGF1*) gene. While FUMA could not be utilized for *indel*

annotation, individual searches for the proxy SNP, rs74911483 in RegulomeDB[55] (score of 5), Combined Annotation Dependent Depletion[56] (maximum CADD score of 5), and GTEx[57] (no eQTLs) did not provide persuasive support for regulatory effects of this variant. However, chromatin interactions were noted with neighboring genes (**Supplemental Figure S3**).

The strongest signal on chromosome 13 was from rs75168521, a non-coding intergenic SNP downstream of *SLITRK5* (SLIT and NTRK like family member 5); there was no evidence that rs75168521 is an eQTL for *SLITRK5* or any other gene. rs75168521 is a perfect LD proxy for rs78886294 which was also genome-wide significant. Both SNPs were also in high D', but low r2 (D'=1; r2=0.33) with numerous SNPs in the 3' region of *MIR4500HG* with the closest SNP being 108 bp from rs75168521. The SNP rs75168521 made chromatin contact with *MIR4500HG* in bladder, liver and the left ventricle (although the gene is only appreciably expressed in the liver). Several additional distal points of contact were also identified (**Supplemental Figure S4**). Conditional analyses of the lead variants on chromosomes 3, 5 and 13 indicated that the remaining genome-wide significant SNPs did not represent additional independent loci on each chromosome (**Supplemental Figure S5A – S5C**). However, additional SNPs in the region did show p-values indicative of potential independent signals that might be clarified with increase in sample size.

The trans-ancestral (EA+AA) analyses identified rs1890881 as genome-wide significant (**Figure 4**). The SNP is in an intron of *RABGAP1L*; however, it is also an eQTL for several neighboring genes. According to the GTEx (v7) database, rs1890881 is an eQTL for 7 genes (48 signals), that included several signals in brain tissue (**Supplemental Table S2**). rs1890881 also made

chromatin contact with several of these genes, including *SERPINC1* (**Supplemental Figure S6;** genes identified using eQTL and chromatin interaction mapping are in red) as well as other distal contacts across tissues. There was no support for independent loci in the region in either ancestral group (**Supplemental Figure S5D and S5E**).

Gene-based and gene-set analyses: Gene-based analyses in the AA data identified 3 genes that surpassed genome-wide correction ($P_{significance}$ = 2.76E-06, corrected for 18,125 protein coding genes; **Supplemental Figure S7; Supplemental Table S3**). The genes were SH3 Domain Binding Kinase Family Member 3 (*SBK3*; chromosome 19), DAZ Interacting Zinc Finger Protein 3 (*DZIP3*; chromosome 3) and CRK Like Proto-Oncogene, Adaptor Protein (*CRKL*; chromosome 22). *DZIP3* and *CRKL* are ubiquitously expressed with appreciable expression in brain regions, while *SBK3* is expressed in cardiac tissue (**Supplemental Figure S8**). Gene-set analyses did not identify any GO terms that surpassed multiple testing correction. We also performed gene function analyses with 26 genes that mapped to the region of 2 of the GWS loci (including, but not limited to *SLITRK5, NPHP3* and *NPHP3-AS1, LINC00433*) on chromosomes 3 and 13. Two positional gene sets (MSigDB_c1) on chromosome 3 and one on chromosome 13 were significantly enriched for prioritized genes (chr3q22, $P_{adjusted}$ = 8.3E-10: *NPHP3, NPHP3-AS1, BFSP2-AS1, SRPRB, C3orf36*; chr3q21, $P_{adjusted}$ = 1.6E-4: *TMEM108, BFSP2, TF*; and chr13q31, $P_{adjusted}$ = 1.4E-09: *SLITRK5, PEX12P1, KRT18P27*). Of these, Transferrin (*TF*) in particular, showed higher average differential expression in brain tissue (**Supplemental Figure S9**).

Despite no individual SNP being genome-wide significant in the EA GWAS, one gene was genome-wide significant (*P2RX6*, p=7.11E-07; threshold $P_{significance}$ = 2.82E-06 for 17,757 coding genes; **Supplemental Table S4**). The Purinergic receptor P2X 6 (*P2RX6*) is expressed in brain tissue, although less robustly than in musculo-skeletal tissue (**Supplemental Figure S10**). In addition, 4 curated gene-set terms were statistically significant (**Supplemental Table S5** for individual genes in the set). These gene sets were derived from Reactome and reflected gene sets involved in signal transduction (sets 1 and 2) and hemostasis (sets 3 and 4). When the correlation between *ANYDEP* and imputed, genetically-regulated gene expression was tested in the EA sample using S-PrediXcan, no genes met the multiple testing corrections; however, *P2RX6* was the most significant gene, with *p* = 3.91e-05 in the putamen basal ganglia tissue model.

Replication: Despite these promising findings, the individual loci did not replicate in any of the replication samples (**Supplemental Table S6;** lowest p =0.06 for chr5:141988181 in Yale-Penn) and meta-analysis across COGA and the replication samples did not retain their genome-wide significance, although findings for the chromosome 3 locus were in the same direction in the AA samples, and for chromosome 1 in the EA samples. In the UK Biobank, there was weak evidence of association between alcohol intake frequency and rs1890881 (beta=-0.010, p=0.026). In addition, rs1890881 was marginally associated with cannabis use (beta=0.029, p=0.048) in the current largest GWAS of the trait[27]. There was evidence of association between alcohol intake and rs74611272 (beta=0.016, p=0.007;[26]), the strongest signal in the EA GWAS, in the UK Biobank.

Extension of Neuroimaging to Significant Loci:

As shown in **Table 5**, carriers of the minor (C) allele of rs75168521 (chr 13), which was associated with increased likelihood of *ANYDEP* in COGA, had blunted right (beta=-0.073, p=0.004), but not left (beta=-0.062, p=0.143), VS activation in the AA subsample. However, rs34066662 genotype (chr 3) was not associated with VS activity in either hemisphere among AA, despite a similar directional pattern (betas>-0.062, ps >0.103). As expected, among EA there was no association between rs75168521 or rs34066662 genotype and VS activation in either hemisphere (|betas|<0.025, ps >0.247). Carriers of the minor (T) allele of rs1890881 (chr 1), which was associated with decreased likelihood of *ANYDEP* in the trans-ancestral meta-analysis (effect driven by alcohol dependence), were characterized by blunted reactivity of the left VS among AA (beta=-0.134, p=0.001). Nominally significant associations that were not robust to Bonferroni correction were observed between rs1890881 genotype and right VS activation in AA (betas<-0.026, ps <0.030) and left (beta=-0.036, p=0.029), but not right (beta=-0.004, p=0.801), VS activation among EA.

DISCUSSION

Alcohol and drug dependence tend to co-aggregate in families[58,59]. Based on a prior GWAS in a smaller subset of 118 COGA families[23], we developed an *ANYDEP* phenotype that represented a diagnosis of dependence on alcohol or any illicit drug. GWAS in the COGA AA families identified novel loci on chromosomes 3, 5, and 13 while the trans-ancestral EA+AA analysis identified a locus on chromosome 1. However, these signals failed to replicate in independent samples. In addition, across the AA and EA GWAS, a total of 4 genes (AA and EA) and 4 gene sets (EA only) survived correction for multiple testing. These findings underscore the feasibility

of using an aggregate substance dependence phenotype to identify underlying shared heritable influences for locus discovery.

Broadly, two categories of loci were identified. First, in the COGA AA families, loci on chromosome 3, 5, and 13 were GWS and appeared to be attributable to contributions from each individual illicit drug, as well as alcohol dependence. Exclusion of alcohol dependence diagnosis (*drug_noalc*) resulted in nominal significance in all three regions, despite the substantially reduced sample size and power. Thus, these loci may represent genetic liability that is common to alcohol and illicit drug dependence that cannot be entirely attributed to alcohol dependence. On the other hand, the locus on chromosome 1 which was GWS in the COGA EA+AA analysis and was supported by signals in both the EA and AA subsamples, showed nearly no evidence for association in the *drug_noalc* analyses (p=0.04; and only in the AA families), suggesting that this signal is primarily due to alcohol dependence. In EAs, this genome-wide significant SNP (rs1890881) is in perfect LD with rs61826952 which was genome-wide significant in the COGA GWAS of DSM-IV alcohol dependence (COGA EA+AA p=8.4E-11; see accompanying paper by Lai et al.). The r^2 in the AFR reference populations is 0.48, potentially indicating an independent signal in the AAs. However, conditional analyses did not support an effect of rs61826952, independent of rs1890881, on *ANYDEP* (**Supplemental Figures 5D and 5E**). Similarly, rs1890881 was not associated with alcohol dependence (Lai et al., accompanying paper), independent of rs61826952 (AA p=0.118; EA p=0.559).

For *ANYDEP*, rs1229984 in *ADH1B*, the most well-replicated signal for alcohol dependence[12] was not GWS. Even relative to findings from the companion paper by Lai et al., where

rs1229984 was GWS in the EA+AA analysis (beta=-0.86, p=1.72E-8), the association between rs1229984 and *ANYDEP* was considerably weaker both in magnitude and significance (**Supplemental Table S7**), despite a considerably larger analytic sample. Additional loci identified for alcohol dependence diagnosis, symptom count and individual criteria in Lai et al., also did not achieve GWS for *ANYDEP*, although all signals were nominally significant with effects in an identical direction. An important distinction between these two companion studies is noteworthy. While Lai et al., excluded individuals with ≥ 2 alcohol or drug abuse/dependence criteria from their unaffected population, the current study allowed these individuals to remain in the unaffected group. Thus, variations across findings in the two studies might be due, not only to differing definitions of affecteds, but also the definition of unaffecteds. Finally, the current study did not identify the same SNPs as were noted in our prior study of *ANYDEP* and its quantitative equivalent in a much smaller subset of these data (N=1,170 – 2,183; Supplemental **Table S7**), which is not uncommon with the substantially increased sample size used here.

In addition to sources of genomic variation (e.g., allele frequencies, LD), distinctions in findings across the ancestral groups are possibly attributable to the pattern of comorbidities in these groups, which may be genetic and environmental in nature. Notably, a fair proportion of AA qualified for a diagnosis of *ANYDEP* due to cannabis or cocaine dependence, whereas the preponderance of EAs primarily endorsed alcohol dependence. In addition, cannabis and cocaine dependence diagnoses in AA were relatively more severe (i.e., more criteria were endorsed). Furthermore, *drug_noalc* (*h2*=0.63) was more heritable than *ANYDEP* (*h2*=0.37) and alcohol dependence (*h2*=0.27) itself in the AA but not the EA families. Thus, despite the smaller sample size, the AA subsample may have been better powered to identify loci more closely

related to drug dependence. These patterns of individual and comorbid drug use disorders are also quite consistent with the broader epidemiological literature[3,21]. For instance, AA are more likely to initiate use of cannabis prior to alcohol and are more likely to escalate to problem use[60,61]. Similarly, AA are at nearly 3.5 increased odds of transitioning from cocaine use to dependence than their EA counterparts, even after adjustment for sociodemographic features and psychiatric comorbidity[62]. However, these population differences may reflect socio-cultural trends or represent barriers to access to prevention programs among minority populations, thus increasing rates of lifetime drug dependence[63]. While we might speculate that the three loci identified in the AA GWAS are more likely to relate to liability to both alcohol and drug dependence, this observation may merely be an artifact of cultural effects on the expression of genetic susceptibility.

Alternatively, the AA findings might be false positives. Due to sparser LD in AA relative to EA, the application of a uniform threshold of $p < 5E-08$ for attributing GWS in AAs may not be sufficiently stringent[64,65]. However, gene-based tests were also successful at identifying three significant genes. Thus, at least for the gene-based tests, even after correction for differences in LD, significant findings in the smaller AA sample were identified.

Despite several findings at the level of individual loci, genes and even gene-set terms (for EA), none of the biological units identified in this GWAS were related to genes previously linked to alcohol or drug related phenotypes. The gene sets, for instance, were broadly related to hemostasis and signal transduction. Prior gene set enrichment analyses have identified other gene sets related to signal transduction more broadly but not specifically via our gene-set terms[66,67].

However, as shown in Supplemental **Figure S11**, loci on chromosome 1, 3, 13 and 7, as well as the 4 genes that surpassed genome-wide correction (*CRKL, DZIP3, SBK3, P2RX6*) did show associations ($p < 0.05$) with other psychiatric, cognitive and behavioral traits. For instance, based on comparisons with other published and unpublished GWAS across multiple phenotypes, *CRKL* variants have been linked to age at smoking cessation (p=0.002) while *DZIP3* variants have been related to bipolar disorder (p=3.3E-05). Nonetheless, the nature of the effect of these variants on alcohol and illicit drug dependence remains unknown.

The relatively small sample size for the replication cohorts, especially when examining AA individuals, may have contributed to our limited evidence for replication. However, the neuroimaging extension provides evidence for an interesting, albeit preliminary, link between GWS loci and ventral striatum (VS) reactivity. For rs75168521, African American carriers of the C allele, which was associated with increased likelihood of *ANYDEP* in the COGA GWAS, had blunted response to positive versus negative feedback. Decreased reward sensitivity to rewarding, non-drug stimuli has been well documented within addiction, with evidence that this may arise following persistent drug exposure[68]. However, as the DNS sample is characterized by relatively low levels of substance use and related problems (other than alcohol[31]), the association between genotype and blunted VS response to reward may plausibly be viewed as a predisposing factor. One might speculate that individuals, particularly adolescents, with a blunted response to rewarding stimuli, in general, may require larger drug amounts or more potent drugs for reinforcement, and thus, be more susceptible to the development of severe addiction[69]. Such an interpretation is consistent with evidence that unaffected offspring of alcohol dependent individuals and adolescents who later develop problematic drug use have reduced VS response to

anticipatory cues of monetary reward[70,71] as well as evidence that individuals who are less sensitive to the intoxicating effects of alcohol are at greater risk for dependence[72].

In direct contrast to the results for rs75168521, rs1890881 (chromosome 1) major C allele homozygotes, who were at increased risk for *ANYDEP* (driven by the association with alcohol dependence) in the COGA GWAS, had elevated reward-related VS response (identical to Lai et al.). While these findings may, on the surface, appear to be inconsistent with one another, literature suggests that both relatively reduced and heightened VS response to reward may be associated with substance involvement and dependence liability according to unique and shared mechanisms[31]. For example, evidence that reward-related VS activity is positively coupled with problematic drinking[73] as well as behavioral and self-reported impulsivity[74], converges with stage-based theories of addiction postulating that elevated impulsivity may lead to greater substance use exposure and experimentation that may lay the foundation for the development of problematic substance use. On the other hand, a parallel literature has also linked relative hypoactivity of the VS to drug-seeking behaviors, which has often been theorized to reflect compensation for blunted reactivity to reward[71,75]. Thus, it is plausible that blunted VS reward response associated with rs75168521 may confer susceptibility to extreme and generalized forms of drug dependence. On the other hand, the finding for rs1890881 might typify individuals at high neurobiological susceptibility for substance use engagement, particularly with alcohol which is easily accessible and socially accepted. Given the heterogeneity of substance use-related phenotypes, it is plausible that different genetic risk markers may impact disease risk through distinct mechanisms, and that these seemingly divergent theoretical models (e.g., impulsivity vs reward deficiency) may not be mutually exclusive.

Lastly, laterality differences were apparent following multiple testing correction (i.e., rs75168521 genotype was associated with blunted right, but not left, VS response while rs1890881 genotype was associated with blunted left, but not right VS response in AA). While prior reports have found evidence of lateralized associations with reward-related processing in the VS[76,77], the directionality of associations in our study were consistent across hemispheres and in some cases reached nominal levels of significance (**Table 5**). As such, it is possible that our lateralized findings resulted from limited power. Overall, our neuroimaging findings, while preliminary, showed ancestral specificity consistent with the GWAS, and suggest a putative role for the *ANYDEP*-associated variants in general reward responsiveness. However, despite correction for multiple testing, it is also plausible that these findings represent a false positive given our small sample, which also prevented us from testing potential quadratic effects which might be expected given that both relatively reduced and heightened VS response to reward were associated with genetic risk for substance use phenotypes.

Several limitations are worth noting. First, despite interesting findings and the partially high risk sample design, our sample size is underpowered to detect the modest effect sizes typically associated with substance use disorders[12]. For instance, in the AA subsample, 80% power to detect a common variant (MAF ≥ 35%) is only expected for genotype relative risks≥1.28[78], which is fairly high for psychiatric disorders. Second, due to the high degree of relatedness in our data, ascertainment, and the relatively small number of "cases", methods such as GCTA[79] or LD Score regression[80] that are typically used to assess SNP-heritability were not appropriate. Instead, we report heritability using familial relatedness. Third, we elected to derive diagnoses

based on the DSM-IV nomenclature for dependence instead of DSM-5 definitions for substance use disorders. Even though the DSM-5 definition of substance use disorders is more contemporary, it relies on a lower symptom burden (e.g., ≥2 of 11 vs ≥3 of 7 criteria for DSM-IV dependence), which may dilute identification of genetic effects on more severe forms of the disorder. However, when we examined the association between GWS variants and a count of DSM-5 criteria across alcohol and illicit drugs, these SNPs were associated with that count but not at GWS levels (p>5E-5). Finally, 39% of the controls met criteria for substance abuse – we elected to include these individuals to maintain the sample size. Consistent with this, results for chromosome 3 (AA p=2E-8) remained GWS while those for chromosomes 1 (EA p=3E-4; AA p=6E-5), 5 (AA p=7E-8) and 13 (AA p=4E-5) were attenuated in statistical significance but not in magnitude upon exclusion of individuals with abuse from the controls. Finally, the lack of replication was discouraging, although it is noteworthy that sample sizes for the replication cohorts were modest, and they may not have had sufficient power to replicate findings with these effect sizes. Nonetheless, upon meta-analysis, these variants did not show consistent genome-wide support indicating considerable heterogeneity across-samples, low power, or raising the possibility that the current findings are false positives.

In conclusion, we leveraged the high degree of comorbid substance dependence in COGA to identify novel loci that may confer risk for both alcohol and drug dependence and parse them from those variants that relate more specifically to alcohol dependence liability (Lai et al., accompanying paper). Our results provide preliminary evidence for ancestrally-specific effects of loci that undergird addiction to alcohol and illicit drugs. Further, we find preliminary ancestry-specific evidence that GWS loci associated with dependence liability are also associated with

reward-related VS response providing a compelling putative neural mechanism through which genetic risk might influence dependence liability. Notably, large scale GWAS of psychiatric disorders, with the exception of substance use disorders, have traditionally focused on populations of European ancestry. While the genetics of substance use disorders has been examined in AAs (e.g., [81,82]), sample sizes remain fairly modest, especially given the potential expectation of a higher burden of multiple testing. To delineate the role of genetic influences on substance use disorders in such minority populations, who also may further suffer due to restricted access to treatments, targeted data collection is needed.

References

1. Peacock A, Leung J, Larney S, et al. Global statistics on alcohol, tobacco and illicit drug use: 2017 status report. *Addiction (Abingdon, England)*. 2018;113(10):1905-1926.

2. Grant BF, Saha TD, Ruan WJ, et al. Epidemiology of DSM-5 Drug Use Disorder: Results From the National Epidemiologic Survey on Alcohol and Related Conditions-III. *JAMA psychiatry*. 2016;73(1):39-47.

3. Hasin DS, Stinson FS, Ogburn E, Grant BF. 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

4. Degenhardt L, Hall W. Extent of illicit drug use and dependence, and their contribution to the global burden of disease. *Lancet (London, England)*. 2012;379(9810):55-70.

5. Stanaway JD, Afshin A, Gakidou E, et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet.* 2018;392(10159):1923-1994.

6. Kendler KS, Myers J, Prescott CA. Specificity of genetic and environmental risk factors for symptoms of cannabis, cocaine, alcohol, caffeine, and nicotine dependence

7. Vrieze SI, Feng S, Miller MB, et al. Rare nonsynonymous exonic variants in addiction and behavioral disinhibition. *Biol Psychiatry*. 2014;75(10):783-789.

8. Kendler KS, Jacobson KC, Prescott CA, Neale MC. Specificity of genetic and environmental risk factors for use and abuse/dependence of cannabis, cocaine, hallucinogens, sedatives, stimulants, and opiates in male twins. *Am J Psychiatry*. 2003;160(4):687-695.

9. Tsuang MT, Bar JL, Harley RM, Lyons MJ. The Harvard Twin Study of Substance Abuse: what we have learned. *Harv Rev Psychiatry.* 2001;9(6):267-279.

10. Verhulst B, Neale MC, Kendler KS. The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. *Psychol Med.* 2015;45(5):1061-1072.

11. Pasman JA, Verweij KJH, Gerring Z, et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat Neurosci.* 2018;21(9):1161-1170.

12. Walters RK, Polimanti R, Johnson EC, et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat Neurosci.* 2018;21(12):1656-1669.

13. Hancock DB, Reginsson GW, Gaddis NC, et al. Genome-wide meta-analysis reveals common splice site acceptor variant in CHRNA4 associated with nicotine dependence. *Transl Psychiatry.* 2015;5:e651. doi: 10.1038/tp.2015.149.:e651.

14. Sherva R, Wang Q, Kranzler H, et al. Genome-wide Association Study of Cannabis Dependence Severity, Novel Risk Variants, and Shared Genetic Risks. *JAMA Psychiatry.* 2016:10.

15. Agrawal A, Chou YL, Carey C, et al. Genomewide association study identifies a novel locus for cannabis dependence. *Mol Psychiatry.* 2016;revise and resubmit.

16. Nelson EC, Agrawal A, Heath AC, et al. Evidence of CNIH3 involvement in opioid dependence. *Mol Psychiatry.* 2015:10.

17. Martin AR, Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., Daly, M. Hidden 'risk" in polygenic scores: clinical use today could exacerbate health disparities. *biorxivorg (preprint).* 2018;https://www.biorxiv.org/content/early/2018/10/11/441261.

18. Need AC, Goldstein DB. Next generation disparities in human genomics: concerns and remedies. *Trends Genet.* 2009;25(11):489-494.

19. Duncan LE, Ratanatharathorn A, Aiello AE, et al. Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. *Molecular psychiatry.* 2018;23(3):666-673.

20. Marquez-Luna C, Loh PR, South Asian Type 2 Diabetes C, Consortium STD, Price AL. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet Epidemiol.* 2017;41(8):811-823.

21. Luczak SE, Khoddam R, Yu S, Wall TL, Schwartz A, Sussman S. A review of the prevalence and co-occurrence of addictions in US ethnic/racial groups: Implications for genetic research. *The American Journal on Addictions.* 2016:n/a-n/a.

22. Falk D, Yi HY, Hiller-Sturmhofel S. An epidemiologic analysis of co-occurring alcohol and drug use and disorders: findings from the National Epidemiologic Survey of Alcohol and Related Conditions (NESARC). *Alcohol research & health : the journal of the National Institute on Alcohol Abuse and Alcoholism.* 2008;31(2):100-110.

23. Wetherill L, Agrawal A, Kapoor M, et al. Association of substance dependence phenotypes in the COGA sample. *Addiction biology.* 2015;20(3):617-627.

24. Bierut LJ, Agrawal A, Bucholz KK, et al. A genome-wide association study of alcohol dependence. *Proceedings of the National Academy of Sciences of the United States of America.* 2010;107(11):5082-5087.

25. Gelernter J, Kranzler HR, Sherva R, et al. Genome-wide association study of alcohol dependence:significant findings in African- and European-Americans including novel risk loci. *Molecular psychiatry.* 2014;19(1):41-49.

26. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nature genetics.* 2018;50(11):1593-1599.

27. Pasman JA, Verweij KJH, Gerring Z, et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nature neuroscience.* 2018;21(9):1161-1170.

28. Heitzeg MM, Cope LM, Martz ME, Hardee JE. Neuroimaging Risk Markers for Substance Abuse: Recent Findings on Inhibitory Control and Reward System Functioning. *Curr Addict Rep.* 2015;2(2):91-103.

29. Koob GF, Volkow ND. Neurobiology of addiction: a neurocircuitry analysis. *Lancet Psychiatry.* 2016;3(8):760-773.

30. Nikolova YS, Ferrell RE, Manuck SB, Hariri AR. Multilocus genetic profile for dopamine signaling predicts ventral striatum reactivity. *Neuropsychopharmacology.* 2011;36(9):1940-1947.

31. Nikolova YS, Knodt AR, Radtke SR, Hariri AR. Divergent responses of the amygdala and ventral striatum predict stress-related problem drinking in young adults: possible differential markers of affective and impulsive pathways of risk for alcohol use disorder. *Molecular psychiatry.* 2016;21(3):348-356.

32. Edenberg HJ. The collaborative study on the genetics of alcoholism: an update. *Alcohol research & health : the journal of the National Institute on Alcohol Abuse and Alcoholism.* 2002;26(3):214-218.

33. Reich T, Edenberg HJ, Goate A, et al. Genome-wide search for genes affecting the risk for alcohol dependence. *Am J Med Genet.* 1998;81(3):207-215.

34. Bucholz KK, Cadoret R, Cloninger CR, et al. 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.* 1994;55(2):149-158.

35. Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V. A validity study of the SSAGA--a comparison with the SCAN. *Addiction (Abingdon, England).* 1999;94(9):1361-1370.

36. Association AP. *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR).* American Psychiatric Association; 2000.

37. Grucza RA, Bucholz KK, Rice JP, Bierut LJ. Secular trends in the lifetime prevalence of alcohol dependence in the United States: a re-evaluation. *Alcoholism, clinical and experimental research.* 2008;32(5):763-770.

38. Muthen LK, Muthen BO. *Mplus User's Guide.* Los Angeles CA: Muthen & Muthen; 1998-2011.

39. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *American journal of human genetics.* 1998;62(5):1198-1211.

40. Edenberg HJ, Koller DL, Xuei X, et al. Genome-wide association study of alcohol dependence implicates a region on chromosome 11. *Alcoholism, clinical and experimental research.* 2010;34(5):840-852.

41. Meyers JL, Zhang J, Wang JC, et al. An endophenotype approach to the genetics of alcohol dependence: a genome wide association study of fast beta EEG in families of African ancestry. *Molecular psychiatry.* 2017.

42. Wetherill L, Kapoor M, Agrawal A, et al. Family-based association analysis of alcohol dependence criteria and severity. *Alcoholism, clinical and experimental research.* 2014;38(2):354-366.

43. Delaneau O, Howie B, Cox AJ, Zagury JF, Marchini J. Haplotype estimation using sequencing reads. *American journal of human genetics.* 2013;93(4):687-696.

44. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. *Nature genetics.* 2016;48(10):1284-1287.

45. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *American journal of human genetics.* 1998;63(1):259-266.

46. Chen MH, Yang Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics (Oxford, England).* 2010;26(4):580-581.

47. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England).* 2010;26(17):2190-2191.

48. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nature communications.* 2017;8(1):1826.

49. Gamazon ER, Segre AV, van de Bunt M, et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nature genetics.* 2018;50(7):956-967.

50. Barbeira AN, Dickinson SP, Bonazzola R, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nature communications.* 2018;9(1):1825.

51. Zhou H, Polimanti R, Yang BZ, et al. Genetic Risk Variants Associated With Comorbid Alcohol Dependence and Major Depression. *JAMA psychiatry.* 2017;74(12):1234-1241.

52. Sherva R, Wang Q, Kranzler H, et al. Genome-wide Association Study of Cannabis Dependence Severity, Novel Risk Variants, and Shared Genetic Risks. *JAMA psychiatry.* 2016;73(5):472-480.

53. Delgado MR, Miller MM, Inati S, Phelps EA. An fMRI study of reward-related probability learning. *NeuroImage.* 2005;24(3):862-873.

54. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics (Oxford, England).* 2005;21(2):263-265.

55. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome research.* 2012;22(9):1790-1797.

56. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic acids research.* 2018.

57. Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. *Nature.* 2017;550(7675):204-213.

58. Bierut LJ, Dinwiddie SH, Begleiter H, et al. Familial transmission of substance dependence: alcohol, marijuana, cocaine, and habitual smoking: a report from the Collaborative Study on the Genetics of Alcoholism. *Archives of general psychiatry.* 1998;55(11):982-988.

59. Merikangas KR, Stolar M, Stevens DE, et al. Familial transmission of substance use disorders. *Archives of general psychiatry.* 1998;55(11):973-979.

60. Sartor CE, Agrawal A, Lynskey MT, et al. Cannabis or alcohol first? Differences by ethnicity and in risk for rapid progression to cannabis-related problems in women. *Psychological medicine.* 2013;43(4):813-823.

61. Wu L-T, Zhu H, Swartz MS. Trends in cannabis use disorders among racial/ethnic population groups in the United States. *Drug and alcohol dependence.* 2016;165:181-190.

62. Lopez-Quintero C, Perez de los Cobos J, Hasin DS, et al. Probability and predictors of transition from first use to dependence on nicotine, alcohol, cannabis, and cocaine: results of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *Drug and alcohol dependence.* 2011;115(1-2):120-130.

63. Dick DM, Barr P, Guy M, Nasim A, Scott D. (Invited 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 / American Academy of Psychiatrists in Alcoholism and Addictions.* 2017.

64. Kanai M, Tanaka T, Okada Y. Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. *Journal of human genetics.* 2016;61(10):861-866.

65. Sobota RS, Shriner D, Kodaman N, et al. Addressing population-specific multiple testing burdens in genetic association studies. *Annals of human genetics.* 2015;79(2):136-147.

66. Biernacka JM, Geske J, Jenkins GD, et al. Genome-wide gene-set analysis for identification of pathways associated with alcohol dependence. *The international journal of neuropsychopharmacology.* 2013;16(2):271-278.

67. Han S, Yang BZ, Kranzler HR, et al. Integrating GWASs and human protein interaction networks identifies a gene subnetwork underlying alcohol dependence. *American journal of human genetics.* 2013;93(6):1027-1034.

68. Volkow ND, Wang G-J, Fowler JS, Tomasi D, Telang F, Baler R. Addiction: decreased reward sensitivity and increased expectation sensitivity conspire to overwhelm the brain's control circuit. *BioEssays : news and reviews in molecular, cellular and developmental biology.* 2010;32(9):748-755.

69. Spear LP. The adolescent brain and age-related behavioral manifestations. *Neuroscience and biobehavioral reviews.* 2000;24(4):417-463.

70. Andrews MM, Meda SA, Thomas AD, et al. Individuals family history positive for alcoholism show functional magnetic resonance imaging differences in reward sensitivity that are related to impulsivity factors. *Biological psychiatry.* 2011;69(7):675-683.

71. Büchel C, Peters J, Banaschewski T, et al. Blunted ventral striatal responses to anticipated rewards foreshadow problematic drug use in novelty-seeking adolescents. *Nature communications.* 2017;8:14140.

72. Schuckit MA. Low level of response to alcohol as a predictor of future alcoholism. *The American journal of psychiatry.* 1994;151(2):184-189.

73. Carey CE, Knodt AR, Conley ED, Hariri AR, Bogdan R. Reward-related ventral striatum activity links polygenic risk for attention-deficit/hyperactivity disorder to problematic alcohol use in young adulthood. *Biological psychiatry Cognitive neuroscience and neuroimaging.* 2017;2(2):180-187.

74. Hariri AR, Brown SM, Williamson DE, Flory JD, de Wit H, Manuck SB. Preference for immediate over delayed rewards is associated with magnitude of ventral striatal activity. *The Journal of neuroscience : the official journal of the Society for Neuroscience.* 2006;26(51):13213-13217.

75. Blum K, Braverman ER, Holder JM, et al. Reward deficiency syndrome: a biogenetic model for the diagnosis and treatment of impulsive, addictive, and compulsive behaviors. *Journal of psychoactive drugs.* 2000;32 Suppl:i-iv, 1-112.

76. Martin-Soelch C, Szczepanik J, Nugent A, et al. Lateralization and gender differences in the dopaminergic response to unpredictable reward in the human ventral striatum. *The European journal of neuroscience.* 2011;33(9):1706-1715.

77. Pizzagalli DA, Holmes AJ, Dillon DG, et al. Reduced caudate and nucleus accumbens response to rewards in unmedicated individuals with major depressive disorder. *The American journal of psychiatry.* 2009;166(6):702-710.

78. Johnson J AG. GAS Power Calculator: web-based power calculator for genetic association studies. *https://wwwbiorxivorg/content/101101/164343v1*. 2017.

79. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88(1):76-82.

80. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47(3):291-295.

81. Gelernter J, Kranzler HR, Sherva R, et al. Genome-wide association study of nicotine dependence in American populations: identification of novel risk loci in both African-Americans and European-Americans. *Biological psychiatry.* 2015;77(5):493-503.

82. Gelernter J, Sherva R, Koesterer R, et al. Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. *Molecular psychiatry.* 2014;19(6):717-723.

Acknowledgements

COGA: 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, B. Cho (Virginia Commonwealth University); and Mark Kos (University of Texas Rio Grande Valley). A. Parsian and H. Chen 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).

The Duke Neurogenetics Study is supported by Duke University and the National Institutes of Health (NIDA DA033369).

This research is also supported by K02DA32573 (AA), MH109532 & DA040411 (AA, ECJ).

Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268201200008I.

A. Agrawal receives additional funding support from NIDA (DA032573).

A.P. Anokhin receives additional funding support from NIH (HD083614, DA040716, DA038834).

R. Bogdan receives additional funding support from NIH (AG052564, HD083614, AG045231).

D. Dick receives additional funding from K02AA018755.

A. Hariri receives additional funding support from NIH (Grant Nos. R01-DA031579 and R01-AG049789).

J. Meyers receives additional funding support from NIDA (K01DA037914).

Conflict of Interest:

Alison Goate is listed as an inventor on Issued U.S. Patent 8080,371, "Markers for Addiction" covering the use of certain variants in determining the diagnosis, prognosis, and treatment of addiction.

John Nurnberger receives support as an investigator for Janssen.

No other authors report biomedical financial interests or potential conflicts of interest.

Author Contributions: LW, DL, ECJ, RB and AAgrawal conducted analyses, drafted and revised the manuscript. TF, RB and AAgrawal supervised all aspects of the study. TF and AMG provided oversight with genomic analyses and interpretation. ARH collected the DNS sample. KKB, DMD, VH, CK, JK, SK, JLM, JIN, MS, DMS, RET, JT, BP and HJE provided feedback on development of phenotypes and in statistical analyses. AAnokhin, LB, ARH, JLM and BP provided support with the neural extension. All authors reviewed and approved the submission.

LEGENDS

Figure 1: Manhattan plots for the (PANEL **a**) European-American (EA), (Panel **b**) African-American (AA), and (PANEL **c**) trans-ancestral (EA+AA).

Figure 2: Regional association plots for genome-wide significant loci on (**a**) chromosome 3 and (**b**) chromosome 13 in analyses of AA families

Figure 3: Regional association plots for genome-wide significant loci on chromosome 1 in the trans-ancestral (EA+AA) analysis shown with (**a**) LD based on CEU population, and (**b**) LD based on AFR population

FIGURE 1.

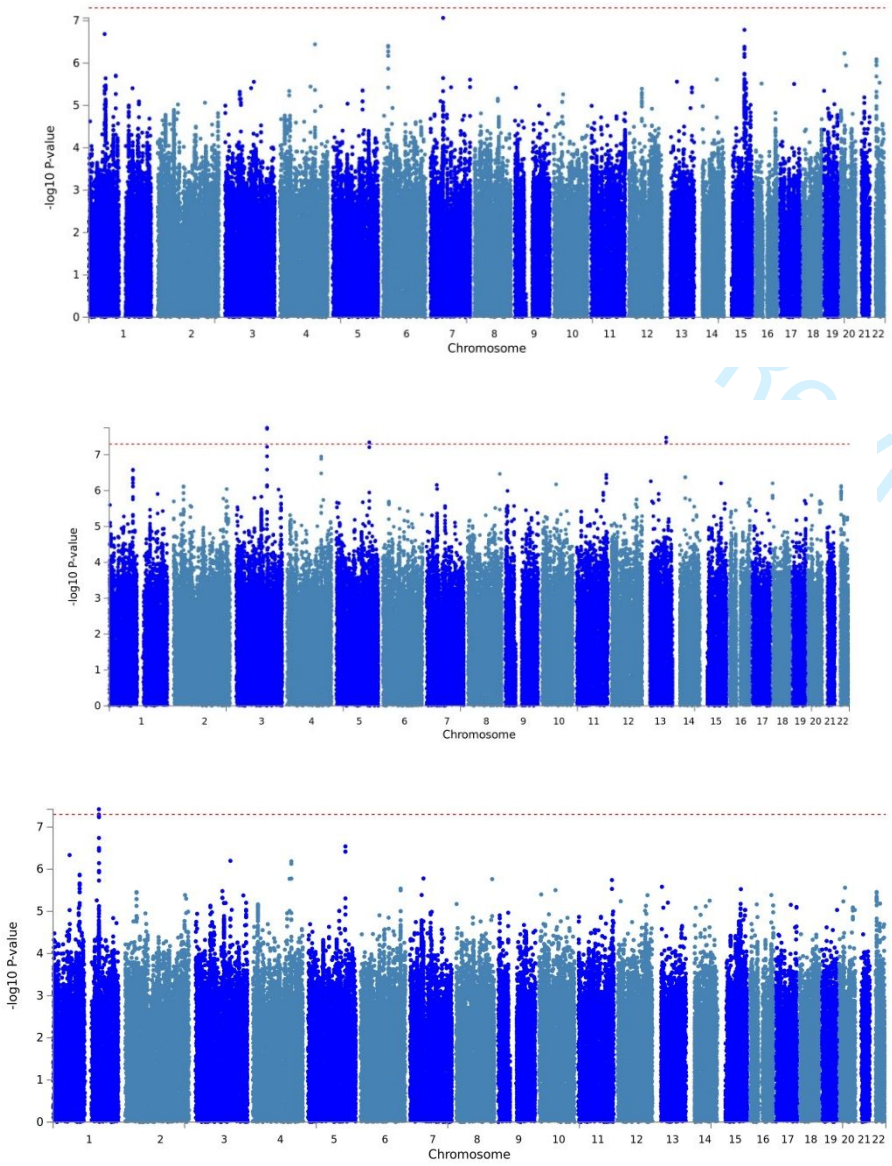


FIGURE 2a:

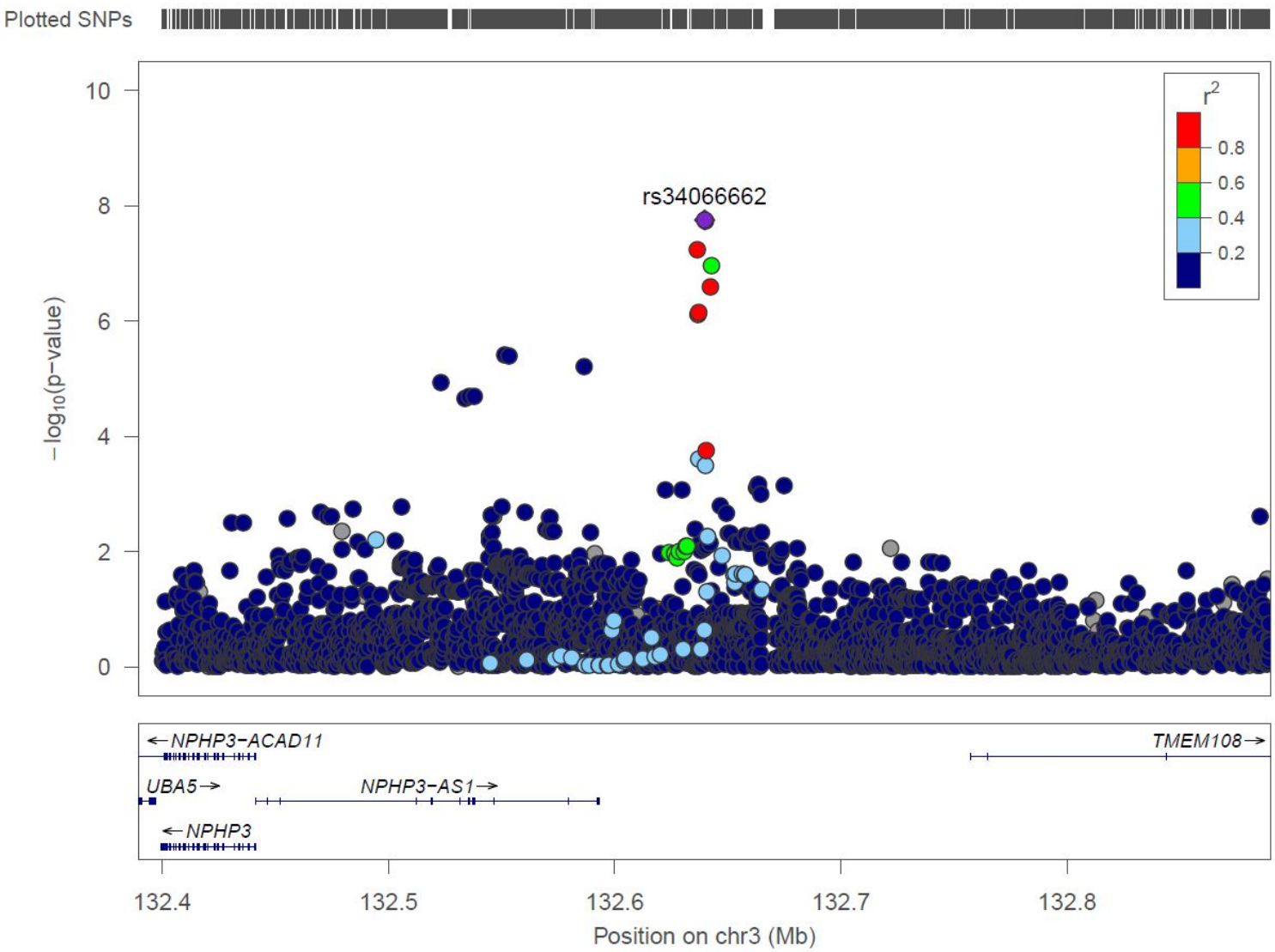

Plotted SNPs
rs34066662
r2
0.8
0.6
0.4
0.2
−log10(p−value)
10
8
6
4
2
0
←NPHP3−ACAD11
TMEM108→
UBA5→
NPHP3−AS1→
←NPHP3
132.4
132.5
132.6
132.7
132.8
Position on chr3 (Mb)

FIGURE 2b:

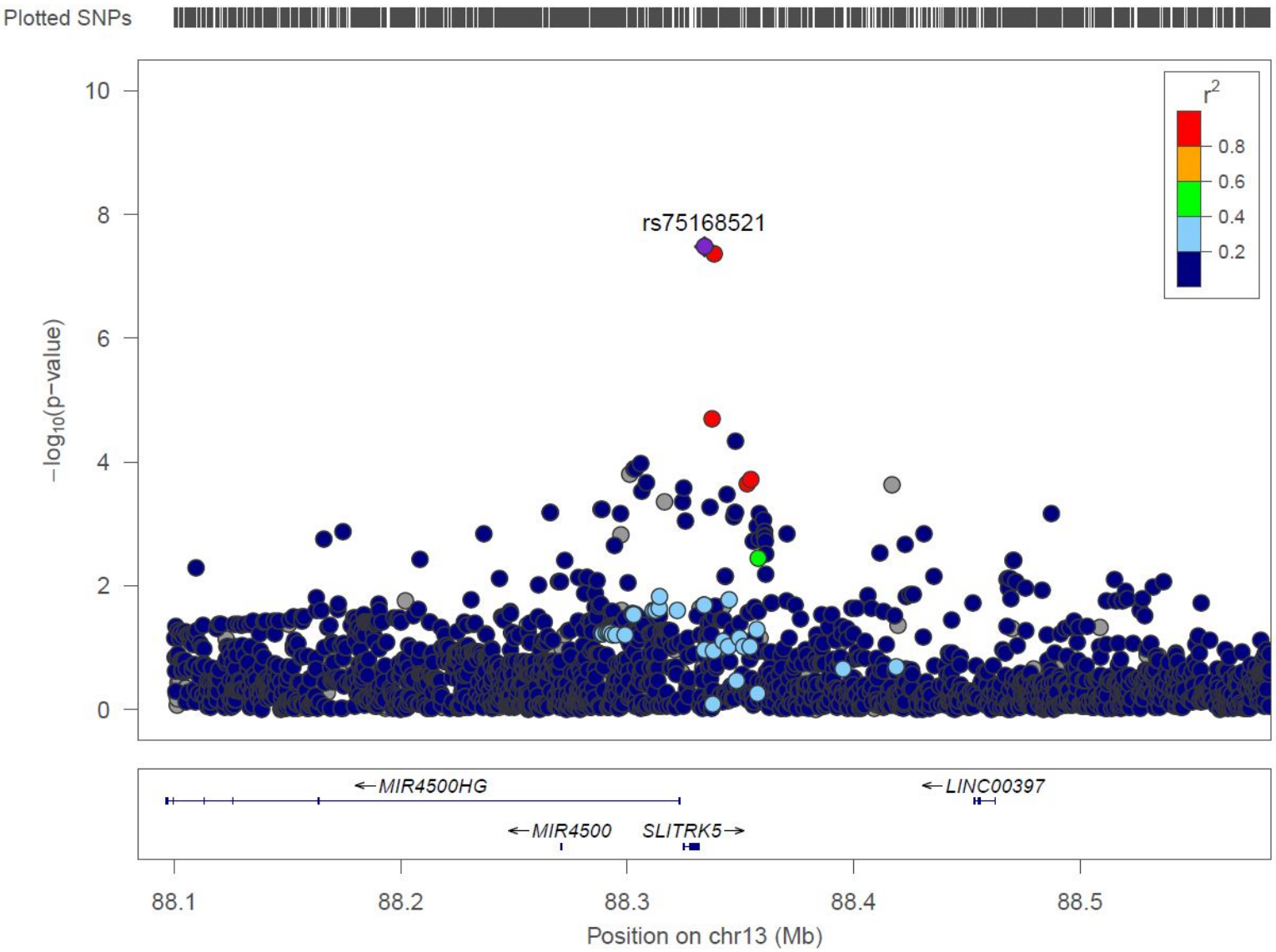

Plotted SNPs
rs75168521
r2
0.8
0.6
0.4
0.2
−log10(p−value)
10
8
6
4
2
0
MIR4500HG
LINC00397
MIR4500
SLITRK5
88.1
88.2
88.3
88.4
88.5
Position on chr13 (Mb)

FIGURE 3a:

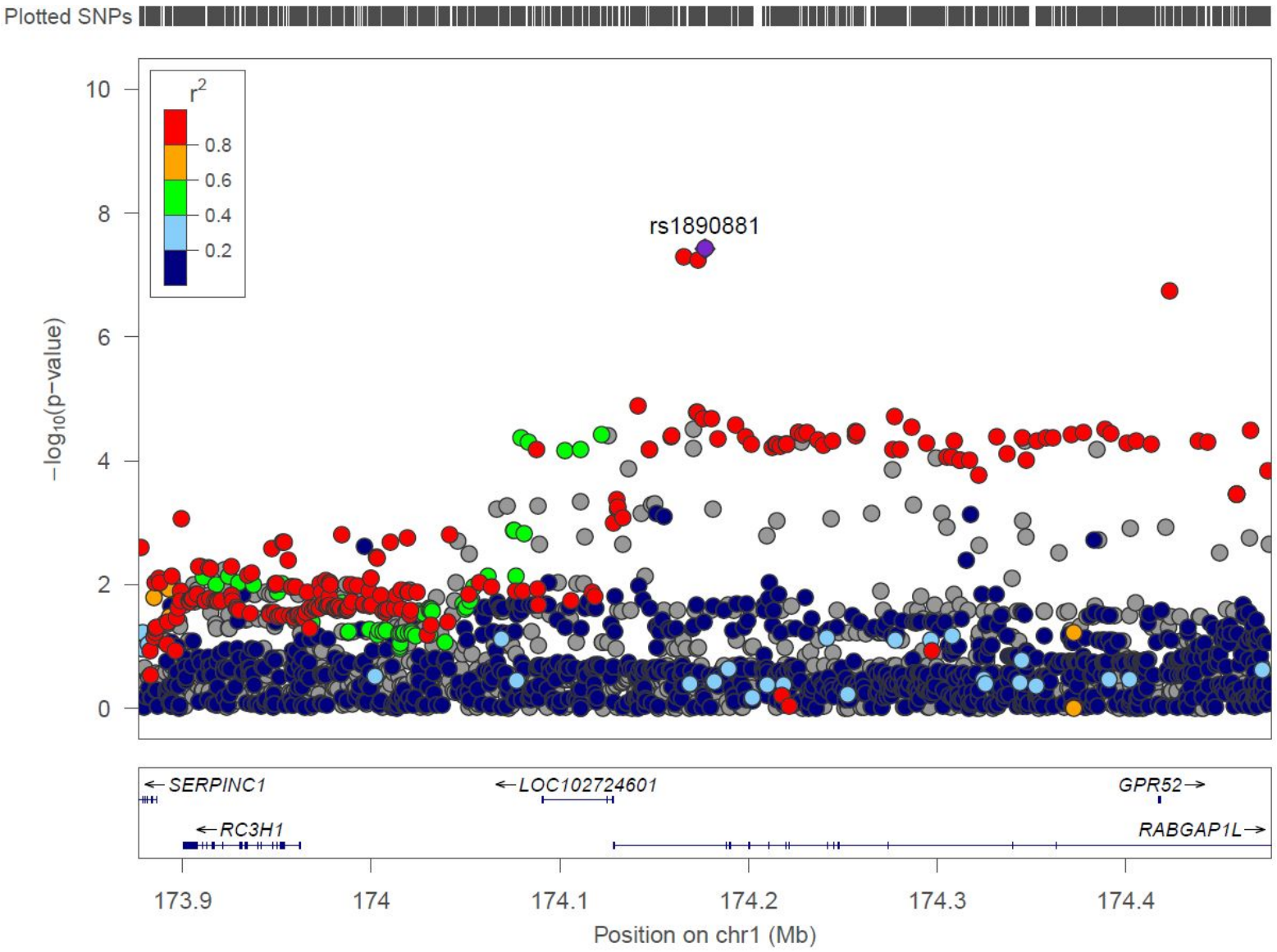

Plotted SNPs
r2
0.8
0.6
0.4
0.2
rs1890881
−log10(p−value)
10
8
6
4
2
0
SERPINC1
LOC102724601
GPR52
RC3H1
RABGAP1L
173.9
174
174.1
174.2
174.3
174.4
Position on chr1 (Mb)

FIGURE 3b:

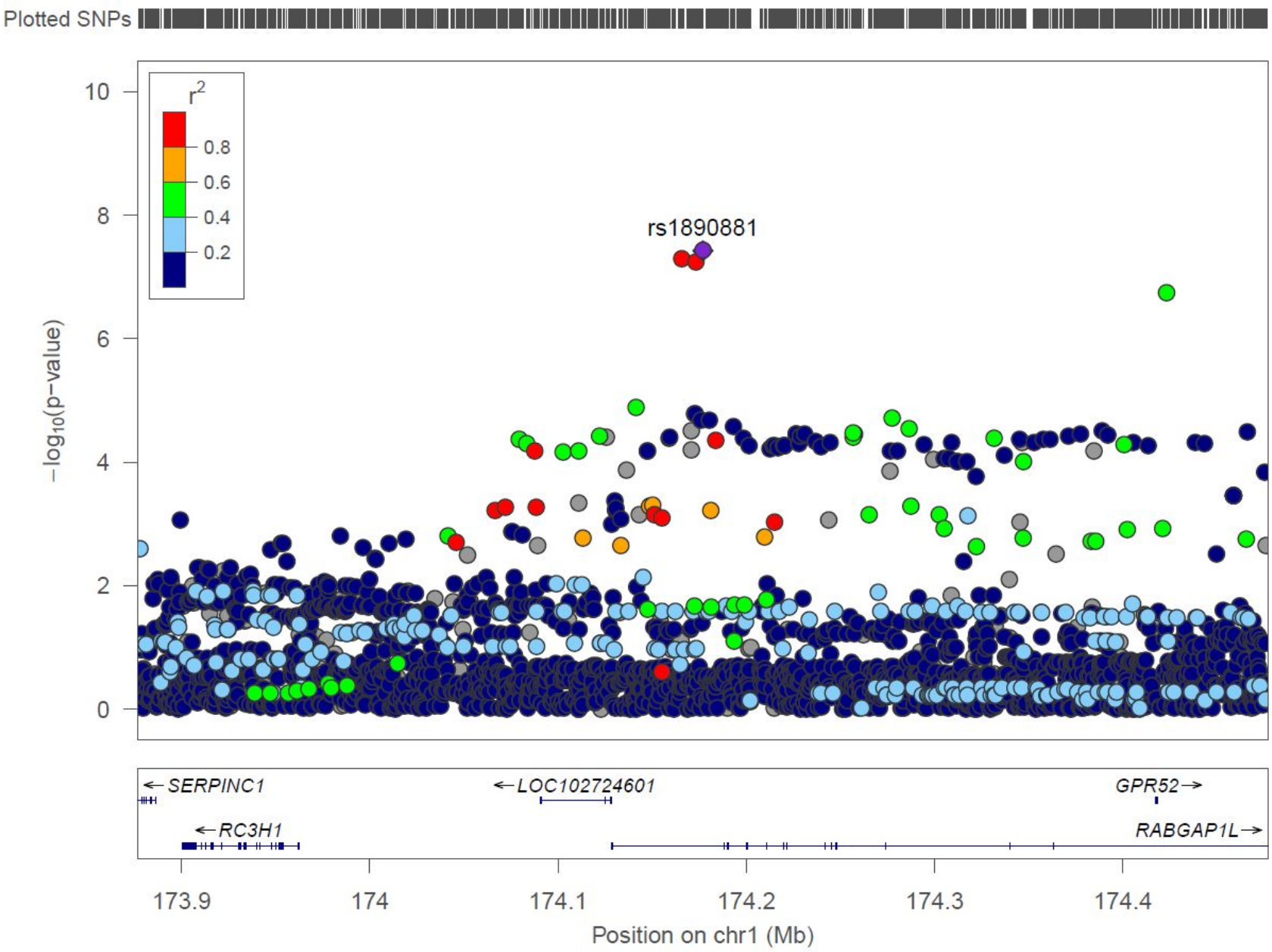


Table 1. Sample size for primary analysis of

dependence on alcohol or any illicit drug (ANYDEP), as well as for post-hoc analyses of top loci for alcohol and individual drug phenotypes.

	European-American (EA)		**African-American (AA)**	
	Cases (#)	Controls (#)	Cases (#)	Controls (#)
ANYDEP	2,927	4,364	1,315	1,817
Alcohol	2,351	4,364	901	1.817
Cannabis	1,228	4,364	667	1,817
Cocaine	765	4,364	581	1,817
Sedatives	267	4,364	31	1,817
Stimulants	530	4,364	53	1,817
Opioids	334	4,364	142	1,817
drug_noalc	563	4,364	422	1,817

Table 2: Results for most significant SNPs associated with *ANYDEP*.

	Chr:bp	Effect allele: alternate allele	Effect allele frequency (EA:AA)	EA: beta, SE, p-value	AA beta, SE, p-value	Trans-ancestral (EA+AA) beta, SE, p-value
rs34066662	3:132639776	T:C	0.06:0.19	0.094 (0.082), 0.249	-0.398 (0.071), 1.77E-08	-0.181 (0.056), 1.33E-03
rs58801820	3:132640091	T:G	0.06:0.19	0.094 (0.082), 0.249	-0.393 (0.082), 1.89E-08	-0.180 (0.056), 1.28E-03
rs75168521	13:88334193	T:C	0.91:0.85	-0.037 (0.073), 0.615	-0.428 (0.078), 3.31E-08	-0.217 (0.056), 1.11E-04
rs78886294	13:88338399	T:C	0.91:0.85	-0.035 (0.073), 0.626	-0.426 (0.078), 4.38E-08	-0.212 (0.056), 1.41E-04
rs527904740	5:141988181	GAA:GAAA	NA:0.95	-	-0.694 (0.127), 4.48E-08	-
rs1890881	1:174176923	T:C	0.93:0.85	0.327 (0.083), 8.95E-05	0.350 (0.082), 1.94E-05	0.339 (0.062), 3.77E-08
rs74611272	7:51850533	T:C	0.96:0.99	-0.520 (0.097), 8.6E-08	0.118 (0.292), 0.687	-0.50 (0.096), 1.65E-06

Legend: Covariates included sex, birth cohorts, ancestral principal components, array;

Chr = chromosome; Bp = base pair; EA = European American; AA = African American; SE = standard error;

Table 3. Results for individual drugs for top variants associated with *ANYDEP*.

	Chr:bp	Alcohol Beta, SE, p-value	Cannabis beta, SE, p-value	Cocaine beta, SE, p-value	Stimulant beta, SE, p-value	Sedative beta, SE, p-value	Opioid beta, SE, p-value
In African American (AA) GWAS							
rs34066662	3:132639776	-0.396 (0.08), 1.70E-06	-0.347 (0.087), 8.05E-05	-0.305 (0.091), 8.24E-04	-0.612 (0.313), 0.051	-1.40 (0.534), 0.009	-0.733 (0.227), 0.0013
rs58801820	3:132640091	-0.393 (0.08), 1.74E-06	-0.343 (0.087), 8.19E-05	-0.303 (0.091), 8.55E-04	-0.586 (0.313), 0.051	-1.36 (0.535), 0.011	-0.734 (0.228), 0.0013
rs527904740	5: 141988181	-0.730 (0.144), 4.24E-07	-0.645 (0.154), 2.90E-05	-0.507 (0.184), 5.88E-03	-0.852 (0.414), 0.04	-0.759 (0.558), 0.17	-0.516 (0.321), 0.11
rs75168521	13:88334193	-0.403 (0.09), 4.08E-06	-0.370 (0.091), 4.49E-05	-0.281 (0.112), 0.012	-0.593 (0.254), 0.019	-0.562 (0.313), 0.07	-0.154 (0.208), 0.459
rs78886294	13:88338399	-0.400 (0.09), 4.57E-06	-0.368 (0.090), 4.71E-05	-0.287 (0.110), 0.009	-0.581 (0.252),0.021	-0.547 (0.313), 0.08	-0.180 (0.204), 0.380
rs1890881	1:174176923	0.366 (0.09), 5.36E-05	0.171 (0.111), 0.125	0.330 (0.105), 1.6E-03	-0.450 (0.245), 0.065	0.167 (0.342), 0.626	0.161 (0.189), 0.393
In European American (EA) GWAS							

rs1890881	1:174176923	0.184 (0.16), 0.26	0.245 (0.110), 0.026	0.344 (0.130), 0.008	0.197 (0.140), 0.156	0.222 (0.185),0.231	0.100 (0.169), 0.357
rs74611272	7:51850533	-0.073 (0.018), 5.10E-05	-0.597 (0.122), 9.59E-07	-0.492 (0.152), 1.24E-03	-0.411 (0.167), 0.014	-0.463 (0.258), 0.07	-0.513 (0.223), 0.022

Legend: Covariates included sex, birth cohorts, ancestral principal components, array; for some SNPs and individual drugs, statistical convergence issues arose when including birth cohorts, and thus, age was substituted for birth cohort.
Chr = chromosome; Bp = base pair; SE = standard error; GWAS = genome-wide association study; SNP = single nucleotide polymorphism

Table 4: Results for alcohol dependence and for *drug_noalc* (alcohol dependent individuals excluded from cases and controls) for top SNPs.

	Chr:bp	Effect allele: alternate allele	Alcohol dependence (Lai et al*)	*drug_noalc*
African-American (AA) GWAS				
rs34066662	3:132639776	T:C	-0.454 (0.092), 8.92E-07	-0.396 (0.110), 3.14E-04
rs58801820	3:132640091	T:G	-0.451 (0.092), 9.00E-07	-0.393 (0.109), 3.01E-04
rs527904740	5:141988181	GAA: GAAA	-0.756 (0.756), 2.65E-05	-0.628 (0.169), 2.06E-04
rs75168521	13:88334193	T:C	-0.486 (0.112), 1.56E-05	-0.371 (0.107), 4.88E-04
rs78886294	13:88338399	T:C	-0.484 (0.113), 1.77E-05	-0.368 (0.106), 5.04E-04
rs1890881	1:174176923	T:C	0.554 (0.121), 4.70E-06	0.235 (0.116), 0.04
European-American (EA) GWAS				
rs1890881	1:174176923	T:C	0.471 (0.108), 1.29E-05	0.184 (0.16), 0.26
rs74611272	7:51850533	T:C	-0.442 (0.121), 2.61E-04	-0.073 (0.018), 5.10E-05

*Refers to accompanying paper by Lai et al which examined alcohol dependent cases (DSM-IV dependent) and controls (drank at least one drink of alcohol and endorsed at most 1 criterion for alcohol or any drug dependence).
Chr = chromosome; Bp = base pair; SNP = single nucleotide polymorphism

Table 5. Associations between response of the ventral striatum to positive > negative feedback and genotype in the Duke Neurogenetic Sample

		African-Americans (AA): beta [95% C.I.], p-value		European-Americans (AA): beta [95% C.I.], p-value	
SNP	Effect allele	**RIGHT VS**	**LEFT VS**	**RIGHT VS**	**LEFT VS**
rs34066662	T	-0.073 [-0.164 - 0.019], 0.103	-0.062 [-0.143 - 0.018], 0.132	0.002 [-0.053 - 0.057], 0.932	-0.008 [-0.057 - 0.043], 0.762
rs75168521	C	-0.111 [-0.19 - -0.036], 0.004	-0.065 [-0.153 - -0.24], 0.143	0.025 [-0.017 - 0.068], 0.247	0.011 [-0.027 - 0.048], 0.565
rs1890881	T	-0.098 [-0.175 - -0.020], 0.013	-0.134 [-0.213- -0.057], 0.001	-0.004 [-0.039 - 0.030], 0.801	-0.036 [-0.068 - -0.004], 0.029

SNP = single nucleotide polymorphism; chr5:141988181 was not available.

SUPPORTING INFORMATION

Supplemental Text

COGA

COGA Genotyping: COGA samples were genotyped on arrays including the Illumina Human1M array (Illumina, San Diego, CA, USA) at the Center for Inherited Disease Research (CIDR), Johns Hopkins University[1], the Illumina Human OmniExpress 12V1 array (Illumina, San Diego, CA, USA) at the Genome Technology Access Center, Washington University School of Medicine in St. Louis[2,3], the Illumina 2.5M array (Illumina, San Diego, CA, USA) at CIDR[4], and the Smokescreen genotyping array (Biorealm LLC, Walnut, CA, USA) at Rutgers University. About 91.5% of the African-American participants were genotyped on the Illumina 2.5M array, with the remainder predominantly on the Smokescreen (save 52 individuals on the Illumina Human1M). Of the European-Americans, 59% were genotyped on the Smokescreen, 25% on the Illumina OmniExpress 12V1, and the remainder on the Illumina Human1M. Genotyping concordance for 2-127 samples genotyped on at least two different arrays was > 99.18%. As outlined in the **Main Text**, a subset of 47,000 high quality SNPs were used for quality control, pedigree checks and computation of ancestral principal components. F_{HET} was predominantly lower than 0.20, with the exception of 3 individuals with $F_{HET} < 0.25$. Pedigree structures were reconciled, and genotypes were evaluated for Mendelian errors (Pedcheck;[5]) with the revised pedigree structure when necessary. All remaining genotype inconsistencies were set to missing. SNPs or individuals with high rates of Mendelian errors were excluded. The same set of common, independent and high quality SNPs was used in EIGENSTRAT[6] to calculate principal components. Reference data from the 1000 Genomes (Phase 3, version 5) YRI, CEU, JPT and CHB populations were utilized as anchors, and individual genetic ancestry was assigned accordingly. Families were assigned a family-based ancestry of EA, AA, or Other, according to

the majority of individuals in that family, and to the more heterogeneous population when there were equal proportions. SHAPEIT version 2[7], which utilizes pedigree information, was used to impute, followed by Minimac3[8], and imputation was performed within each array using the cosmopolitan reference panel (Phase 3, version 5, http://www.internationalgenome.org/). SNPs with missing rates < 5%, MAF > 3%, and HWE p values > 0.0001 that were non A/T or C/G alleles were used for imputation.

Duke Neurogenetics Study (DNS)

Extension of Neuroimaging to Significant Loci: Neuroimaging and genetic data that were fully processed by 02/01/2016 were available from 118 non-Hispanic African American and 481 non-Hispanic European American participants who completed the Duke Neurogenetics Study (DNS)[9]. Ancestry was determined by self-report and confirmed by ancestry informative principal components (see Genotyping below; i.e., no individuals were ±6 standard deviations from the mean on the top 10 components). Each participant provided informed written consent prior to participation in accord with the guidelines of the Duke University Medical Center Institutional Review Board. All participants were in good general health and free of DNS exclusion criteria: (1) medical diagnosis of cancer, stroke, diabetes requiring insulin treatment, chronic kidney or liver disease or lifetime psychotic symptoms; (2) use of psychotropic, glucocorticoid or hypolipidemic medication, and (3) conditions affecting cerebral blood flow and metabolism (e.g., hypertension). Current DSM-IV Axis I and select Axis II disorders (Antisocial Personality Disorder and Borderline Personality Disorder) were assessed with the electronic Mini International Neuropsychiatric Interview[10] and Structured Clinical Interview for the DSM-IV Axis II (structured clinical interview for DSM disorders-II)[11] (First et al., 1997). These disorders

were not exclusionary as the DNS sought to establish broad variability in multiple behavioral phenotypes related to psychopathology.

Genotyping

DNA was isolated from saliva derived from Oragene DNA self-collection kits (DNA Genotek) customized for 23andMe (https://www.23andme.com/). DNA extraction and genotyping were performed by the National Genetics Institute (NGI), a Clinical Laboratory Improvement Amendments-certified clinical laboratory and subsidiary of Laboratory Corporation of America. One of two Illumina arrays with custom content was used to provide genome-wide genotyping data for each participant: the HumanOmniExpress or HumanOmniExpress-24[12].

Genotype imputation was run separately for all DNS participants with genome-wide chip data using the pre-phasing/imputation stepwise approach implemented in SHAPEIT[7] and IMPUTE2[13] using only biallelic variants and the default value for effective size of the population (20,000), and chunk sizes of 3 megabase (Mb) and 5Mb for the respective arrays. Within each array batch, genotyped variants used for imputation were required to have missingness<0.02, Hardy-Weinberg equilibrium p>10-6, and MAF>0.01. The imputation reference set consisted of 2,504 phased haplotypes from the full 1000 Genomes Project Phase 3 dataset (May 2013, over 70 million variants, release "v5a").

Ten principal components (PCs) were generated using identity-by-state analysis in PLINK[14] of whole-genome variants within the AA and EA subsamples independently as well as in the combined sample. The number of PCs to include as covariates (i.e., AA: PC1-PC3; EA: PC1-PC2; combined sample PC1-PC4) in analyses was determined by examining the amount of

variance captured by each component and selecting the components occurring prior to the point of inflection on the scree plot.

Duke Neurogenetics Study Methods

Ventral Striatum Reactivity Paradigm. As previously described[15], a blocked-design number-guessing paradigm consisting of a pseudorandom presentation of three blocks each of predominantly positive (80% correct guess) or negative (20% correct guesses) feedback linked to monetary gains and losses, respectively, was used to probe ventral striatum activity associated with gains and losses. These blocks were interleaved with three control blocks. Participants were led to believe that their performance would determine their monetary gain at the end of scanning sessions. However, outcome probabilities were fixed, and all participants received $10 in winnings regardless of performance.

Each feedback block was composed of 5 trials. Each feedback trail began with the presentation of the back of a card and participants were given 3 seconds to guess, via button press, whether the value of the card is lower (right index finger) or higher (right middle finger) than 5 (face cards exclude). The numerical value of the card was then presented for 500 milliseconds, followed by appropriate feedback (green upward-facing arrow for positive feedback; red downward-facing arrow for negative feedback) for an additional 500 milliseconds. A crosshair was then presented for 3 seconds, for a total trial length of 7 seconds. During control blocks, participants saw an "x" for 3 seconds, during which they were instructed to push a button, which was followed by an asterisk (500 milliseconds) and a yellow circle (500 milliseconds). Each block was preceded by an instruction of "Guess Number" (positive or

negative feedback blocks) or "Press Button" (control blocks) for 2 seconds resulting in a total block length of 38 seconds and a total task length of 342 seconds.

Blood-oxygen-level dependent (BOLD) functional magnetic resonance imaging (fMRI) Data Acquisition. Each participant was scanned using one of two identical research-dedicated GE MR750 3 T scanner equipped with high-power high-duty-cycle 50-mT/m gradients at 200 T/m/s slew rate, and an eight-channel head coil for parallel imaging at high bandwidth up to 1MHz at the Duke-UNC Brain Imaging and Analysis Center. A semi-automated high-order shimming program was used to ensure global field homogeneity. A series of 34 interleaved axial functional slices aligned with the anterior commissure-posterior commissure plane were acquired for full-brain coverage using an inverse-spiral pulse sequence to reduce susceptibility artifacts (TR/TE/flip angle=2000 ms/30 ms/60; FOV=240mm; 3.75×3.75×4mm voxels; interslice skip=0). Four initial radiofrequency excitations were performed (and discarded) to achieve steady-state equilibrium. To allow for spatial registration of each participant's data to a standard coordinate system, high-resolution three-dimensional structural images were acquired in 34 axial slices coplanar with the functional scans (TR/TE/flip angle=7.7 s/3.0 ms/12; voxel size=0.9×0.9×4mm; FOV=240mm, interslice skip=0).

BOLD fMRI Data Pre-Processing. Preprocessing was conducted using Statistical Parametric Mapping 8 (SPM8; www.fil.ion.ucl.ac.uk/spm). Images for each subject were realigned to the first volume in the time series to correct for head motion, spatially normalized into a standard stereotactic space (Montreal Neurological Institute template) using a 12-parameter affine model (final resolution of functional images=2mm isotropic voxels), and smoothed to minimize noise

and residual difference in gyral anatomy with a Gaussian filter, set at 6-mm full-width at half-maximum. Voxel-wise signal intensities were ratio normalized to the whole-brain global mean. Variability in single-subject whole-brain functional volumes was determined using the Artifact Recognition Toolbox (http://www.nitrc.org/projects/artifact_detect). Individual whole-brain BOLD fMRI volumes meeting at least one of two criteria were assigned a lower weight in determination of task-specific effects: (1) significant mean-volume signal intensity variation (i.e. within volume mean signal greater or less than 4 standard deviation of mean signal of all volumes in time series) and (2) individual volumes where scan-to-scan movement exceeded 2mm translation or 2° rotation in any direction.

fMRI Quality Assurance Criteria. Quality control criteria for inclusion of a participant's imaging data were: <5% volumes exceed artifact detection criteria for motion or signal intensity outliers and ≥90% coverage of signal within 5mm bilateral ventral striatum spheres centered at [±12, 10, -10]. Participants were excluded from imaging analyses for a large number of movement outliers in fMRI data (n=30), insufficient task engagement (n=8), scanner-related artifacts (n=2), equipment malfunction (n=3), incidental structural brain abnormalities (n=2), study non-completion or incomplete data (n=3), and falling asleep (n=1).

BOLD fMRI Analysis Procedures. Consistent with prior studies, we focused on ventral striatum activity resulting from the contrast of all positive feedback blocks relative to all negative feedbac blocks (i.e., Positive feedback > Negative feedback) as an index of reward-related ventral striatum reactivity[9,16,17]. Following preprocessing, linear contrasts employing canonical hemodynamic response functions were used to estimate the differential effects of feedback (i.e., reward) from the contrast of Positive Feedback > Negative Feedback for each individual using

SPM8. Parameter estimates were extracted from the top voxels (Montreal Neurological Institute (MNI) coordinates=[-12, 8, -8], t=14.802, p<0.05 family-wise error (FWE), ke=10) and right (MNI coordinates=[12, 10, -8], t=14.067, p<0.05 FWE) within a left and right ventral striatum region of interest (a 10 mm sphere centered on MNI x=±12 y=12 z=-10) defined from a prior study[17]. These parameter estimates were subsequently used for all statistical analyses. To maintain variability but constrain the influence of extreme outliers, prior to analyses all variables were Winsorized (i.e., outliers more than mean ± 3 standard deviations were set at ± 3 standard deviations from the mean).

1. Edenberg HJ, Koller DL, Xuei X, et al. Genome-wide association study of alcohol dependence implicates a region on chromosome 11. *Alcoholism, clinical and experimental research.* 2010;34(5):840-852.

2. Wetherill L, Agrawal A, Kapoor M, et al. Association of substance dependence phenotypes in the COGA sample. *Addiction biology.* 2015;20(3):617-627.

3. Wetherill L, Kapoor M, Agrawal A, et al. Family-based association analysis of alcohol dependence criteria and severity. *Alcoholism, clinical and experimental research.* 2014;38(2):354-366.

4. Meyers JL, Zhang J, Wang JC, et al. An endophenotype approach to the genetics of alcohol dependence: a genome wide association study of fast beta EEG in families of African ancestry. *Molecular psychiatry.* 2017.

5. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *American journal of human genetics.* 1998;63(1):259-266.

6. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics.* 2006;38(8):904-909.

7. Delaneau O, Howie B, Cox AJ, Zagury JF, Marchini J. Haplotype estimation using sequencing reads. *American journal of human genetics.* 2013;93(4):687-696.

8. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. *Nature genetics.* 2016;48(10):1284-1287.

9. Nikolova YS, Knodt AR, Radtke SR, Hariri AR. Divergent responses of the amygdala and ventral striatum predict stress-related problem drinking in young adults: possible differential markers of affective and impulsive pathways of risk for alcohol use disorder. *Molecular psychiatry.* 2016;21(3):348-356.

10. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of clinical psychiatry.* 1998;59 Suppl 20:22-33;quiz 34-57.

11. First MB, Gibbon, M., Spitzer, R.L., Williams, J.B., Benjamin, L.S. *Structured Clinical Interview for DSM-IV Axis II Personality Disorders, (SCID-II).* Washington, DC: American Psychiatric Press; 1997.

12. Hu Y, Shmygelska A, Tran D, Eriksson N, Tung JY, Hinds DA. GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nature communications.* 2016;7:10448.

13. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007;39(7):906-913.

14. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics.* 2007;81(3):559-575.

15. Forbes EE, Brown SM, Kimak M, Ferrell RE, Manuck SB, Hariri AR. Genetic variation in components of dopamine neurotransmission impacts ventral striatal reactivity associated with impulsivity. *Molecular psychiatry.* 2009;14(1):60-70.

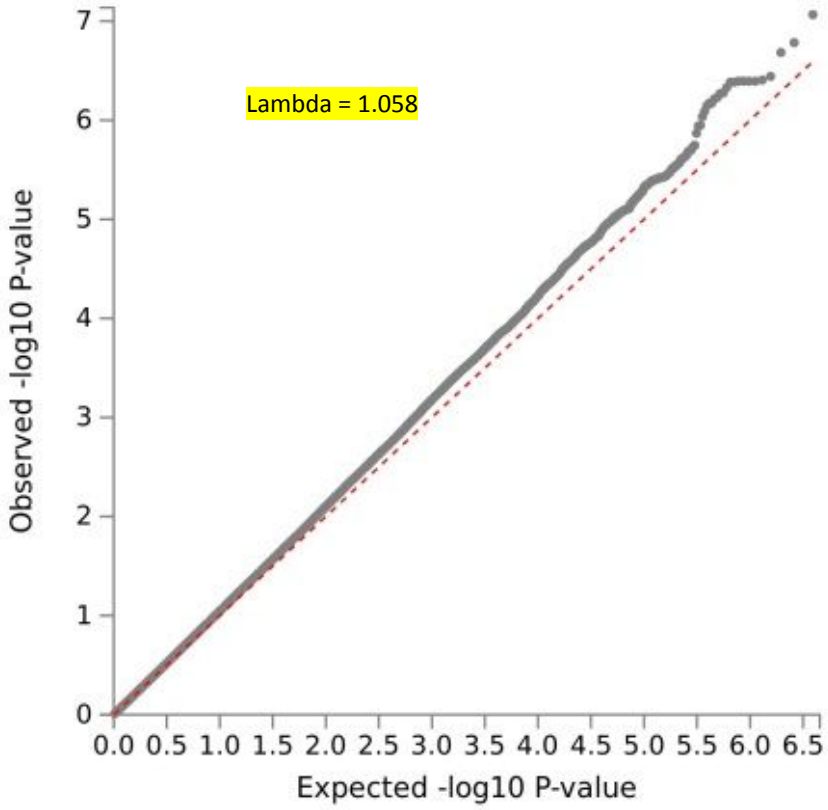
16. Carey CE, Knodt AR, Conley ED, Hariri AR, Bogdan R. Reward-related ventral striatum activity links polygenic risk for attention-deficit/hyperactivity disorder to problematic alcohol use in young adulthood. *Biological psychiatry Cognitive neuroscience and neuroimaging.* 2017;2(2):180-187.

17. Hariri AR, Brown SM, Williamson DE, Flory JD, de Wit H, Manuck SB. Preference for immediate over delayed rewards is associated with magnitude of ventral striatal activity. *The Journal of neuroscience : the official journal of the Society for Neuroscience.* 2006;26(51):13213-13217.

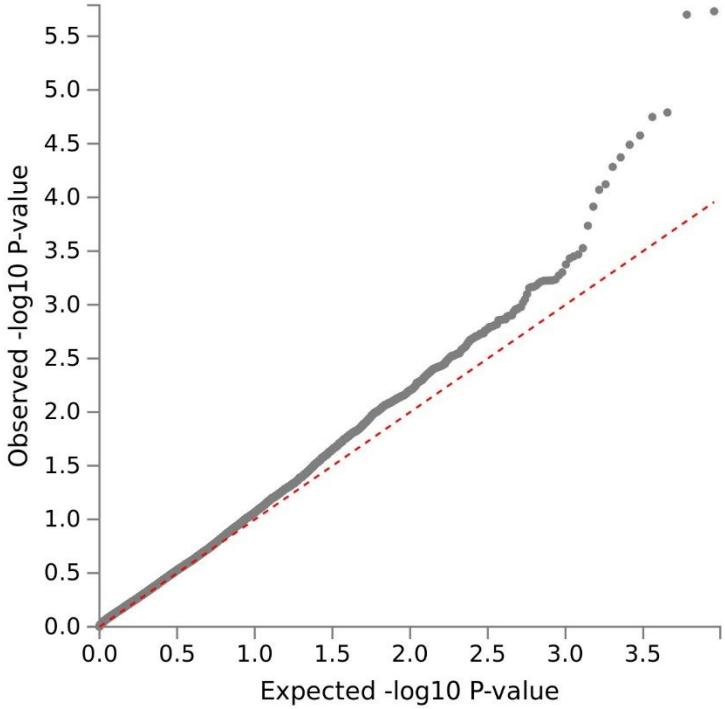
Supplemental Figures

Supplemental Figure S1: QQ plots (EA, AA, EA+AA)

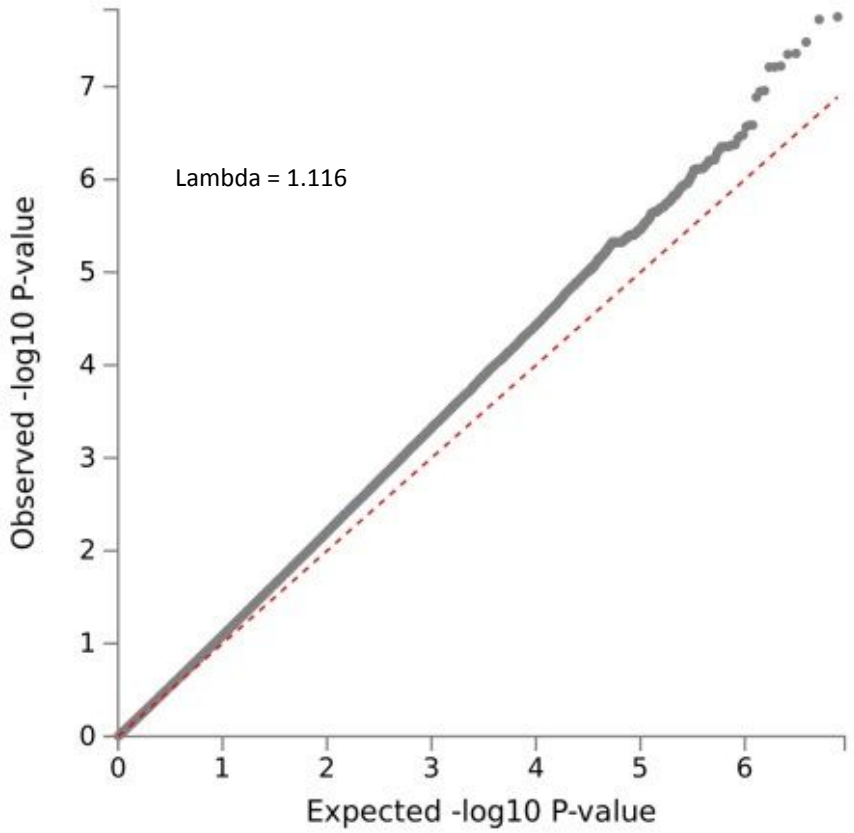
S1A: GWAS QQ Plot for EA families



S1B: Gene-based QQ Plot for EA families

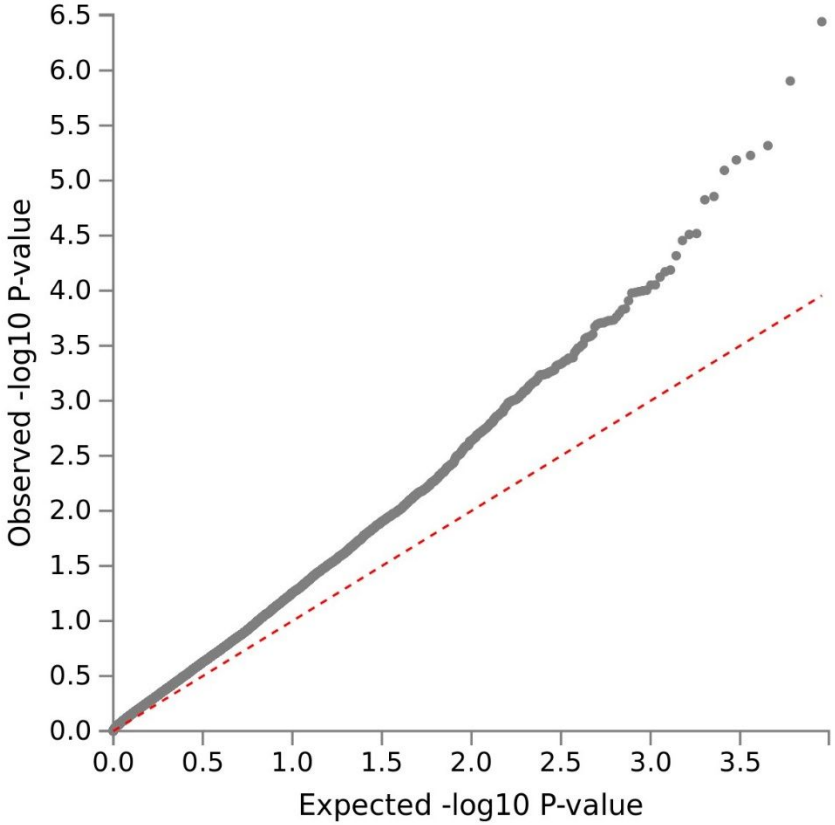


S1C. GWAS QQ Plot for AA families

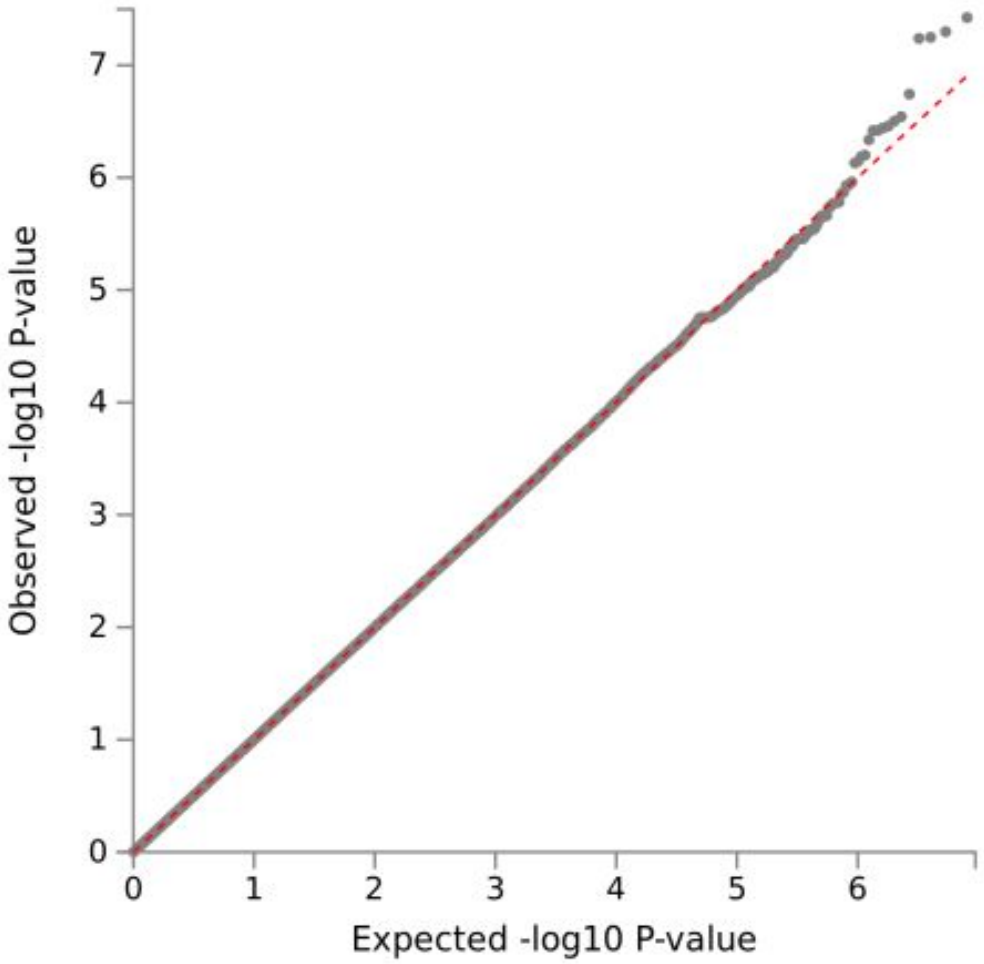


Legend: To account for the elevated lambda in the AA GWAS, GWAS was rerun in the AA sample using 6 and 10 principal components; lambdas remained elevated at 1.12. Exclusion of individuals with an individual genetic ancestry that differed from their family assigned ancestry (n=47) did not modify the lambda (1.12). We conclude that this inflation represents true polygenic signal and not cryptic admixture

S1D: Gene-based QQ Plot for AA families

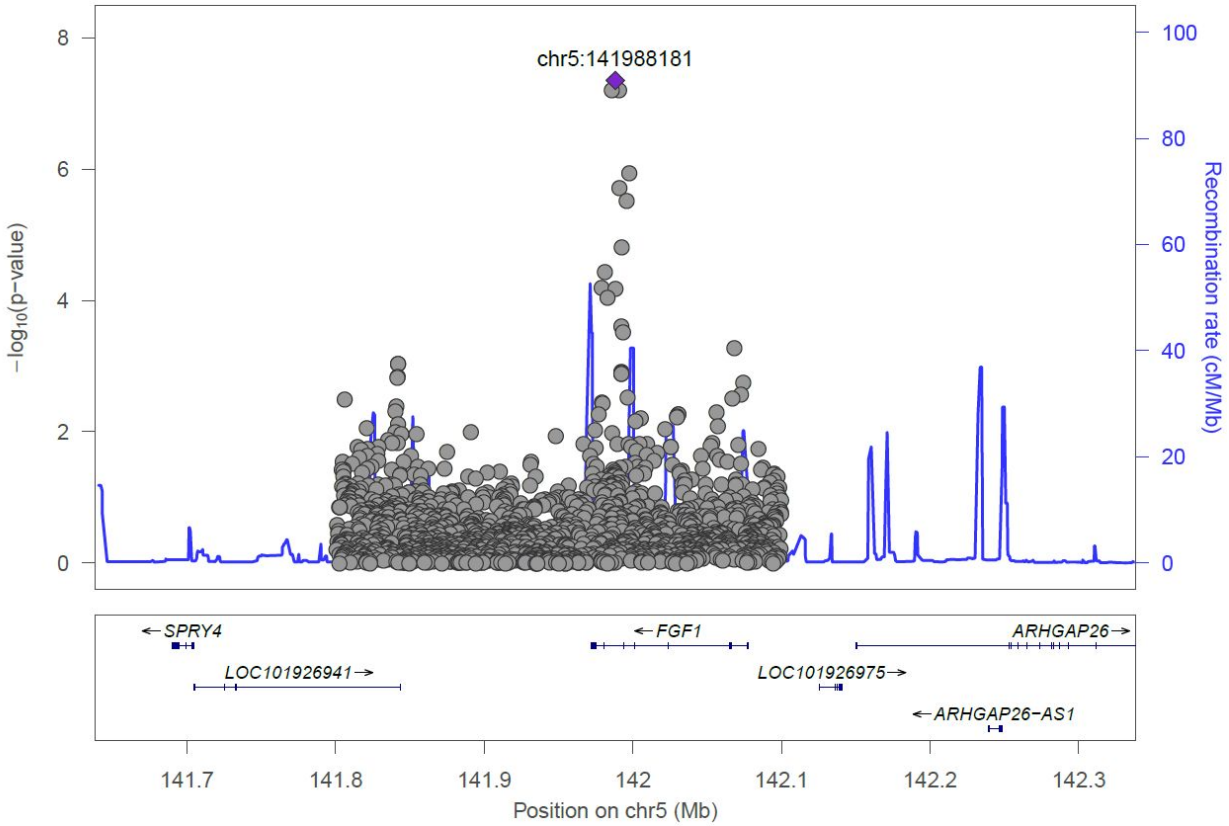


S1E. QQ Plot for EA + AA meta-analysis



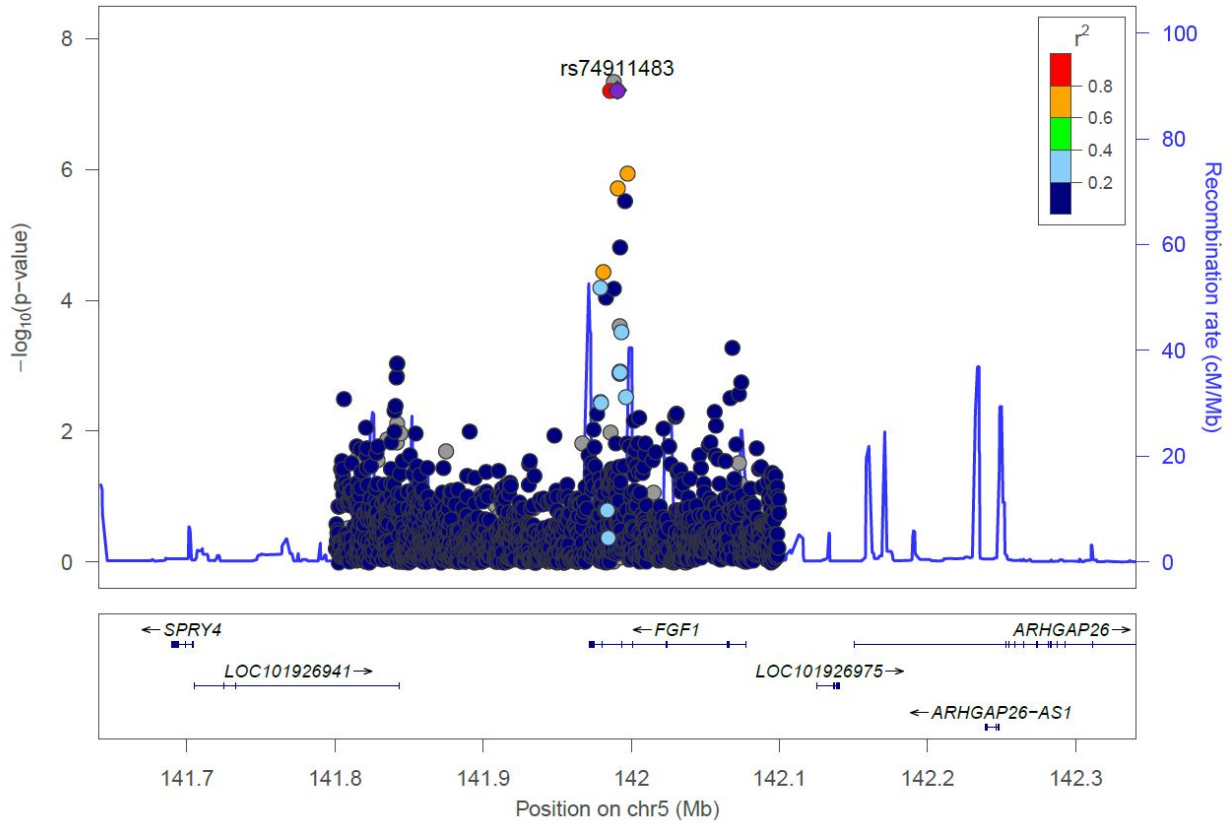
Supplemental Figure S2: Regional association plots for chromosome 5 region.

S2A: Regional association plot for chr5:141988181, rs32414870

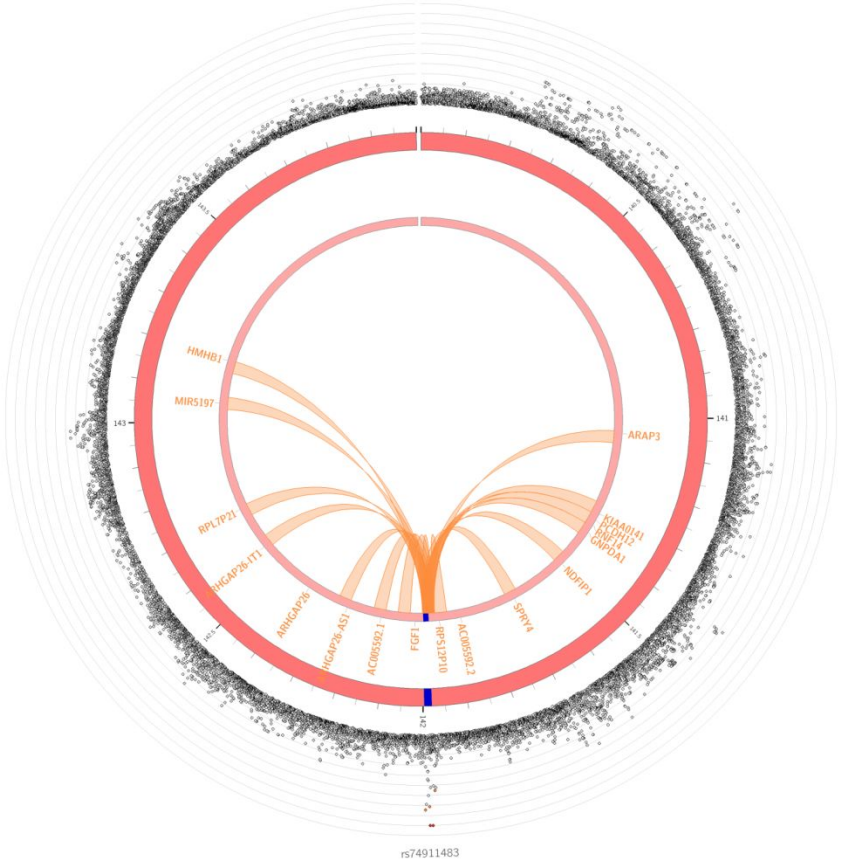


Legend: LD is not depicted as the target variant is an indel and locuszoom does not estimate LD between SNPs and indels.

S2B: Regional association plot for rs74911483, second most significant variant in chromosome 5 region (chr5: 141990602). AFR reference population used for LD.

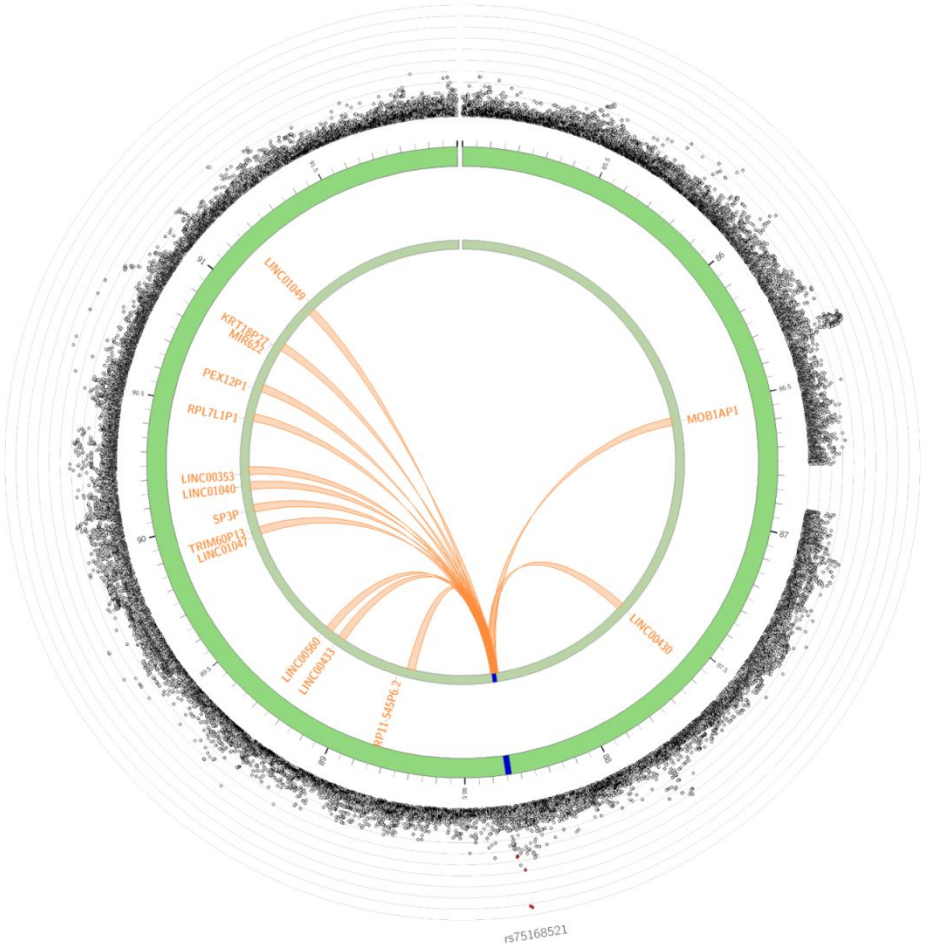


Supplemental Figure S3. Chromatin interaction mapping for rs74911483, second most significant variant in chromosome 5 region (chr5: 141990602).



Legend: The circos plot shows the regional association along the chromosome (outer ring), chromosomal position, (inner red band) and chromatin interactions, as indicated with orange connecting swirls.

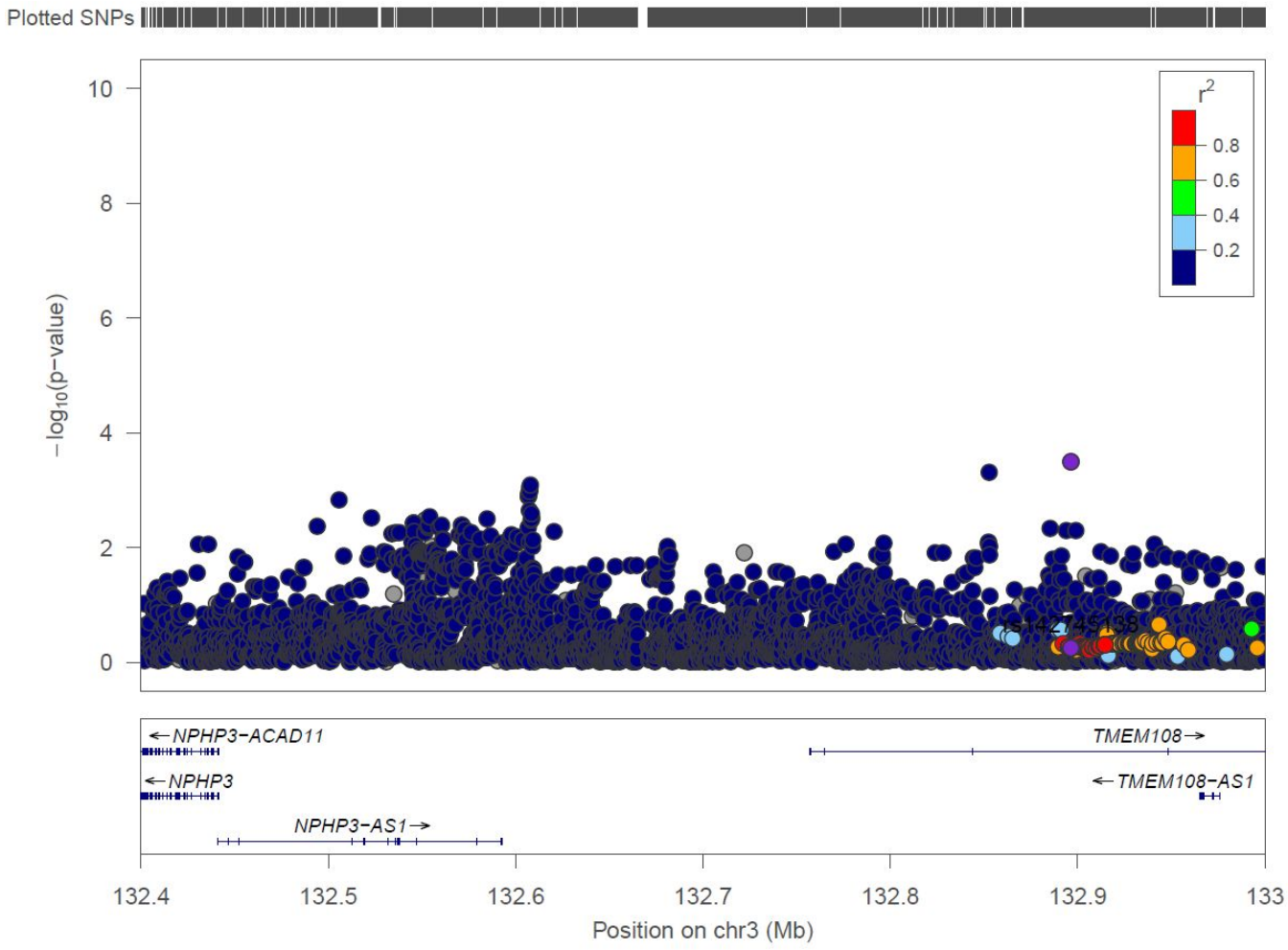
Supplemental Figure S4. Chromatin interaction mapping for rs75168521 on chromosome 13



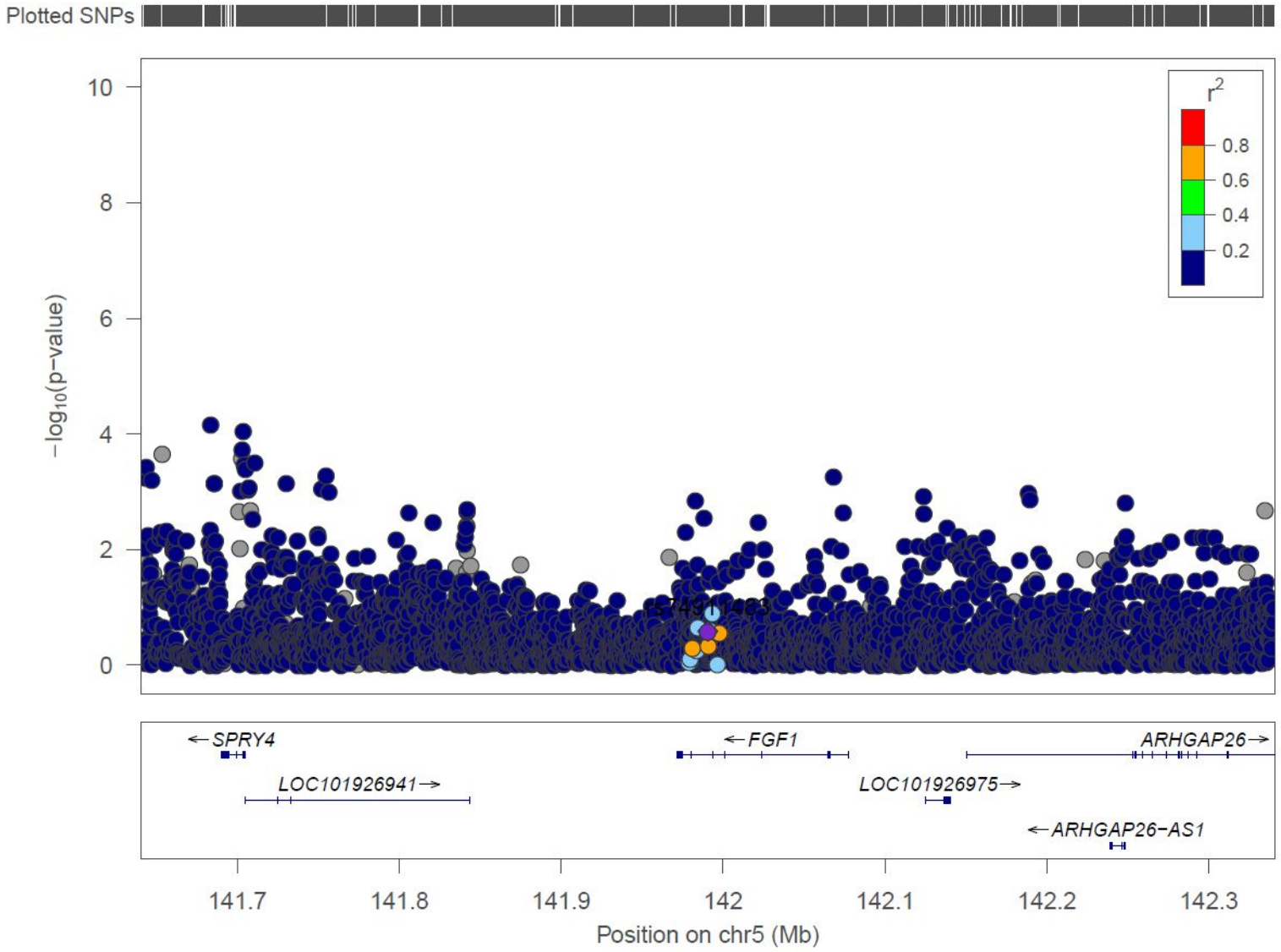
Legend: The circos plot shows the regional association along the chromosome (outer ring), chromosomal position, (inner green band) and chromatin interactions, as indicated with orange connecting swirls.

Supplemental Figure S5: regional association for conditional analyses

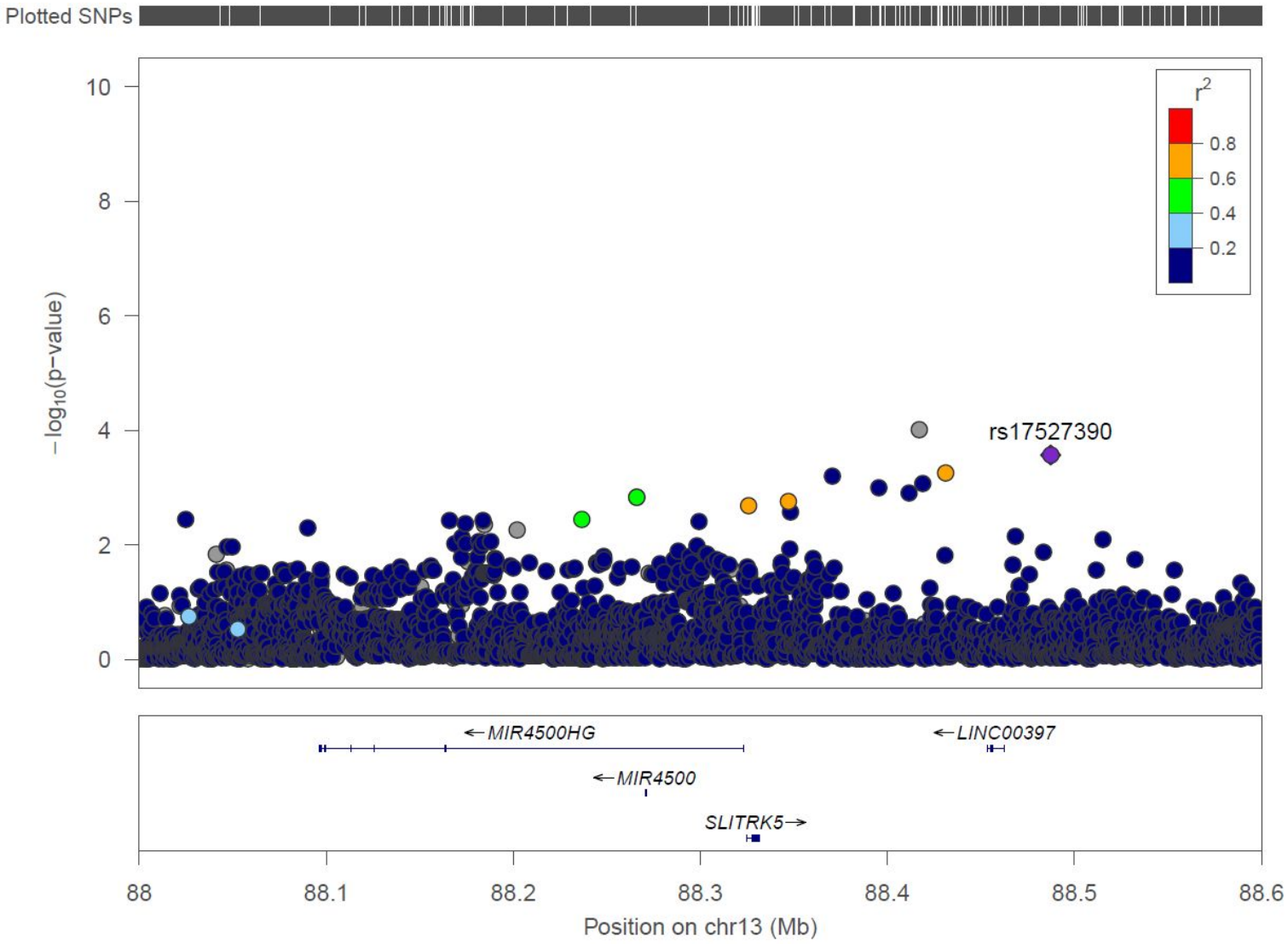
S5A: Regional association plot for conditional analyses in the Chromosome 3 GWS region



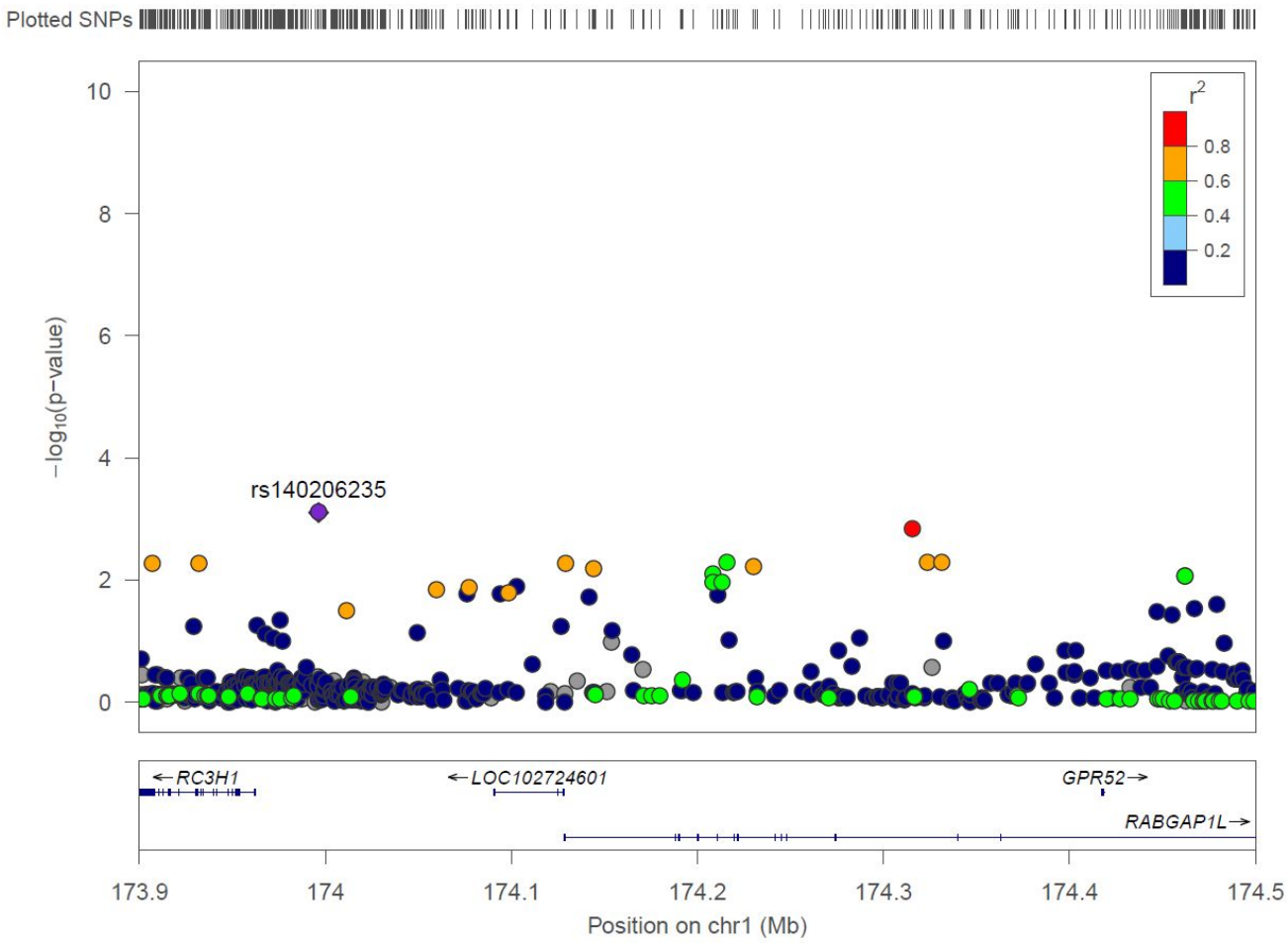
S5B. Regional association plot for conditional analyses in the Chromosome 5 GWS region



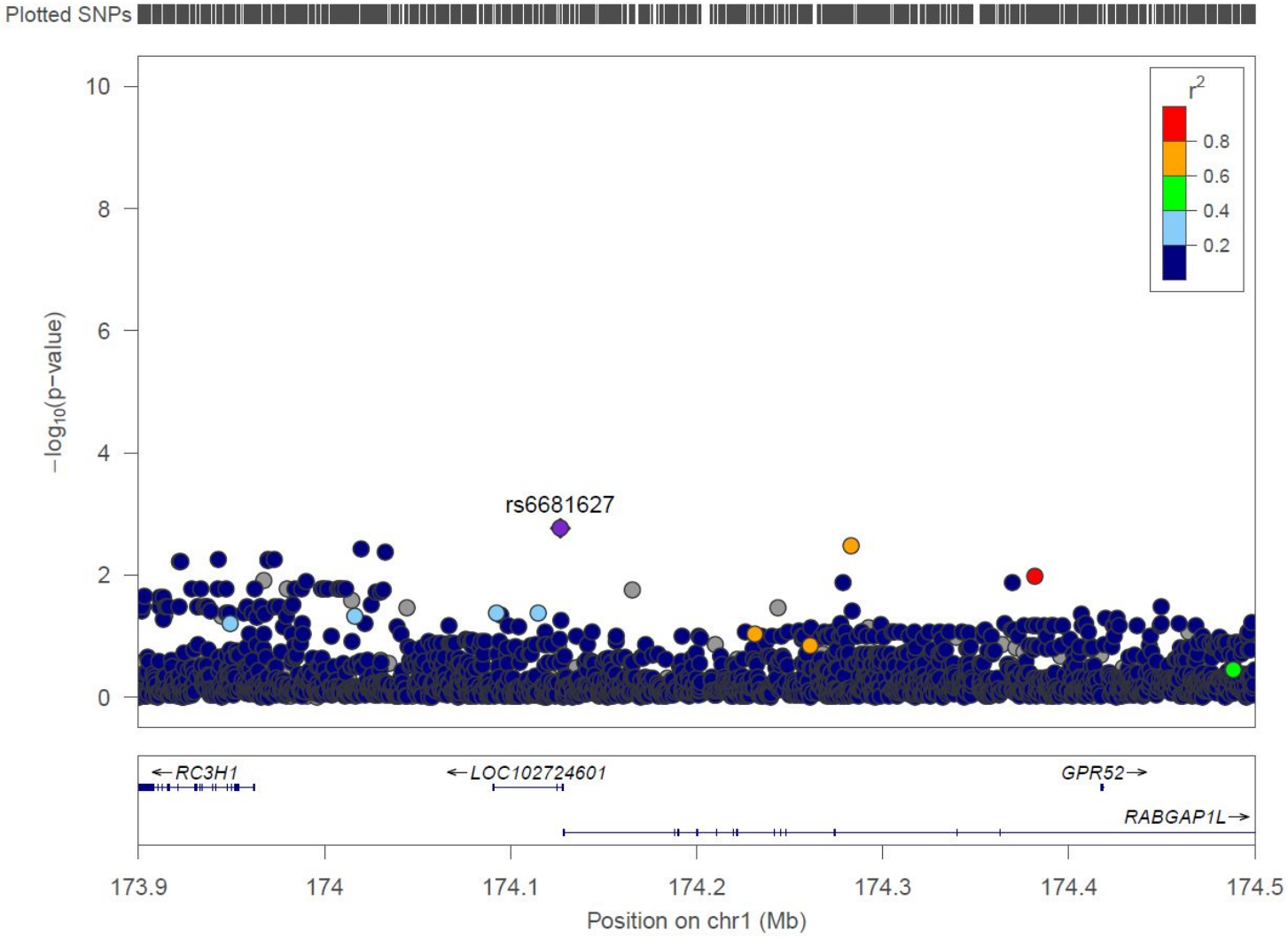
S5C. Regional association plot for conditional analyses in the Chromosome 13 GWS region



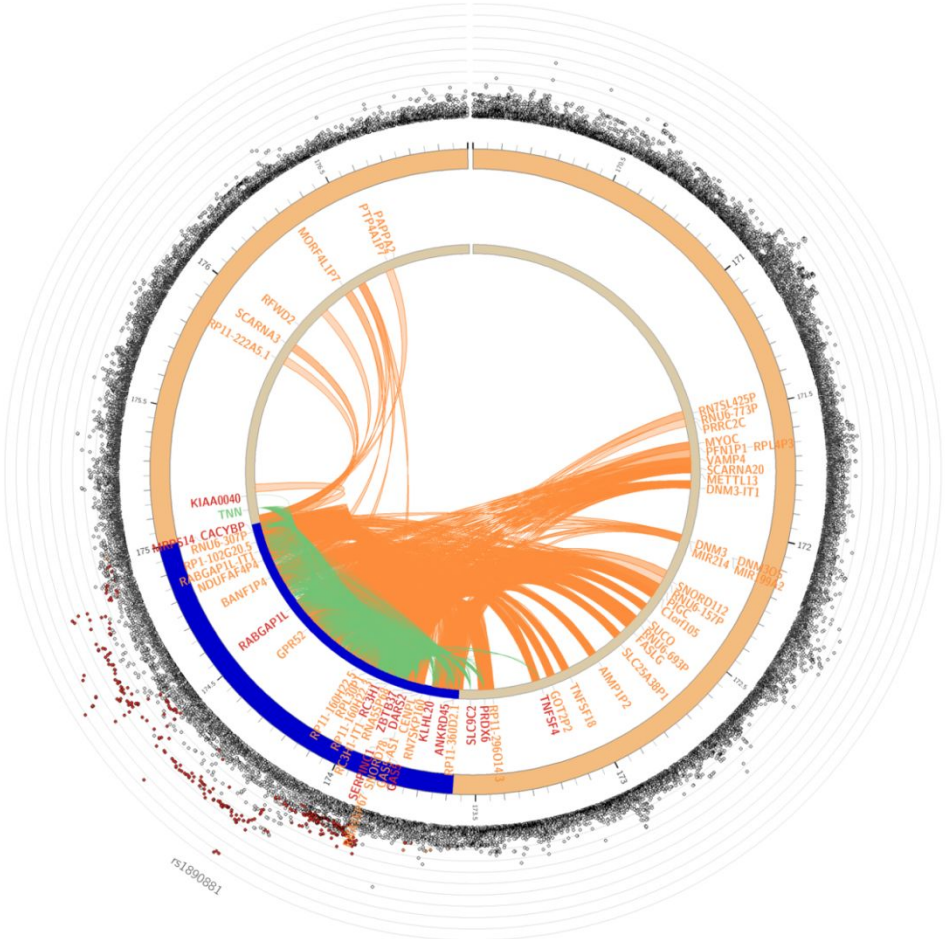
S5D. Regional association plot for conditional analyses in the Chromosome 1 GWS region in European-Americans.



S5E. Regional association plot for conditional analyses in the Chromosome 1 GWS region in African-Americans.



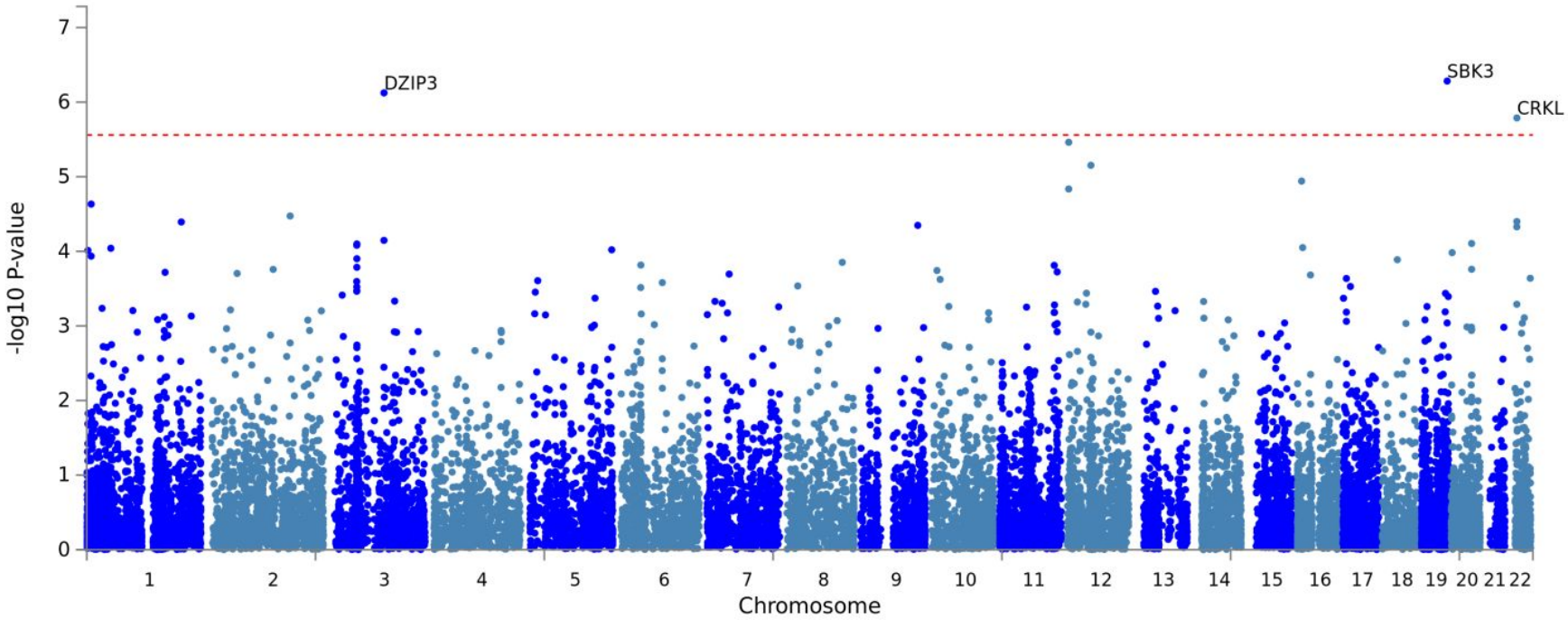
Supplemental Figure S6: Chromatin and eQTL mapping for rs1890881



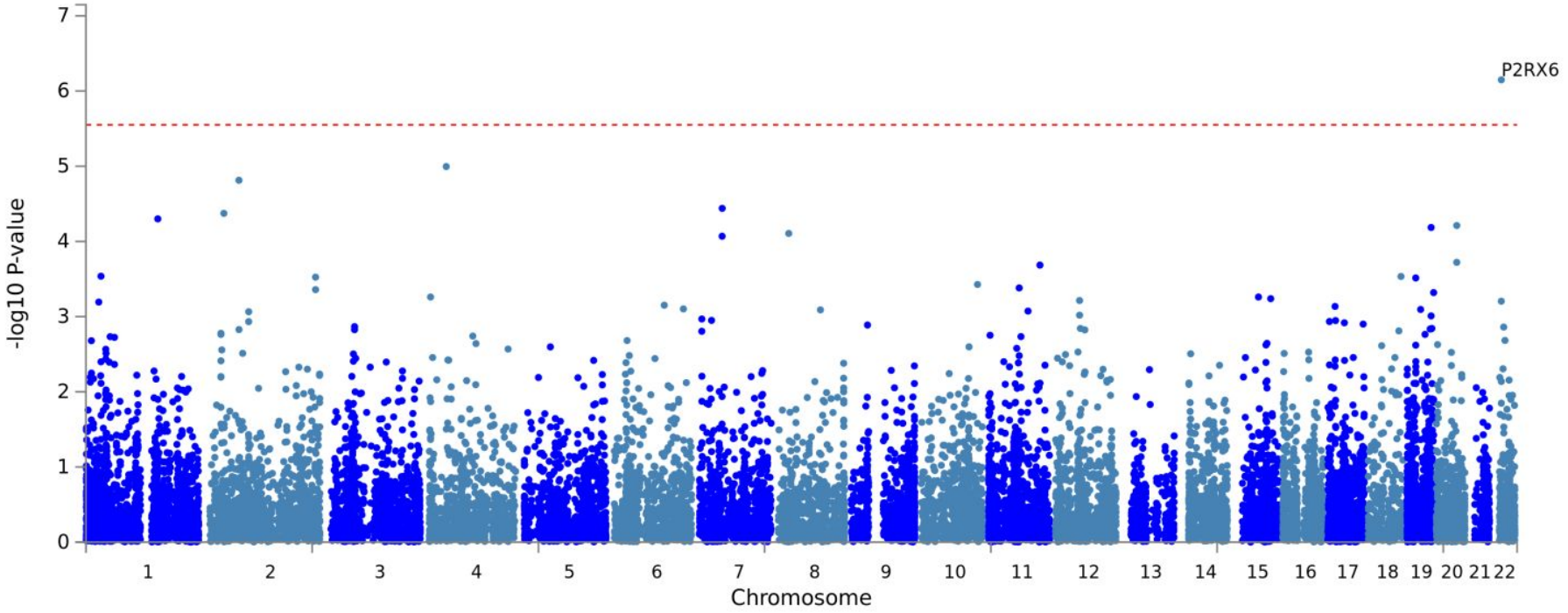
Legend: The circos plot shows the regional association along the chromosome (outer ring), chromosomal position, (inner blue band denoting region of assocation) and chromatin interactions, as indicated with orange connecting swirls. The SNP is an eQTL for genes shown in red, and green connectors indicate evidence for chromatin interactions with genes related to the eQTL.

Supplemental Figure S7: Gene based Manhattan plots in African American and European American families

S7A. Gene-based association analysis showing genome-wide significant genes in African-American families

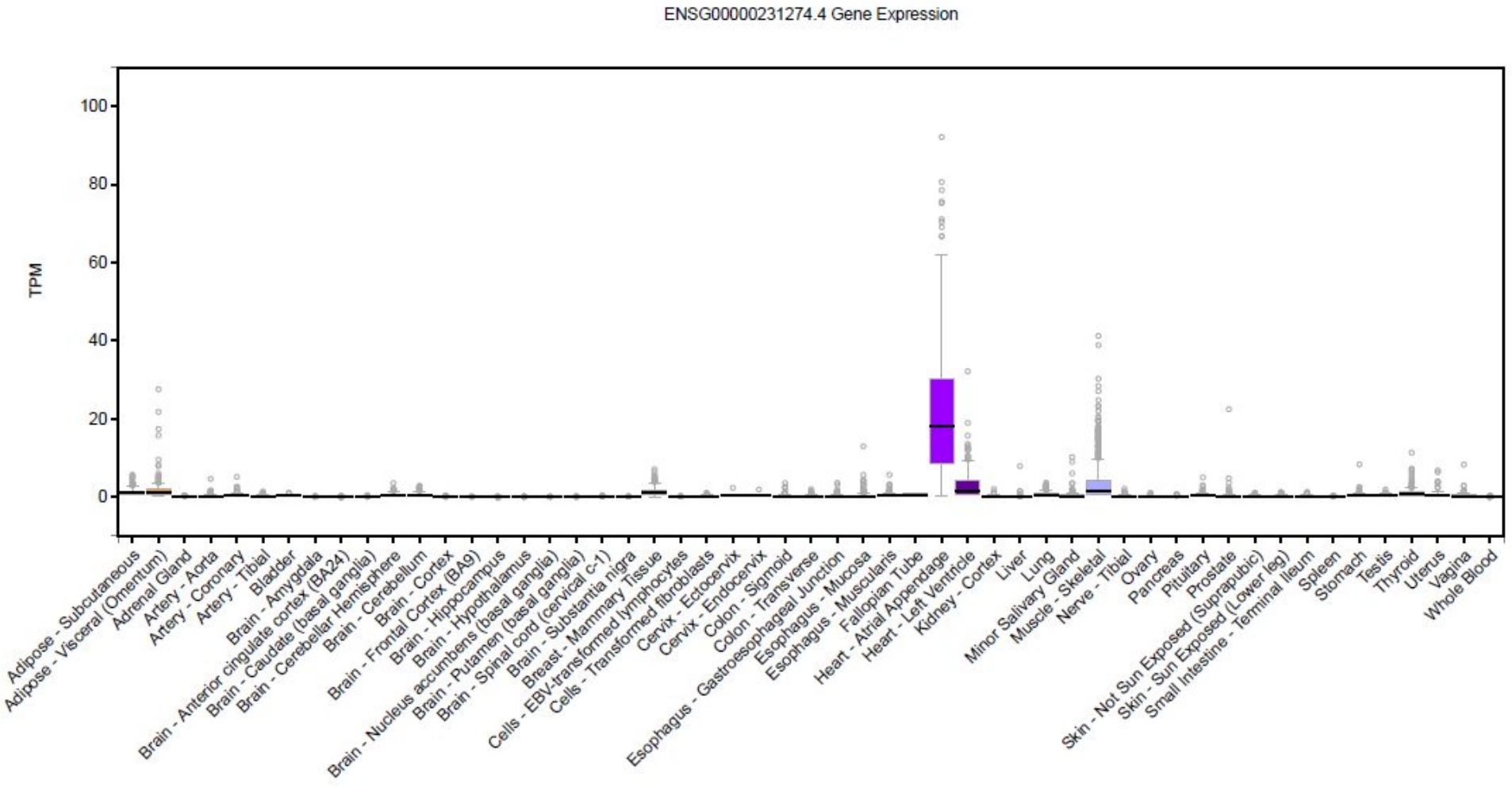


S7B. Gene-based association analysis showing genome-wide significant genes in European-American families

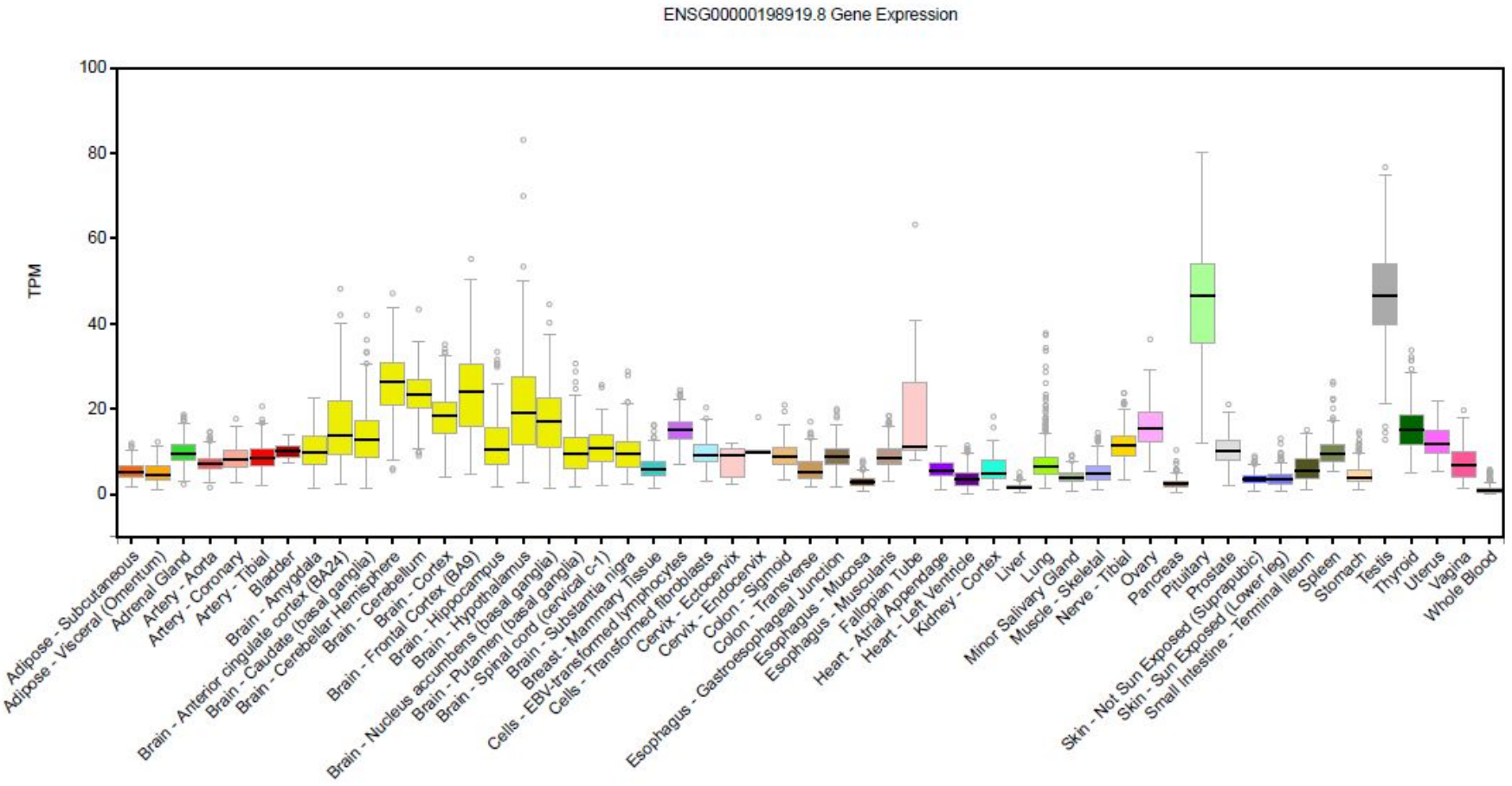


Supplemental Figure S8: Expression of genes from AA families: *SBK3, DZIP3, CRKL*

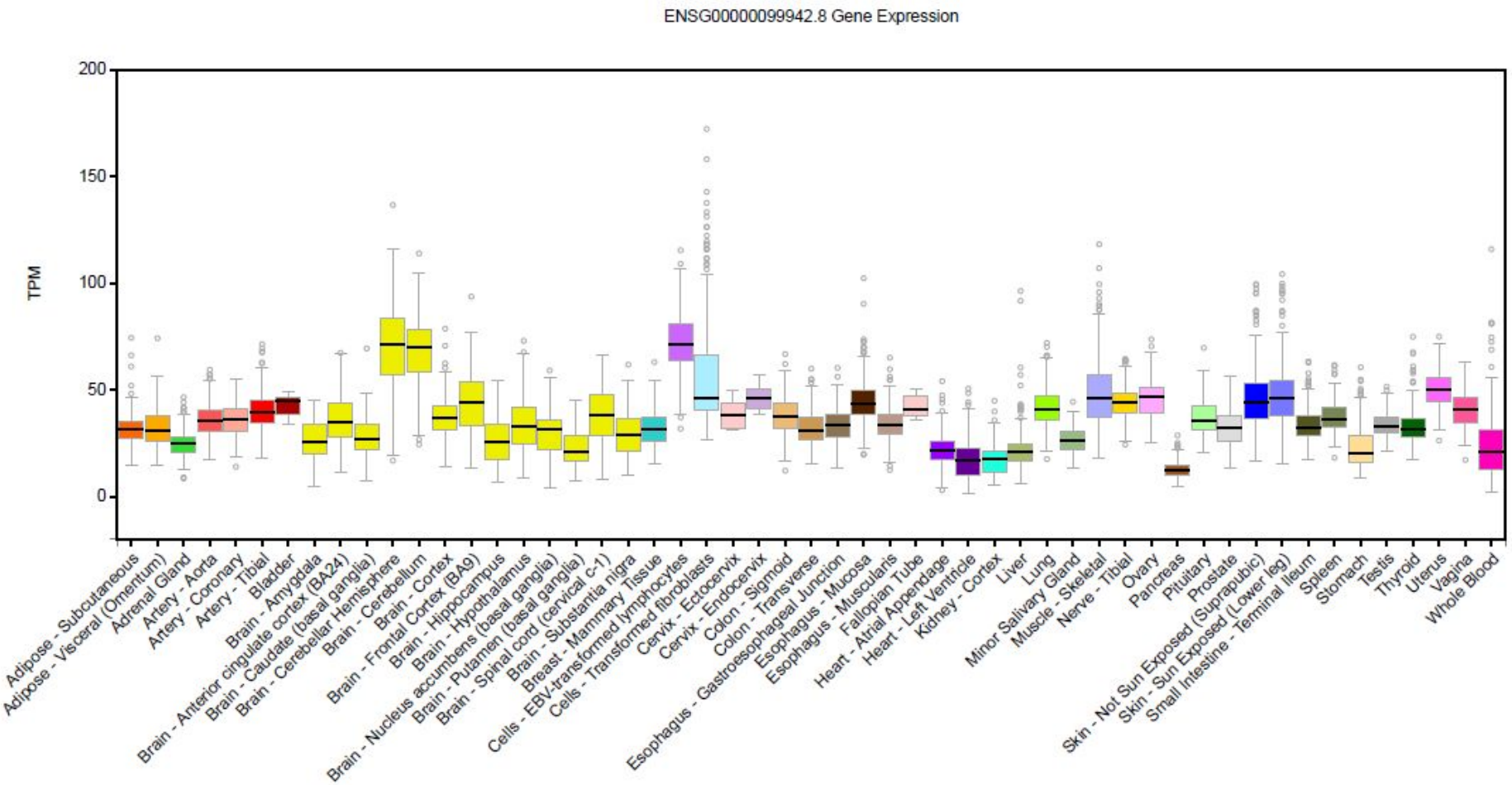
S8A. Expression of SBK3 across tissues in GTEx.



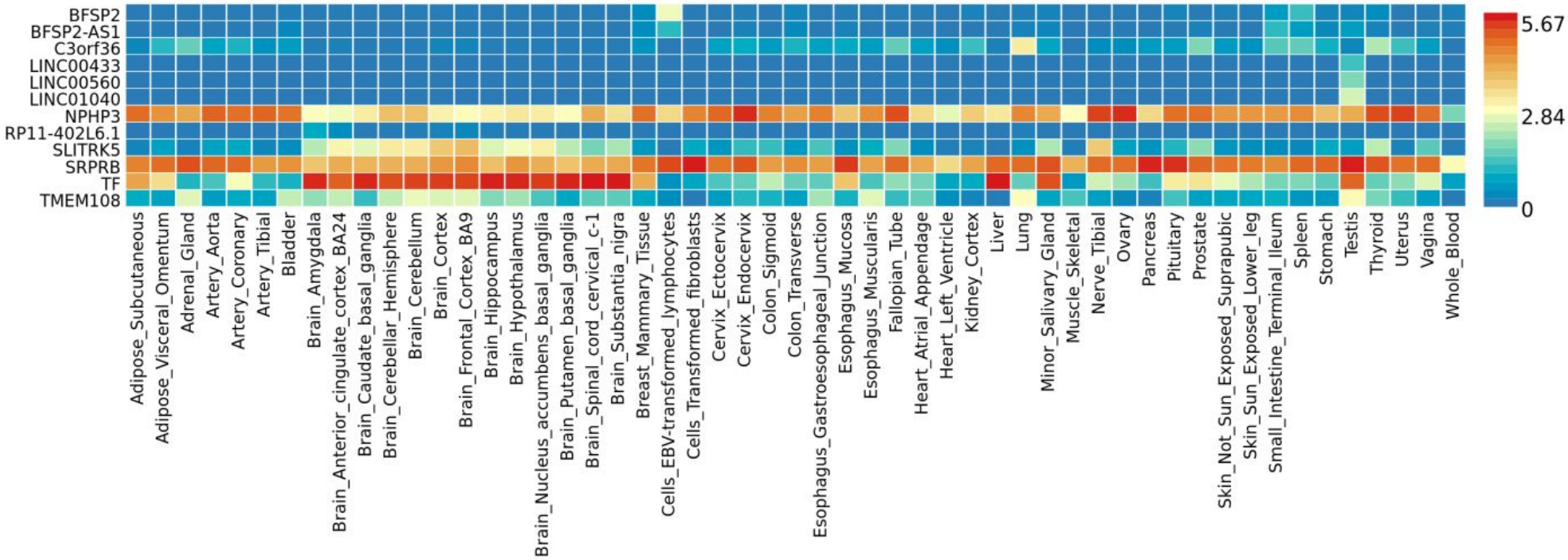
S8B. Expression of DZIP3 across tissues in GTEx.



S8C. Expression of CRKL across tissues in GTEx.

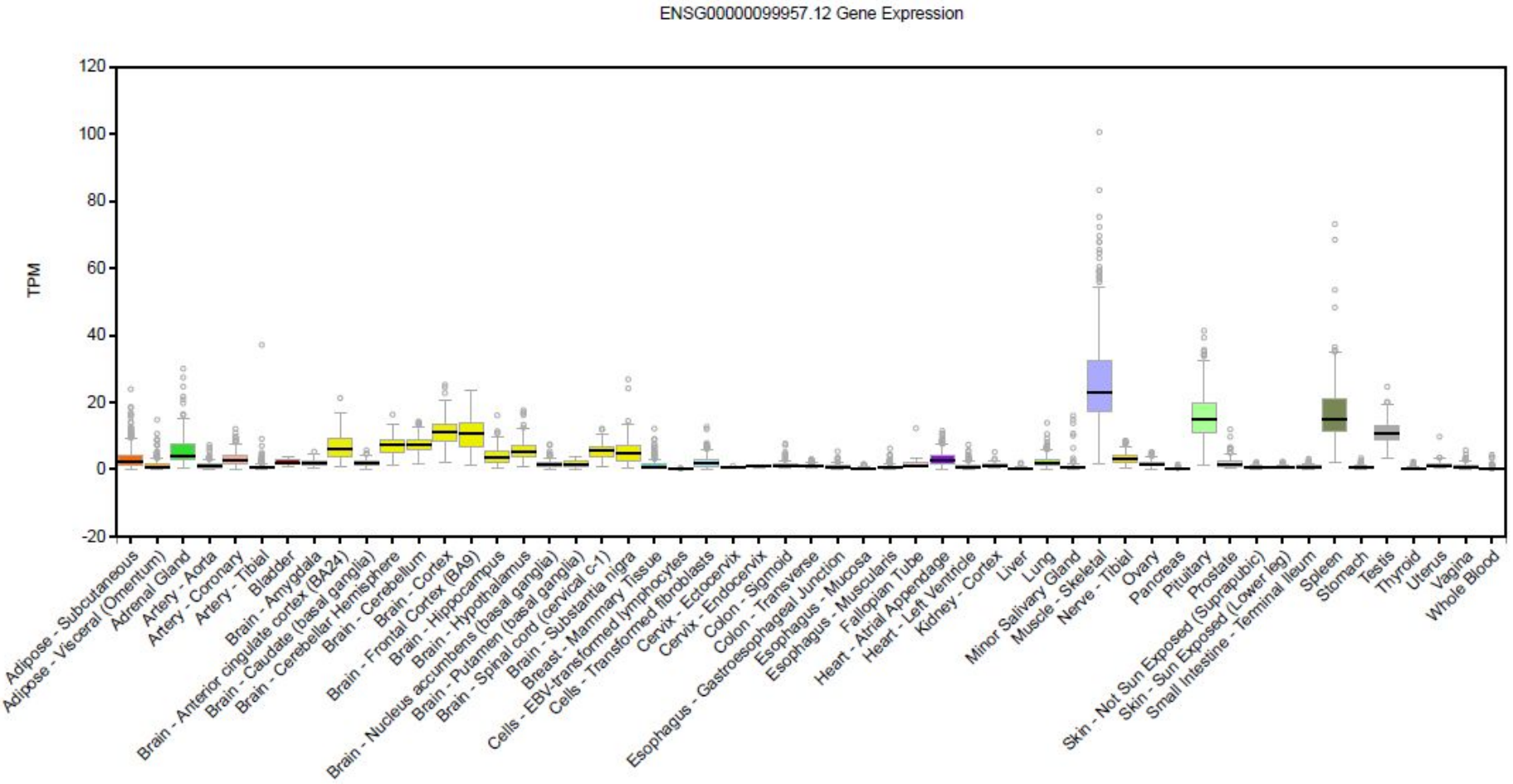


Supplemental Figure S9: Enrichment of genes identified within the African-American GWAS with differential expression across 53 tissue types from GTEx.



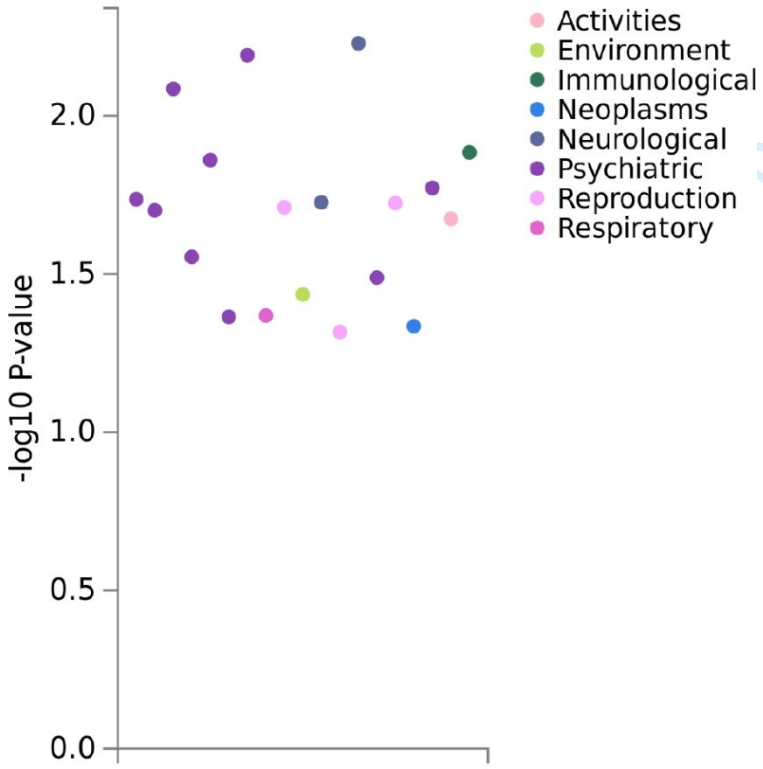
Legend: Genes identified using positional, eQTL and chromatin interaction mapping were selected (n=26 genes, contrasted with 35808 background genes). Twelve genes with differential expression are shown. The average of the log2 transformed expression value is shown; darker red indicates higher expression of that gene in the tissue, relative to darker blue color.

Supplemental Figure S10: Expression of *P2RX6,* the genome-wide significant gene, in tissue from GTEx.

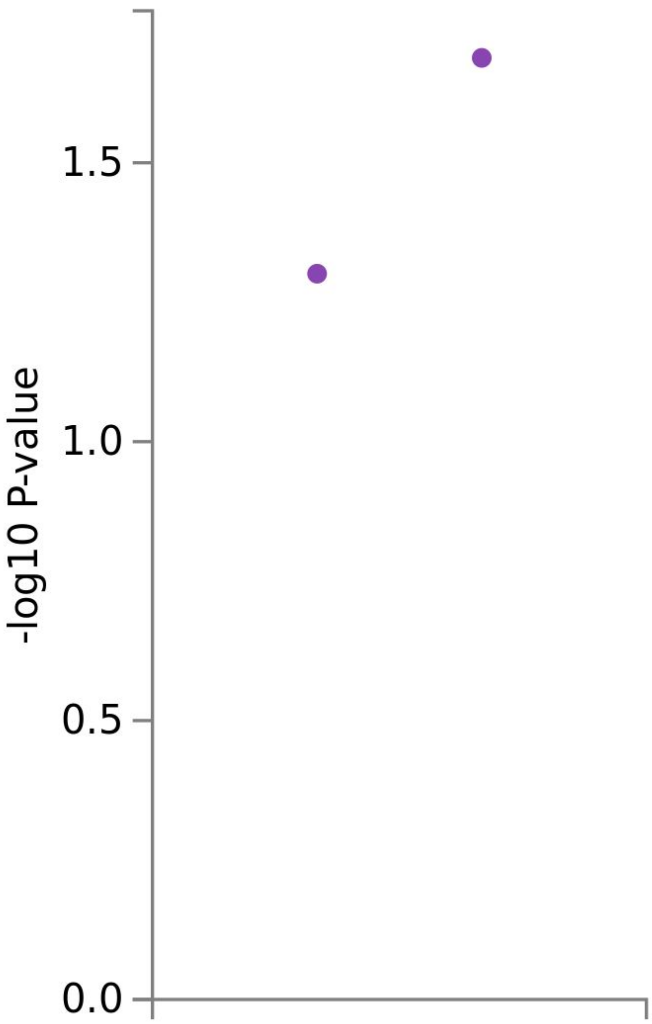


Supplemental Figure S11. Plots showing results for genome-wide significant SNPs and genes in the African-Americans and European-American families from COGA using summary statistics from GWAS conducted on other traits and disorders, from published studies, pre-prints sharing summary statistics and from phenotypes in the UK Biobank. Results drawn from 3,798 GWAS across 217 studies of 2,824 unique traits from http://atlas.ctglab.nl/

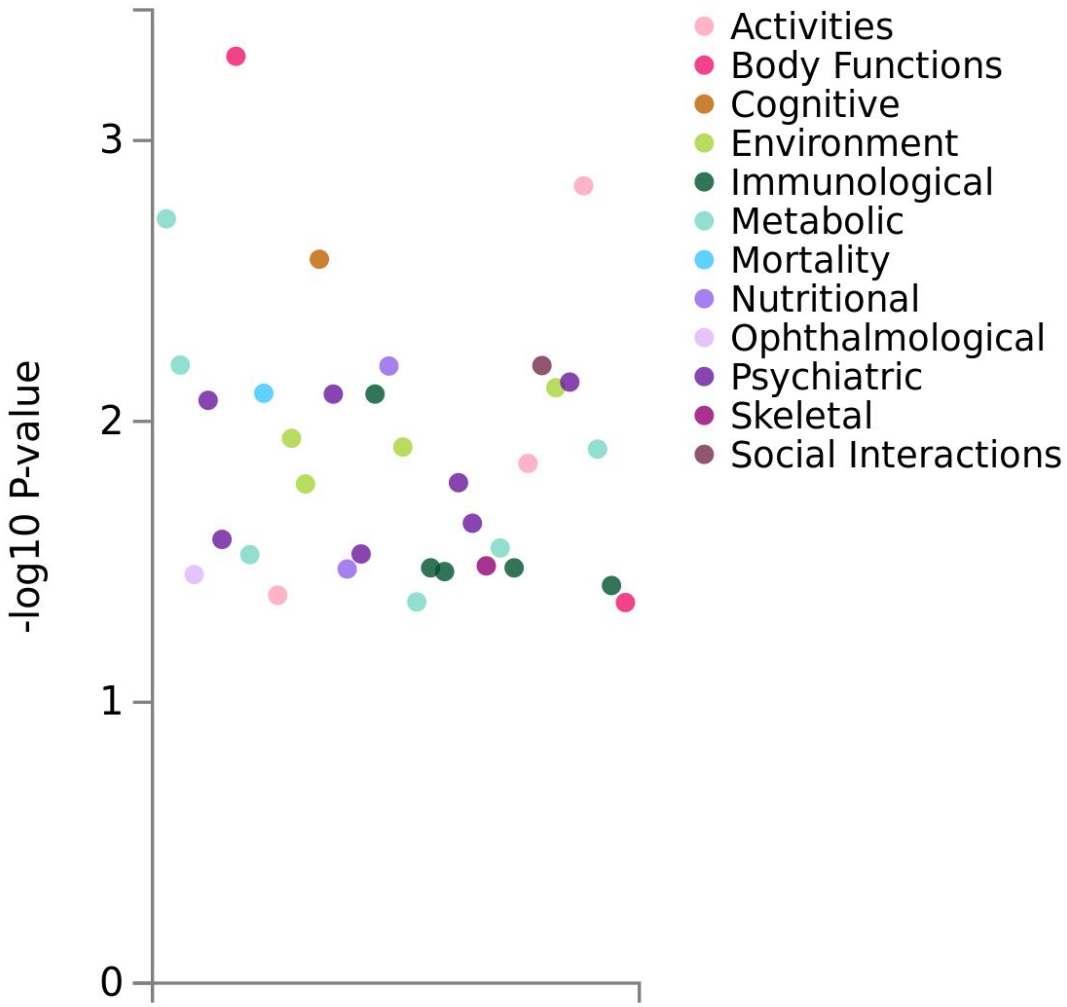
S11A. Chromosome 3 (rs34066662)



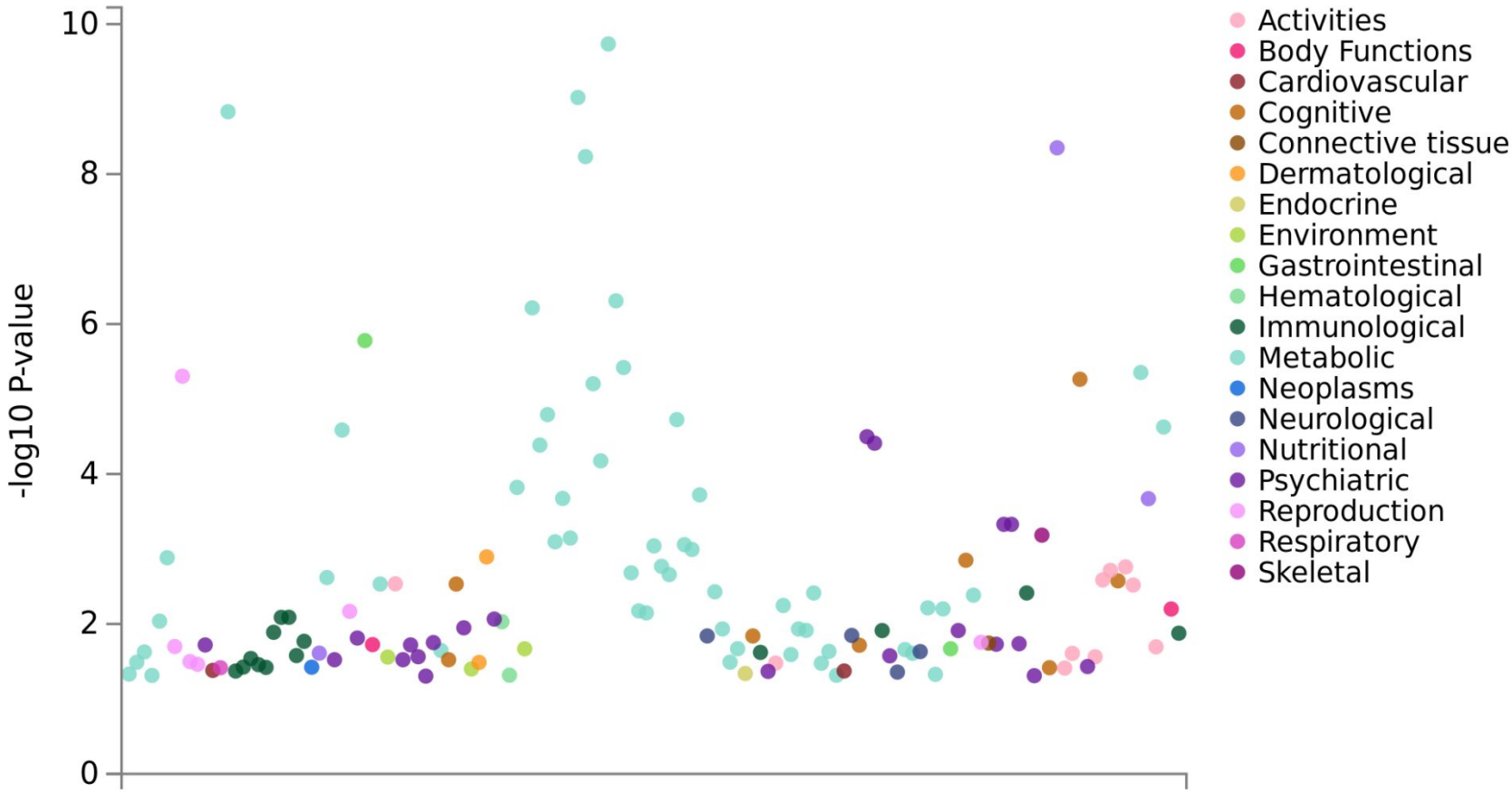
S11B. Chromosome 5 (rs34204870)



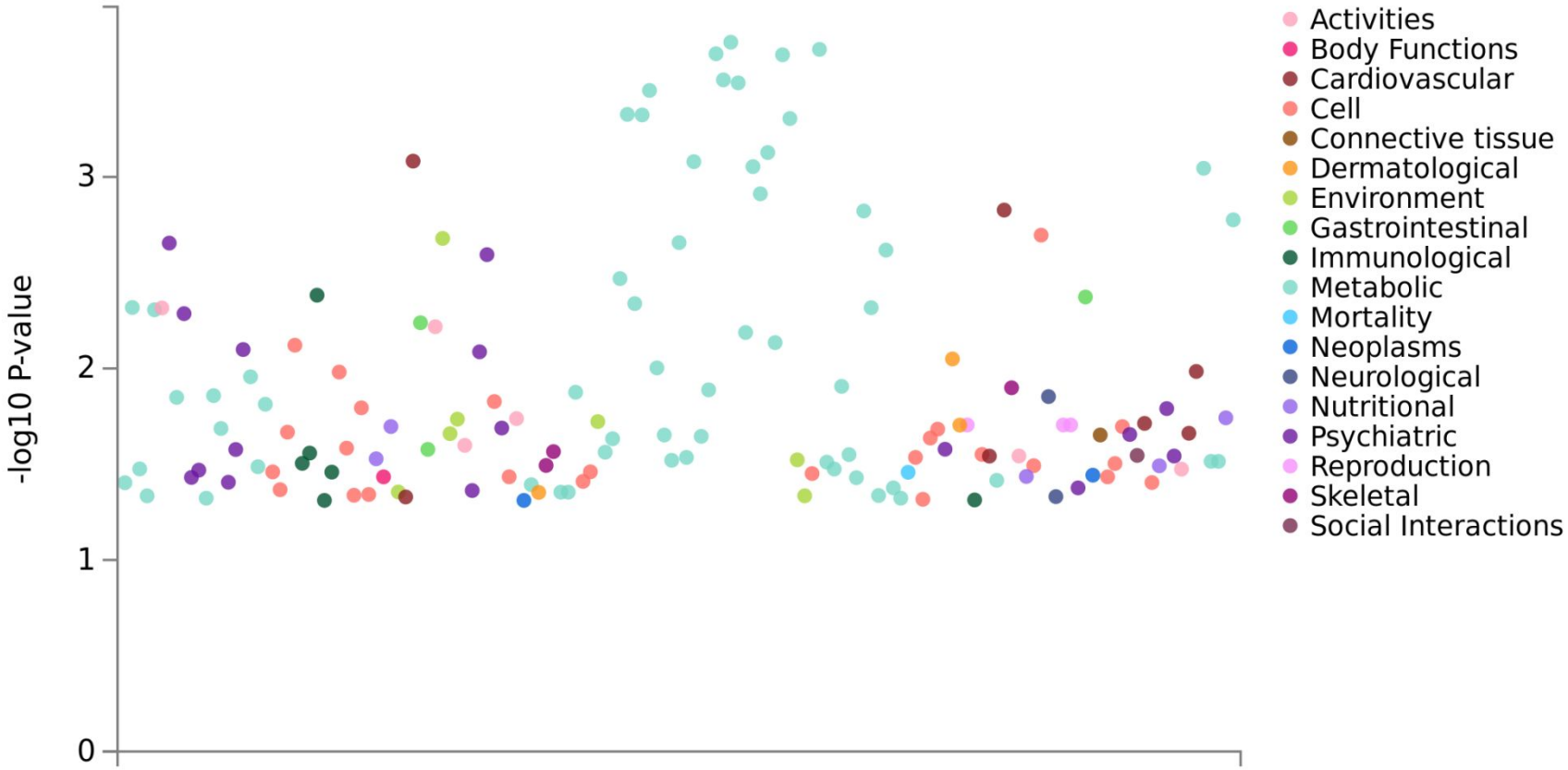
S11C. Chromosome 13 (rs75168521)



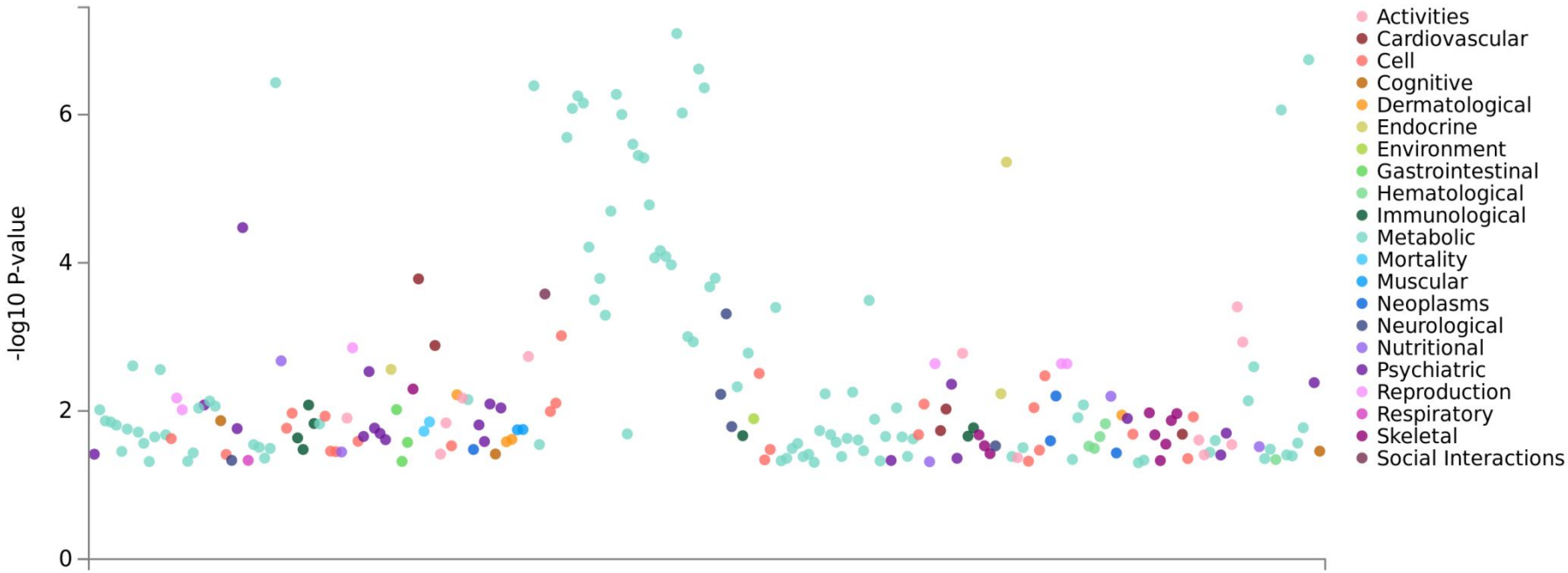
S11D. Chromosome 1 (rs1890881)


10
8
6
4
2
0
-log10 P-value
Activities
Body Functions
Cardiovascular
Cognitive
Connective tissue
Dermatological
Endocrine
Environment
Gastrointestinal
Hematological
Immunological
Metabolic
Neoplasms
Neurological
Nutritional
Psychiatric
Reproduction
Respiratory
Skeletal

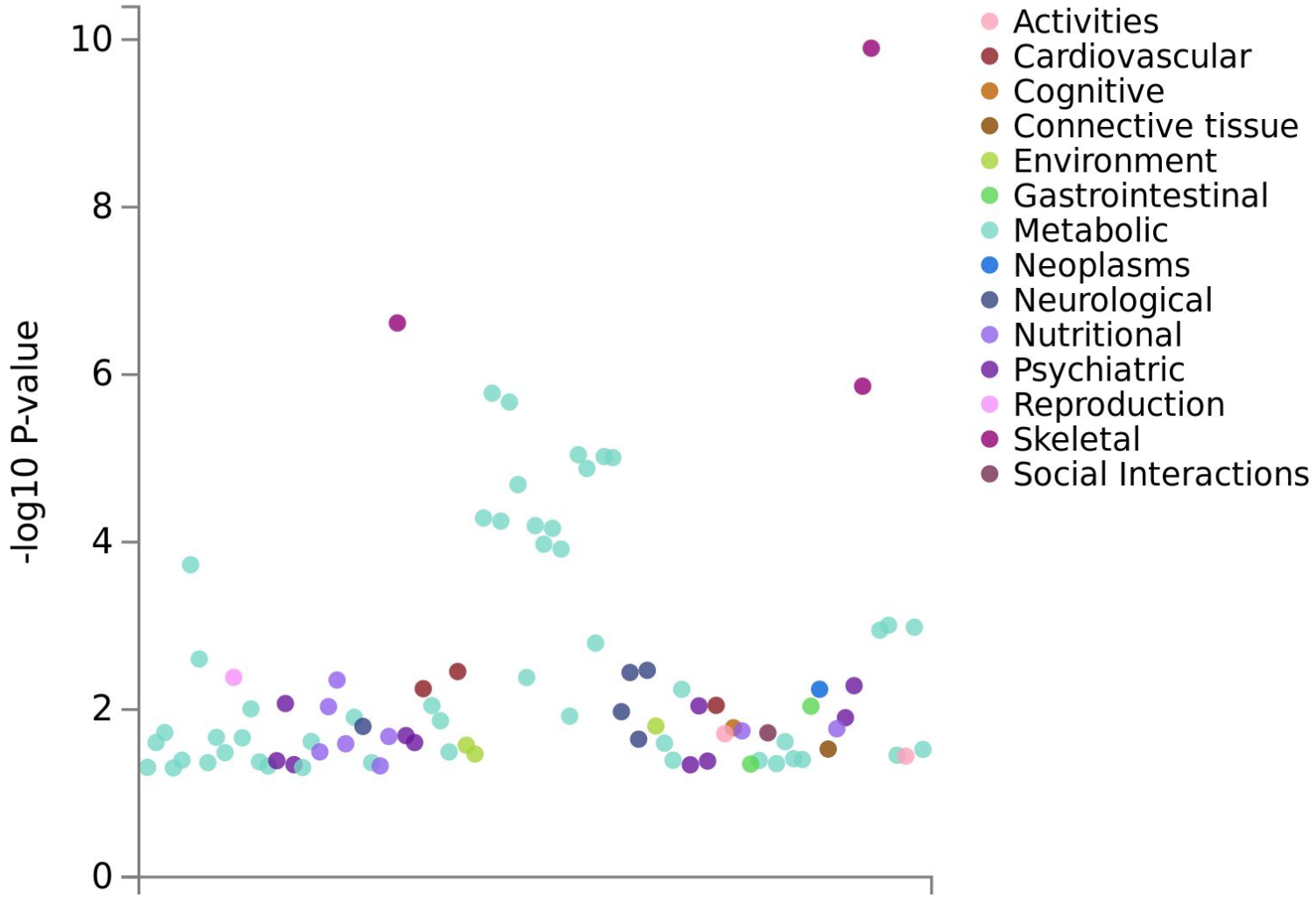
S11E. *CRKL* gene



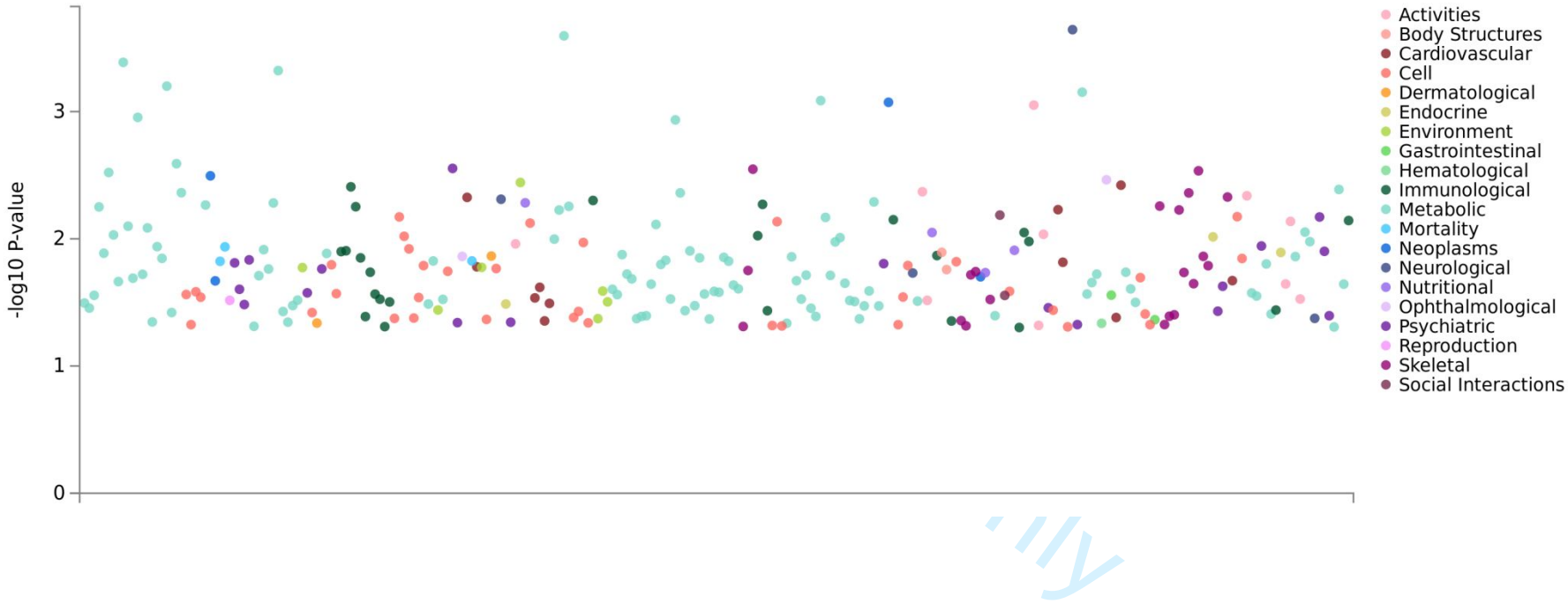
S11F. *DZIP3* gene



S11G. *SBK3* gene



S11H. *P2RX6* gene



Supplemental Tables

Table S1:Cross drug dependence correlations [95% Confidence Intervals]; Correlations for African-Americans shown above diagonal European-Americans shown below diagonal

	Marijuana	Cocaine	Sedatives	Stimulants	Opioids	Alcohol
Marijuana	1	0.45 [0.39-0.51]	0.53 [0.40-0.66]	0.51 [0.39-0.62]	0.33 [0.23-0.43]	0.49 [0.43-0.54]
Cocaine	0.65 [0.61-0.68]	1	0.47 [0.31-0.62]	0.50 [0.38-0.62]	0.67 [0.60-0.73]	0.74 [0.70-0.78]
Sedatives	0.57 [0.52-0.63]	0.62 [0.57-0.67]	1	0.77 [0.67-0.88]	0.67 [0.54-0.97]	0.55 [0.40-0.71]
Stimulants	0.63 [0.59-0.68]	0.62 [0.56-0.67]	0.66 [0.61-0.72]	1	0.56 [0.43-0.68]	0.48 [0.36-0.61]
Opioids	0.58 [0.52-0.63]	0.66 [0.61-0.71]	0.82 [0.78-0.86]	0.63 [0.57-0.68]	1	0.52 [0.44-0.59]
Alcohol	0.57 [0.54-0.61]	0.68 [0.65-0.71]	0.66 [0.60 -0.71]	0.61 [0.56-0.65]	0.54 [0.49-0.60]	1

Table S2: rs1890881 eQTL genes in GTEx v7

Gencode Id	Gene Symbol	P-Value	NES	Tissue
ENSG00000183831.6	ANKRD45	2.90E-24	-0.59	Testis
ENSG00000183831.6	ANKRD45	2.30E-06	-0.89	Brain - Anterior cingulate cortex (BA24)
ENSG00000183831.6	ANKRD45	3.10E-05	-0.32	Brain - Nucleus accumbens (basal ganglia)
ENSG00000117593.8	DARS2	1.10E-09	-0.35	Esophagus - Muscularis
ENSG00000117593.8	DARS2	3.30E-09	-0.28	Muscle - Skeletal
ENSG00000117593.8	DARS2	1.00E-08	-0.26	Heart - Atrial Appendage
ENSG00000117593.8	DARS2	2.80E-08	-0.24	Heart - Left Ventricle
ENSG00000117593.8	DARS2	6.60E-06	-0.20	Adipose - Subcutaneous
ENSG00000117593.8	DARS2	7.20E-05	-0.13	Lung
ENSG00000152061.17	RABGAP1L	1.60E-13	-0.44	Esophagus - Muscularis
ENSG00000152061.17	RABGAP1L	2.00E-12	0.45	Cells - Transformed fibroblasts
ENSG00000152061.17	RABGAP1L	2.00E-08	-0.39	Esophagus - Gastroesophageal Junction
ENSG00000152061.17	RABGAP1L	9.80E-06	0.22	Esophagus - Mucosa
ENSG00000152061.17	RABGAP1L	3.50E-05	0.25	Testis
ENSG00000152061.17	RABGAP1L	5.00E-05	0.21	Lung
ENSG00000152061.17	RABGAP1L	1.50E-04	0.20	Adipose - Subcutaneous
ENSG00000135870.7	RC3H1	3.40E-05	-0.27	Cells - Transformed fibroblasts
ENSG00000117601.9	SERPINC1	1.90E-16	0.78	Esophagus - Muscularis
ENSG00000117601.9	SERPINC1	9.10E-14	0.55	Muscle - Skeletal
ENSG00000117601.9	SERPINC1	1.90E-12	0.75	Esophagus - Mucosa
ENSG00000117601.9	SERPINC1	2.30E-12	0.56	Lung
ENSG00000117601.9	SERPINC1	7.20E-12	0.73	Adipose - Visceral (Omentum)
ENSG00000117601.9	SERPINC1	1.00E-11	0.73	Heart - Left Ventricle
ENSG00000117601.9	SERPINC1	2.70E-11	0.64	Nerve - Tibial
ENSG00000117601.9	SERPINC1	5.00E-10	0.59	Adipose - Subcutaneous
ENSG00000117601.9	SERPINC1	9.80E-10	0.58	Thyroid
ENSG00000117601.9	SERPINC1	2.10E-08	0.53	Skin - Sun Exposed (Lower leg)

ENSG00000117601.9	SERPINC1	7.30E-08	0.84	Pituitary
ENSG00000117601.9	SERPINC1	6.40E-07	0.66	Esophagus - Gastroesophageal Junction
ENSG00000117601.9	SERPINC1	8.10E-07	0.50	Artery - Tibial
ENSG00000117601.9	SERPINC1	1.50E-06	0.54	Heart - Atrial Appendage
ENSG00000117601.9	SERPINC1	1.80E-06	0.82	Brain - Cortex
ENSG00000117601.9	SERPINC1	2.70E-06	0.73	Artery - Coronary
ENSG00000117601.9	SERPINC1	4.60E-06	0.69	Prostate
ENSG00000117601.9	SERPINC1	5.10E-06	0.53	Artery - Aorta
ENSG00000117601.9	SERPINC1	7.00E-06	0.83	Brain - Hippocampus
ENSG00000117601.9	SERPINC1	8.90E-06	0.70	Brain - Caudate (basal ganglia)
ENSG00000117601.9	SERPINC1	1.40E-05	0.64	Colon - Sigmoid
ENSG00000117601.9	SERPINC1	1.80E-05	0.77	Brain - Nucleus accumbens (basal ganglia)
ENSG00000117601.9	SERPINC1	3.60E-05	0.64	Adrenal Gland
ENSG00000117601.9	SERPINC1	4.10E-05	0.93	Brain - Anterior cingulate cortex (BA24)
ENSG00000162753.10	SLC9C2	1.70E-08	0.53	Thyroid
ENSG00000162753.10	SLC9C2	2.50E-08	0.82	Brain - Cortex
ENSG00000162753.10	SLC9C2	2.90E-08	0.68	Pituitary
ENSG00000162753.10	SLC9C2	2.40E-06	1.10	Brain - Putamen (basal ganglia)
ENSG00000162753.10	SLC9C2	5.30E-06	0.23	Lung
ENSG00000162753.10	SLC9C2	2.70E-05	0.54	Brain - Nucleus accumbens (basal ganglia)
ENSG00000120332.11	TNN	2.80E-09	-0.61	Liver

Table S3: Gene-based association using MAGMA in African-Americans

GENE	SYMBOL	P	CHR	START	STOP	NSNPS	NPARAM	N	ZSTAT
ENSG00000142606	MMEL1	9.74E-05	1	2,522,078	2,564,481	223	20	3059	3.73
ENSG00000188807	TMEM201	2.33E-05	1	9,648,932	9,674,935	130	32	3063	4.07
ENSG00000134717	BTF3L4	9.11E-05	1	52,521,797	52,556,388	153	27	2990	3.74
ENSG00000143466	IKBKE	4.05E-05	1	206,643,791	206,670,223	151	52	2999	3.94
ENSG00000163093	BBS5	3.36E-05	2	170,335,688	170,382,432	268	37	3040	3.99
ENSG00000114302	PRKAR2A	8.30E-05	3	48,782,030	48,885,279	240	31	2978	3.77
ENSG00000198218	QRICH1	7.98E-05	3	49,067,140	49,131,796	162	33	2970	3.78
ENSG00000163507	KIAA1524	7.13E-05	3	108,268,716	108,308,491	233	42	2970	3.80
ENSG00000198919	**DZIP3**	7.52E-07	3	108,308,529	108,413,693	509	41	2980	4.81
ENSG00000234284	ZNF879	9.54E-05	5	178,450,753	178,462,065	74	11	3103	3.73
ENSG00000185681	MORN5	4.49E-05	9	124,922,190	124,962,367	234	55	3054	3.92
ENSG00000111203	ITFG2	3.46E-06	12	2,921,788	2,968,957	238	42	2990	4.50
ENSG00000053702	NRIP2	1.46E-05	12	2,934,514	2,944,710	48	14	2995	4.18
ENSG00000110925	CSRNP2	7.03E-06	12	51,454,990	51,477,447	107	26	3040	4.34
ENSG00000187555	USP7	1.15E-05	16	8,985,951	9,058,371	366	40	3045	4.23
ENSG00000175643	RMI2	8.92E-05	16	11,343,476	11,445,619	870	79	3071	3.75
ENSG00000231274	**SBK3**	5.21E-07	19	56,052,023	56,056,909	21	11	3057	4.88
ENSG00000168612	ZSWIM1	7.85E-05	20	44,509,866	44,513,905	15	8	3026	3.78
ENSG00000241973	PI4KA	4.70E-05	22	21,061,979	21,213,705	656	35	3013	3.91
ENSG00000099940	SNAP29	3.98E-05	22	21,213,271	21,245,506	120	22	2974	3.95
ENSG00000099942	**CRKL**	1.63E-06	22	21,271,714	21,308,037	140	18	3013	4.65

Table S4 **MAGMA Gene-based in EA**									
GENE	**SYMBOL**	**P**	**CHR**	**START**	**STOP**	**NSNPS**	**NPARAM**	**N**	**ZSTAT**
ENSG00000099957	**P2RX6**	**7.11E-07**	22	21,364,097	21,383,119	62	15	6410	4.82
ENSG00000154274	C4orf19	1.01E-05	4	37,455,563	37,625,117	789	65	6952	4.26
ENSG00000169764	UGP2	1.54E-05	2	64,068,074	64,118,696	122	18	6993	4.17
ENSG00000106070	GRB10	3.64E-05	7	50,657,760	50,861,159	706	32	7055	3.97
ENSG00000172954	LCLAT1	4.24E-05	2	30,670,092	30,867,091	722	20	7085	3.93
ENSG00000163563	MNDA	5.02E-05	1	158,801,107	158,819,296	24	5	6838	3.89
ENSG00000062598	ELMO2	6.15E-05	20	44,994,688	45,061,704	61	6	6686	3.84
ENSG00000204604	ZNF468	6.53E-05	19	53,341,261	53,360,902	30	3	6002	3.83
ENSG00000173530	TNFRSF10D	7.83E-05	8	22,993,101	23,021,543	62	8	6526	3.78
ENSG00000132437	DDC	8.53E-05	7	50,526,134	50,633,154	578	24	7212	3.76

Table S5: Significant Genes in curated gene sets in EA

Curated gene sets -reactome g beta gamma signalling through plc beta; p=1.36E-06									
GENE	GENE SYMBOLS	CHR	START	STOP	NSNPS	NPARAM	N	ZSTAT	P
ENSG00000127588	G protein subunit gamma 13(GNG13)	16	848,041	850,733	10	2	6434	2.02	0.022
ENSG00000127920	G protein subunit gamma 11(GNG11)	7	93,551,011	93,557,922	15	5	6716	1.84	0.033
ENSG00000078369	G protein subunit beta 1(GNB1)	1	1,716,729	1,822,495	143	13	7102	1.71	0.044
ENSG00000168243	G protein subunit gamma 4(GNG4)	1	235,710,987	235,814,054	298	27	6645	1.48	0.069
ENSG00000167083	G protein subunit gamma transducin 2 (GNGT2)	17	47,280,153	47,287,936	25	7	7133	1.06	0.144
ENSG00000162188	G protein subunit gamma 3(GNG3)	11	62,475,130	62,476,673	1	1	6908	1.06	0.145
ENSG00000176533	G protein subunit gamma 7(GNG7)	19	2,511,217	2,702,707	738	88	6741	1.04	0.150
ENSG00000186469	G protein subunit gamma 2(GNG2)	14	52,292,913	52,446,060	654	62	7177	1.03	0.153
ENSG00000114450	G protein subunit beta 4(GNB4)	3	179,116,990	179,169,378	113	15	6729	0.98	0.163
ENSG00000149782	phospholipase C beta 3(PLCB3)	11	64,018,995	64,036,622	27	9	7176	0.96	0.168
ENSG00000127928	G protein subunit gamma transducin 1 (GNGT1)	7	93,220,885	93,540,577	823	46	6898	0.89	0.187
ENSG00000172380	G protein subunit gamma 12(GNG12)	1	68,167,149	68,299,150	298	24	6949	0.86	0.194
ENSG00000182621	phospholipase C beta 1(PLCB1)	20	8,112,824	8,949,003	2659	164	6923	0.83	0.202
ENSG00000069966	G protein subunit beta 5(GNB5)	15	52,413,117	52,483,566	207	18	6914	0.58	0.280
ENSG00000167414	G protein subunit gamma 8(GNG8)	19	47,137,333	47,137,942	3	2	6856	0.29	0.387

ENSG00000242616	G protein subunit gamma 10(GNG10)	9	114,423,615	114,432,526	37	7	6913	0.27	0.393
ENSG00000174021	G protein subunit gamma 5(GNG5)	1	84,964,008	84,972,248	27	7	7041	0.15	0.439
ENSG00000172354	G protein subunit beta 2(GNB2)	7	100,271,154	100,276,797	6	3	7019	-0.03	0.512
ENSG00000137841	phospholipase C beta 2(PLCB2)	15	40,570,377	40,600,136	53	8	6521	-0.79	0.785
ENSG00000111664	G protein subunit beta 3(GNB3)	12	6,949,118	6,956,557	7	2	6684	-1.03	0.847
Curated gene sets-reactome glucagon type ligand receptors; p=2.6E-06									
GENE	GENE SYMBOLS	CHR	START	STOP	NSNPS	NPARAM	N	ZSTAT	P
ENSG00000127588	G protein subunit gamma 13(GNG13)	16	848,041	850,733	10	2	6434	2.02	0.022
ENSG00000106018	vasoactive intestinal peptide receptor 2(VIPR2)	7	158,820,866	158,937,649	423	22	7089	1.94	0.026
ENSG00000010310	gastric inhibitory polypeptide receptor(GIPR)	19	46,171,502	46,186,982	34	6	7056	1.85	0.032
ENSG00000127920	G protein subunit gamma 11(GNG11)	7	93,551,011	93,557,922	15	5	6716	1.84	0.033
ENSG00000078369	G protein subunit beta 1(GNB1)	1	1,716,729	1,822,495	143	13	7102	1.71	0.044
ENSG00000168243	G protein subunit gamma 4(GNG4)	1	235,710,987	235,814,054	298	27	6645	1.48	0.069
ENSG00000115263	glucagon(GCG)	2	162,999,392	163,008,914	10	1	6671	1.14	0.127
ENSG00000146469	vasoactive intestinal peptide(VIP)	6	153,071,933	153,080,900	12	4	6976	1.11	0.133
ENSG00000080293	secretin receptor(SCTR)	2	120,197,419	120,282,070	279	19	6953	1.10	0.136
ENSG00000167083	G protein subunit gamma transducin 2(GNGT2)	17	47,280,153	47,287,936	25	7	7133	1.06	0.144

ENSG00000162188	G protein subunit gamma 3(GNG3)	11	62,475,130	62,476,673	1	1	6908	1.06	0.145
ENSG00000176533	G protein subunit gamma 7(GNG7)	19	2,511,217	2,702,707	738	88	6741	1.04	0.150
ENSG00000186469	G protein subunit gamma 2(GNG2)	14	52,292,913	52,446,060	654	62	7177	1.03	0.153
ENSG00000114450	G protein subunit beta 4(GNB4)	3	179,116,990	179,169,378	113	15	6729	0.98	0.163
ENSG00000127928	G protein subunit gamma transducin 1(GNGT1)	7	93,220,885	93,540,577	823	46	6898	0.89	0.187
ENSG00000172380	G protein subunit gamma 12(GNG12)	1	68,167,149	68,299,150	298	24	6949	0.86	0.194
ENSG00000114812	vasoactive intestinal peptide receptor 1(VIPR1)	3	42,530,791	42,579,059	105	21	6819	0.81	0.209
ENSG00000069966	G protein subunit beta 5(GNB5)	15	52,413,117	52,483,566	207	18	6914	0.58	0.280
ENSG00000112164	glucagon like peptide 1 receptor(GLP1R)	6	39,016,574	39,055,519	176	21	6780	0.53	0.299
ENSG00000070031	secretin(SCT)	11	626,431	627,143	1	1	6611	0.30	0.381
ENSG00000167414	G protein subunit gamma 8(GNG8)	19	47,137,333	47,137,942	3	2	6856	0.29	0.387
ENSG00000242616	G protein subunit gamma 10(GNG10)	9	114,423,615	114,432,526	37	7	6913	0.27	0.393
ENSG00000174021	G protein subunit gamma 5(GNG5)	1	84,964,008	84,972,248	27	7	7041	0.15	0.439
ENSG00000118702	growth hormone releasing hormone(GHRH)	20	35,879,489	35,890,238	23	4	6938	0.10	0.462
ENSG00000172354	G protein subunit beta 2(GNB2)	7	100,271,154	100,276,797	6	3	7019	-0.03	0.512
ENSG00000141433	adenylate cyclase activating polypeptide 1(ADCYAP1)	18	904,944	912,173	17	7	6638	-0.04	0.516

ENSG00000106128	growth hormone releasing hormone receptor(GHRHR)	7	30,978,284	31,032,869	168	28	6726	-0.52	0.698
ENSG00000159224	gastric inhibitory polypeptide(GIP)	17	47,035,916	47,045,958	27	6	6576	-0.62	0.733
ENSG00000065325	glucagon like peptide 2 receptor(GLP2R)	17	9,725,523	9,795,419	231	16	6788	-0.78	0.784
ENSG00000111664	G protein subunit beta 3(GNB3)	12	6,949,118	6,956,557	7	2	6684	-1.03	0.847
ENSG00000087460	GNAS complex locus(GNAS)	20	57,414,773	57,486,247	200	38	7118	-1.27	0.898
ENSG00000078549	ADCYAP receptor type I(ADCYAP1R1)	7	31,092,076	31,151,089	140	27	6599	-1.55	0.939
Curated gene sets-reactome prostacyclin signalling through prostacyclin receptor; p=3.6E-06									
GENE	GENE SYMBOLS	CHR	START	STOP	NSNPS	NPARAM	N	ZSTAT	P
ENSG00000127588	G protein subunit gamma 13(GNG13)	16	848,041	850,733	10	2	6434	2.02	0.022
ENSG00000127920	G protein subunit gamma 11(GNG11)	7	93,551,011	93,557,922	15	5	6716	1.84	0.033
ENSG00000078369	G protein subunit beta 1(GNB1)	1	1,716,729	1,822,495	143	13	7102	1.71	0.044
ENSG00000168243	G protein subunit gamma 4(GNG4)	1	235,710,987	235,814,054	298	27	6645	1.48	0.069
ENSG00000167083	G protein subunit gamma transducin 2(GNGT2)	17	47,280,153	47,287,936	25	7	7133	1.06	0.144
ENSG00000162188	G protein subunit gamma 3(GNG3)	11	62,475,130	62,476,673	1	1	6908	1.06	0.145
ENSG00000176533	G protein subunit gamma 7(GNG7)	19	2,511,217	2,702,707	738	88	6741	1.04	0.150
ENSG00000186469	G protein subunit gamma 2(GNG2)	14	52,292,913	52,446,060	654	62	7177	1.03	0.153
ENSG00000114450	G protein subunit beta 4(GNB4)	3	179,116,990	179,169,378	113	15	6729	0.98	0.163

ENSG00000127928	G protein subunit gamma transducin 1(GNGT1)	7	93,220,885	93,540,577	823	46	6898	0.89	0.187
ENSG00000172380	G protein subunit gamma 12(GNG12)	1	68,167,149	68,299,150	298	24	6949	0.86	0.194
ENSG00000069966	G protein subunit beta 5(GNB5)	15	52,413,117	52,483,566	207	18	6914	0.58	0.280
ENSG00000167414	G protein subunit gamma 8(GNG8)	19	47,137,333	47,137,942	3	2	6856	0.29	0.387
ENSG00000242616	G protein subunit gamma 10(GNG10)	9	114,423,615	114,432,526	37	7	6913	0.27	0.393
ENSG00000174021	G protein subunit gamma 5(GNG5)	1	84,964,008	84,972,248	27	7	7041	0.15	0.439
ENSG00000172354	G protein subunit beta 2(GNB2)	7	100,271,154	100,276,797	6	3	7019	-0.03	0.512
ENSG00000160013	prostaglandin I2 (prostacyclin) receptor (IP)(PTGIR)	19	47,123,725	47,128,375	6	1	7076	-0.79	0.785
ENSG00000111664	G protein subunit beta 3(GNB3)	12	6,949,118	6,956,557	7	2	6684	-1.03	0.847
ENSG00000087460	GNAS complex locus(GNAS)	20	57,414,773	57,486,247	200	38	7118	-1.27	0.898

Curated gene sets-reactome adp signalling through p2ry12; p=4.33E-06

GENE	GENE SYMBOLS	CHR	START	STOP	NSNPS	NPARAM	N	ZSTAT	P
ENSG00000127588	G protein subunit gamma 13(GNG13)	16	848,041	850,733	10	2	6434	2.02	0.022
ENSG00000127920	G protein subunit gamma 11(GNG11)	7	93,551,011	93,557,922	15	5	6716	1.84	0.033
ENSG00000078369	G protein subunit beta 1(GNB1)	1	1,716,729	1,822,495	143	13	7102	1.71	0.044
ENSG00000168243	G protein subunit gamma 4(GNG4)	1	235,710,987	235,814,054	298	27	6645	1.48	0.069
ENSG00000167083	G protein subunit gamma transducin 2(GNGT2)	17	47,280,153	47,287,936	25	7	7133	1.06	0.144

ENSG00000162188	G protein subunit gamma 3(GNG3)	11	62,475,130	62,476,673	1	1	6908	1.06	0.145
ENSG00000176533	G protein subunit gamma 7(GNG7)	19	2,511,217	2,702,707	738	88	6741	1.04	0.150
ENSG00000186469	G protein subunit gamma 2(GNG2)	14	52,292,913	52,446,060	654	62	7177	1.03	0.153
ENSG00000114450	G protein subunit beta 4(GNB4)	3	179,116,990	179,169,378	113	15	6729	0.98	0.163
ENSG00000127928	G protein subunit gamma transducin 1(GNGT1)	7	93,220,885	93,540,577	823	46	6898	0.89	0.187
ENSG00000065135	G protein subunit alpha i3(GNAI3)	1	110,091,233	110,136,975	74	11	7125	0.87	0.191
ENSG00000172380	G protein subunit gamma 12(GNG12)	1	68,167,149	68,299,150	298	24	6949	0.86	0.194
ENSG00000069966	G protein subunit beta 5(GNB5)	15	52,413,117	52,483,566	207	18	6914	0.58	0.280
ENSG00000114353	G protein subunit alpha i2(GNAI2)	3	50,263,724	50,296,787	36	9	6926	0.37	0.354
ENSG00000167414	G protein subunit gamma 8(GNG8)	19	47,137,333	47,137,942	3	2	6856	0.29	0.387
ENSG00000242616	G protein subunit gamma 10(GNG10)	9	114,423,615	114,432,526	37	7	6913	0.27	0.393
ENSG00000174021	G protein subunit gamma 5(GNG5)	1	84,964,008	84,972,248	27	7	7041	0.15	0.439
ENSG00000127955	G protein subunit alpha i1(GNAI1)	7	79,763,271	79,848,718	212	28	6907	0.11	0.455
ENSG00000172354	G protein subunit beta 2(GNB2)	7	100,271,154	100,276,797	6	3	7019	-0.03	0.512
ENSG00000111664	G protein subunit beta 3(GNB3)	12	6,949,118	6,956,557	7	2	6684	-1.03	0.847
ENSG00000169313	purinergic receptor P2Y12(P2RY12)	3	151,055,168	151,102,600	136	18	6577	-1.59	0.944

Table S6: Replication in three independent samples: beta (standard error), p-value, and direction of effect sizes for meta-analyses across COGA and replication samples

SNP	Effect.allele	Yale-Penn AA	SAGE AA	AA meta	SAGE EA	EA meta	EA+AA meta
rs1890881	T	-0.13 (0.10), 0.20	-0.20 (0.17), 0.25	0.11 (0.06), 0.07, +--	0.09 (0.13), 0.49	0.26 (0.07), 4.94E-4, ++	0.18 (0.05), 2.15E-4, ++--+,
rs34066662	T	-0.03 (0.09), 0.78	-0.07 (0.15), 0.67	-0.23 (0.05), 2.46E-5, ---	n/a	n/a	n/a
rs58801820	T	-0.03 (0.09,), 0.74	-0.06 (0.15), 0.67	-0.23 (0.05), 2.07E-5, ---	n/a	n/a	n/a
chr5:141988181	GAA	-0.34 (0.18), 0.06	0.12 (0.28), 0.66	-0.30 (0.10), 2.77E-3, -+-	n/a	n/a	n/a
rs75168521	T	0.03 (0.11), 0.78	-0.03 (0.17), 0.85	-0.24 (0.06), 1.23E-4, -+-	n/a	n/a	n/a
rs78886294	T	0.03 (0.11), 0.79	-0.03 (0.17), 0.84	-0.24 (0.06), 1.42E-4, -+-	n/a	n/a	n/a

SNP = single nucleotide polymorphism
SAGE = Study of Addiction: Genes and Environment, dbGaP phs000092.v1.p1
Yale-Penn = dbGaP phs000425.v1.p1
EA = European American
AA = African American
AA meta-analysis includes COGA, SAGE and Yale-Penn AA
EA meta-analysis includes COGA and SAGE EA
EA+AA meta-analysis includes COGA and SAGE (EA and AA) as well as Yale-Penn AA
For meta-analyses across COGA and replication samples, direction is denoted by + or – for sign of effect size in COGA AA, COGA EA Yale-Penn AA, SAGE AA, SAGE EA

Table S7: *ANYDEP* Findings in the EA, AA and EA+AA analysis for replicated alcohol dependence loci, findings from Lai et al., (accompanying paper) and a prior study of "ANYDEP" and quantitative count of alcohol and drug dependence (quantdep) in a subset of COGA.

SNP	Reference Allele	EA (beta, se, p)	AA (beta, se, p)	EA+AA (beta, se, p)
From Lai et al. (accompanying paper)				
Chr 1 (rs61826952)	A	0.33 (0.08), 7.37E-5	0.41 (0.10), 3.18E-5	0.37 (0.07), 5.78E-5
Chr 2 (rs7597960)	A	0.15 (0.05), 3.15E-3	0.05 (0.07), 0.47	0.12 (0.04), 6.55E-3
Chr 4 (rs1229984)	T	-0.042 (0.12), 5.93E-4	-0.58 (0.26), 0.03	-0.45 (0.12), 1.1E-4
Chr 8 (rs188227250)	A	0.45 (0.11) 5.89E-5	-0.17 (0.23), 0.45	0.34 (0.10), 1.33E-3
Chr 15 (rs1912461)	T	-0.63 (0.16), 9.2E-5	0.18 (0.22), 0.42	-0.36 (0,14), 8.41E-3
From Wetherill et al. (2015; n=1,770-2,183)				
Chr 2 (rs2952621)	T	0.09 (0.04), 0.03	-0.08 (0.06), 0.19	0.04 (0.04), 0.25
Chr 18 (rs2567261)	T	0.13 (0.08), 0.10	0.13 (0.08), 0.09	0.13 (0.06), 0.02

Note: All effect sizes for *ANYDEP* were in the same direction as effects in Lai et al., and [23]