HHS Public Access

Author manuscript

Psychol Addict Behav. Author manuscript; available in PMC 2018 May 01.

Published in final edited form as:

Psychol Addict Behav. 2017 May ; 31(3): 354–366. doi:10.1037/adb0000270.

Genetic Variation in the Exome: Associations with Alcohol and Tobacco Co-Use

Jacqueline M. Otto[1], **Ian R. Gizer**[1], **Jarrod M. Ellingson**[1], and **Kirk C. Wilhelmsen**[2,3]

[1]Department of Psychological Sciences, University of Missouri–Columbia

[2]Department of Genetics and Neurology, University of North Carolina at Chapel Hill

[3]Renaissance Computing Institute (RENCI)

Abstract

Shared genetic factors represent one underlying mechanism thought to contribute to high rates of alcohol and tobacco co-use and dependence. Common variants identified by molecular genetic studies tend to confer only small disease risk, and rare protein-coding variants are posited to contribute to disease risk, as well. However, given that genotyping technologies allowing for their inclusion in association studies have only recently become available, the magnitude of their contribution is poorly understood. The current study examined genetic variation in protein-coding regions (i.e., the exome) for associations with measures of lifetime alcohol and tobacco co-use. Participants from the UCSF Family Alcoholism Study (N = 1,862) were genotyped using an exome-focused genotyping array, and assessed for DSM-IV diagnoses of alcohol and tobacco dependence and quantitative consumption measures using a modified version of the Semi-Structured Assessment for the Genetics of Alcoholism. Analyses included single variant, gene-based, and pathway-based tests of association. One *EMR3* variant and a pathway related to genes upregulated in mesenchymal stem cells during the late phase of adipogenesis met criteria for statistical significance. Suggestive associations were consistent with previous findings from studies of substance use and dependence, including variants in the *CHRNA5 - CHRNA3 - CHRNB4* gene cluster with cigarettes smoked per day. Further, several variants and genes demonstrated suggestive association across phenotypes, suggesting that shared genetic factors may underlie risk for increased levels of alcohol and tobacco use, as well as psychopathology more broadly, providing insight into our understanding of the genetic architecture underlying these traits.

Keywords

exome; alcohol; tobacco; genetic association; gene variants

Alcohol and tobacco are frequently used in combination, and rates for past-year co-use and dependence in nationally representative samples are relatively high. Estimates from the 2001–2002 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)

Correspondence concerning this article should be directed to: Jacqueline M. Otto, Department of Psychological Sciences, University of Missouri–Columbia. 210 McAlester Hall, Columbia, MO 65211. Contact: jmohfb@mail.missouri.edu.

indicated that 21.7% of all U.S. individuals aged 18 and older reported past-year alcohol and tobacco use, and 2.9% of U.S. adults met criteria for both an alcohol use disorder and nicotine dependence (Falk, Yi, & Hiller-Sturmhofel, 2006). Environmental, psychosocial, and biological factors contribute to the high rates of co-use and dependence, and substantial research has examined their independent contributions and the interplay among them.

Behavioral genetics studies can provide insight into the approximate influence of environmental and heritable factors that contribute to these traits. For example, studies of this kind have demonstrated that shared and unique environmental factors account for 10% and 39%, respectively, of the variation in liability to alcohol dependence (Verhulst, Neale, & Kendler, 2015), while these estimates can differ by gender and ethnicity for nicotine dependence (Li, 2006; Sartor et al., 2015). Further, 49% of the variation in liability to alcohol dependence can be attributed to heritable influences (Verhulst, et al., 2015), while heritability estimates (h^2) for nicotine dependence range from 0.46 for females and 0.59 for males (Li, 2006), though it should be noted that the relative proportions of genetic and shared and unique environmental influences vary and change from adolescence to adulthood (e.g., Kendler, Schmitt, Aggen, & Prescott, 2008). In addition to univariate behavior genetic analyses examining the proportions of genetic and environmental influences on a single trait, multivariate behavior genetic analyses examining the etiologic overlap among multiple traits suggest common or shared genetic and environmental factors may be contributing to the increased consumption levels and dependence rates across substances within families (e.g., Palmer et al., 2012). The overlap in environmental and genetic factors implicated in the liability for alcohol and tobacco use and dependence is thus indicative of a shared etiology and common mechanisms of action.

Despite the described findings from twin studies, molecular genetic studies, which focus on the identification of the specific genes or genomic regions involved in the etiology of a trait, have been less consistent in their conclusions about specific causal regions or single gene variants implicated in alcohol and tobacco use and dependence. Genome-wide association studies (GWASs) and meta-analytic GWASs conducted on number of cigarettes smoked per day and nicotine dependence have identified multiple common variants within the nAChR genes on chromosome 15q24–q25 (e.g., Chen et al., 2012; Thorgeirsson et al., 2008; Tobacco and Genetics Consortium, 2010). In contrast, meta-analyses conducted on GWA studies of alcohol consumption and dependence have been less informative, largely due to smaller sample sizes (Kapoor et al., 2013; Wang et al., 2011). Studies on comorbid alcohol and tobacco dependence have also been inconsistent. GWASs have identified some risk loci associated with comorbid DSM-IV alcohol and nicotine dependence, including variants in *KIAA1409* and near *MARK1* and *DDX6* (Lind et al., 2010) and within the *SH3BP5-NR2C2* region on chromosome 3 (Zuo et al., 2012). However, these relations have been difficult to replicate and the overall amount of variance explained by individual regions or variants (typically less than 2%) falls short of heritability estimates from twin studies.

A potential limitation of GWAS and linkage studies originate from the types of genetic variants that they are designed to capture in analysis. GWAS were originally designed to identify common variation in the genome (i.e., variants with a minor allele frequency [MAF] ≥0.05) associated with a trait of interest. As a result, GWAS are ideal for testing whether

complex disease—disease caused by many genes, none of which are necessary nor sufficient to cause the disease—can be attributed to commonly-occurring variants. Variants with lower frequency (0.005 < MAF < 0.05) can be detected by linkage studies, but only if their effect size is large enough. However, many types of allelic variation, including low-frequency point mutations and structural variation, are thought to influence disease risk (Manolio et al., 2009). With respect to the former, which is the focus of the present report, it has been suggested that numerous rare variants (MAF < 1%) of moderate to small effect may be contributing, in part, to the discrepancy between the additive effects of individual common variants and twin heritability estimates, i.e., the 'missing heritability' of complex disease (Bodmer & Bonilla, 2008; Manolio et al., 2009).

Population genetics theories describe numerous reasons as to why rare variants—specifically rare variants in protein-coding regions (exons) of the genome—are considered to be important in explaining disease risk, although it should be noted that both coding and non-coding (e.g., regulatory) genetic variation is likely to contribute to these phenotypes (Schork et al., 2013). The majority of single-nucleotide variants (SNVs) within coding regions are rare (MAF < 0.05), rather than common (Nelson et al., 2012), and more likely to be functional (Marth et al., 2011). Functional variants include, among others, nonsynonymous mutations or polymorphisms that result in amino acid sequence change and affect protein function, compared to synonymous mutations whose amino acid product is the same. Up to 70% of rare variants are associated with reduced survival, and thus are subject to strong purifying selection (Kryukov, Pennacchio, & Sunyaev, 2007). Therefore, rare variants of large effect are not often observed for common, complex (i.e., non-Mendelian) traits, and are unlikely to play a major role in their etiology. Rather, it is likely that together with other forms of genetic variation, rare variants with low to moderate effect sizes likely function in an additive fashion to increase disease risk (Pritchard, 2001; Pritchard & Cox, 2002). Given theoretical arguments that rare variants may contribute to the “missing heritability” problem, analyzing rare variant associations in coding regions may provide valuable insight about complex disease etiology.

Until recently, DNA sequencing was the only method available for evaluating the effects of rare exonic variation on complex phenotypes. Because of the expense involved, this method is limited in terms of the sample sizes that can be achieved. Although genotyping strategies are only capable of measuring typed variation (in contrast to sequencing), the methodology is less expensive given that it examines only a subset of the genome. To this end, an exome chip genotyping array was developed in order to allow for larger sample sizes and increase power to detect associations with rare variants. The approach of exome chip genotyping is similar to that used for GWASs, which tests anywhere from 500,000 to 7,000,000 markers with MAF > 0.05 across the entire genome, and therefore includes non-coding DNA (Attia et al., 2009). In contrast, exome chip genotyping arrays contain markers exclusively from protein-coding regions (~180,000 exons) of approximately 20,000 genes, which comprise about 1% of the total genome. Given that many of the rare variants associated with Mendelian disease are found in protein-coding regions, genotyping the exome lends itself as a novel approach to further investigating the genetic etiology of complex disease and more specifically, alcohol and tobacco co-use.

To date, two studies have been conducted on the relationship between exonic variation and substance use traits (Vrieze et al., 2014; Zuo et al., 2013), although only one of these utilized an exome chip genotyping approach (Vrieze et al., 2014). Both studies restricted their analyses to nonsynonymous variants in the exome, but failed to yield any significant findings. Analysis of exonic variants obtained from a genome-wide genotyping array reported 22 nominal associations with alcohol dependence (Zuo et al., 2013), while analysis of exonic variants obtained from an exome chip genotyping array reported no associations with alcohol, tobacco, and other drug use phenotypes (Vrieze et al., 2014). While there appears to be some basis for disease-causing loci in the exome, findings have been inconclusive and methods have been limited in terms of their focus on the types of variants tested.

The current study sought to extend findings from previous research and determine whether rare variation within protein-coding regions is associated with lifetime measures of alcohol and tobacco co-use and dependence using an exome chip genotyping microarray. Analyses were conducted at the level of single variants, genes, and pathways (i.e., gene sets), without *a priori* hypotheses regarding relevant genes and pathways, as the specific genetic etiology of alcohol and tobacco use disorders is still poorly understood, and has been shown to involve biological processes both in and outside the central nervous system. Therefore, a wide range of genes and pathways were selected for inclusion, as a primary focus of the current study was to ascertain whether low-frequency variants of modest effect could be detected, and these associations, if present, should be fairly robust to large corrections for multiple testing.

Method

Sample

Participants with both genotype and phenotype data ($N = 2,524$) from a larger study on the genetics of alcohol dependence susceptibility, the UCSF Family Alcoholism Study, were included in analyses for the current study (see Vieten, Seaton, Feiler, & Wilhelmsen, 2004 for a detailed description of the original study). The sample was composed of 1,218 small family pedigrees, ranging in size from 3–20 individuals. Individuals were recruited from the community and invited to participate if they met criteria for a DSM-IV lifetime alcohol dependence diagnosis, and had parents and/or siblings who agreed to take part in the study. Recruitment from a community population ensured that findings and interpretation would not be limited to a treatment-seeking sample. Exclusion criteria for individuals sampled from the community consisted of: (1) diagnosis of a past 6-month psychiatric condition other than depression or anxiety disorder; (2) past 6-month drug dependence diagnosis for substances other than alcohol, tobacco, or marijuana; (3) serious medical illness; and/or (4) being a non-English speaker. Family members were not subject to any strict inclusion or exclusion criteria, and many of these individuals reported varying levels of substance use, as well. Institutional IRB committees approved all study procedures, and individuals provided informed consent prior to participation.

Measures

Semi-Structured Assessment for the Genetics of Alcoholism—A modified version of the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA; Bucholz et al., 1994) was administered to assess for alcohol abuse and dependence, nicotine and other drug use, demographic information, and medical and psychiatric diagnoses. The SSAGA is an empirically validated polydiagnostic semi-structured psychiatric interview with high test-retest reliability, ranging from 0.70 to 0.90 for specific substance dependence diagnoses (Bucholz et al., 1994).

Phenotypes—The following phenotypes were used in analyses to test for association with rare variation in the exome: (1) lifetime DSM-IV alcohol and tobacco dependence diagnoses (*AlcDep* and *NicDep*, respectively), (2) maximum number of drinks consumed per day during heaviest period of use (*MaxDrinks*), (3) number of cigarettes smoked per day during heaviest period of use (*CPD*), and (4) an interaction product term for a continuous measure of alcohol and tobacco co-use *(DrinksCPD),* as well as a dichotomous measure of lifetime alcohol and tobacco co-dependence (*AlcNicDep*). The alcohol dependence diagnoses were created using a combination of data from the SSAGA and best estimate procedures implemented by the study investigators (Vieten et al., 2004). The remaining phenotypes were created strictly from the SSAGA, which uses numerous skip-out items in assessing heavier levels of use. Participants who responded negatively to whether they have smoked 100 cigarettes in their lifetime are designated as non-smokers, not administered the full tobacco use section, and classified as not meeting criteria for NicDep. Thus, while NicDep diagnoses could be assigned to all participants, regardless of smoking history, CPD was only assessed for individuals who endorsed smoking at least 100 cigarettes in their lifetime (68.6% of total sample, $n = 1,277$). Similarly, MaxDrinks was assessed only for individuals who endorsed drinking every day for a week or more (69.3% of total sample, $n = 1,290$), but AlcDep diagnoses could be assigned to all participants.

Genotyping

Quality control—The Affymetrix Axiom Exome Genotyping Array (Affymetrix, Inc.) was used for rare variant genotyping and analysis, capturing rare variants with a MAF > 0.005%. This chip contains more than 300,000 coding SNVs, including synonymous (no change in protein sequence), non-synonymous (change in amino acid protein sequence), splice (joining of exons during or following transcription), and stop codon (nucleotide triplet that signals termination of translation) variants. The array also contains approximately 30,000 simple and complex indels (i.e., insertions and deletions) corresponding to the draft Phase 1 1,000 Genomes Project (The 1000 Genomes Project Consortium, 2012), and a variety of non-coding annotations, including variants in intergenic regions, introns, variants upstream (toward 5’ end of the DNA strand) and downstream (toward 3’ end of the DNA strand) from the gene transcript, and variants in untranslated regions (UTRs), both upstream from the start codon (UTR-5) and downstream from termination codon (UTR-3). There are also over 5,000 SNPs that have shown significant associations with a variety of complex traits in one or more GWA studies. Assays and genotype calls were made according to protocols provided by Affymetrix (Affymetrix, Inc.).

A number of standard genotyping quality control steps were conducted on the initial dataset of 295,988 SNVs (Anderson et al., 2010) using PLINK 1.07 (Purcell et al., 2007) in order to assess genotyping sample quality and accuracy, as well as sample identities. Degree of relatedness estimations were conducted using the Pedigree Relationship Statistical Test software (PREST; Sun, Wilder, & McPeek, 2002), which ensures that the reported familial relationships are accurate, and also identifies unreported familial relationships. Discrepant self-reported and genetic identities (pedigree errors) were resolved if their familial relation could be established; 36 unresolved discrepant individuals were excluded. In addition, gender checks of sample identities were evaluated, and 6 individuals with unresolved discrepant sex codes were excluded.

Five individuals and 11,504 SNVs were removed due to low genotype call rates (individuals or SNVs with call rates < 95%), and 207,980 monomorphic SNVs (MAF = 0.0; all individuals carried the same genotype at these loci) were excluded. Following tests for deviations from Hardy-Weinberg equilibrium, 481 SNVs with a *p*-value less than the cutoff of 1e-05 were removed. Calculation of Mendelian errors and heterozygosity rates per individual resulted in the removal of 10 individuals. There were 207 individuals with duplicate samples; the overall concordance rate for genotype calls across duplicate samples was approximately 98%, and SNVs with discordant genotype calls across these duplicate samples were set to missing. All duplicate markers, mitochondrial markers and markers on sex chromosomes were excluded. Allele frequencies were calculated with a subset of unrelated European-ancestry individuals and cross-referenced with the European sample for the 1,000 Genomes Project (The 1000 Genomes Project Consortium, 2012). This resulted in the exclusion of 1,108 SNVs, whose allele frequencies differed more than 0.20 from this reference panel. Following the completion of quality control assessment, 1,862 individuals and 72,884 SNVs remained in the final dataset used for analyses.

The final sample of 1,862 individuals from 778 families was predominantly female (62%; n = 1153) and self-identified as Caucasian (93%), with a mean age of 49.2 (SD = 13.2) years for all family members. The mean reported education level was 14.5 years (SD = 2.9), and median annual income was approximately $48,000. Fifteen percent (n = 369) had been diagnosed with DSM-IV tobacco dependence only, 18% (n = 464) with DSM-IV alcohol dependence only, and 35% (n = 880) with both. The average number of maximum drinks consumed per day during the heaviest period of use was 12.2 (SD = 11.3) and 14.3 (SD = 11.5), for the total sample, and those diagnosed with DSM-IV alcohol dependence, respectively. The average number of cigarettes smoked per day during the heaviest period of use was 22.2 (SD = 14.7) and 25.0 (SD = 14.8), for the total sample, and those diagnosed with DSM-IV tobacco dependence, respectively.

Data Analysis

Ancestry estimations were calculated from variants (MAF≥0.01) using principal components analysis (Price, Patterson, Plenge, Weinblatt, Shadick, & Reich, 2006) within the Genome-wide Complex Trait Analysis software (GCTA; Yang, Lee, Goddard, & Visscher, 2011). The resulting ancestry estimates (i.e., first four eigenvectors) were used as covariates in subsequent association tests to control for possible population substructure. Examination of

a scree plot (Supplementary Figure 1) and visual inspection of scatterplots for the first four eigenvectors confirmed the existence of four significant components. These eigenvectors correlated highly with self-reported ancestry. The first eigenvector was correlated with European ancestry: r = 0.718, and the second eigenvector was correlated with African ancestry (excluding European ancestry individuals): r = 0.792. The third eigenvector was correlated with East Asian ancestry: r = 0.936, as estimated within the 1,000 Genomes Project due to the low proportion of East Asian ancestry in the UCSF sample. Finally, the fourth eigenvector was moderately correlated with admixed American ancestry: r = 0.316, with the relatively lower correlation reflecting the high levels of European ancestry admixture in these populations.

Single variant association tests—Single variant association tests were conducted for SNVs with MAF≥0.01, which included the approximate 5,000 tag SNPs included from previous GWAS and additional consortia efforts. Analyses were conducted using the Efficient Mixed Model Association eXpedited (EMMAX; Kang et al., 2010) software package, which uses a variance component mixed-model approach in order to account for possible population stratification due to ancestry and familial relatedness; these analyses included sex, age, age-squared, and the first four eigenvectors generated from the principal components analysis as covariates for the four univariate phenotypes: AlcDep, NicDep, MaxDrinks, and CPD. The interaction term DrinksCPD was modeled with sex, age, age-squared, and the first four eigenvectors as covariates, as well as the main effects of MaxDrinks and CPD. The dichotomous AlcNicDep interaction phenotype measured presence or absence of both dependence diagnoses. Main effects were not included in the latter model, as each perfectly predicted the co-dependence phenotype, but analyses included sex, age, age-squared, and the first four eigenvectors as covariates (Price et al., 2006). Given that the model assumptions for single variant association tests depend on allele count, such tests are typically underpowered to find significant associations with rare variants, and evaluating the significance of rare variant associations with a dependent variable must be conducted in the context of a single gene or pathway.

Gene-based association tests—Gene-based association tests expand on single variant approaches by evaluating the influence of multiple variants within a single gene on a phenotype. The SKAT-O test (Lee et al., 2012) capitalizes on the strengths of both the burden test (Asimit & Zeggini, 2010) and the non-burden sequence kernel association test (SKAT; Wu et al., 2011) to create an optimally powerful association test for use in gene-based analyses. The burden test assumes causal variants with identical effects, while SKAT allows for non-causal variants or mixed causal variant effects. As such, the SKAT-O test statistic is a weighted combination of the SKAT and burden test statistics, with the optimal weight derived from the correlations of the regression coefficients. Thus, the SKAT-O test statistic was used for gene-based analyses within EMMAX as described for the single variant analyses. Phenotypes and covariates were modeled in an identical manner as for single variant analyses, and tests were conducted on all variants, regardless of MAF or annotation (e.g., coding vs. non-coding variants).

Pathway-based association tests—As an extension of gene-based tests, pathway-based tests consider the effects of multiple variants in groups of genes within a single biological pathway or gene set. Biological pathways can be broadly defined as groups of biologically-related genes, that are either (1) organized to represent a common direction and regulation towards a specified outcome, such as the metabolic pathway involved in gluconeogenesis, or (2) organized to represent shared relationships among elements such as genes or gene products, and may lack a common, specific outcome (Ramanan, Shen, Moore, & Saykin, 2012; Wang, Li, & Hakonarson, 2010). Online databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa & Goto, 2000) provide complete graphical diagrams of the biological processes associated with single genes and how genes interact within a pathway to produce specific outcomes. The Gene Ontology project (GO; Ashburner et al., 2000; The Gene Ontology Consortium, 2015) uses empirical data derived from experimental or computational analysis (e.g., protein levels ascertained through Western blot) to annotate and group individual genes into sets (i.e., gene sets) based on components of their biological function. Biological functions belong to one of three ontologies: cellular component (where gene products are active, e.g., nuclear inner membrane), molecular function (the functions or activities of a gene product, e.g., transporter activity), or biological process (pathways and larger processes made up of the functions/activities of multiple gene products, e.g., signal transduction).

One method for conducting pathway analyses begins with the raw genotype data for genes within a biological pathway or gene set, and then collapses across all variants contained within these genes. The SKAT-O test statistic can be extended to this approach, and thus, pathway analyses were conducted using SKAT-O within the EMMAX framework in an identical manner as described for the single variant and gene-based analyses.

For the purposes of the present study, the units of analysis for all pathway-based tests were annotated gene sets belonging to the C2: curated gene sets (n = 4,722) and C5: GO gene sets (n = 1,454) included in the Molecular Signatures Database (MSigDB; Subramanian et al., 2005). MSigDB excludes certain GO gene sets if (1) they belong to a very broad category, (2) contain fewer than 10 genes, or (3) if their members are identical to members of another gene set. The C2 gene set collection includes metabolic and signaling pathways from online databases such as KEGG (Kanehisa & Goto, 2000), as well as expression signatures of chemical and genetic perturbations. The C5 gene set collection includes pathways grouped by their GO project term belonging to one of three ontologies (cellular component, molecular function, or biological process) and association to human genes. Each of the C2 and C5 (N = 6,176) pathways were tested for association with the six alcohol and tobacco use phenotypes.

Results

Genomic inflation factor (lambda) values were evaluated to assess for deviations from normality of the test statistics. Lambda values ([1-median p-value across all tests conducted]/0.5) ranged from 1.00 to 1.02 for all univariate phenotypes, with the exception of MaxDrinks (λ = 1.04). A square-root transformation of MaxDrinks was used to achieve normality, yielding a lambda of 1.02, and analyses were conducted with this transformed

variable. Lambda values for DrinksCPD and AlcNicDep were 0.98 and 1.01, respectively. Phenotypic correlations ranged from 0.183 (CPD vs. MaxDrinks) to 0.384 (MaxDrinks vs. AlcDep). For all bivariate correlations among variables at the phenotypic level, see Supplementary Table 1.

Single-variant Association Tests

Single variant association tests were conducted for all variants with MAF≥0.01 (n = 58,652). The Genetic type I error calculator (GEC; Li, Yeung, Cherny, & Sham, 2012) was used to estimate the effective number of independent tests (n = 41,378) by accounting for correlations (i.e., linkage disequilibrium) among variants. The GEC program was also used to compute the significance threshold necessary to control the type I error rate at 0.05. The critical p-value was 1.21 x 10^{-6} (0.05/41,378), and this was set as the threshold for statistical significance. Consistent with genome-wide approaches to relax this critical value by approximately three orders of magnitude (e.g., from 5.0 x 10^{-8} to 5.0 x 10^{-5}) for determining suggestive association, a critical value of $p < 1.21 \times 10^{-3}$ was set as the threshold for suggestive association, both for individual phenotypes as well as cross-phenotypes (i.e., associated with three or more phenotypes). Despite the moderate degree of correlation across phenotypes, a conservative p-value cut-off was retained in order to minimize false positive findings, given the large number of single variants tested. A low-frequency missense variant in *EMR3* was significantly associated with DrinksCPD (rs117374816[C]), MAF = 0.019, b = -23.227, standard error (SE) = 4.73, $p = 1.05 \times 10^{-6}$), and accounted for 2.4% of the variation in DrinksCPD. Two variants demonstrated suggestive association ($p < 1.21 \times 10^{-3}$) with three or more phenotypes (Table 1): rs138707300 and rs10867752. In addition, a synonymous *IREB2* variant near the cholinergic nicotinic receptor subunit gene cluster on chromosome 15q24-q25 showed suggestive association with CPD (rs13180[T]), MAF = 0.433, b = 2.537, SE = 0.578, $p = 1.24 \times 10^{-5}$; Supplementary Table 2). Additional suggestive associations in this region (see Supplementary Table 2) included two variants in *CHRNA3* (rs938682[A], MAF = 0.321, b = 2.856, SE = 0.673, $p = 2.33 \times 10^{-5}$; and rs1051730[G], MAF = 0.302, $b = -2.251$, SE = 0.618, $p = 2.79 \times 10^{-4}$), as well as rs16969968[G] in *CHRNA5* (MAF = 0.302, $b = -2.267$, SE = 0.618, $p = 2.57 \times 10^{-4}$), which is in strong linkage disequilibrium with rs1051730 and therefore highly correlated with the *CHRNA3* signal.

Gene-based Association Tests

Gene-based association tests were conducted for all variants in the gene regardless of MAF, and were restricted to those genes with more than one variant available for analysis (n = 12,240). Critical p-value thresholds of 4.08 x 10^{-6} (0.05/12,240) and 4.08 x 10^{-3} were set to determine statistical significance and suggestive evidence of association, respectively. The number of variants per gene ranged from 2 to 209. Many genes demonstrated suggestive association with each of the phenotypes (see Supplementary Table 3), although none met criteria for statistical significance. Twelve genes showed suggestive association ($p < 0.01$) with three or more phenotypes (Table 2). Current approaches for conducting gene-based tests of association do not allow for valid estimation of effect sizes; therefore, the analyses in the current study do not include estimates of the effect for a single gene.

Pathway-based Association Tests

Pathway-based association tests were conducted by including all variants in a given pathway, regardless of MAF (n = 6,176). Critical p-value thresholds of 8.10×10^{-6} (0.05/6,176) and 8.10×10^{-3} were set to determine statistical significance and suggestive evidence of association, respectively. The number of variants per pathway ranged from 3 to 7,533. One pathway (C2:NAKAMURA_ADIPOGENESIS_LATE_UP; 'Genes up-regulated in mesenchymal stem cells during late phase of adipogenesis, defined as days 7 to 14 of culturing with adipogenic hormones') was significantly associated with AlcDep ($p = 2.62 \times 10^{-6}$), and numerous pathways demonstrated suggestive association with each phenotype ($p < 8.10 \times 10^{-3}$; see Supplementary Table 4). Twenty pathways showed suggestive association ($p < 0.01$) with three or more phenotypes (Table 3). Similar to gene-based tests, current approaches for conducting pathway-based tests of association do not allow for valid estimation of effect sizes, and thus are not reported.

Discussion

This study examined the effects of genetic variation in protein-coding regions of the genome on alcohol and tobacco co-use and dependence. Three sets of results emerged from the study analyses. First, a missense variant (rs117374816) in *ERM3* was significantly associated with DrinksCPD. Suggestive associations were also observed for single variants in genes identified in previous studies of substance use and dependence. Second, a number of the top signals in the single variant and gene-based tests showed suggestive association with multiple phenotypes, and may implicate a more general liability to psychopathology. Finally, multiple biological pathways of interest emerged in analyses that grouped variants across sets of genes, furthering our understanding of the mechanisms involved in alcohol and tobacco use and dependence. These included a significant association of AlcDep with a pathway related to increased gene expression during the late phase of adipogenesis, and a second pathway containing genes with high-CpG-density promoters that lack histone H3 methylation marks in the brain that demonstrated suggestive association with multiple phenotypes. Each of these three sets of results will be discussed in turn, as well as the implications of specific genetic influences and the limitations of functional exonic variants' ability to explain the missing heritability of alcohol and tobacco use phenotypes with current sample sizes.

As stated, the first set of results corresponds to the suggestive associations for relevant variants in candidate genes that have been previously associated with substance use traits, demonstrating the potential of an exome-focused analytic approach if larger samples can be obtained. Variants in *CHRNA3*, *CHRNA5* and *IREB2* (Bierut, 2009) were among the top associations with CPD and implicate loci from previous meta-analyses of smoking behaviors and lung cancer risk (Supplementary Table 2). These included rs938682 in *CHRNA3*, a variant previously correlated with a risk locus associated with age of first regular tobacco use (Stephens et al., 2013), as well as rs16969968 in *CHRNA5* and rs1051730 in *CHRNA3*, which are in strong linkage disequilibrium and thus highly correlated. The latter two variants corroborate associations with smoking phenotypes in previous studies (Chen et al., 2012; Saccone et al., 2010; Tobacco and Genetics Consortium, 2010; Ware, van den Bree, &

Munafo, 2011) and provide further evidence to support a causal role in smoking quantity. Though not statistically significant, the range of effect sizes observed for these variants (e.g., $R^2 < 0.02$) approached those of past studies. As such, these trend-level findings provide validation of the exome chip genotyping microarray approach for evaluating the effects of rare protein-coding variation on alcohol and tobacco use and dependence by identifying several relations between known variants robustly associated with these phenotypes in previous studies.

Although not previously associated with substance use phenotypes, one variant (rs117374816) in *EMR3* (also referred to as *ADGRE3*) was significantly associated with DrinksCPD. *EMR3* encodes a member of the class B seven-span transmembrane (TM7) receptor family expressed predominantly by immune system cells, and may play a role in myeloid-myeloid interactions during immune and inflammatory responses. Previous research has demonstrated a relationship between psychiatric disorders and immune functioning (e.g., Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium, 2015; Wang, Yang, Gelernter, & Zhao, 2015), warranting future investigations for *EMR3* with larger samples.

Similar to the results for single variant analyses, many genes demonstrated suggestive association with each of the phenotypes. The relative lack of significant gene-based tests is consistent with previous studies that tested for associations of rare and common variants within genes with alcohol and tobacco use phenotypes (Vrieze et al., 2014; Zuo et al., 2013). Notably, these earlier studies restricted their gene-based tests to include only nonsynonymous variants, whereas the current study expanded these analyses to include a large number of different variant types. Despite a wider scope of putatively functional variants within genes, the results from this study were consistent with those of the earlier studies, and suggest that much larger sample sizes will be needed to detect the effects of exonic variants on alcohol and tobacco use. Further, findings from both studies provide a useful examination of the genetic architecture of alcohol and tobacco co-use and dependence by incorporating the analysis of rare coding variation.

The absence of strong effects at the gene level suggests that there may not be a collection of multiple variants within one specific gene of large effect that underlies risk for alcohol and tobacco co-use and dependence. Rather, consistent with findings from GWAS of complex traits, including easily measurable traits that strongly influenced by genetic factors, such as height, the current study suggests that a highly polygenic architecture underlies alcohol and tobacco use and misuse. Therefore, associated loci, be they common or rare, are likely to be located across hundreds of genes, similar to recent reports of 697 genome-wide significant SNPs in more than 400 gene regions associated with variation in human height (Wood et al., 2014). As a result, any single variant or gene is going to have a relatively small effect at the population level in terms of conferring risk for alcohol and tobacco use phenotypes.

The second set of results that emerged was the cross-phenotype associations observed for the most highly associated single variants and genes, several of which corresponded to genes that have also been associated with a broad range of psychiatric phenotypes. At the single variant level, a missense variant in *SYNE1* (rs138707300) and one intergenic variant on

chromosome 9 (rs10867752) showed cross-phenotype suggestive associations with three or more phenotypes: AlcDep, NicDep, MaxDrinks, and AlcNicDep. The synonymous variant rs138707300 belongs to the *SYNE1* gene, which has been associated with psychiatric phenotypes in past research, including alcohol dependence (Edenberg et al., 2010), depression (Green et al., 2013; Y. Liu et al., 2011), and bipolar disorder (e.g., Cross-Disorder Group of the Psychiatric Genomics Consortium & Consortium, 2013; Green et al., 2013; Y. Liu et al., 2011; Sklar et al., 2011).

At the gene-level, twelve genes showed cross-phenotype suggestive associations with three or more phenotypes (Table 2). These included the *LIMK2* gene, which demonstrated suggestive association with AlcDep, CPD, and AlcNicDep, and the *MYOCD* gene with AlcDep, MaxDrinks, and AlcNicDep. *LIMK2* has been associated with treatment response for methamphetamine addiction (Li et al., 2014), as well as other psychiatric phenotypes in past research (Datta, Arion, Corradi, & Lewis, 2015; Zhao et al., 2015), while a variant in *MYOCD* was the most strongly associated variant with heroin addiction in a separate study (Nielsen et al., 2008). Many genes demonstrated cross-phenotype suggestive associations and have shown association with a variety of non-substance use-related psychiatric phenotypes across independent studies (e.g., *BOC:* Terwisscha et al., 2013; *PTX3*: Drexhage et al., 2010, Haarman et al., 2014; *C5ORF42*: Fisher et al., 2015; *OR4E2*: Aragam, Wang, Anderson, & Liu, 2013).

These findings suggest that variants within these genes may underlie a more general predisposition to psychopathology, rather than acting as substance-specific risk factors. This approach is consistent with arguments that while subfactors such as the internalizing and externalizing dimensions can be identified, an overarching super factor may cut across all aspects of psychopathology (Caspi et al., 2014). In this way, substance use behaviors could result from transdiagnostic endophenotypes (both genetic and environmental in nature), and the observation of manifest co-occurring disorders would therefore be the result of these shared transdiagnostic endophenotypes, rather than an indication of independent underlying constructs. Consistent with this interpretation, the number of variants identified in the present report that had demonstrated suggestive association with multiple psychiatric phenotypes in previous studies adds to evidence from recently published studies indicating that several common variants conferred risk for psychopathology in a manner that cut across diagnostic boundaries (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013). As such, the findings from this study expand on prior research by incorporating the effects of rare protein-coding variation in cross-phenotype and cross-disorder analyses. In addition to the variants that have shown robust association with specific substances, there may exist non-specific genetic factors in the risk for increased levels of alcohol and tobacco use and dependence, as suggested by previous studies (e,g., Palmer et al., 2012; Zuo et al., 2012). Although shared environmental factors also contribute to alcohol and tobacco use (Do et al., 2015; Verhulst et al., 2015), the current findings extend multiple lines of research by evaluating the effects of both rare and common protein-coding variants across independent substance use traits.

Finally, numerous pathways demonstrated suggestive association with each of the phenotypes, and one pathway (C2:NAKAMURA_ADIPOGENESIS_LATE_UP) was

significantly associated with AlcDep (Supplementary Table 3). This pathway includes genes up-regulated in mesenchymal stem cells (which are multipotent progenitor cells that can give rise to several types of cells belonging to human skeletal tissues, such as cartilage, bone and fat) during the late phase of adipogenesis. Ethanol exposure has been associated with adipogensis in human bone marrow mesenchymal stem cells (Wezeman & Gong, 2004), thus providing a potential explanation for the observed overlap. The top suggestive associations included pathways involved in hormone action (Holmans et al., 2009), neural structures and neurotransmitter systems (Jia, Wang, Meltzer, & Zhao, 2010), immune signaling (Jia et al., 2010; Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium, 2015), and cell-adhesion and structure proteins (O'Dushlaine et al., 2010), all of which are consistent with previous GWAS pathway analyses of psychiatric phenotypes.

In addition, genes belonging to the Class C/3 metabotropic glutamate and pheromone receptors pathway (C2: REACTOME CLASS C3 METABOTROPIC GLUTAMATE PHEROMONE RECEPTORS) showed suggestive association with AlcDep, which replicates existing research that suggests a direct effect of alcohol on the glutamatergic neurotransmitter system (Ponomarev, Wang, Zhang, Harris, & Mayfield, 2012). Previous studies have also provided evidence of differential brain methylation levels between alcohol-dependent individuals and controls, such that individuals with alcohol dependence exhibit DNA hypomethylation (Ponomarev et al., 2012). Consistent with this hypothesis, the top association for AlcNicDep showed suggestive association with a pathway containing genes with high-CpG-density promoters that lack histone H3 methylation marks in the brain (C2: MEISSNER BRAIN HCP WITH H3 UNMETHYLATED), which also suggests a more general effect for alcohol and other drug dependence on gene expression via epigenetic mechanisms such as DNA methylation. This pathway also demonstrated suggestive association with multiple phenotypes, including AlcDep, NicDep, and AlcNicDep (Supplementary Table 4), and thus warrants further investigation of epigenetic mediation of genetic risk for alcohol and tobacco use behaviors.

Some study limitations regarding the phenotypes, tested variants, and sample characteristics warrant consideration. Inspection of lambda values for initial association tests of MaxDrinks indicated that observed test statistics deviated from normality of the expected -log(p) values, which was corrected by employing a square-root transformation of the MaxDrinks variable. Although this correction resulted in the removal of possible systematic bias in the association analyses, it necessarily has the added effect of also transforming the nature of the variable, and thus its interpretation in the context of these results. In addition, the small sample size and the relatively large number of tests conducted (41,378 effective independent SNV tests, 12,240 gene-based tests, and 6,176 pathway-based tests) limited power to detect effects, especially at the single variant level. This limitation is somewhat mitigated for the gene- and pathway gene-set analyses, given the smaller number of tests conducted relative to the single variant tests. Furthermore, the advantage of focusing on protein-coding regions, relative to conventional GWAS, is the ability to select and interrogate variants within these regions that lead to direct changes in the amino acid sequence of the encoded protein, and thus prioritize these variants as more likely to have an influence on the phenotype. Outside of gene regions, the potential consequence of a given base pair substitution at any one site is

much more difficult to predict. However, the small number of statistically significant associations suggests that, similar to conventional GWAS, larger samples will be needed to examine the association of rare and common variants with alcohol and tobacco use phenotypes.

The number of individual variants included in tests of association also varied greatly across genes and pathways, and was positively skewed, especially for pathways, representing another potential limitation. The SKAT-O test was developed as a more robust alternative to the burden or non-burden kernel tests. In this way, the SKAT-O should account for the number of rare causal variants with the same direction of effect versus the number of noncausal or causal variants with different association directions, but the large range of variants across tests still merits consideration as power to detect true associations will decrease with the addition of many noncausal variants. In addition, future research should consider the inclusion of variants within regulatory regions that might influence the expression of specific genes, given evidence of their role in complex disease etiology and the inability to account for their effects in the current study (Schork et al., 2013). In terms of the study sample, it should also be noted that the majority of the participants were European, and although analyses controlled for population stratification, the generalizability of the results across different ancestral groups is still relatively limited. Analyses were not conducted separately in the largest ancestral group (i.e., European ancestry), as doing so would limit the sample size further, and given the small sample sizes of non-European ancestral groups, there was no ability to conduct comparisons across ancestral groups. Finally, the sample was originally recruited based on alcohol dependence status, and so the power to detect rare variant associations with tobacco use and dependence may be limited.

To date, only one other study has used an exome chip genotyping approach to examine the effects of rare variants in protein-coding regions of the genome on alcohol and tobacco use and dependence (Vrieze et al., 2014). Though rare exonic variants are thought to be important influences in complex disease etiology, the absence of statistically significant associations from both Vrieze et al. (2014) and the current study indicate that limitations exist in these variants' ability to explain much of the missing heritability of alcohol and tobacco use phenotypes with current sample sizes. It is also unclear to what extent these findings might change with a sample that differed in the relative amount of exposure to alcohol and tobacco. While all participants could be assigned a diagnosis of AlcDep or NicDep, quantitative measures of alcohol and tobacco use were only available for individuals who endorsed smoking at least 100 cigarettes and drinking every day for a week or more in their lifetime. Although shared as well as nonshared environmental factors are a significant contributor to the timing (i.e., age) of substance use initiation, there is also evidence of shared genetic factors that influence both age of substance use initiation and symptoms of a substance use disorder (Richmond-Rakerd et al., 2016). Therefore, while there is evidence that genetic factors contribute to substance use initiation and later disorder, we were unable to compare whether the suggestive associations identified in the current study are found at different levels of exposure and alcohol and tobacco use, given the small number of individuals who denied lifetime smoking or regular alcohol use.

Nonetheless, the results provide further support and replicate robust findings in the literature on the association of variants in the nicotinic acetylcholine receptor genes with tobacco use phenotypes, as well as a number of variants that have been linked to psychiatric traits more broadly. Suggestive findings from the present report also suggest that variants in pathways related to hormone action, neural structures and neurotransmitter systems, immune signaling, and cell-adhesion and structure proteins, metabotropic glutamate receptors, and DNA methylation may be related to alcohol and nicotine dependence, as well as higher levels of alcohol and tobacco use. Future investigations might attempt to address the described limitations, as well as validate the single variants, genes, and pathways of interest using whole-genome sequence methods and measures of epigenetic influences on alcohol and tobacco co-use and dependence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by a grant from NIDA R01DA030976. Additional funding was provided by the State of California and the Ernest Gallo Clinic and Research Center for Medical Research on Alcohol and Substance Abuse through the University of California at San Francisco. Selected data and findings were presented at the 37th and 38th Annual Scientific Meetings of the Research Society on Alcoholism.

References

Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nature Protocols. 2010; 5(9):1564–1573. DOI: 10.1038/nprot.2010.116 [PubMed: 21085122]

Aragam N, Wang KS, Anderson JL, Liu X. TMPRSS9 and GRIN2B are associated with neuroticism: a genome-wide association study in a European sample. Journal of Molecular Neuroscience. 2013; 50(2):250–256. DOI: 10.1007/s12031-012-9931-1 [PubMed: 23229837]

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, … Sherlock G. Gene Ontology: tool for the unification of biology. Nature Genetics. 2000; 25(1):25–29. DOI: 10.1038/75556 [PubMed: 10802651]

Asimit J, Zeggini E. Rare variant association analysis methods for complex traits. Annual Review of Genetics. 2010; 44(1):293–308. DOI: 10.1146/annurevgenet-102209-163421

Attia J, Ioannidis JP, Thakkinstian A, McEvoy M, Scott RJ, Minelli C, … Guyatt G. How to use an article about genetic association: A: Background concepts. JAMA: The Journal of the American Medical Association. 2009; 301(1):74–81. DOI: 10.1001/jama.2008.901 [PubMed: 19126812]

Bierut LJ. Nicotine dependence and genetic variation in the nicotinic receptors. Drug and Alcohol Dependence. 2009; 104(S1):S64–S69. DOI: 10.1016/j.drugalcdep.2009.06.003 [PubMed: 19596527]

Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nature Genetics. 2008; 40(6):695–701. DOI: 10.1038/ng.f.136 [PubMed: 18509313]

Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI, … Schuckit MA. New semi-structured psychiatric interview for use in genetic linkage studies-report on reliability of SSAGA. Journal of Studies on Alcohol. 1994; 55(2):1–10. DOI: 10.15288/jsa.1994.55.149

Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, … Moffitt TE. The p factor: One general psychopathology factor in the structure of psychiatric disorders? Clinical Psychological Science. 2014; 2(2):119–137. DOI: 10.1177/2167702613497473 [PubMed: 25360393]

Chen LS, Saccone NL, Culverhouse RC, Bracci PM, Chen CH, Dueker N, … Bierut LJ. Smoking and genetic risk variation across populations of European, Asian, and African American ancestry - A meta-analysis of chromosome 15q25. Genetic Epidemiology. 2012; 36(4):340–351. DOI: 10.1002/gepi.21627 [PubMed: 22539395]

Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. The Lancet. 2013; 381(9875):1371–1379. DOI: 10.1016/S0140-6736(12)62129-1

Datta D, Arion D, Corradi JP, Lewis DA. Altered expression of CDC42 signaling pathway components in cortical layer 3 pyramidal cells in schizophrenia. Biological Psychiatry. 2015; 78(11):775–785. org/10.1016/j.biopsych.2015.03.030. [PubMed: 25981171]

Do EK, Prom-Wormley EC, Eaves LJ, Silberg JL, Miles DR, Maes HH. Genetic and environmental influences on smoking behavior across adolescence and young adulthood in the Virginia Twin Study of Adolescent Behavioral Development and the Transitions to Substance Abuse follow-up. Twin Research and Human Genetics. 2015; 18(1):43–51. DOI: 10.1017/thg.2014.78 [PubMed: 25662421]

Drexhage RC, van der Heul-Nieuwenhuijsen L, Padmos RC, van Beveren N, Cohen D, Versnel MA, … Drexhage HA. Inflammatory gene expression in monocytes of patients with schizophrenia: overlap and difference with bipolar disorder. A study in naturalistically treated patients. International Journal of Neuropsychopharmacology. 2010; 13(10):1369–1381. DOI: 10.1017/S1461145710000799 [PubMed: 20633309]

Edenberg HJ, Koller DL, Xuei X, Wetherill L, McClintick JN, Almasy L, … Foroud T. Genome-wide association study of alcohol dependence implicates a region on chromosome 11. Alcoholism: Clinical and Experimental Research. 2010; 34(5):840–852. DOI: 10.1111/j.1530-0277.2010.01156.x

Falk D, Yi H, Hiller-Sturmhofel S. An epidemiologic analysis of co-occurring alcohol and tobacco use and disorders: Findings from the National Epidemiological Survey on Alcohol and Related Conditions. Alcohol Research & Health. 2006; 29(3):162–171. [PubMed: 17373404]

Fisher HL, Murphy TM, Arseneault L, Caspi A, Moffitt TE, Viana J, … Wong CC. Methylomic analysis of monozygotic twins discordant for childhood psychotic symptoms. Epigenetics. 2015; 10(11):1014–1023. DOI: 10.1080/15592294.2015.1099797 [PubMed: 26479702]

Green EK, Grozeva D, Forty L, Gordon-Smith K, Russell E, Farmer A, … Craddock N. Association at SYNE1 in both bipolar disorder and recurrent major depression. Molecular Psychiatry. 2013; 18(5):614–617. DOI: 10.1038/mp.2012.48 [PubMed: 22565781]

Haarman BCM, Riemersma-Van der Lek RF, Burger H, Netkova M, Drexhage RC, Bootsman F, … Nolen WA. Relationship between clinical features and inflammation-related monocyte gene expression in bipolar disorder – towards a better understanding of psychoimmunological interactions. Bipolar Disorders. 2014; 16(2):137–150. DOI: 10.1111/bdi.12142 [PubMed: 24286609]

Holmans P, Green EK, Pahwa JS, Ferreira MAR, Purcell SM, Sklar P, … Craddock N. Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. American Journal of Human Genetics. 2009; 85(1):13–24. DOI: 10.1016/j.ajhg.2009.05.011 [PubMed: 19539887]

Jia P, Wang L, Meltzer HY, Zhao Z. Common variants conferring risk of schizophrenia: A pathway analysis of GWAS data. Schizophrenia Research. 2010; 122(1–3):38–42. DOI: 10.1016/j.schres.2010.07.001 [PubMed: 20659789]

Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. 2000; 28(1):27–30. DOI: 10.1093/nar/28.1.27 [PubMed: 10592173]

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, … Eskin E. Variance component model to account for sample structure in GWAS. Nature Genetics. 2010; 42(4):348–354. DOI: 10.1038/ng.548 [PubMed: 20208533]

Kapoor M, Wang JC, Wetherill L, Le N, Bertelsen S, Hinrichs AL, … Goate A. A meta-analysis of two genome-wide association studies to identify novel loci for maximum number of alcoholic drinks. Human Genetics. 2013; 132(10):1141–1151. DOI: 10.1007/s00439-013-1318-z [PubMed: 23743675]

Kendler KS, Schmitt E, Aggen SH, Prescott CA. Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. Archives of General Psychiatry. 2008; 65(6):674–682. DOI: 10.1001/archpsyc.65.6.674 [PubMed: 18519825]

Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: Implications for complex disease and association studies. The American Journal of Human Genetics. 2007; 80(4):727–739. DOI: 10.1086/513473 [PubMed: 17357078]

Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, … Lin X. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. The American Journal of Human Genetics. 2012; 91(2):224–237. DOI: 10.1016/j.ajhg.2012.06.007 [PubMed: 22863193]

Li MD. The genetics of nicotine dependence. Current Psychiatry Reports. 2006; 8(2):158–64. DOI: 10.1007/s11920-006-0016-0 [PubMed: 16539894]

Li MD, Wang J, Niu T, Ma JZ, Seneviratne C, Ait-Daoud N, … Johnson BA. Transcriptome profiling and pathway analysis of genes expressed differentially in participants with or without a positive response to topiramate treatment for methamphetamine addiction. BMC Medical Genomics. 2014; 7(1):1–15. doi.org/10.1186/s12920-014-0065-x. [PubMed: 24397966]

Li MX, Yeung JMY, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant *p*-value thresholds in commercial genotyping arrays and public imputation reference datasets. Human Genetics. 2012; 131(5):747–756. DOI: 10.1007/s00439-011-1118-2 [PubMed: 22143225]

Lind PA, Macgregor S, Vink JM, Pergadia ML, Hansell NK, de Moor MHM, … Madden PAF. A genomewide association study of nicotine and alcohol dependence in Australian and Dutch populations. Twin Research and Human Genetics. 2010; 13(1):10–29. DOI: 10.1375/twin.13.1.10 [PubMed: 20158304]

Liu Y, Blackwood DH, Caesar S, de Geus EJ, Farmer A, Ferreira MAR, … Heutink P. Meta-analysis of genome-wide association data of bipolar disorder and major depressive disorder. Molecular Psychiatry. 2011; 16(1)doi: 10.1038/mp.2009.107

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, … Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009; 461(7265):747–753. DOI: 10.1038/nature08494 [PubMed: 19812666]

Marth GT, Yu F, Indap AR, Garimella K, Gravel S, Leong WF. … The 1000 Genomes Project. The functional spectrum of low-frequency coding variation. Genome Biology. 2011; 12(9):R84.doi: 10.1186/gb-2011-12-9-r84 [PubMed: 21917140]

Nielsen D, Ji F, Yuferov V, Ho A, Chen A, Levran O, … Kreek M. Genotype patterns that contribute to increased risk for or protection from developing heroin addiction. Molecular Psychiatry. 2008; 13(4):417–428. DOI: 10.1038/sj.mp.4002147 [PubMed: 18195715]

Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, Verzilli C, … Mooser V. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science. 2012; 337(6090):100–104. DOI: 10.1126/science.1217876 [PubMed: 22604722]

Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. Nature Neuroscience. 2015; 18(2):199–209. DOI: 10.1038/nn.3922 [PubMed: 25599223]

O'Dushlaine C, Kenny E, Heron E, Donohoe G, Gill M, Morris D, Corvin A. Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. Molecular Psychiatry. 2010; 16(3):286–292. DOI: 10.1038/mp.2010.7 [PubMed: 20157312]

Palmer RHC, Button TM, Rhee SH, Corley RP, Young SE, Stallings MC, … Hewitt JK. Genetic etiology of the common liability to drug dependence: Evidence of common and specific mechanisms for DSM-IV dependence symptoms. Drug and Alcohol Dependence. 2012; 123:S24–S32. DOI: 10.1016/j.drugalcdep.2011.12.015 [PubMed: 22243758]

Ponomarev I, Wang S, Zhang L, Harris RA, Mayfield RD. Gene coexpression networks in human brain identify epigenetic modifications in alcohol dependence. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience. 2012; 32(5):1884–1897. DOI: 10.1523/JNEUROSCI.3136-11.2012 [PubMed: 22302827]

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. DOI: 10.1038/ng1847 [PubMed: 16862161]

Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? The American Journal of Human Genetics. 2001; 69:124–137. DOI: 10.1086/321272 [PubMed: 11404818]

Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease–common variant… or not? Human Molecular Genetics. 2002; 11:2417–2423. DOI: 10.1093/hmg/11.20.2417 [PubMed: 12351577]

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, … Sham PC. PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics. 2007; 81(3):559–575. DOI: 10.1086/519795 [PubMed: 17701901]

Ramanan VK, Shen L, Moore JH, Saykin AJ. Pathway analysis of genomic data: Concepts, methods, and prospects for future development. Trends in Genetics. 2012; 28(7):323–332. DOI: 10.1016/j.tig.2012.03.004 [PubMed: 22480918]

Richmond-Rakerd LS, Slutske WS, Lynskey MT, Agrawal A, Madden PAF, Bucholz KK, … Martin NG. Age at first use and later substance use disorder: Shared genetic and environmental pathways for nicotine, alcohol, and cannabis. Journal of Abnormal Psychology. 2016; 125(7):946–59. DOI: 10.1037/abn0000191 [PubMed: 27537477]

Saccone NL, Culverhouse RC, Schwantes-An TH, Cannon DS, Chen X, Cichon S, … Kong X. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. PLoS Genetics. 2010; 6(8):e1001053.doi: 10.1371/journal.pgen.1001053 [PubMed: 20700436]

Sartor CE, Grant JD, Agrawal A, Sadler B, Madden PAF, Heath AC, Bucholz KK. Genetic and environmental contributions to initiation of cigarette smoking in young African-American and European-American women. Drug and Alcohol Dependence. 2015; 157:54–59. DOI: 10.1016/j.drugalcdep.2015.10.002 [PubMed: 26482091]

Schork AJ, Thompson WK, Pham P, Torkamani A, Roddey JC, Sullivan PF, … Dale AM. All SNPs are not created equal: Genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. PLOS Genetics. 2013; 9(4):e1003449. org/10.1371/journal.pgen.1003449.s032. [PubMed: 23637621]

Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N, … Purcell SM. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nature Genetics. 2011; 43(10):977–983. DOI: 10.1038/ng.943 [PubMed: 21926972]

Stephens SH, Hartz SM, Hoft NR, Saccone NL, Corley RC, Hewitt JK, … Ehringer MA. Distinct loci in the CHRNA5/CHRNA3/CHRNB4 gene cluster are associated with onset of regular smoking. Genetic Epidemiology. 2013; 37(8):846–859. DOI: 10.1002/gepi.21760 [PubMed: 24186853]

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, … Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. PNAS: Proceedings of the National Academy of Sciences. 2005; 102(43):15545–15550. DOI: 10.1073/pnas.0506580102

Sun L, Wilder K, McPeek MS. Enhanced Pedigree Error Detection. Human Heredity. 2002; 54(2):99–110. DOI: 10.1159/000067666 [PubMed: 12566741]

Terwisscha van Scheltinga A, Bakker SC, van Haren NEM, Derks EM, Buizer-Voskamp JE, Boos HBM, … Kahn R. Genetic schizophrenia risk variants jointly modulate total brain and white matter volume. Biological Psychiatry. 2013; 73(6):525–531. DOI: 10.1016/j.biopsych.2012.08.017 [PubMed: 23039932]

The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 490:56–65. DOI: 10.1038/nature11632

The Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Research. 2015; 43(Database issue):D1049–D1056. DOI: 10.1093/nar/gku1179 [PubMed: 25428369]

Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, … Stefansson K. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature. 2008; 452(7187):638–642. DOI: 10.1038/nature06846 [PubMed: 18385739]

Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nature Genetics. 2010; 42(5):441–447. DOI: 10.1038/ng.571 [PubMed: 20418890]

Verhulst B, Neale MC, Kendler KS. The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. Psychological Medicine. 2015; 45(5):1061–1072. DOI: 10.1017/S0033291714002165 [PubMed: 25171596]

Vieten C, Seaton KL, Feiler HS, Wilhelmsen KC. The University of California, San Francisco Family Alcoholism Study. I. Design, Methods, and Demographics. Alcoholism: Clinical and Experimental Research. 2004; 28(10):1509–1516. DOI: 10.1097/01.ALC.0000142261.32980.64

Vrieze SI, Feng S, Miller MB, Hicks BM, Pankratz N, Abecasis G, … McGue M. Rare nonsynonymous exonic variants in addiction and behavioral disinhibition. Biological Psychiatry. 2014; 75(10):783–789. DOI: 10.1016/j.biopsych.2013.08.027 [PubMed: 24094508]

Wang K, Li M, Hakonarson HH. Analysing biological pathways in genome-wide association studies. Nature Reviews Genetics. 2010; 11(12):843–854. DOI: 10.1038/nrg2884

Wang KS, Liu X, Zhang Q, Pan Y, Aragam N, Zeng M. A meta-analysis of two genome-wide association studies identifies 3 new loci for alcohol dependence. Journal of Psychiatric Research. 2011; 45(11):1419–1425. DOI: 10.1016/j.jpsychires.2011.06.005 [PubMed: 21703634]

Wang Q, Yang C, Gelernter J, Zhao H. Pervasive pleiotropy between psychiatric disorders and immune disorders revealed by integrative analysis of multiple GWAS. Human Genetics. 2015; 134(0): 1195–1209. DOI: 10.1007/s00439-015-1596-8 [PubMed: 26340901]

Ware JJ, van den Bree MBM, Munafo MR. Association of the CHRNA5-A3- B4 gene cluster with heaviness of smoking: A meta-analysis. Nicotine & Tobacco Research. 2011; 13(12):1167–1175. DOI: 10.1093/ntr/ntr118 [PubMed: 22071378]

Wezeman FH, Gong Z. Adipogenic effect of alcohol on human bone marrow-derived mesenchymal stem cells. Alcoholism: Clinical and Experimental Research. 2004; 28(7):1091–1101. DOI: 10.1097/01.ALC.0000130808.49262.F5

Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, … Frayling TM. Defining the role of common variation in the genomic and biological architecture of adult human height. Nature Genetics. 2014; 46(11):1173–1186. DOI: 10.1038/ng.3097 [PubMed: 25282103]

Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the Sequence Kernel Association Test. The American Journal of Human Genetics. 2011; 89(1):82–93. DOI: 10.1016/j.ajhg.2011.05.029 [PubMed: 21737059]

Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. The American Journal of Human Genetics. 2011; 88(1):76–82. DOI: 10.1016/j.ajhg.2010.11.011 [PubMed: 21167468]

Zhao Z, Xu J, Chen J, Kim S, Reimers M, Bacanu SA, … Chen X. Transcriptome sequencing and genome-wide association analyses reveal lysosomal function and actin cytoskeleton remodeling in schizophrenia and bipolar disorder. Molecular Psychiatry. 2015; 20(5):563–572. org/10.1038/mp.2014.82. [PubMed: 25113377]

Zuo L, Saba L, Wang K, Zhang X, Krystal JH, Tabakoff B, Luo X. Exome-wide association study of replicable nonsynonymous variants conferring risk for alcohol dependence. Journal of Studies on Alcohol and Drugs. 2013; 74(4):622–625. DOI: 10.1038/ng.835 [PubMed: 23739027]

Zuo L, Zhang F, Zhang H, Zhang XY, Wang F, Li CSR, … Luo X. Genome-wide search for replicable risk gene regions in alcohol and nicotine co-dependence. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2012; 159B(4):437–444. DOI: 10.1002/ajmg.b.32047

Table 1

Cross-phenotype single variant genetic association tests (suggestive associations $p < 1.21 \times 10^{-3}$ bolded).

Chr	Position	dbSNP rsID	Gene	Variant annotation	MAF	AlcDep p- value	NicDep p- value	MaxDrinks p-value	CPD p-value	AlcNicDep p-value	DrinksCPD p-value
6	152565758	rs138707300	SYNE1	G/A_Missense	0.048	**1.17E-03**	**5.60E-04**	1.72E-02	1.74E-01	**3.62E-04**	8.87E-01
9	84015037	rs10867752	---	T/C_Intergenic	0.173	**1.81E-05**	9.06E-02	**1.62E-04**	1.26E-02	**3.24E-04**	4.99E-01

Note. Chr = chromosome. dbSNP rsID = the Single Nucleotide Polymorphism Database (dbSNP) reference SNP ID. MAF = minor allele frequency observed in study sample. AlcDep = alcohol dependence diagnosis. NicDep = nicotine dependence diagnosis. MaxDrinks = maximum number of drinks consumed in a 24-hour period. CPD = cigarettes smoked per day. AlcNicDep = both alcohol and nicotine dependence diagnoses. DrinksCPD = product term of MaxDrinks and CPD.

Table 2

Cross-phenotype gene-based genetic association tests (suggestive associations $p < 0.01$ bolded).

Chr	Position start	Position end	Gene	Number of variants	AlcDep p-value	NicDep p-value	MaxDrinks p-value	CPD p-value	AlcNicDep p-value	DrinksCPD p-value
1	32259791	32280610	SPOCD1	8	**3.28E-03**	7.84E-01	1.00E+00	**4.44E-03**	1.11E-01	**7.38E-04**
2	48873662	48898798	GTF2A1L	6	1.40E-02	8.67E-02	**6.28E-03**	3.37E-02	**9.77E-04**	**9.32E-03**
2	86433240	86439550	MRPL35	2	**2.12E-05**	5.21E-02	**4.82E-04**	2.88E-02	1.42E-02	**2.20E-03**
3	112974641	113003273	BOC	3	2.39E-02	1.93E-02	**2.14E-03**	**8.83E-04**	6.27E-02	**4.23E-04**
3	157155314	157161305	PTX3	3	**3.38E-03**	**2.25E-03**	5.92E-01	1.52E-01	**3.36E-04**	4.37E-01
5	37142583	37196028	C5orf42	8	**6.81E-03**	3.75E-02	1.00E-01	**6.52E-05**	6.48E-02	**1.04E-04**
10	115959034	115980422	TDRD1	5	**8.16E-05**	**8.66E-03**	4.32E-02	3.75E-01	**5.67E-04**	6.66E-01
14	22133648	22134018	OR4E2	4	**3.48E-03**	**5.96E-03**	1.52E-02	7.94E-02	**1.15E-03**	1.41E-02
15	97326853	97327394	SPATA8	3	**4.41E-03**	3.66E-02	1.00E+00	**5.05E-03**	**3.57E-03**	6.19E-01
17	12620711	12666691	MYOCD	8	**2.02E-03**	9.95E-02	**2.43E-03**	2.01E-02	**4.11E-03**	**3.89E-04**
19	43013363	43023255	CEACAM1	2	1.00E+00	6.01E-01	**3.24E-03**	**8.82E-03**	8.39E-01	**5.85E-03**
22	31621792	31674324	LIMK2	7	**2.15E-03**	2.46E-02	6.65E-01	**8.96E-03**	**4.20E-03**	2.43E-01

Note. Chr = chromosome. AlcDep = alcohol dependence diagnosis. NicDep = nicotine dependence diagnosis. MaxDrinks = maximum number of drinks consumed in a 24-hour period. CPD = cigarettes smoked per day. AlcNicDep = both alcohol and nicotine dependence diagnoses. DrinksCPD = product term of MaxDrinks and CPD.

Table 3

Cross-phenotype pathway-based genetic association tests (all suggestive associations $p < 0.01$).

M SigDB Gene Set	Pathway Description	Number of variants	AlcDep p-value	NicDep p-value	M ax Drinks p-value	CPD p-value	AlcNic Dep p-value	Drinks CPD p-value
C2:GRADE_COLON_CANCER_UP	Up-regulated genes in colon carcinoma tumors compared to the matched normal mucosa samples	2,205	**1.69E-03**	**3.10E-03**	**8.55E-03**	3.73E-02	1.04E-02	**8.29E-03**
C2:GAVIN_FOXP3_TARGETS_CLUSTER_P3	Cluster P3 of genes with similar expression profiles in peripheral T lymphocytes after X-linked transcription factor FOXP3 loss of function	595	1.15E-01	7.15E-01	**6.61E-03**	**4.12E-03**	8.74E-01	**3.17E-03**
C2:LIM _MAMMARY_LUMINAL_MATURE_DN	Genes consistently down-regulated in mature mammary luminal cells both in mouse and human species	397	**4.28E-03**	3.57E-02	4.04E-02	**6.55E-03**	6.09E-02	**7.98E-03**
C2:LOPEZ_MESOTELIOMA_SURVIVAL_TIME_DN	Top genes higher expressed in long term mesothelioma survivors	14	**6.30E-03**	**1.17E-03**	4.05E-02	5.61E-02	**1.02E-03**	2.77E-02
C2:FEVR_CTNNB1_TARGETS_DN	Genes down-regulated in intestinal crypt cells upon deletion of CTNNB1	1,648	**4.49E-03**	**6.16E-03**	1.16E-01	6.99E-02	**2.53E-03**	5.59E-02
C5:CELL_DEVELOPMENT	The process whose specific outcome is the progression of the cell over time, from its formation to the mature structure.	1,938	**3.39E-03**	**6.30E-05**	1.36E-02	7.92E-02	**9.71E-05**	6.31E-02
C2:KYNG_RESPONSE_TO_H2O2_VIA_ERCC6	Genes down-regulated in CS-B cells (Cockayne syndrome fibroblast, CS) with deficient ERCC6 in response to hydrogen peroxide	45	**9.61E-03**	3.50E-02	**3.90E-03**	4.67E-01	**6.35E-03**	1.34E-01
C2:MEISSNER_BRAIN_HCP_WITH_H3_UNMETHYLATED	Genes with high-CpG-density promoters (HCP) that have no histone H3 methylation marks in brain	111	**4.21E-04**	**2.88E-03**	4.88E-01	7.06E-02	**5.03E-05**	1.44E-01
C5:METAL_ION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	Enables the transfer of metal ions from one side of a membrane to the other.	580	**3.12E-03**	**3.54E-04**	2.86E-01	**7.16E-03**	1.70E-02	2.74E-01
C5:CATION_CHANNEL_ACTIVITY	Enables the energy-independent passage of cations across a lipid bilayer down a concentration gradient	480	**5.09E-03**	**2.59E-04**	4.01E-01	2.24E-02	**8.27E-03**	3.33E-01
C5:APOPTOSIS_GO	---	1,190	**9.21E-03**	**1.79E-03**	6.49E-01	1.01E-01	**1.42E-03**	3.42E-01
C5:PROGRAMMED_CELL_DEATH	---	1,190	**9.21E-03**	**1.79E-03**	6.49E-01	1.01E-01	**1.42E-03**	3.42E-01
C5:REGULATION_OF_APOPTOSIS	---	924	**5.91E-03**	**2.44E-03**	7.42E-01	1.16E-01	**1.28E-03**	4.29E-01
C5:REGULATION_OF_PROGRAMMED_CELL_DEATH	---	924	**5.91E-03**	**2.44E-03**	7.42E-01	1.16E-01	**1.28E-03**	4.29E-01
C5:REGULATION_OF_DEVELOPMENTAL_PROCESS	---	1,417	**8.27E-03**	**4.50E-03**	5.75E-01	9.62E-02	**9.80E-03**	5.17E-01
C5:NEGATIVE_REGULATION_OF_DEVELOPMENTAL_PROCESS	Any process that stops, prevents or reduces the rate or extent of development, the biological process whose specific outcome is the progression of an organism over time from an initial condition to a later condition	630	**4.58E-03**	**3.67E-03**	2.83E-01	6.68E-01	**4.39E-03**	5.71E-01
C5:SECRETION_BY_CELL	The controlled release of a substance by a cell	333	**3.15E-03**	**3.09E-03**	4.22E-01	5.15E-01	**3.87E-03**	6.97E-01
C2:KEGG_ALANINE_ASPARTATE_AND_GLUTAMATE_METABOLISM	Alanine, aspartate and glutamate metabolism	92	**8.58E-03**	**3.29E-03**	7.90E-02	3.16E-01	**5.17E-03**	7.54E-01
C2:CHEOK_RESPONSE_TO_HD_MTX_DN	Genes specifically down-regulated in pediatric acute lymphoblastic leukemia patients by high-dose methotrexate	52	**1.40E-04**	7.57E-02	**4.47E-03**	3.97E-01	**9.52E-03**	7.77E-01
C2:KEGG_AXON_GUIDANCE	Axon guidance	486	**5.85E-03**	**5.36E-03**	6.58E-02	7.31E-01	**4.20E-03**	1.00E+00

Note. AlcDep = alcohol dependence diagnosis. NicDep = nicotine dependence diagnosis. MaxDrinks = maximum number of drinks consumed in a 24-hour period. CPD = cigarettes smoked per day. AlcNicDep = both alcohol and nicotine dependence diagnoses. DrinksCPD = product term of MaxDrinks and CPD. MSigDB C2 and C5 gene sets may be accessed at http://software.broadinstitute.org/gsea/msigdb/collections.jsp