

**Title:** Exome chip meta-analysis elucidates the genetic architecture of rare coding variants in smoking and drinking behavior.

**Keywords:** Tobacco, Alcohol, Addiction, GWAS, Exome, Behavioral Genetics

**Authors, in order with affiliation:**

<b>Dajiang J. Liu*</b>	Institute of Personalized Medicine, Penn State College of Medicine
<b>David M. Brazel*</b>	Institute for Behavioral Genetics, University of Colorado Boulder Department of Molecular, Cellular, and Developmental Biology, University of Colorado Boulder
<b>Valérie Turcot</b>	Montreal Heart Institute, Montreal, Quebec, H1T 1C8, Canada Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, Quebec, H3T 1J4, Canada
<b>Xiaowei Zhan</b>	Department of Clinical Science, Center for Genetics of Host Defense, University of Texas Southwestern
<b>Jian Gong</b>	Public Health Sciences Division, Fred Hutchinson Cancer Research Center
<b>Daniel R. Barnes</b>	Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
<b>Sarah Bertelsen</b>	Department of Neuroscience, Icahn School of Medicine at Mount Sinai
<b>Yi-Ling Chou</b>	Washington University
<b>A. Mesut Erzurumluoglu</b>	Department of Health Sciences, University of Leicester
<b>Jessica D. Faul</b>	Survey Research Center, Institute for Social Research, University of Michigan
<b>Jeff Haessler</b>	Public Health Sciences Division, Fred Hutchinson Cancer Research Center
<b>Anke R. Hammerschlag</b>	Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam
<b>Chris Hsu</b>	University of Southern California
<b>Manav Kapoor</b>	Department of Neuroscience, Icahn School of Medicine at Mount Sinai
<b>Dongbing Lai</b>	Department of Medical and Molecular Genetics, Indiana University School of Medicine
<b>Nhung Le</b>	Department of Medical Microbiology, Immunology and Cell Biology, Southern Illinois University School of Medicine
<b>Christiaan A de Leeuw</b>	Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam

**Anu Loukola** Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; Department of Public Health, University of Helsinki, Helsinki, Finland

**Massimo Mangino** Twin Research & Genetic Epidemiology Unit, Kings College, London

**Carl A. Melbourne** Department of Health Sciences, University of Leicester

**Giorgio Pistis** Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche (CNR), Monserrato, Cagliari, Italy.

**Beenish Qaiser** Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; Department of Public Health, University of Helsinki, Helsinki, Finland

**Rebecca Rohde** University of North Carolina, Chapel Hill

**Yaming Shao** University of North Carolina, Chapel Hill

**Heather Stringham** Department of Biostatistics, University of Michigan

**Leah Wetherill** Department of Medical and Molecular Genetics, Indiana University School of Medicine

**Wei Zhao** Department of Epidemiology, School of Public Health, University of Michigan

**Arpana Agrawal** Department of Psychiatry, Washington University School of Medicine

**Laura Bierut** Department of Psychiatry, Washington University School of Medicine

**Chu Chen** Public Health Sciences Division, Fred Hutchinson Cancer Research Center  
Department of Epidemiology and Department of Otolaryngology; Head and Neck Surgery, University of Washington, Seattle, WA

**Charles B. Eaton** Department of Family Medicine, Brown University, Providence, RI

**Alison Goate** Department of Neuroscience, Icahn School of Medicine at Mount Sinai

**Christopher Haiman** Department of Preventative Medicine, Keck School of Medicine, University of Southern California

**Andrew Heath** Department of Psychiatry, Washington University

**William G. Iacono** Department of Psychology, University of Minnesota

**Nicholas G. Martin** Queensland Institute for Medical Research

**Tinca J. Polderman** Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam

**CHD Exome+ Consortium<sup>††</sup>**

**Consortium for Genetics of Smoking Behaviour<sup>††</sup>**

**Alex Reiner** Public Health Sciences Division, Fred Hutchinson Cancer Research Center

<b>John Rice</b>	Department of Epidemiology, University of Washington, Seattle, WA
<b>David Schlessinger</b>	Departments of Psychiatry and Mathematics, Washington University St. Louis
<b>H Steven Scholte</b>	National Institute on Aging, National Institutes of Health
<b>Jennifer A. Smith</b>	Department of Psychology, University of Amsterdam & Amsterdam Brain and Cognition, University of Amsterdam
<b>Jean-Claude Tardif</b>	Department of Epidemiology, School of Public Health, University of Michigan
<b>Hilary A. Tindle</b>	Montreal Heart Institute, Montreal, Quebec, H1T 1C8, Canada
<b>Andreis R van der Leij</b>	Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, Quebec, H3T 1J4, Canada
<b>Michael Boehnke</b>	Department of Medicine, Vanderbilt University, Nashville, TN
<b>Jenny Chang-Claude</b>	Department of Psychology, University of Amsterdam & Amsterdam Brain and Cognition, University of Amsterdam
<b>Francesco Cucca</b>	Department of Biostatistics, School of Public Health, University of Michigan
<b>Sean P. David</b>	Division of Cancer Epidemiology, German Cancer Research Center
<b>Tatiana Foroud</b>	Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche (CNR), Monserrato, Cagliari, Italy.
<b>Sharon L.R. Kardia</b>	Department of Medicine, Stanford University, Stanford, CA
<b>Charles Kooperberg</b>	Department of Medical and Molecular Genetics, Indiana University School of Medicine
<b>Markku Laakso</b>	Department of Epidemiology, School of Public Health, University of Michigan
<b>Guillaume Lettre</b>	Public Health Sciences Division, Fred Hutchinson Cancer Research Center
<b>Pamela Madden</b>	University of Eastern Finland, Finland
<b>Matt McGue</b>	Montreal Heart Institute, Montreal, Quebec, H1T 1C8, Canada
<b>Kari North</b>	Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, Quebec, H3T 1J4, Canada
<b>Danielle Posthuma</b>	Department of Psychiatry, Washington University
<b>Timothy Spector</b>	Department of Psychology, University of Minnesota
<b>Daniel Stram</b>	Department of Epidemiology, University of North Carolina, Chapel Hill
	Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam
	Department of Clinical Genetics, VU University Medical Centre Amsterdam, Amsterdam Neuroscience
	Department of Genetic Epidemiology, Kings College, London
	Department of Preventative Medicine, Keck School of Medicine, University of Southern California

**David R. Weir** Survey Research Center, Institute for Social Research, University of Michigan

**Jaakko Kaprio** Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; Department of Public Health, University of Helsinki, Helsinki, Finland

**Gonçalo R. Abecasis** Department of Biostatistics, School of Public Health, University of Michigan

**Scott Vrieze**<sup>†</sup> Department of Psychology, University of Minnesota

\*These authors contributed equally to the work.

†Address correspondence to Scott Vrieze, Department of Psychology, University of Minnesota. 75 East River Road, Minneapolis, MN 55455. E-mail: [vrie0006@umn.edu](mailto:vrie0006@umn.edu)

††See supplement for a list of authors associated with the replication consortia.

## **Abstract:**

**Background:** Smoking and alcohol use behaviors in humans have been associated with common genetic variants within multiple genomic loci. Investigation of rare variation within these loci holds promise for identifying causal variants impacting biological mechanisms in the etiology of disordered behavior. Microarrays have been designed to genotype rare nonsynonymous and putative loss of function variants. Such variants are expected to have greater deleterious consequences on gene function than other variants, and significantly contribute to disease risk.

**Methods:** In the present study, we analyzed ~250,000 rare variants from 17 independent studies. Each variant was tested for association with five addiction-related phenotypes: cigarettes per day, pack years, smoking initiation, age of smoking initiation, and alcoholic drinks per week. We conducted single variant tests of all variants, and gene-based burden tests of nonsynonymous or putative loss of function variants with minor allele frequency less than 1%.

**Results:** Meta-analytic sample sizes ranged from 70,847 to 164,142 individuals, depending on the phenotype. Known loci tagged by common variants replicated, but there was no robust evidence for individually associated rare variants, either in gene based or single variant tests. Using a modified method-of-moment approach, we found that all low frequency coding variants, in aggregate, contributed 1.7% to 3.6% of the phenotypic variation for the five traits ( $p < .05$ ).

**Conclusions:** The findings indicate that rare coding variants contribute to phenotypic variation, but that much larger samples and/or denser genotyping of rare variants will be required to successfully identify associations with these phenotypes, whether individual variants or gene-based associations.

## Introduction

Tobacco and alcohol use together account for more morbidity and mortality in Western cultures than any other single risk factor or health outcome<sup>1</sup>. These preventable and modifiable behaviors are heritable<sup>2</sup>, have been associated previously in human and model organism research with multiple common genetic variants<sup>3-7</sup>, and most prominently feature genes involved in alcohol/nicotine metabolism and nicotinic receptors.

Advances in sequencing technology have led to cost-effective “exome arrays”, which affordably genotype a few hundred thousand rare (minor allele frequency [MAF]<1%), putatively functional exonic variants. Compared to common SNPs (MAF>1%) used in genome-wide association studies (GWAS), rare exonic variants may have greater potential to elucidate the biological mechanisms of addiction and other complex traits<sup>8,9</sup>. Loss of function (LoF) variants result in the loss of normal function of a protein, and may have greater phenotypic impact than other variants that do not have obvious biological consequences<sup>10,11</sup>. One well-known example is rare LoF mutations in *PCSK9* that greatly reduce risk of cardiovascular disease with no apparent negative effects, encouraging the development of a new class of *PCSK9* inhibitor drugs<sup>12,13</sup>.

The analysis of any rare event, including rare genetic variants, presents analytical challenges. First, statistical power is a function of MAF, such that rare variants of small to moderate effect require very large samples to achieve adequate statistical power<sup>14</sup>. Statistical association techniques have been developed to mitigate this issue, including tests that aggregate information across many low-frequency variants<sup>15</sup>. These “burden” tests can improve power under certain assumptions, such as that a large proportion of the aggregated variants are independently associated with the phenotype of interest. Here, we use novel methods to implement a variety of genetic association tests, in the largest sample currently available, to test the effect of rare and low-frequency exonic variants on tobacco and alcohol use behaviors.

The vast majority of existing addiction-related rare variant studies use targeted sequencing of known addiction-associated loci to discover and test for association. This has led to intriguing new leads, especially within nicotinic receptor gene clusters<sup>16-25</sup> and alcohol metabolism genes<sup>26-28</sup> for alcohol and nicotine dependence. This strategy has also produced rare variant associations with alcohol dependence in genes not previously implicated in addiction. In one case, burden testing was used to find an association with rare variants in *SERINC2*<sup>28</sup>. In another case, a burden test across *PTP4A1*, *PHF3*, and *EYS* showed an association with alcohol dependence<sup>29</sup>. Single variant tests did not reach significance after multiple-testing corrections in either case. In part due to the nature of burden tests, especially when conducted across multiple genes, these findings do not have simple biological interpretations, and no rare variant results have been replicated.

Some studies also leverage information about predicted functional consequences of rare mutations to increase the power of association analyses. For example, one study of nicotine dependence found significant rare single-variant associations in *CHRNA4*, but only when variants were weighted by their effect on the cellular response to nicotine and acetylcholine<sup>30</sup>. Such positive findings benefit from replication, which has not always been straightforward. For example, all rare variant associations in addiction are, to our knowledge, candidate gene analyses with type I error thresholds based only on tests within that region. Historically, such analyses have tended to produce overly optimistic estimates of the number of associated loci<sup>31</sup>. Genome-wide analyses with more conservative type I error thresholds have reported null rare variant findings across an array of phenotypes relevant to addiction<sup>32-34</sup>. Precisely because genome-wide analyses are conducted on many variants across the genome, they are in principle able to discover novel rare variant associations within new or known loci. One way to improve power in genome-wide analyses is through genetic association meta-analysis, which entails the aggregation of results across many studies to achieve large sample sizes. We present here such a meta-analysis, aggregating studies with rare variant genotype arrays and

measured alcohol and nicotine use, to arrive at a highly powered test of the hypothesis that rare exonic variants affect addiction-related outcomes.

In addition to single variant and gene-level tests, we also conducted tests of the contribution of rare nonsynonymous variants to the heritability of our alcohol and tobacco use phenotypes. Twin studies, as well as studies of the aggregate effects of common variants, have found both alcohol use and tobacco use to be heritable behaviors<sup>34-39</sup>. Research on the aggregate contribution of rare variants, however, has been scarce, with previous work on related phenotypes in smaller samples failing to detect aggregate effects for smoking and alcohol consumption<sup>32</sup>. In this study, we implemented a novel method-of-moments approach to analyze heritability and genetic correlations due to variants genotyped on the exome array. We used meta-analytic summary statistics to quantify the contribution to heritability of variants in various functional categories and frequency bins, estimated the genetic correlation between smoking and drinking traits, and evaluated the contribution of rare coding variants to the phenotypic variation of smoking and alcohol use behavior.

## **Methods and Materials**

Seventeen studies contributed summary statistics for meta-analysis. These studies, their sample sizes, and available phenotypes are listed in the online supplement (Tables S1 and S2). Two studies (HRS and COGA) provided results for individuals of European and African ancestry separately. One study (the UK Biobank of European ancestry) was stratified into two samples according to ascertainment protocol and genotyping method. Thus, in the end, 20 independent sets of results from 17 independent studies were submitted for meta-analysis.

### **Ancestry**

All analyses were stratified by ancestry. Eighteen datasets were on individuals of European ancestry and two datasets on individuals of African ancestry.

### **Phenotypes**



Phenotypes were selected to be relevant to prior GWAS of smoking and alcohol use, common in psychological, medical, and epidemiological data sets, and known to be correlated with measures of substance dependence. Five phenotypes were selected based on their inclusion in previous successful GWAS studies<sup>4, 40-42</sup> and availability among large exome chip studies.

1. Cigarettes per day. The average number of cigarettes smoked in a day among current and former smokers. Studies with binned responses retained their existing bins. Studies that recorded an integer value binned responses into one of four categories: 1=1-10, 2=11-20, 3=21-30, 4=31 or more. Anyone reporting 0 cigarettes per day was coded as missing. This phenotype is a component of commonly used measures of nicotine dependence such as the Fagerstrom Test for Nicotine Dependence.

2. Pack Years. Defined in the same way as cigarettes per day but not binned, divided by 20 (cigarettes in a pack), and multiplied by number of years smoking. This yields a measure of total overall exposure to tobacco and is relevant to cancer and chronic obstructive pulmonary disease risk.

3. Age of Initiation of Smoking. A measure of early cigarette use. Defined as the age at which a participant first started smoking regularly.

4. Smoking Initiation. A binary variable of whether the individual had ever been a regular smoker (1) or not (0), and often defined as having smoked at least 100 cigarettes during one's lifetime.

5. Drinks per week. A measure of drinking frequency/quantity. The average number of drinks per week in current or former drinkers.

### **Genotypes**

Fifteen of the seventeen studies were genotyped with the Illumina HumanExome BeadChip, which contains ~250,000 low-frequency nonsynonymous variants, variants from the GWAS catalog, and a small number of variants selected for other purposes. Two studies were genotyped on the Illumina Human Core Exome, which includes an additional ~250,000 tag SNPs.



Finally, the present study used the initial release of 150,000 UK Biobank participants, which comprised two cohorts: 1) the UK BiLEVE cohort of ~50,000 heavy and never smokers genotyped on the UK BiLEVE array and 2) ~100,000 participants genotyped on the UK Axiom array. These arrays are highly similar and described elsewhere (<http://www.ukbiobank.ac.uk/scientists-3/uk-biobank-axiom-array>). They both include >800,000 variants including ~630,000 genome-wide tagging markers, ~110,000 rare coding variants that are largely a subset of variants also genotyped on the Illumina HumanExome Beadchip, and an additional ~107,000 variants chosen for the study of specific medical conditions.

### **Generation of Summary Association Statistics**

Twenty sets of results from 17 independent studies (Table S1) with smoking and drinking phenotypes were included in the discovery phase. Summary statistics were adjusted by local analysts for age, sex, any study-specific covariates, and ancestry principal components (see Table S2 for genomic controls). For studies with related individuals (see Table S1), relatedness was accounted for in linear mixed models using empirically estimated kinships from common SNPs<sup>43</sup>. Residuals were inverse-normalized to help ensure well-behaved test statistics for rare variant tests.

Quality control of per-study summary statistics included evaluation and correction of strand flips and allele flips through systematic comparison of alleles and allele frequencies against reference datasets ExAC v2.0, 1000 Genomes Phase 3, and dbSNP. Variants with call rates < 0.9, and Hardy Weinberg  $p < 1 \times 10^{-7}$ , and polymorphic in < 3 studies, were also removed. The latter filter was meant to avoid findings that could not be broadly replicated across the 17 studies.

Variants were annotated against RefSeq 1.9<sup>44</sup>. The allelic spectrum of all nonsynonymous, start loss/gain, stop loss/gain, or splice acceptor/donor is displayed in Figure 1 for cigarettes per day, stratified by whether the variant exists only in the UK Biobank, only in

studies genotyped with the Illumina exome chip, or in both. More details on the allelic spectra within functional classes are available in Table S3 and Figure S1.

### **Meta-analysis**

We performed meta-analysis in rareMETALS version 5.8<sup>45</sup> using the Mantel-Haenszel method<sup>45</sup>. For gene-level burden tests, we selected variants predicted to be nonsynonymous, start loss, start gain, stop loss, stop gain, or splice donor/acceptor within each gene from RefSeq 1.9<sup>44</sup>. Two complementary gene-level association tests were performed: the sequence kernel association test (SKAT; <sup>46, 47</sup>) with MAF cutoff 1% and a variable MAF threshold test (VTMC; <sup>48</sup>) with a maximum MAF=1%. We chose variants with MAF≤1% as we were interested in the contribution of variants with a frequency lower than that which has been reliably imputed and tested in past GWAS meta-analyses. Exceedingly rare variants, with minor allele counts less than five, were excluded from single variant analyses due to extremely low expected power. These rare variants were included in all gene-based tests.

There exist known genetic associations between common variants and smoking or drinking phenotypes, including variants within the nicotinic receptor gene cluster on chromosome 15 with cigarettes per day<sup>4, 5</sup>; *CYP2A6* and *CYP2B6* with cigarettes per day<sup>42</sup>; *AUTS2*, *KLB*, *ADH1B*, *ALDH2*, and *GCKR* with alcohol use<sup>40, 41</sup>; and *NCAM1* and *TEX41* with smoking initiation<sup>49</sup>. We conducted sequential forward selection association tests, as implemented in rareMETALS, for rare variants within these regions, controlling for any common variant associations in these regions.

Association testing was done in stages. First, we tested common variants within all known loci associated with these phenotypes, as listed above. To these variants we applied the standard genome-wide significance threshold of  $p < 5e^{-8}$ . Second, for rare variants with MAF < 1% and minor allele count  $\geq 5$ , we applied a Bonferroni correction for the number of such variants tested, resulting in p-value thresholds from  $2.1 \times 10^{-7}$  to  $2.2 \times 10^{-7}$  depending on the phenotype.

This threshold was applied to both marginal (unconditional) analyses and forward selection conditional analyses of rare variants within known loci.

Third and finally, for each known and previously validated locus associated with these and related traits, we explored a relaxed multiple-testing threshold based only on the number of rare variants within a 1MB region around the most highly significant (usually common variant) association within that region. Each locus-wide p-value threshold is provided in Table 1. This approach is meant to mimic the typical candidate gene or targeted sequencing approach, where one or a few known loci are analyzed separately from the rest of the genome. While this threshold is overly liberal, it allows a more direct comparison between our results and existing publications of rare variants described in the introduction.

Finally, we attempted to replicate previous rare variant associations referenced in the introduction and listed in Table S6. The prior studies were of alcohol or nicotine dependence. We attempted replication in our phenotypes for any single variant when that variant was included on the exome array (5 of 23 variants were available) and, if not, we took the variant with the smallest p-value within the same gene as the original finding (16 of remaining 18 variants), or any gene-based burden test for which content existed on the array (27 of 27 prior associated genes had content on the array). We applied a liberal threshold that corrected only for the number of tests conducted for this replication exercise ( $.05/52=.00096$ ).

### **Replication Data**

We replicated any novel exome-wide significant rare variant ( $MAF < 1\%$ ) in two additional exome chip smoking meta-analysis efforts, the CHD Exome+ Consortium ( $N=17,789$ ) and the Consortium for Genetics of Smoking Behaviour ( $N=28,583$ ). Both consortia defined their phenotypes similarly and corrected for sex, age, principal components (and/or genetic relatedness, as appropriate), and inverse-normalized prior to association analysis.

### **Genetic Architecture Analysis**

We performed heritability and genetic correlation analyses using a modified method-of-moment estimator, adapted to the analysis of sparsely genotyped rare variants. The method calculates covariate-adjusted LD scores from summary statistics based upon partial correlations and quantifies the uncertainty of LD scores with a bootstrap procedure that uses multiple contributing studies. The estimation of heritability follows established methods. Detailed descriptions of the approach can be found in the supplement.

## Results

Conditionally independent, significant single-variant association results are displayed in Table 1. We discovered a single novel association signal for a single rare variant, only for cigarettes per day, rs36015615 (N=30,030; Beta=1.3;  $p=9.5 \times 10^{-9}$ ), a nonsynonymous SNP in the gene *STARD3*. This novel variant did not replicate in either of two replication consortium datasets, the CHD Exome+ Consortium (N=17,789, Beta=-.01,  $p=.94$ ) or the Consortium for Genetics of Smoking Behaviour (N=28,583, Beta=.056,  $p=.84$ ).

Two known common variants within the *CHRNA5-CHRNA3-CHRNA4* locus (rs16969968 and rs938682) were independently associated with cigarettes per day and pack years. Conditional tests of rare nonsynonymous variants within these genes were non-significant. We verified at  $p < 5 \times 10^{-8}$  a common variant association near *CYP2A6* for pack years. For drinks per week, we replicated at  $p < 5 \times 10^{-8}$  a variant in *GCKR* and a known low-frequency association for a nonsynonymous variant in *AHD1B*, but did not replicate at  $p < 5 \times 10^{-8}$  prior genome-wide associations around *AUTS2*. In *GCKR* we discovered a common nonsynonymous SNP, rs1260326, associated with drinks per week. This SNP is 10,047 base pairs from, and in high LD ( $r^2=.97$  in 1000 Genomes Phase 3 data) with the intronic *GCKR* SNP rs780094 that almost reached statistical significance ( $p=1.6 \times 10^{-7}$ ) in a recent report<sup>40</sup>.

We removed variants that were only present in two or fewer studies to avoid reporting associations that arose solely from the UK Biobank and are essentially unreplicable. This filter removed several genome-wide significantly associated common variants previously reported to

associate with either tobacco or alcohol use. These variants included rs1137115 in *CYP2A6* associated with cigarettes per day as well as some rare variants within that locus that showed evidence of conditionally independent association. Additional UK Biobank-associated variants were rs4144892 in *NCAM1* associated with SI; rs58930260, rs11694518, and rs12619517 associated with SI; rs12648443 in *ADH1C* associated with drinks per week; and rs13146907 in an intron of *KLB* associated with drinks per week. These results are reported in Table S4 of the supplement.

SKAT gene-based tests of nonsynonymous variants with MAF<1% resulted in one significant association with *ADH1C* (SKAT  $p=1.0\times 10^{-8}$ ), although after conditioning this gene-based test on a nearby genome-wide significant nonsynonymous variant in *ADH1B*, the *ADH1C* effect becomes nonsignificant ( $p=.52$ ). Variable threshold gene based burden tests (VTCCM) yielded no significantly associated gene.

Out of four published rare single variant associations with addiction phenotypes, we replicated only one, even after examining other variants in the same gene and applying a relaxed multiple testing threshold (Table S6). The variant was rs56175056 in *CHRNA4* ( $p=9.4e^{-6}$ ), previously identified in an Icelandic population<sup>50</sup>. Out of twenty six genes that have been associated with alcohol or nicotine dependence in published rare variant burden tests, we found a significant association for one, *ADH1C* (Table S6), described in the previous paragraph.

Heritability was estimated for each trait and partitioned by annotation category. First, we annotated variants on the exome chip based upon gene definitions in RefSeq 1.9, using SEQMINER version 6.0<sup>51</sup>. Sixteen functional categories were considered, including downstream, essential splice site, noncoding exon, intergenic, intron, common nonsynonymous (MAF>0.01), rare nonsynonymous (MAF<0.01), normal splice site, start gain/loss, stop gain/loss, synonymous, and 3'/5' untranslated regions. We fitted the baseline model with 16 categories, and estimated phenotypic variance explained by each category (Table S5).

Significant phenotypic variance was explained from rare nonsynonymous variants for all traits ( $p < 0.05$ ), from 1.7%-3.6% (Table 2). We also estimated the phenotypic variance explained by all variants on the exome chip, through aggregating the variance explained by each significant category ( $p < 0.05$ ) listed in Table S5. The total variance explained was highest for cigarettes per day ( $4.6\% \pm 1.3\%$  standard error) and the lowest for drinks per week ( $2.4 \pm 0.8\%$ ).

All pairs of traits are genetically correlated (Table 3) except for cigarettes per day and smoking initiation, and the direction of the genetic correlations are in the expected direction. For instance, cigarettes per day has a positive genetic correlation with drinks per week ( $0.04 \pm 0.0084$ ), consistent with the observation that the increased alcohol consumption is correlated with increased tobacco consumption. Age of initiation has a negative correlation with all other traits, which is consistent with the observation that an earlier age of smoking initiation is correlated with increased tobacco and alcohol consumption in adulthood. The patterns and magnitudes of correlation are highly similar when considering only rare nonsynonymous variants (Table 3).

## Discussion

With a maximum sample size ranging from 70,847 to 164,142, the present study is the largest study to date of low-frequency nonsynonymous and LoF variants in smoking and alcohol use. Our meta-analytic study design allowed us to conduct single variant, gene-based burden tests, and exact conditional analyses accounting for common variants on the Illumina exome chip and UK Biobank arrays. Despite these analytical advantages and a large sample size, we were unable to discover robust, novel associations for nonsynonymous or LoF variants. The one novel associated rare variant in *STARD3* did not replicate in two complementary large exome chip meta-analysis consortia.

We discovered a common nonsynonymous SNP, rs1260326, in *GCKR*, associated with drinks per week. The T→C change results in a nonsynonymous (Leu→Pro) and splice region change in the final codon of the 14th exon in *GCKR*. The mutation is predicted to be possibly

damaging by PolyPhen-2<sup>52</sup>, although the functional significance of this variant is unknown.

Denser genotyping or genotype imputation will verify whether this particular variant has a direct causal relationship to drinks per week, or if the association arises artifactually due to linkage disequilibrium between this variant and other variants in the locus. We replicated one of four previous rare single variant associations and twenty five out of twenty six gene-level associations. Possible explanations include the relatively thin coverage of the exome chip compared to targeted resequencing, phenotypic heterogeneity (previous studies used dependence diagnoses), differences in study population, or overestimation of true effects in the original studies.

We showed that rare nonsynonymous variants on the exome chip explain significant proportions of phenotypic variance. The exome chip was designed to genotype coding variants uncovered in ~12,000 sequenced exomes. By design, it comprehensively ascertained high confidence rare nonsynonymous, splice, and stop variants within those sequences and only sparsely genotypes other classes of variation, including common variants. The use of the exome chip therefore limited our ability to quantify heritability for these other types of variants, or to conduct enrichment tests. Care should also be taken when interpreting those results for which we had substantial coverage on the exome chip. The estimates should be interpreted as “chip heritability”, which is the proportion of heritability that can be tagged by variants on the chip. Even rare nonsynonymous variants may be in linkage disequilibrium with other nearby variants, and thus the percent variance explained by nonsynonymous variants may not be solely attributed to the genotyped variants. Additional fine mapping and denser genotype data is needed to dissect the contribution of any given variant or class of variants. Nonetheless, our results provide preliminary evidence that nonsynonymous variants contribute substantially to the genetic etiology of smoking and drinking.

The exome chip design is an efficient way to accumulate large samples genotyped with a moderate number of low-frequency exonic variants. The effect size spectrum of low-frequency



variants on complex traits is poorly understood and, despite our large sample sizes, it may well be that our meta-analysis was underpowered to detect variants with small effects on smoking and alcohol use behaviors<sup>14</sup>. The maximum sample size for cigarettes per day, for example, was  $N \sim 75,000$ . At this sample size, we had 80% power to detect a variant accounting for  $>.05\%$  of variance. A small effect, but if the variant in question has  $MAF=0.1\%$ , it translates to a standardized regression weight of 0.5. That is, for every risk allele an individual carries, their expected phenotype increases by  $\frac{1}{2}$  of a standard deviation. Such a variant would be highly consequential for the individuals who carry it, and of considerable scientific interest. The result is similar for SI, where  $N \sim 165,000$ , and we had 80% power to detect an odds ratio  $>1.14$  for a variant with  $MAF=1\%$ . We had 80% power to detect an odds ratio  $>1.5$  for a variant with  $MAF=0.1\%$ . The present results indicate there are no rare variants on the exome chip with such effects on smoking and alcohol use in European ancestry individuals.

A similar line of reasoning can be used to put the chip heritability results into context. We found that rare nonsynonymous variants contribute to heritability (e.g.,  $\sim 3\%$ ) in these traits. Rare disease-associated variants are expected to have larger effects than common variants, in the sense that carrying a rare mutation is expected to have a larger phenotypic impact, if only due to purifying selection for deleterious mutations. However, even if the effect is large in that sense, any rare mutation by definition only affects a small number of individuals. Thus, a rare variant with a large effect accounts for a tiny fraction of variation in any common, complex disorder or trait. In the present study, there were  $\sim 130,000$  nonsynonymous variants with  $MAF < 1\%$  (Table S3) and they in aggregate appear to account for substantial variation in the phenotypes. So while the present results provide evidence that rare nonsynonymous variants play a significant role in risk for smoking and alcohol use behavior but that individual rare variants associations remain undetectable even at the sample sizes accumulated here.

**Acknowledgements:** Research reported in this article was supported by the National Institute on Drug Abuse and the National Human Genome Research Institute of the National Institutes of Health under award numbers R01DA037904 (SIV), R21DA040177 (DJL), R01HG008983 (DJL), and 5T3DA017637-13 (DMB), as well as funding sources listed in the Supplementary Note. This research has been conducted using the UK Biobank Resource under Application Number 6395.

**Financial Disclosures:** There are no conflicts to disclose

## References

1. Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJL, Coll CRA. Selected major risk factors and global and regional burden of disease. *Lancet*. 2002;360(9343):1347-1360.
2. Polderman TJ, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, Posthuma D. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet*. 2015.
3. Eng MY, Luczak SE, Wall TL. ALDH2, ADH1B, and ADH1C genotypes in Asians: A literature review. *Alcohol Research & Health*. 2007;30(1):22-27.
4. Furberg H, Kim Y, Dackor J, Boerwinkle E, Franceschini N, Ardissino D, Bernardinelli L, Mannucci PM, Mauri F, Merlini PA, Absher D, Assimes TL, Fortmann SP, Iribarren C, Knowles JW, Quertermous T, Ferrucci L, Tanaka T, Bis JC, Furberg CD, Haritunians T, McKnight B, Psaty BM, Taylor KD, Thacker EL, Almgren P, Groop L, Ladenvall C, Boehnke M, Jackson AU, Mohlke KL, Stringham HM, Tuomilehto J, Benjamin EJ, Hwang SJ, Levy D, Preis SR, Vasani RS, Duan J, Gejman PV, Levinson DF, Sanders AR, Shi JX, Lips EH, McKay JD, Agudo A, Barzan L, Bencko V, Benhamou S, Castellsague X, Canova C, Conway DI, Fabianova E, Foretova L, Janout V, Healy CM, Holcatova I, Kjaerheim K, Lagiou P, Lissowska J, Lowry R, Macfarlane TV, Mates D, Richiardi L, Rudnai P, Szeszenia-Dabrowska N, Zaridze D, Znaor A, Lathrop M, Brennan P, Bandinelli S, Frayling TM, Guralnik JM, Milaneschi Y, Perry JRB, Altshuler D, Elosua R, Kathiresan S, Lucas G, Melander O, O'Donnell CJ, Salomaa V, Schwartz SM, Voight BF, Penninx BW, Smit JH, Vogelzangs N, Boomsma DI, de Geus EJC, Vink JM, Willemssen G, Chanock SJ, Gu FY, Hankinson SE, Hunter DJ, Hofman A, Tiemeier H, Uitterlinden AG, van Duijn CM, Walter S, Chasman DI, Everett BM, Pare G, Ridker PM, Li MD, Maes HH, Audrain-McGovern J, Posthuma D, Thornton LM, Lerman C, Kaprio J, Rose JE, Ioannidis JPA, Kraft P, Lin DY, Sullivan PF. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genetics*. 2010;42(5):441-U134.
5. Saccone NL, Culverhouse RC, Schwantes-An TH, Cannon DS, Chen X, Cichon S, Giegling I, Han S, Han Y, Keskitalo-Vuokko K, Kong X, Landi MT, Ma JZ, Short SE, Stephens SH, Stevens VL, Sun L, Wang Y, Wenzlaff AS, Aggen SH, Breslau N, Broderick P, Chatterjee N, Chen J, Heath AC, Heliovaara M, Hoft NR, Hunter DJ, Jensen MK, Martin NG, Montgomery GW, Niu T, Payne TJ, Peltonen L, Pergadia ML, Rice JP, Sherva R, Spitz MR, Sun J, Wang JC, Weiss RB, Wheeler W, Witt SH, Yang BZ, Caporaso NE, Ehringer MA, Eisen T, Gapstur SM, Gelernter J, Houlston R, Kaprio J, Kendler KS, Kraft P, Leppert MF, Li MD, Madden PA, Nothen MM, Pillai S, Rietschel M, Rujescu D, Schwartz A, Amos CI, Bierut LJ. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet*. 2010;6(8).
6. Bierut LJ, Stitzel JA. Genetic Contributions of the alpha 5 Nicotinic Receptor Subunit to Smoking Behavior. *Nicotinic Receptors*. 2014;26:327-339.
7. Luczak SE, Glatt SJ, Wall TL. Meta-analyses of ALDH2 and ADH1B with alcohol dependence in Asians. *Psychological Bulletin*. 2006;132(4):607-621.
8. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao FM, Zou J, Pierce-Hollman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu DM, Altshuler DM, Ardissino D,

- Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, NcCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Consortium EA. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-+.
9. Minikel E, Lek M, Samocha KE, Karczewski KJ, Marshall JL, Armean I, Ware J, Daly MJ, MacArthur DG, Consortium EA. An early glimpse of saturation mutagenesis in humans: Insights from protein-coding genetic variation in 60,706 people. *Prion*. 2016;10:S107-S107.
  10. Sveinbjornsson G, Albrechtsen A, Zink F, Gudjonsson SA, Oddson A, Masson G, Holm H, Kong A, Thorsteinsdottir U, Sulem P, Gudbjartsson DF, Stefansson K. Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat Genet*. 2016;48(3):314-317.
  11. Marouli E, Graff M, Medina-Gomez C, Lo KS, Wood AR, Kjaer TR, Fine RS, Lu Y, Schurmann C, Highland HM, Rueger S, Thorleifsson G, Justice AE, Lamparter D, Stirrups KE, Turcot V, Young KL, Winkler TW, Esko T, Karaderi T, Locke AE, Masca NG, Ng MC, Mudgal P, Rivas MA, Vedantam S, Mahajan A, Guo X, Abecasis G, Aben KK, Adair LS, Alam DS, Albrecht E, Allin KH, Allison M, Amouyel P, Appel EV, Arveiler D, Asselbergs FW, Auer PL, Balkau B, Banas B, Bang LE, Benn M, Bergmann S, Bielak LF, Bluher M, Boeing H, Boerwinkle E, Boger CA, Bonnycastle LL, Bork-Jensen J, Bots ML, Bottinger EP, Bowden DW, Brandslund I, Breen G, Brilliant MH, Broer L, Burt AA, Butterworth AS, Carey DJ, Caulfield MJ, Chambers JC, Chasman DI, Chen YI, Chowdhury R, Christensen C, Chu AY, Cocca M, Collins FS, Cook JP, Corley J, Galbany JC, Cox AJ, Cuellar-Partida G, Danesh J, Davies G, de Bakker PI, de Borst GJ, de Denus S, de Groot MC, de Mutsert R, Deary IJ, Dedoussis G, Demerath EW, den Hollander AI, Dennis JG, Di Angelantonio E, Drenos F, Du M, Dunning AM, Easton DF, Ebeling T, Edwards TL, Ellinor PT, Elliott P, Evangelou E, Farmaki AE, Faul JD, Feitosa MF, Feng S, Ferrannini E, Ferrario MM, Ferrieres J, Florez JC, Ford I, Fornage M, Franks PW, Frikke-Schmidt R, Galesloot TE, Gan W, Gandin I, Gasparini P, Giedraitis V, Giri A, Girotto G, Gordon SD, Gordon-Larsen P, Gorski M, Grarup N, Grove ML, Gudnason V, Gustafsson S, Hansen T, Harris KM, Harris TB, Hattersley AT, Hayward C, He L, Heid IM, Heikkila K, Helgeland O, Hernesniemi J, Hewitt AW, Hocking LJ, Hollensted M, Holmen OL, Hovingh GK, Howson JM, Hoyng CB, Huang PL, Hveem K, Ikram MA, Ingelsson E, Jackson AU, Jansson JH, Jarvik GP, Jensen GB, Jhun MA, Jia Y, Jiang X, Johansson S, Jorgensen ME, Jorgensen T, Jousilahti P, Jukema JW, Kahali B, Kahn RS, Kahonen M, Kamstrup PR, Kanoni S, Kaprio J, Karaleftheri M, Kardia SL, Karpe F, Kee F, Keeman R, Kiemenev LA, Kitajima H, Kluivers KB, Kocher T, Komulainen P, Kontto J, Kooner JS, Kooperberg C, Kovacs P, Kriebel J, Kuivaniemi H, Kury S, Kuusisto J, La Bianca M, Laakso M, Lakka TA, Lange EM, Lange LA, Langefeld CD, Langenberg C, Larson EB, Lee IT, Lehtimaki T, Lewis CE, Li H, Li J, Li-Gao R, Lin H, Lin LA, Lin X, Lind L, Lindstrom J, Linneberg A, Liu Y, Liu Y, Lophatananon A, Luan J, Lubitz SA, Lyytikainen LP, Mackey DA, Madden PA, Manning AK, Mannisto S, Marenne G, Marten J, Martin NG, Mazul AL, Meidtner K, Metspalu A, Mitchell P, Mohlke KL, Mook-Kanamori DO, Morgan A, Morris AD, Morris AP, Muller-Nurasyid M, Munroe PB, Nalls MA, Nauck M, Nelson CP, Neville M, Nielsen SF, Nikus K, Njolstad PR, Nordestgaard BG, Ntalla I, O'Connel JR, Oksa H, Loohuis LM, Ophoff RA, Owen KR, Packard CJ, Padmanabhan S, Palmer CN, Pasterkamp G, Patel AP, Pattie A, Pedersen O, Peissig PL, Peloso GM, Pennell CE, Perola M, Perry JA, Perry JR, Person TN, Pirie A, Polasek O, Posthuma D, Raitakari OT, Rasheed A, Rauramaa R, Reilly DF, Reiner AP, Renstrom F, Ridker PM, Rioux JD, Robertson N, Robino A, Rolandsson O, Rudan I,

- Ruth KS, Saleheen D, Salomaa V, Samani NJ, Sandow K, Sapkota Y, Sattar N, Schmidt MK, Schreiner PJ, Schulze MB, Scott RA, Segura-Lepe MP, Shah S, Sim X, Sivapalaratnam S, Small KS, Smith AV, Smith JA, Southam L, Spector TD, Speliotes EK, Starr JM, Steinthorsdottir V, Stringham HM, Stumvoll M, Surendran P, t Hart LM, Tansey KE, Tardif JC, Taylor KD, Teumer A, Thompson DJ, Thorsteinsdottir U, Thuesen BH, Tonjes A, Tromp G, Trompet S, Tsfantakis E, Tuomilehto J, Tybjaerg-Hansen A, Tyrer JP, Uher R, Uitterlinden AG, Ulivi S, van der Laan SW, Van Der Leij AR, van Duijn CM, van Schoor NM, van Setten J, Varbo A, Varga TV, Varma R, Edwards DR, Vermeulen SH, Vestergaard H, Vitart V, Vogt TF, Vozzi D, Walker M, Wang F, Wang CA, Wang S, Wang Y, Wareham NJ, Warren HR, Wessel J, Willems SM, Wilson JG, Witte DR, Woods MO, Wu Y, Yaghoobkar H, Yao J, Yao P, Yerges-Armstrong LM, Young R, Zeggini E, Zhan X, Zhang W, Zhao JH, Zhao W, Zhao W, Zheng H, Zhou W, Consortium EP-I, Consortium CHDE, Exome BPC, Consortium TD-G, Go TDGC, Global Lipids Genetics C, ReproGen C, Investigators M, Rotter JI, Boehnke M, Kathiresan S, McCarthy MI, Willer CJ, Stefansson K, Borecki IB, Liu DJ, North KE, Heard-Costa NL, Pers TH, Lindgren CM, Oxvig C, Kutalik Z, Rivadeneira F, Loos RJ, Frayling TM, Hirschhorn JN, Deloukas P, Lettre G. Rare and low-frequency coding variants alter human adult height. *Nature*. 2017;542(7640):186-190.
12. Cohen JC, Boerwinkle E, Mosley TH, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *New England Journal of Medicine*. 2006;354(12):1264-1272.
  13. Hall SS. Genetics: a gene of rare effect. *Nature*. 2013;496(7444):152-155.
  14. Auer PL, Reiner AP, Wang G, Kang HM, Abecasis GR, Altshuler D, Bamshad MJ, Nickerson DA, Tracy RP, Rich SS, Project NGENES, Leal SM. Guidelines for Large-Scale Sequence-Based Complex Trait Association Studies: Lessons Learned from the NHLBI Exome Sequencing Project. *Am J Hum Genet*. 2016.
  15. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet*. 2014;95(1):5-23.
  16. Yang J, Wang S, Yang Z, Hodgkinson CA, Iarikova P, Ma JZ, Payne TJ, Goldman D, Li MD. The contribution of rare and common variants in 30 genes to risk nicotine dependence. *Mol Psychiatry*. 2014.
  17. McClure-Begley TD, Papke RL, Stone KL, Stokes C, Levy AD, Gelernter J, Xie P, Lindstrom J, Picciotto MR. Rare human nicotinic acetylcholine receptor alpha4 subunit (CHRNA4) variants affect expression and function of high-affinity nicotinic acetylcholine receptors. *J Pharmacol Exp Ther*. 2014;348(3):410-420.
  18. Piliguian M, Zhu AZ, Zhou Q, Benowitz NL, Ahluwalia JS, Sanderson Cox L, Tyndale RF. Novel CYP2A6 variants identified in African Americans are associated with slow nicotine metabolism in vitro and in vivo. *Pharmacogenet Genomics*. 2014;24(2):118-128.
  19. Haller G, Druley T, Vallania FL, Mitra RD, Li P, Akk G, Steinbach JH, Breslau N, Johnson E, Hatsukami D, Stitzel J, Bierut LJ, Goate AM. Rare missense variants in CHRNA4 are associated with reduced risk of nicotine dependence. *Hum Mol Genet*. 2012;21(3):647-655.
  20. Haller G, Kapoor M, Budde J, Xuei X, Edenberg H, Nurnberger J, Kramer J, Brooks A, Tischfield J, Almasy L, Agrawal A, Bucholz K, Rice J, Saccone N, Bierut L, Goate A. Rare missense variants in CHRNA3 and CHRNB3 are associated with risk of alcohol and cocaine dependence. *Hum Mol Genet*. 2014;23(3):810-819.
  21. Zuo L, Tan Y, Li C-SR, Wang Z, Wang K, Zhang X, Lin X, Chen X, Zhong C, Wang X, Wang J, Lu L, Luo X. Associations of rare nicotinic cholinergic receptor gene variants to nicotine and alcohol dependence. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 2016.



22. Xie P, Kranzler HR, Krauthammer M, Cosgrove KP, Oslin D, Anton RF, Farrer LA, Picciotto MR, Krystal JH, Zhao H, Gelernter J. Rare Nonsynonymous Variants in Alpha-4 Nicotinic Acetylcholine Receptor Gene Protect Against Nicotine Dependence. *Biological Psychiatry*. 2011;70:528-536.
23. Wessel J, McDonald SM, Hinds Da, Stokowski RP, Javitz HS, Kennemer M, Krasnow R, Dirks W, Hardin J, Pitts SJ, Michel M, Jack L, Ballinger DG, McClure JB, Swan GE, Bergen AW. Resequencing of Nicotinic Acetylcholine Receptor Genes and Association of Common and Rare Variants with the Fagerström Test for Nicotine Dependence. *Neuropsychopharmacology*. 2010;35:2392-2402.
24. Thorgeirsson TE, Steinberg S, Reginsson GW, Bjornsdottir G, Rafnar T, Jonsdottir I, Helgadottir A, Gretarsdottir S, Helgadottir H, Jonsson S, Matthiasson SE, Gislason T, Tyrfinngsson T, Gudbjartsson T, Isaksson HJ, Hardardottir H, Sigvaldason A, Kiemeny LA, Haugen A, Zienolddiny S, Wolf HJ, Franklin WA, Panadero A, Mayordomo JI, Hall IP, Rönmark E, Lundbäck B, Dirksen A, Ashraf H, Pedersen JH, Masson G, Sulem P, Thorsteinsdottir U, Gudbjartsson DF, Stefansson K. A rare missense mutation in CHRNA4 associates with smoking behavior and its consequences. *Molecular Psychiatry*. 2016;21:594-600.
25. Olfson E, Saccone NL, Johnson EO, Chen L-S, Culverhouse R, Doheny K, Foltz SM, Fox L, Gogarten SM, Hartz S, Hetrick K, Laurie CC, Marosy B, Amin N, Arnett D, Barr RG, Bartz TM, Bertelsen S, Borecki IB, Brown MR, Chasman DI, van Duijn CM, Feitosa MF, Fox ER, Franceschini N, Franco OH, Grove ML, Guo X, Hofman A, Kardina SLR, Morrison AC, Musani SK, Psaty BM, Rao DC, Reiner AP, Rice K, Ridker PM, Rose LM, Schick UM, Schwander K, Uitterlinden AG, Vojinovic D, Wang J-C, Ware EB, Wilson G, Yao J, Zhao W, Breslau N, Hatsukami D, Stitzel JA, Rice J, Goate A, Bierut LJ. Rare, low frequency and common coding variants in CHRNA5 and their contribution to nicotine dependence in European and African Americans. *Molecular Psychiatry*. 2016;21:601-607.
26. Peng Q, Gizer IR, Libiger O, Bizon C, Wilhelmsen KC, Schork NJ, Ehlers CL. Association and ancestry analysis of sequence variants in ADH and ALDH using alcohol-related phenotypes in a Native American community sample. *Am J Med Genet B Neuropsychiatr Genet*. 2014;165B(8):673-683.
27. Way M, McQuillin A, Saini J, Ruparelia K, Lydall GJ, Guerrini I, Ball D, Smith I, Quadri G, Thomson AD, Kasiakogia-Worlley K, Cherian R, Gunwardena P, Rao H, Kottalgi G, Patel S, Hillman A, Douglas E, Qureshi SY, Reynolds G, Jauhar S, O'Kane A, Dedman A, Sharp S, Kandaswamy R, Dar K, Curtis D, Morgan MY, Gurling HMD. Genetic variants in or near ADH1B and ADH1C affect susceptibility to alcohol dependence in a British and Irish population. *Addiction Biology*. 2015;20:594-604.
28. Zuo L, Wang K-S, Zhang X-Y, Li C-sR, Zhang F, Wang X, Chen W, Gao G, Zhang H, Krystal JH, Luo X. Rare SERINC2 variants are specific for alcohol dependence in individuals of European descent. *Pharmacogenetics and Genomics*. 2013;23:395-402.
29. Zuo L, Zhang X, Deng H-w, Luo X. Association of rare PTP4A1-PHF3-EYS variants with alcohol dependence. *Journal of Human Genetics*. 2013;58:178-179.
30. Haller G, Li P, Esch C, Hsu S, Goate AM, Steinbach JH. Functional Characterization Improves Associations between Rare Non-Synonymous Variants in CHRNA4 and Smoking Behavior. *PLoS ONE*. 2014;9:e96753.
31. Duncan LE, Keller MC. A Critical Review of the First 10 Years of Candidate Gene-by-Environment Interaction Research in Psychiatry. *American Journal of Psychiatry*. 2011;168(10):1041-1049.
32. Vrieze SI, Feng S, Miller MB, Hicks BM, Pankratz N, Abecasis GR, Iacono WG, McGue M. Rare nonsynonymous exonic variants in addiction and behavioral disinhibition. *Biol Psychiatry*. 2014;75(10):783-789.

33. Vrieze SI, Malone SM, Vaidyanathan U, Kwong A, Kang HM, Zhan X, Flickinger M, Irons D, Jun G, Locke AE, Pistis G, Porcu E, Levy S, Myers RM, Oetting W, McGue M, Abecasis G, Iacono WG. In search of rare variants: preliminary results from whole genome sequencing of 1,325 individuals with psychophysiological endophenotypes. *Psychophysiology*. 2014;51(12):1309-1320.
34. Vrieze SI, Malone SM, Pankratz N, Vaidyanathan U, Miller MB, Kang HM, McGue M, Abecasis G, Iacono WG. Genetic associations of nonsynonymous exonic variants with psychophysiological endophenotypes. *Psychophysiology*. 2014;51(12):1300-1308.
35. Hicks BM, Schalet BD, Malone SM, Iacono WG, McGue M. Psychometric and genetic architecture of substance use disorder and behavioral disinhibition measures for gene association studies. *Behavior Genetics*. 2011;41(4):459-475.
36. Vrieze SI, McGue M, Miller MB, Hicks BM, Iacono WG. Three mutually informative ways to understand the genetic relationships among behavioral disinhibition, alcohol use, drug use, nicotine use/dependence, and their co-occurrence: twin biometry, GCTA, and genome-wide scoring. *Behavior Genetics*. 2013;43(2):97-107.
37. Vink JM, Willemsen G, Boomsma DI. Heritability of smoking initiation and nicotine dependence. *Behav Genet*. 2005;35(4):397-406.
38. Maes HH, Sullivan PF, Bulik CM, Neale MC, Prescott CA, Eaves LJ, Kendler KS. A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use and nicotine dependence. *Psychological Medicine*. 2004;34(7):1251-1261.
39. Swan GE, Carmelli D, Rosenman RH, Fabsitz RR, Christian JC. Smoking and alcohol consumption in adult male twins: genetic heritability and shared environmental influences. *J Subst Abuse*. 1990;2(1):39-50.
40. Schumann G, Liu CY, O'Reilly P, Gao H, Song P, Xu B, Ruggeri B, Amin N, Jia T, Preis S, Lepe MS, Akira S, Barbieri C, Baumeister S, Cauchi S, Clarke TK, Enroth S, Fischer K, Hallfors J, Harris SE, Hieber S, Hofer E, Hottenga JJ, Johansson A, Joshi P, Kaartinen N, Laitinen J, Lemaitre R, Loukola A, Luan J, Lyytikainen LP, Mangino M, Manichaikul A, Mbarek H, Milaneschi Y, Moayyeri A, Mukamal K, Nelson C, Nettleton J, Partinen E, Rawal R, Robino A, Rose L, Sala C, Satoh T, Schmidt R, Schrautz K, Scott R, Smith AV, Starr JM, Teumer A, Trompet S, Uitterlinden AG, Venturini C, Vergnaud AC, Verweij N, Vitart V, Vuckovic D, Wedenoja J, Yengo L, Yu B, Zhang W, Zhao JH, Boomsma DI, Chambers J, Chasman DI, Daniela T, de Geus E, Deary I, Eriksson JG, Esko T, Eulenburger V, Franco OH, Froguel P, Gieger C, Grabe HJ, Gudnason V, Gyllenstein U, Harris TB, Hartikainen AL, Heath AC, Hocking L, Hofman A, Huth C, Jarvelin MR, Jukema JW, Kaprio J, Kooner JS, Kutalik Z, Lahti J, Langenberg C, Lehtimäki T, Liu Y, Madden PAF, Martin N, Morrison A, Penninx B, Pirastu N, Psaty B, Raitakari O, Ridker P, Rose R, Rotter JJ, Samani NJ, Schmidt H, Spector TD, Stott D, Strachan D, Tzoulaki I, van der Harst P, van Duijn CM, Marques-Vidal P, Vollenweider P, Wareham NJ, Whitfield JB, Wilson J, Wolffenbuttel B, Bakalkin G, Evangelou E, Liu Y, Rice KM, Desrivieres S, Kliewer SA, Mangelsdorf DJ, Muller CP, Levy D, Elliott P. KLB is associated with alcohol drinking, and its gene product beta-Klotho is necessary for FGF21 regulation of alcohol preference. *Proceedings of the National Academy of Sciences of the United States of America*. 2016;113(50):14372-14377.
41. Jorgenson E, Thai KK, Hoffmann TJ, Sakoda LC, Kvale MN, Banda Y, Schaefer C, Risch N, Mertens J, Weisner C, Choquet H. Genetic contributors to variation in alcohol consumption vary by race/ethnicity in a large multi-ethnic genome-wide association study. *Mol Psychiatry*. 2017.
42. Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, Sulem P, Rafnar T, Esko T, Walter S, Gieger C, Rawal R, Mangino M, Prokopenko I, Magi R, Keskitalo K, Gudjonsdottir IH, Gretarsdottir S, Stefansson H, Thompson JR, Aulchenko YS, Nelis M, Aben KK, den Heijer M, Dirksen A, Ashraf H, Soranzo N, Valdes AM,



- Steves C, Uitterlinden AG, Hofman A, Tonjes A, Kovacs P, Hottenga JJ, Willemsen G, Vogelzangs N, Doring A, Dahmen N, Nitz B, Pergadia ML, Saez B, De Diego V, Lezcano V, Garcia-Prats MD, Ripatti S, Perola M, Kettunen J, Hartikainen AL, Pouta A, Laitinen J, Isohanni M, Huei-Yi S, Allen M, Krestyaninova M, Hall AS, Jones GT, van Rij AM, Mueller T, Dieplinger B, Haltmayer M, Jonsson S, Mattheiasson SE, Oskarsson H, Tyrffingsson T, Kiemeney LA, Mayordomo JI, Lindholt JS, Pedersen JH, Franklin WA, Wolf H, Montgomery GW, Heath AC, Martin NG, Madden PAF, Giegling I, Rujescu D, Jarvelin MR, Salomaa V, Stumvoll M, Spector TD, Wichmann HE, Metspalu A, Samani NJ, Penninx BW, Oostra BA, Boomsma DI, Tiemeier H, van Duijn CM, Kaprio J, Gulcher JR, McCarthy MI, Peltonen L, Thorsteinsdottir U, ENGAGE Consortium, Stefansson K. Sequence variants at CHRN3-CHRNA6 and CYP2A6 affect smoking behavior. *Nature Genetics*. 2010;42(5):448-U135.
43. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet*. 2010;42(4):348-354.
44. Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, Farrell CM, Hart J, Landrum MJ, McGarvey KM, Murphy MR, O'Leary NA, Pujar S, Rajput B, Rangwala SH, Riddick LD, Shkeda A, Sun H, Tamez P, Tully RE, Wallin C, Webb D, Weber J, Wu W, DiCuccio M, Kitts P, Maglott DR, Murphy TD, Ostell JM. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res*. 2014;42(Database issue):D756-763.
45. Liu DJ, Peloso GM, Zhan X, Holmen OL, Zawistowski M, Feng S, Nikpay M, Auer PL, Goel A, Zhang H, Peters U, Farrall M, Orho-Melander M, Kooperberg C, McPherson R, Watkins H, Willer CJ, Hveem K, Melander O, Kathiresan S, Abecasis GR. Meta-analysis of gene-level tests for rare variant association. *Nat Genet*. 2014;46(2):200-204.
46. Wu MC, Lee S, Cai TX, Li Y, Boehnke M, Lin XH. Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. *American Journal of Human Genetics*. 2011;89(1):82-93.
47. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics*. 2012;13(4):762-775.
48. Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Wei LJ, Sunyaev SR. Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet*. 2010;86(6):832-838.
49. Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Soler Artigas M, Billington CK, Kheirallah AK, Allen R, Cook JP, Probert K, Obeidat M, Bosse Y, Hao K, Postma DS, Pare PD, Ramasamy A, Consortium UKBE, Magi R, Mihailov E, Reinmaa E, Melen E, O'Connell J, Frangou E, Delaneau O, Ox GSKC, Freeman C, Petkova D, McCarthy M, Sayers I, Deloukas P, Hubbard R, Pavord I, Hansell AL, Thomson NC, Zeggini E, Morris AP, Marchini J, Strachan DP, Tobin MD, Hall IP. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med*. 2015;3(10):769-781.
50. Thorgeirsson TE, Steinberg S, Reginsson GW, Bjornsdottir G, Rafnar T, Jonsdottir I, Helgadottir A, Gretarsdottir S, Helgadottir H, Jonsson S, Mattheiasson SE, Gislason T, Tyrffingsson T, Gudbjartsson T, Isaksson HJ, Hardardottir H, Sigvaldason A, Kiemeney LA, Haugen A, Zienolddiny S, Wolf HJ, Franklin WA, Panadero A, Mayordomo JI, Hall IP, Ronmark E, Lundback B, Dirksen A, Ashraf H, Pedersen JH, Masson G, Sulem P, Thorsteinsdottir U, Gudbjartsson DF, Stefansson K. A rare missense mutation in CHRNA4 associates with smoking behavior and its consequences. *Mol Psychiatry*. 2016;21(5):594-600.

51. Zhan X, Liu DJ. SEQMINER: An R-Package to Facilitate the Functional Interpretation of Sequence-Based Associations. *Genet Epidemiol.* 2015.
52. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7(4):248-249.

**Table 1.** Significant results for common and rare (MAF < 1%) variants. Unconditional results and conditional results based on stepwise forward selection are shown.

Phenotype	Known Locus	Chr	Pos	rsID	Ref	Alt	Nearest Gene	Annot.	Sample Size	No. of Studies	ALT AF %	Unconditional Association Tests		Stepwise Forward Selection Conditional Tests			
												Beta (SE)	p-value	Conditioning Variant(s)	Conditional Beta (SE)	Conditional p-value	Locus-wide Bonferroni
Cigarettes per day	<i>CHRNA5-CHRNA3-CHRNA4</i>	15	78806023	rs8034191 <sup>a</sup>	T	C	<i>AGPHD1</i>	Intron	75,493	20	34.1	.09 (.005)	3.7e <sup>-57</sup>	n/a			
		15	78896547	rs938682	G	A	<i>CHRNA3</i>	Intron	75,493	20	77.6	.09 (.006)	2.9e <sup>-46</sup>	rs8034191	.06 (.006)	6.3e <sup>-18</sup>	3.1e <sup>-5</sup>
		-	17	37814687	rs36015615 <sup>b</sup>	G	A	<i>STARD3</i>	Nonsynon	30,030	14	0.03	1.29 (.224)	1.6e <sup>-8</sup>	n/a		
Pack years	<i>CHRNA5-CHRNA3-CHRNA4</i>	15	78806023	rs8034191 <sup>a</sup>	T	C	<i>AGPHD1</i>	Intron	72,909	19	34.0	.08 (.006)	6.6e <sup>-41</sup>	n/a			
		15	78896547	rs938682	G	A	<i>CHRNA3</i>	Intron	72,909	19	77.6	.05 (.006)	1.4e <sup>-31</sup>	15:78806023	.05 (.006)	4.1e <sup>-12</sup>	1.1e <sup>-4</sup>
		<i>CYP2A6</i>	19	41302706	rs7937	C	T	<i>RAB4B</i>	UTR3	41,270	3	56.1	.40 (.007)	1.6e <sup>-8</sup>	n/a		
Drinks per week	<i>ADH1B-ADH1C</i>	4	100239319	rs1229984	T	C	<i>ADH1B</i>	Nonsynon	105,567	9	98.9	.23 (.016)	4.2e <sup>-44</sup>	n/a			
		<i>GCKR</i>	2	27730940	rs1260326	T	C	<i>GCKR</i>	Nonsynon	139,103	20	62.1	.03 (.004)	4.1e <sup>-16</sup>	n/a		

<sup>a</sup>This variant is a proxy for the known common nonsynonymous SNP rs16969968, a known causal variant affecting heaviness of smoking. In our analyses, rs16969968 had the second-most significant association p-value, after rs8034191.

<sup>b</sup>rs36015615 did not replicate in two additional datasets. See main text.

Note: All p-values are corrected for genomic inflation factor based on all variants with MAC ≥ 5 tested for that phenotype. Chr=chromosome, Pos=position (build 37), Ref=reference allele on GRCh37, Alt=alternate allele, N=sample size across all studies that genotyped the variant, ALT AF=allele frequency of the alternate allele estimated in the meta-analysis. All variants that are from only 2 studies were unique to the UK Biobank array. The Bonferroni p-value threshold for all low-frequency (MAF < 1%) variants ranged from 1.8e<sup>-7</sup> to 1.9e<sup>-7</sup>, depending on the phenotype. The full set of summary statistics and additional information about each association, is hosted at <https://genome.psych.umn.edu/index.php/GSCAN>

**Table 2: Estimation of Heritability Explained by Variants on Exome Array.** We estimate the heritability based upon a baseline model with 16 different functional categories. The reported heritability  $\hat{h}^2$  is based upon the cumulative value from the functional categories with significant heritabilities. We also report the its standard deviation ( $se(\hat{h}^2)$ ) and p-values, estimated using jackknife.

Annotation	Phenotype	Heritability Estimates		
		$\hat{h}^2$	$se(\hat{h}^2)$	P-Value
All Variants	Age of Initiation of smoking	.044	.017	0.0048
	Cigarettes Per Day	.046	.013	.00020
	Pack Years	.044	.013	.00040
	Smoking Initiation	.027	.010	.0035
	Drinks Per Week	.024	.0080	.0015
Rare (MAF < .01) Nonsynonymous Variants	Age of Initiation of smoking	.036	.017	.017
	Cigarettes Per Day	.033	.012	.0030
	Pack Years	.032	.013	.0069
	Smoking Initiation	.025	.010	.0062
	Drinks Per Week	.017	.0080	.017

**Table 3: Estimation of Genetic Correlation Between Smoking and Drinking Traits.** We estimate genetic correlations between 5 smoking and drinking traits. Genetic correlation estimates ( $\hat{r}_g$ ), their standard deviation ( $se(\hat{r}_g)$ ) and p-values are reported.

Trait 1	Trait 2	Genetic Correlation		
		$\hat{r}_g$	$se(\hat{r}_g)$	P-value
<b>A. Aggregated Genetic Correlation Induced by All Variants on the Exome Array</b>				
Age of Initiation of Smoking	Cigarettes Per Day	-0.024	0.010	0.020
Age of Initiation of Smoking	Smoking Initiation	-0.037	0.012	0.0017
Age of Initiation of Smoking	Drinks Per Week	-0.023	0.010	0.023
Age of Initiation of Smoking	Pack Years	-0.03	0.010	0.0040
Cigarettes Per Day	Smoking Initiation	0.0027	0.0088	0.76
Cigarettes Per Day	Drinks Per Weeks	0.040	0.0084	$1.6 \times 10^{-6}$
Cigarettes Per Day	Pack Years	0.054	0.011	$1.4 \times 10^{-6}$
Smoking Initiation	Drinks Per Week	0.041	0.0058	$9.4 \times 10^{-13}$
Smoking Initiation	Pack Years	0.018	0.0057	0.0012
Drinks Per Week	Pack Year	0.025	0.0070	0.00038
<b>B. Genetic Correlation Induced by Rare (MAF &lt; 1%) Nonsynonymous Variants</b>				
Age of Initiation of Smoking	Cigarettes Per Day	-0.024	0.010	0.020
Age of Initiation of Smoking	Smoking Initiation	-0.033	0.011	0.0026
Age of Initiation of Smoking	Drinks Per Week	-0.023	0.0094	0.013
Age of Initiation of Smoking	Pack Years	-0.032	0.0088	0.00021
Cigarettes Per Day	Smoking Initiation	0.0025	0.0084	0.76
Cigarettes Per Day	Drinks Per Weeks	0.043	0.0076	$1.10 \times 10^{-8}$
Cigarettes Per Day	Pack Years	0.059	0.010	$1.50 \times 10^{-8}$
Smoking Initiation	Drinks Per Week	0.013	0.0051	0.0084
Smoking Initiation	Pack Years	0.010	0.0044	0.019
Drinks Per Week	Pack Year	0.011	0.0056	0.049

**Figure 1.** Distribution of nonsynonymous and loss of function variant allele frequencies in the Illumina exome array and the UK Biobank arrays, generated from the results for cigarettes per day. (Allelic spectra for other phenotypes may differ slightly). Note there are only 241 variants that were present only in the UK Biobank and not on the Illumina Exome Chip; these 241 variants are not displayed in the figure. MAF = minor allele frequency estimated in the meta-analysis.

