



Full length article

Exploring the role of low-frequency and rare exonic variants in alcohol and tobacco use



Andries T. Marees^{a,b,*,1}, Anke R. Hammerschlag^{c,1}, Lisa Bastarache^d, Hilde de Kluiver^e, Florence Vorspan^{f,g}, Wim van den Brink^a, Dirk J. Smit^a, Damiaan Denys^a, Eric R. Gamazon^{h,i}, Ruifang Li-Gao^j, Elemi J. Breetvelt^{k,1}, Mark C.H. de Groot^m, Tessel E. Galeslootⁿ, Sita H. Vermeulenⁿ, Jan L. Poppelaars^{o,p}, Patrick C. Souverein^q, Renske Keeman^r, Renée de Mutsert^j, Raymond Noordam^s, Frits R. Rosendaal^j, Najada Stringa^p, Dennis O. Mook-Kanamori^{j,t}, Ilonca Vaartjes^u, Lambertus A. Kiemenyⁿ, Martin den Heijer^{j,v}, Natasja M. van Schoor^p, Olaf H. Klungel^q, Anke H. Maitland-Van der Zee^w, Marjanka K. Schmidt^r, Tinca J.C. Polderman^c, Andries R. van der Leij^x, Danielle Posthuma^{c,y}, Eske M. Derks^{a,b}

^a Department of Psychiatry, Amsterdam Neuroscience, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

^b QIMR Berghofer, Translational Neurogenomics Group, Brisbane, Australia

^c Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

^d Center for Precision Medicine, Vanderbilt University Medical Center, Nashville, TN 37232, United States

^e GGZ inGeest and Department of Psychiatry, Amsterdam Public Health Research Institute, VU University Medical Center, Amsterdam, The Netherlands

^f Assistance Publique—Hôpitaux de Paris, Hôpital Fernand Widal, Département de Psychiatrie et de Médecine Addictologique, 200 Rue du Faubourg Saint-Denis, Paris, France

^g Inserm umr-s 1144, Université Paris Descartes, Université Paris Diderot, 4 Avenue de l'Observatoire, Paris, France

^h Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, United States

ⁱ Clare Hall, University of Cambridge, Cambridge, CB3 9AL, United Kingdom

^j Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

^k The Dalglish Family 22q Clinic, Toronto General Hospital, University Health Network, Toronto, Canada

^l Clinical Genetics Research Program, Centre for Addiction and Mental Health, Toronto, Canada

^m Department of Clinical Chemistry and Haematology, Division of Laboratory and Pharmacy, University Medical Center Utrecht, Utrecht, The Netherlands

ⁿ Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands

^o Department of Sociology, VU University, Amsterdam, The Netherlands

^p Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, VU University Medical Center, Amsterdam, The Netherlands

^q Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute of Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

^r Division of Molecular Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands

^s Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands

^t Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands

^u Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

^v Department of Internal Medicine, VU University Medical Center, Amsterdam, The Netherlands

^w Department of Respiratory Medicine, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

^x Department of Brain and Cognition, University of Amsterdam, Amsterdam, The Netherlands

^y Clinical Genetics, Vrije Universiteit Medical Center, Amsterdam, The Netherlands

ARTICLE INFO

Keywords:

Addiction
Exome
Rare variants
Tobacco
Nicotine
Alcohol
PRS

ABSTRACT

Background: Alcohol and tobacco use are heritable phenotypes. However, only a small number of common genetic variants have been identified, and common variants account for a modest proportion of the heritability. Therefore, this study aims to investigate the role of low-frequency and rare variants in alcohol and tobacco use. **Methods:** We meta-analyzed ExomeChip association results from eight discovery cohorts and included 12,466 subjects and 7432 smokers in the analysis of alcohol consumption and tobacco use, respectively. The ExomeChip interrogates low-frequency and rare exonic variants, and in addition a small pool of common variants. We investigated top variants in an independent sample in which ICD-9 diagnoses of “alcoholism” (N = 25,508) and

* Corresponding author at: Translational Neurogenomics Group, QIMR Berghofer, Bancroft Centre, B719, 300 Herston Road, Heston, Queensland, Australia, 4006.

E-mail address: andriestm@hotmail.com (A.T. Marees).

¹ Shared first author.

Pathway analysis

“tobacco use disorder” (N = 27,068) had been assessed. In addition to the single variant analysis, we performed gene-based, polygenic risk score (PRS), and pathway analyses.

Results: The meta-analysis did not yield exome-wide significant results. When we jointly analyzed our top results with the independent sample, no low-frequency or rare variants reached significance for alcohol consumption or tobacco use. However, two common variants that were present on the ExomeChip, rs16969968 ($p = 2.39 \times 10^{-7}$) and rs8034191 ($p = 6.31 \times 10^{-7}$) located in *CHRNA5* and *AGPHD1* at 15q25.1, showed evidence for association with tobacco use.

Discussion: Low-frequency and rare exonic variants with large effects do not play a major role in alcohol and tobacco use, nor does the aggregate effect of ExomeChip variants. However, our results confirmed the role of the *CHRNA5-CHRNA3-CHRNA4* cluster of nicotinic acetylcholine receptor subunit genes in tobacco use.

1. Introduction

Alcohol and tobacco use belong to the world’s leading health risks and are responsible for the premature death of 3.3 million and 6 million people each year, respectively (World Health Organization, 2014; World Health Organization, 2015). Overall 5.1% of the global burden of disease and injury, measured in disability-adjusted life years (DALYs), is attributable to alcohol, and 3.7% is attributable to smoking (World Health Organization, 2009, 2014). Alcohol and tobacco use initiation and severity are influenced by a combination of genetic and environmental risk factors. Examples of environmental exposures that impact on substance use outcomes are social stress, traumatic life events, peer pressure, inadequate parenting, insufficient social control, and low socio-economic status (De Bellis and Zisk, 2014; Kendler et al., 2011; Lijffijt et al., 2014; Loke and Mak, 2013; Van Ryzin et al., 2012; Young-Wolff et al., 2011). Heritability estimates (i.e., the proportion of phenotypic variance attributable to genetic variance) of 40–60% and 45–86% (Broms et al., 2006; Mbarek et al., 2015; Verhulst et al., 2015; Vink et al., 2005) for alcohol and tobacco use-related traits, respectively, indicate a strong genetic component influencing these behaviors. Elucidating which genetic variants contribute to alcohol and tobacco use is an important step in unraveling the underlying biological mechanisms. Although current pharmacological treatments for alcohol and tobacco use disorders have demonstrated positive treatment outcome, the effects are moderate at best, and many patients do not benefit from these treatments (Goh and Morgan, 2017; Stead and Lancaster, 2012). An improved understanding of the biological mechanisms will aid the development of novel medications and prevention methods for substance use-related problems.

Various methods have been applied to identify genetic variants associated with substance use. During the last decade, Genome-Wide Association Studies (GWAS) have been the preferred study design due to the capacity to study single nucleotide polymorphisms (SNPs) in a genome-wide manner without the need for *a priori* hypotheses. GWAS usually captures common variants (i.e., variants with a minor allele frequency (MAF) larger than 5% in the population), but recent large-scale GWAS also included low-frequency variants (MAF 1–5%) and even variants that occur in 0.1% of the population.

The latest GWAS for alcohol consumption comprised 112,117 individuals and identified 14 significant loci (Clarke et al., 2017), including variants in the gene *KLB*, which had been identified previously by Schumann et al. (2016). The gene product of *KLB*, β -klotho, controls alcohol use in mice (Schumann et al., 2016). Other significant loci were found in the alcohol dehydrogenase (ADH) gene cluster, which has consistently been associated with alcohol consumption (Gelernter et al., 2014; Macgregor et al., 2009). SNPs identified in these genes alter alcohol metabolism (Harada et al., 1983; Thomasson et al., 1991; Yoshida et al., 1991). Furthermore, the authors reported multiple gene-based associations including *DRD2*, encoding a dopamine receptor, and *PDE4B*, which plays a role in signal transduction.

For tobacco use, SNPs located at the chromosomal region 15q25.1, which includes the *CHRNA5-CHRNA3-CHRNA4* cluster of nicotinic acetylcholine receptor subunit genes, show the most robust associations

(David et al., 2012; Liu et al., 2010; Saccone et al., 2010a; Tobacco and Genetics, 2010). Other genes implicated in tobacco use are *HECTD2-AS1* (*LOC100188947*), *EGLN2*, *BDNF*, and *DBH* (Tobacco and Genetics, 2010). Although the identified loci and genes for tobacco use are different from alcohol consumption, twin studies suggest that some proportion of the heritability of alcohol and tobacco use is shared between the two traits (Koopmans et al., 1997).

Individual SNPs identified by GWAS explain only a modest proportion of the variation in complex traits (Manolio et al., 2009). In addition, the aggregate effect of all SNPs included in a GWAS does not fully capture the family-based heritability of traits. For alcohol consumption and tobacco use, GWAS SNPs generally capture less than 50% of the twin-based heritability (Lubke et al., 2012; Mbarek et al., 2015; Vrieze et al., 2013). The proportion of the genetic liability which has not yet been explained by the studied common variants is referred to as “the missing heritability” (Manolio et al., 2009). Part of the missing heritability may be explained by gene-environment (GxE) interaction, structural variants, and epistasis. Furthermore, rare genetic variation (MAF < 0.01), which is not captured in conventional GWAS arrays, may contribute substantially to the heritability of complex traits (Speed et al., 2017).

Genome-wide interrogations of rare variants that contribute to alcohol and tobacco use are scarce. Studies that instead investigated the role of rare variants in a limited number of candidate genes showed significant associations with substance dependence. For example, variants within the gene *CHRNA3* are associated with alcohol dependence (Haller et al., 2014) while rare variants in the genes *CHRNA4*, *NRXN1*, *CHRNA9*, *TAS2R38*, *CHRNA2*, *NTRK2*, *GABBR2*, *GRIN3A*, *DNM1*, *DBH*, *NRXN2*, *ANKK1/DRD2*, *NRXN3*, *CDH13*, and *ARRB2* were shown to influence the risk of nicotine dependence (Haller et al., 2012; Yang et al., 2015). However, the associations of these rare variants with alcohol and nicotine dependence have not been replicated, and are therefore suggestive in nature. The relevance of investigating rare variants is further demonstrated by an observed inverse relationship between a variant’s effect size and its frequency in the population (Park et al., 2011). This relationship indicates that a genome-wide exploration of the impact of rare genetic variants may provide additional insights into the genetic etiology of substance use phenotypes.

The introduction of the Illumina Infinium HumanExome BeadChip array (Illumina Inc., San Diego, CA) enabled the study of rare genetic variants in a genome-wide and cost-effective way (Cirulli and Goldstein, 2010; Grove et al., 2013; Lee et al., 2014; Vrieze et al., 2014). The “ExomeChip” contains a dense amount of low-frequency and rare nonsynonymous SNPs, and an additional smaller set of common SNPs that aid quality control of the genotypes. The nonsynonymous SNPs change the protein sequence by coding for different amino acids. Their interrogation facilitates functional interpretation of causal variants or genes due to a clear mechanism through which they exert their effect on traits and diseases. Combined with potentially large effect sizes, rare nonsynonymous SNPs are attractive drug targets. For alcohol and nicotine dependence, a single ExomeChip study has been conducted in 7181 individuals. No significant SNP or gene associations were detected, probably due to the limited sample size (Vrieze et al.,

2014).

The current study seeks to increase our knowledge on the role of rare exonic variants in alcohol and tobacco use and to improve our understanding of the biological mechanisms underlying these phenotypes. In a genome-wide meta-analysis, we aggregated the genetic association results for alcohol and tobacco use from eight Dutch cohorts (Brandtsma et al., 2012). The discovery sample includes 12,466 individuals for alcohol consumption and 7432 for tobacco use, which makes it the largest study of its kind to date. In addition, we tested our top findings in a case-control replication sample including 25,508 individuals for tobacco use disorder and 27,068 individuals for alcoholism. The aims of this study are: i) to investigate the association of rare exonic variants with alcohol consumption and tobacco use; ii) to investigate the aggregate effect of multiple rare variants within genes using gene-based tests; iii) to investigate pathways involved in alcohol consumption and tobacco use by performing gene-set analyses; and iv) to investigate the proportion of the total phenotypic variance in alcohol consumption and tobacco use explained by common and rare variants using Polygenic Risk Score (PRS) analysis.

2. Methods

2.1. Participants

The discovery sample included eight Dutch cohorts which were all part of the Biobanking and Biomolecular Research Infrastructure for The Netherlands (BBMRI-NL) (Brandtsma et al., 2012). The total sample, after genotype quality control (see below), consisted of 12,466 subjects for alcohol consumption and 7432 subjects for tobacco use, the sample contains both males and females. Only subjects between 18 and 85 years old were included. Detailed information about the individual cohorts can be found in Supplementary information section 2.

The replication sample consisted of a single cohort from BioVU: Vanderbilt University's DNA biobank (Rodén et al., 2008). Top SNPs for alcohol consumption were replicated using data from 367 cases with an ICD-9 diagnosis of "Alcoholism" and 25,141 controls. Top SNPs for tobacco use were replicated using data from 1927 cases with an ICD-9 diagnosis of "Tobacco use disorder" and 25,141 controls. Subjects were between 13 and 90 years old. Detailed information about the replication cohort can be found in Supplementary information section 2.

2.2. Phenotypes

Alcohol consumption and tobacco use were assessed with self-reports in the discovery cohorts. Alcohol consumption was defined as the number of standard alcoholic drinks per week. Tobacco use was defined by the number of cigarettes smoked per day (CPD) in current and past smokers (i.e., ever-smokers). Subjects who smoked fewer than 100 cigarettes during their life were not included in the analyses (i.e., "non-smokers"). To diminish the effect of outliers due to the skewed nature of the phenotype data, logarithmic transformations were performed. Detailed information about the phenotypes and their recruitment can be found in Supplementary Table 1 and Supplementary Sections 2 and 3. The phenotypes "Tobacco use disorder" and "Alcoholism", from the independent replication sample BioVU, were assessed with the International Classification of Diseases, Ninth Revision (ICD-9 codes).

2.3. Genotypes

All individuals of the discovery cohorts, as well as the replication cohort, were genotyped using the Illumina Human Exome BeadChip (Illumina inc., San Diego, CA). This ExomeChip interrogates 247,870 markers, and mainly includes low-frequency and rare variants (i.e., 90% of the variants had a MAF lower than 5%) (Grove et al., 2013). The remaining variants are common to aid in the genotype quality control. Detailed information about the specifications of the Illumina Infinium

Exome-24 v1.1 BeadChip can be found in Supplementary Table 2. Genotype intensities for the discovery samples were called with GenomeStudio and zCall to obtain genotype data. Initially, we used GenomeStudio (i.e., GenCall) for common variant calling and for the exclusion of individuals with low quality genotypes (Grove et al., 2013; Guo et al., 2014). Subsequently, we used zCall to call rare variants (Goldstein et al., 2012). The genotype intensities for the replication cohort were called with Genotyping module v1.9.4 from GenomeStudio v2011.1.

2.4. Quality control of genotype data

Quality control (QC) was performed using PLINK (Purcell et al., 2007). Each discovery cohort used the same analysis plan for QC in order to reach consistency (see Supplementary information section 3). We included only autosomal variants in the analysis. Individuals with $\geq 5\%$ missingness and SNPs with $\geq 5\%$ missingness were excluded, as well as individuals with ambiguous sex information and a heterozygosity rate outside the 0.35–0.45 range. In non-family based cohorts, one of each pair of individuals who were related ($\text{pihat} > 0.2$) were excluded. In addition, based on multidimensional scaling (MDS) and 1000 Genomes data (Abecasis et al., 2012), all non-Caucasian subjects were excluded. All SNPs with a minor allele count (MAC) of < 5 were removed. Finally, when combining the eight discovery cohorts, SNPs which deviated from HWE with a p -value larger than 1.0×10^{-5} and a call rate lower than 0.95 were excluded. A detailed overview of the QC per cohort can be found in Supplementary Table 3 and Supplementary information section 3. An overview of the remaining number of subjects and SNPs after QC by cohort can be found in Supplementary Table 3. The QC of the replication sample was performed using PLINK 1.9 (Chang et al., 2015), and followed the recommendation of Guo et al. (2014).

2.5. Single-variant and gene-based analyses

For each discovery cohort, association analyses were performed using a linear regression model in RVTESTS (Zhan et al., 2016), in which the inverse normal transformed trait residuals were analyzed. Ten principal components were included as covariates to control for population stratification. In family-based cohorts, relatedness was accounted for using the kinship matrix in RVTESTS (Lippert et al., 2011; Zhan et al., 2016). The association analysis on the independent replication sample was conducted using PLINK 1.9, using either a Fisher's exact test for SNPs with a MAF $< 1\%$ and a logistic regression for all SNPs with a MAF $\geq 1\%$.

To combine the association results of the individual discovery cohorts, we meta-analyzed the results with the software package RAREMETAL (Feng et al., 2014). The meta-analysis was performed using the single variant score statistics of each cohort, which summarize evidence for association, together with the covariance matrix, which summarizes linkage disequilibrium relationships among variants. To determine significance for the single-variant analyses, a Bonferroni correction for the number of SNPs tested was used. Due to the exclusion of SNPs with a MAC lower than 5, fewer SNPs were tested in the CPD sample than in the alcohol consumption sample, since the CPD sample included fewer subjects. For alcohol consumption and CPD, the Bonferroni corrected significance thresholds were set at 6.33×10^{-7} and 7.29×10^{-7} , respectively.

In addition to single-variant analysis, we performed gene-based tests using the sequence kernel association test (SKAT) (Wu et al., 2011). The significance threshold for the SKAT was determined by a Bonferroni correction for the number of genes tested; the thresholds for alcohol consumption and CPD were 5.02×10^{-6} and 5.75×10^{-6} , respectively. In the CPD sample fewer genes were tested than in the alcohol consumption sample, since the CPD sample included fewer SNPs, as explained in the preceding paragraph. The gene-based tests included

only variants with a MAF < 5%, and genes that contained two variants or more.

2.6. Joint analyses of single variants with an independent replication sample

To examine the robustness of our top findings, we also tested their association in BioVU (Roden et al., 2008). For the top 10 SNPs (i.e., SNPs with lowest p-values) that we identified, we performed an association analysis in PLINK (Purcell et al., 2007) using “Tobacco use disorder” and “Alcoholism” of ICD-9 codes. Following the recommendations of Skol et al. (2006) (Skol et al., 2006), we meta-analyzed these results with our discovery meta-analysis using the weighted Z-score method of METAL (Willer et al., 2010). Significance was determined using the Bonferroni corrected threshold from the discovery analysis (6.33×10^{-7} for alcohol consumption and 7.29×10^{-7} for CPD). We used the effective sample sizes to weight the BioVU samples, which were 1447 and 7159 individuals for Alcoholism and Tobacco use disorder, respectively (calculated using the following formula: $((4 \times \text{number of cases} \times \text{number of controls}) / (\text{number of cases} + \text{number of controls}))$), as recommended by the Broad Institute’s Ricopili pipeline).

2.7. Pathway analysis

We investigated whether pathways related to alcohol and nicotine addiction were associated with alcohol consumption and CPD by performing a gene-set analysis in MAGMA (de Leeuw et al., 2015). We tested two Gene Ontology (GO) pathways, namely the ‘response to nicotine’ and the ‘response to ethanol’ (Gene Ontology Consortium, 2015). The gene-based p-values obtained in the SKAT analyses were used as input. We performed competitive gene-set analyses which test whether the genes in the gene-set are more strongly associated with the phenotype of interest than the other genes included in our genotype data ($n = 13,413$). Our dataset enabled us to test 84 genes from the ‘response to ethanol’ and 36 from the ‘response to nicotine’ pathways. The associations were corrected for dependencies between genes (with a gene correlation matrix based on the ExomeChip SNPs) and confounding effects of gene size and gene density.

2.8. Polygenic risk score analysis

To determine whether ExomeChip variants explain a significant proportion of the phenotypic variance when aggregated, we performed a PRS analysis for alcohol consumption and CPD using PRSice (Euesden et al., 2015). The PRS is an individual-level score that is calculated based on the number of risk variants that a person carries, weighted by SNP effect sizes that are derived from an independent large-scale discovery GWAS. As such, the score is an indication of the total genetic risk

of a specific individual for a particular trait, and can be used to estimate genetic overlap between traits (Dudbridge, 2013; Smoller, 2013; Stringer et al., 2014).

We split our total sample of eight cohorts between a discovery (or training) sample and a target sample. The discovery sample consisted of data from seven BBMRI-NL cohorts, ensuring a sufficiently large sample to detect SNP effects for alcohol consumption and CPD. The remaining BBMRI-NL cohort (UHP), which was the second largest sample out of the eight cohorts, was selected as the target sample.

3. Results

3.1. Single-variant and gene-based tests in the discovery and joint analyses

In the discovery sample, the single-variant meta-analysis of the number of drinks per week did not reveal any significant associations (Fig. 1). The top ten SNPs are displayed in Table 1. The lowest p-value was found for a rare variant in *IQSEC1* (MAF = 0.005, $p = 1.22 \times 10^{-6}$). The second most strongly associated SNP was located in *AKAP13* (MAF = 0.017, $p = 9.19 \times 10^{-6}$). The gene-based analysis did not reveal any genes significantly associated with alcohol consumption (Supplementary Fig. 1). The top 10 genes are displayed in Table 2. Similar to the single variant analysis, *IQSEC1* was most strongly associated with alcohol consumption in the gene-based analysis ($p = 8.87 \times 10^{-6}$). Supplementary Table 4 presents the results of the gene-based analysis for genes previously associated with alcohol dependence by rare variant studies.

The single-variant analysis of CPD also did not result in any significant associations after Bonferroni correction for the number of SNPs tested (see Table 1 and Fig. 2). The strongest association was found for a rare variant in *HSPG2* (MAF = 0.003, $p = 1.99 \times 10^{-6}$). In addition, we did not identify any genes that were significantly associated with CPD in the gene-based analysis (Supplementary Fig. 2 and Table 2). The lowest p-value was found for *LETMD1* ($p = 8.68 \times 10^{-5}$). Supplementary Table 4 presents the results of the gene-based analysis for genes previously associated with nicotine dependence by rare variant studies.

We subsequently investigated whether the 10 most strongly associated variants for alcohol consumption and CPD showed an effect on alcohol and tobacco use disorders in BioVU. In this cohort, genotype data on nine alcohol use top SNPs and seven tobacco use top SNPs were present. When jointly analyzing our meta-analysis associations with BioVU results, we identified two genome-wide significant SNPs for tobacco use (disorder) (Table 3). The two common variants for tobacco use (disorder) are an exonic and an intronic SNPs located in *CHRNA5* and *AGPHD*, respectively, located at the chromosomal region at 15q25.1. The most significant variant was rs16969968 in the gene *CHRNA5* ($p = 2.39 \times 10^{-7}$).

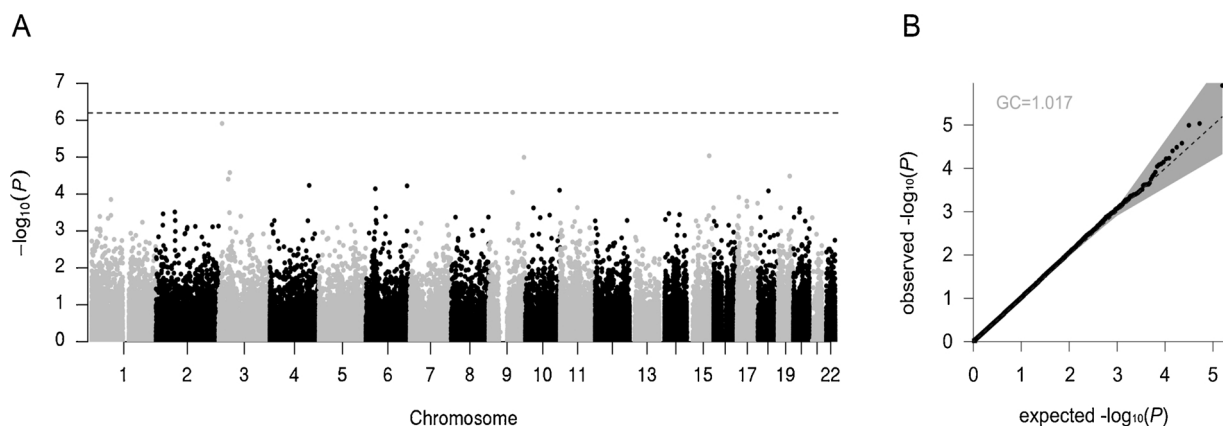


Fig. 1. Single variant association analysis results of alcohol consumption, A) Manhattan plot, B) quantile-quantile (QQ)-plot.

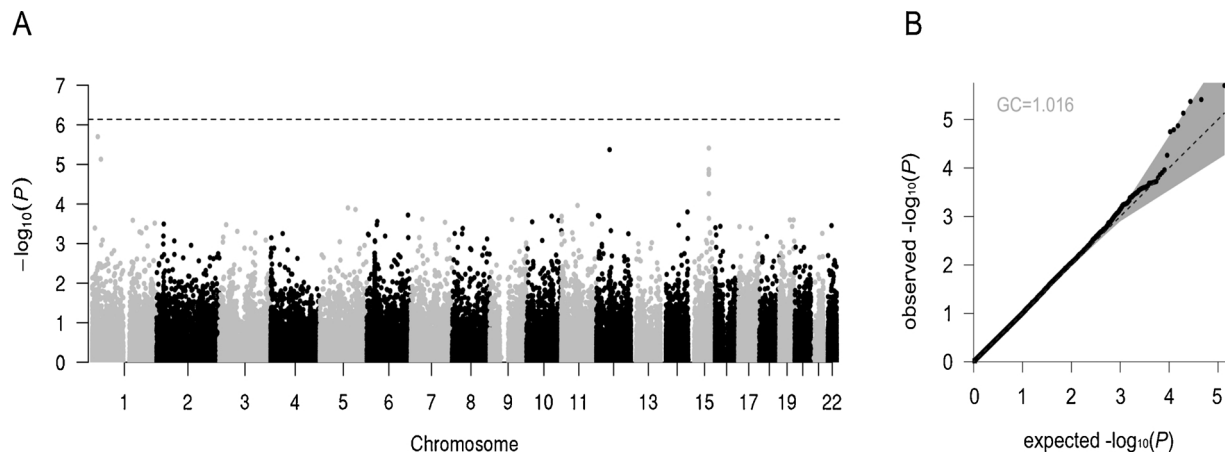


Fig. 2. Single variant association analysis results of CPD, A) Manhattan plot, B) quantile-quantile (QQ)-plot.

3.2. Pathway analyses

The ‘response to ethanol’ and ‘response to nicotine’ pathways from the GO database ([Gene Ontology Consortium, 2015](#)) were not significantly involved in alcohol consumption and CPD (p -values of 0.29 and 0.18, respectively).

3.3. Polygenic risk score analysis

We performed PRS analyses to investigate the aggregate contribution of ExomeChip variants to substance use taking seven of our BBMRI-NL samples as discovery sample, and the remaining BBMRI-NL sample (UHP) as target sample. The PRS analyses did not show a significant contribution of rare variants to the phenotypic variance for alcohol consumption ($p = 0.124$ for $P_T < 0.02$) or CPD ($p = 0.067$ for $P_T < 0.45$). (Supplementary information section 1 and Supplementary Figs. 3 and 4).

4. Discussion

The primary aim of this study was to determine the role of rare genetic variation in alcohol and tobacco use. Conducting a meta-analysis on ExomeChip data of eight cohorts allowed us to study the role of rare and low-frequency exonic variants, and in addition a small pool of common variants, in a large sample. We performed both single-variant and gene-based tests. Despite our large sample size, we were not able to identify novel variants or genes that were significantly associated with alcohol consumption or tobacco use in our discovery sample. However, we identified several potentially interesting signals that we investigated for their robustness in an independent sample with data on alcohol and tobacco use disorders. For tobacco use, the joint analysis of this sample with our initial meta-analysis showed genome-wide significant association for two common genetic variants in the genes *AGPHD1* and *CHRNA5*, both located at chromosomal region 15q25.1. These genes belong to the *IREB2-CHRNA3-CHRNA5-CHRNA4-HYKK(AGPHD1)-PSMA4* gene-cluster that has repeatedly been associated with smoking severity ([Barrie et al., 2017](#); [Liu et al., 2010](#); [Saccone et al., 2010b](#); [Tobacco and Genetics, 2010](#)). The significant non-synonymous variant rs16969968 in *CHRNA5*, a gene which encodes for a nicotinic acetylcholine receptor, is the strongest risk variant for nicotine dependence ([Saccone et al., 2010b](#)). *AGPHD1* was previously reported to be associated with chronic obstructive pulmonary disease (COPD) ([Siedlinski et al., 2013](#)). Our joint analysis of alcohol consumption/alcoholism did not reveal any significant variants.

In addition to the single variant and gene-based analyses, we performed the pathway analysis of two known sets of genes involved in response to nicotine and ethanol. No significant associations were

found, suggesting that if rare variants in these pathways play a role, their effects are small and the statistical power of our study was probably not sufficient to detect them. The PRS analysis did not show a significant role for ExomeChip variants, meaning that variation in alcohol and tobacco use could not be predicted by (primarily) rare and low-frequency variants. In contrast, previously PRS has been successfully used to predict substance use related phenotypes based on common variants ([Vink et al., 2014](#)). Thus, rare exonic variants explain less phenotypic variation than common variants, although the differential predictive ability of common and rare variants could also be explained by insufficient power to estimate effect sizes of rare variants reliably.

Summarizing, we did not find significant associations between rare exonic variants and alcohol and tobacco use. Although somewhat disappointing, it does provide important insight into the genetic architecture of these traits, as it suggests that these phenotypes are not influenced by rare variants with large effect sizes. A power analysis conducted with the Genetic Power Calculator ([Purcell et al., 2003](#); [Visscher et al., 2017](#)) suggests that our discovery samples for alcohol consumption and CPD had sufficient statistical power (> 80%) to detect low-frequency variants (MAF = 0.01) with effect sizes of Cohen’s $d \geq 0.37$ for alcohol consumption and ≥ 0.48 for CPD. In contrast, for the detection of rare variants with a MAF of 0.001, we had sufficient power to detect variants with large effect sizes (i.e., Cohen’s $d \geq 1.17$ for alcohol consumption and Cohen’s $d \geq 1.5$ for tobacco use). Although our findings coupled with the power analysis suggest that rare genetic variants with large effect sizes do not play an important role, we cannot rule out the possibility that rare exonic variants with smaller effects contribute to the phenotypic variation in alcohol consumption and CPD, as previously suggested by candidate gene studies ([Haller et al., 2012, 2014](#); [Olsson et al., 2016](#); [Thorgeirsson et al., 2016](#); [Xie et al., 2011](#); [Yang et al., 2015](#)). Even though not all previously reported rare variants were tagged on the ExomeChip or passed our QC, several of these variants were included in our analyses, and several of the associated genes were analyzed with our gene-based tests. Hence, our observation of the absence of significant associations of rare variants with alcohol consumption or tobacco use, suggests that some of these rare variants may have been false positive findings or that the effect sizes are small. Based on robust findings from the current study, a study by Vrieze et al., the above-mentioned candidate gene studies, and the largest conducted ExomeChip study (investigating adult human height) to date ([Marouli et al., 2017](#)), we estimate that effect sizes up to Cohen’s d of 0.4 can be expected for rare variants contributing to the phenotypic variance in alcohol consumption and CPD ([Marouli et al., 2017](#); [Vrieze et al., 2013](#)). This estimate indicates that extremely large sample sizes would be required for the detection of significant rare variant associations ([Visscher et al., 2017](#)).

We cannot generalize our findings to other substance use traits as the power to detect rare genetic variants depends strongly on the genetic architecture of the trait under investigation. For example, rare and low-frequency variants have been shown to influence other complex traits, including height, ADHD, breast cancer, type 2 diabetes, and schizophrenia (Cheung et al., 2017; Marouli et al., 2017; Richards et al., 2016; Zayats et al., 2016; Zhou et al., 2017). In some of these studies, sample sizes were smaller than the sample size in the present study. A possible explanation is that a larger proportion of the genetic variation may be attributable to rare variants for diseases that directly affect the fitness of individuals, because of inherent negative selection on these variants (Simons and Sella, 2016).

Despite the lack of significant findings, it seems likely that at least some of our top variants are true associations since the majority of our top variants showed a consistent direction of effect over the different cohorts that were included. Interestingly, directions of effect were also fairly consistent in the replication sample in which ICD-9 diagnoses of alcoholism and tobacco use disorder were obtained. This observation suggests that quantitative measurements of alcohol and tobacco use and clinical diagnoses have a shared genetic liability. Furthermore, previous meta-analyses have combined different trait measures into a single phenotypic trait and generally found high genetic correlations between the measures (Okbay et al., 2016; Wray and Sullivan, 2017).

The current findings should be interpreted in view of some key limitations. The ExomeChip only tags variants located in the exomes of genes, whereas most GWAS trait-associated hits are located outside exonic regions (Maurano et al., 2012), suggesting that most trait-associated SNPs act through the regulation of gene expression rather than by altering the protein sequence. It is possible that this characteristic also holds true for rare variants. Furthermore, the phenotypes measured in the discovery and the replication sample were not identical, and this discordance may have reduced the statistical power to detect significant associations in the joint analysis. In addition, most rare variants were not present in all cohorts of our study, which decreased the statistical power to detect an effect. Also, the prevalence rates of alcoholism and tobacco use disorder in BioVU were lower than expected from the general population (Breslau et al., 2001; Hasin et al., 2007), which may be explained by misclassification among the controls. Diagnosis of substance use disorders within BioVU is obtained from medical record data. It is likely that a proportion of the cases (i.e., subjects who meet criteria for alcoholism or tobacco use disorder) were misclassified as controls due to a lack of systematic screening. Furthermore, BioVU is a hospital-based sample, and it may therefore not be representative of the general population.

5. Conclusions

We have shown that low-frequency and rare exonic variants do not have a large impact on inter-individual variability in alcohol and tobacco use. Our large sample size allowed us to rule out, with high confidence, the impact of low-frequency or rare exonic variants with large effect on these phenotypes. Our findings confirmed the influence of common genetic variants located in the *CHRNA5-CHRNA3-CHRNA4* cluster of nicotinic acetylcholine receptor subunit genes on tobacco use.

Role of funding source

ATM and EMD are supported by the Foundation Volksbond Rotterdam, FV is supported by the Investissement d'Avenir program managed by the ANR under reference ANR-11-IDEX-0004-02. B. ARH is supported by The Netherlands Organization for Scientific Research (NWO Brain and Cognition 433-09-228). DOMK is supported by Dutch Science Organization (ZonMW-VENI Grant 916.14.023). DP acknowledges funding from The Netherlands Organization for Scientific Research (NWO VICI 453-14-005). The exome chip data for the NBS were generated in a research project that was financially supported by

BBMRI-NL, a Research Infrastructure financed by the Dutch government (NWO 184.021.007). The Longitudinal Aging Study Amsterdam (LASA) is largely supported by a grant from the Netherlands Ministry of Health Welfare and Sports, Directorate of Long-Term Care. ABCS was funded by the Dutch cancer society project grant NKI2009-4363. Funding for the project was provided by the Netherlands Organization for Scientific Research under award number 184021007, dated July 9, 2009 and made available as a Rainbow Project of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL). The funding sources had no involvement in study design; in the collection, analysis and interpretation of the data; in the writing of the report or the decision to submit for publication.

Contributors

ATM performed a pilot study, wrote the protocol, performed the analyses, and prepared the first draft of the manuscript. ARH performed the analyses and prepared the first draft of the manuscript. LB performed analyses of the replication cohort and contributed in critical revision of the manuscript. HdeK performed part of the analyses and contributed in critical revision of the manuscript. FV, DS, WvdB, and DD have contributed to the clinical interpretation of the results and critical revision of the manuscript. AM, AvdL, DOMK, DP, EG, EJB, FRR, IV, JLP, LAK, MdH, MG, MKS, NMvs, NS, OHK, PCS, RdeM, RK, RN, RLG, SHV, TEG, and TJP all had a role in the data collection process for the current study and contributed in critical revision of the manuscript. EMD designed the methodology, contributed in the writing process, and supervised the project. All authors have reviewed and approved the final manuscript.

Conflict of interest

No conflict declared.

Acknowledgements

Statistical analyses were carried out on the Genetic Cluster Computer (<http://www.geneticcluster.org>) hosted by SURFsara and financially supported by the Netherlands Scientific Organization (NWO 480-05-003 PI: Posthuma) along with a supplement from the Dutch Brain Foundation and the VU University Amsterdam.

ABCS: We are grateful to Frans Hogervorst, Annetien Broeks, and all patients and clinicians involved. The study was funded by the Dutch cancer society project grant NKI2009-4363.

LASA: We are very grateful to all participants of the Longitudinal Aging Study Amsterdam for their participation, and to all interviewers, fieldwork coordinators, supporting staff and researchers for their contribution.

NBS: Principal investigators of the Nijmegen Biomedical Study are L.A. Kiemeny, M. den Heijer, A.L.M. Verbeek, D.W. Swinkels and B. Franke.

NEO: We thank all individuals who participated in the Netherlands Epidemiology in Obesity study, all participating general practitioners for inviting eligible participants and all research nurses for collection of the data. We thank the NEO study group, Pat van Beelen, Petra Noordijk and Ingeborg de Jonge for the coordination, lab and data management of the NEO study. The genotyping in the NEO study was supported by the Centre National de Génotypage (Paris, France), headed by Jean-Francois Deleuze. The NEO study is supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Center, and by the Leiden University, Research Profile Area Vascular and Regenerative Medicine. www.neostudie.nl.

NESSCOG: We thank all participating subjects. This research was part of Science Live, the innovative research program of science center NEMO that enables scientists to carry out real, publishable, peer-

reviewed research using NEMO visitors as volunteers. Dr. D. Posthuma acknowledges funding from The Netherlands Organization for Scientific Research (NWO VICI 453-14-005).

UHP: Leidsche Rijn Julius Gezondheidscentra (LRJG) <http://www.lrjg.nl/>

UCP: acknowledges all participants and the researchers of the UCP study, Mark C.H. de Groot, Olaf H. Klungel, Anke-Hilse Maitland-Van der Zee, Patrick C. Souverein.

Funding for the ExomeChip project was provided by the Netherlands Organization for Scientific Research under award number 184021007, dated July 9, 2009 and made available as a Rainbow Project of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.drugalcdep.2018.03.026>.

References

- 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M. A., Durbin, R.M., Hadsaker, R.E., Kang, H.M., Marth, G.T., McVean, G.A., 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65.
- Barrie, E.S., Hartmann, K., Lee, S.H., Frater, J.T., Seweryn, M., Wang, D.X., Sadee, W., 2017. The CHRNA5/CHRNA3/CHRNA4 nicotinic receptor regulome genomic architecture, regulatory variants, and clinical associations. *Hum. Mutat.* 38, 112–119.
- Brandsma, M., Baas, F., Bakker, P.I.W., Beem, E.P., Boomsma, D.I., Bovenberg, J., Bueno-de Mesquita, B., van Duijn, C.M., Kiemerney, L.A., Klungel, O., de Leeuw, P.W., van Leeuwen, F., Ridder-Numan, J., Stolk, R.P., Slagboom, P.E., Swetz, M., Wijmenga, C., van Ommen, G.J.B., 2012. How to kickstart a national biobanking infrastructure—experiences and prospects of BBMRI-NL. *Nord. Psychol.* 21, 143–148.
- Breslau, N., Johnson, E.O., Hiripi, E., Kessler, R., 2001. Nicotine dependence in the United States—prevalence, trends, and smoking persistence. *Arc. Gen. Psychiatry* 58, 810–816.
- Broms, U., Silventoinen, K., Madden, P.A., Heath, A.C., Kaprio, J., 2006. Genetic architecture of smoking behavior: a study of Finnish adult twins. *Twin Res. Hum. Genet.* 9, 64–72.
- Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.
- Cheung, C.Y., Tang, C.S., Xu, A., Lee, C.H., Au, K.W., Xu, L., Fong, C.H., Kwok, K.H., Chow, W.S., Woo, Y.C., Yuen, M.M., Hai, J.S., Jin, Y.L., Cheung, B.M., Tan, K.C., Cherny, S.S., Zhu, F., Zhu, T., Thomas, G.N., Cheng, K.K., Jiang, C.Q., Lam, T.H., Tse, H.F., Sham, P.C., Lam, K.S., 2017. Exome-chip association analysis reveals an Asian-specific missense variant in PAX4 associated with type 2 diabetes in Chinese individuals. *Diabetologia* 60, 107–115.
- Cirulli, E.T., Goldstein, D.B., 2010. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genet.* 11, 415–425.
- Clarke, T.K., Adams, M.J., Davies, G., Howard, D.M., Hall, L.S., Padmanabhan, S., Murray, A.D., Smith, B.H., Campbell, A., Hayward, C., Porteous, D.J., Deary, I.J., McIntosh, A.M., 2017. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N = 112 117). *Mol. Psychiatry* 22, 1376–1384.
- de Leeuw, C.A., Mooij, J.M., Heskes, T., Posthuma, D., 2015. MAGMA: generalized gene-Set analysis of GWAS data. *PLoS Comput. Biol.* 11, e1004219.
- David, S.P., Hamidovic, A., Chen, G.K., Bergen, A.W., Wessel, J., Kasberger, J.L., Brown, W.M., Petruzella, S., Thacker, E.L., Kim, Y., Nalls, M.A., Tranah, G.J., Sung, Y.J., Ambrosone, C.B., Arnett, D., Bandera, E.V., Becker, D.M., Becker, L., Berndt, S.I., Bernstein, L., Blot, W.J., Broeckel, U., Buxbaum, S.G., Caporaso, N., Casey, G., et al., 2012. Genome-wide meta-analyses of smoking behaviors in African Americans. *Transl. Psychiatry* 2, e119.
- De Bellis, M.D., Zisk, A., 2014. The biological effects of childhood trauma. *Child Adolesc. Psychiatr. Clin. N. Am.* 23, 185–222.
- Dudbridge, F., 2013. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* 9.
- Euesden, J., Lewis, C.M., O'Reilly, P.F., 2015. PRSice: polygenic risk score software. *Bioinformatics* 31, 1466–1468.
- Feng, S., Liu, D., Zhan, X., Wing, M.K., Abecasis, G.R., 2014. RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics* 30, 2828–2829.
- Gelernter, J., Kranzler, H.R., Sherva, R., Almsy, L., Koesterer, R., Smith, A.H., Anton, R., Preuss, U.W., Ridinger, M., Rujescu, D., Wodarz, N., Zill, P., Zhao, H., Farrer, L.A., 2014. Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. *Mol. Psychiatry* 19, 41–49.
- Gene Ontology Consortium, 2015. Gene ontology consortium: going forward. *Nucleic Acids Res.* 43, D1049–D1056.
- Goh, E.T., Morgan, M.Y., 2017. Review article: pharmacotherapy for alcohol dependence—the why, the what and the wherefore. *Aliment. Pharmacol. Ther.* 45, 865–882.
- Goldstein, J.I., Crenshaw, A., Carey, J., Grant, G.B., Maguire, J., Fromer, M., O'Dushlaine, C., Moran, J.L., Chambert, K., Stevens, C., Sklar, P., Hultman, C.M., Purcell, S., McCarroll, S.A., Sullivan, P.F., Daly, M.J., Neale, B.M., 2012. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* 28, 2543–2545.
- Grove, M.L., Yu, B., Cochran, B.J., Haritunians, T., Bis, J.C., Taylor, K.D., Hansen, M., Borecki, I.B., Cupples, L.A., Fornage, M., Gudnason, V., Harris, T.B., Kathiresan, S., Kraaij, R., Launer, L.J., Levy, D., Liu, Y., Mosley, T., Peloso, G.M., Psaty, B.M., Rich, S.S., Rivadeneira, F., Siscovick, D.S., Smith, A.V., Uitterlinden, A., et al., 2013. Best practices and joint calling of the HumanExome BeadChip: the CHARGE consortium. *PLoS One* 8, e68095.
- Guo, Y., He, J., Zhao, S.L., Wu, H., Zhong, X., Sheng, Q.H., Samuels, D.C., Shyr, Y., Long, J.R., 2014. Illumina human exome genotyping array clustering and quality control. *Nat. Protoc.* 9, 2643–2662.
- Haller, G., Druley, T., Vallania, F.L., Mitra, R.D., Li, P., Akk, G., Steinbach, J.H., Breslau, N., Johnson, E., Hatsukami, D., Stitzel, J., Bierut, L.J., Goate, A., 2012. Rare missense variants in CHRNA4 are associated with reduced risk of nicotine dependence. *Hum. Mol. Genet.* 21, 647–655.
- 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., 2014. Rare missense variants in CHRNA3 and CHRNA5 are associated with risk of alcohol and cocaine dependence. *Hum. Mol. Genet.* 23, 810–819.
- Harada, S., Agarwal, D.P., Goedde, H.W., Ishikawa, B., 1983. Aldehyde dehydrogenase isozyme variation and alcoholism in Japan. *Pharmacol. Biochem. Behav.* 18, 151–153.
- Hasin, D.S., Stinson, F.S., Ogburn, E., Grant, B.F., 2007. Prevalence, correlates, disability, and comorbidity of DSM-IV alcohol abuse and dependence in the United States—results from the national epidemiologic survey on alcohol and related conditions. *Arch. Gen. Psychiatry* 64, 830–842.
- Kendler, K.S., Gardner, C., Dick, D.M., 2011. Predicting alcohol consumption in adolescence from alcohol-specific and general externalizing genetic risk factors, key environmental exposures and their interaction. *Psychol. Med.* 41, 1507–1516.
- Koopmans, J.R., vanDoornen, L.J.P., Boomsma, D.I., 1997. Association between alcohol use and smoking in adolescent and young adult twins: a bivariate genetic analysis. *Alcohol Clin. Exp. Res.* 21, 537–546.
- Lee, S., Abecasis, G.R., Boehnke, M., Lin, X., 2014. Rare-variant association analysis: study designs and statistical tests. *Am. J. Hum. Genet.* 95, 5–23.
- Lijffijt, M., Hu, K., Swann, A.C., 2014. Stress modulates illness-course of substance use disorders: a translational review. *Front. Psychiatry* 5, 83.
- Lippert, C., Listgarten, J., Liu, Y., Kadie, C.M., Davidson, R.I., Heckerman, D., 2011. FaST linear mixed models for genome-wide association studies. *Nat. Methods* 8, 833–835.
- Liu, J.Z., Tozzi, F., Waterworth, D.M., Pillai, S.G., Muglia, P., Middleton, L., Berrettini, W., Knouff, C.W., Yuan, X., Waeber, G., Vollenweider, P., Preisig, M., Wareham, N.J., Zhao, J.H., Loos, R.J., Barroso, I., Khaw, K.T., Grundy, S., Barter, P., Mahley, R., Kesaniemi, A., McPherson, R., Vincent, J.B., Strauss, J., Kennedy, J.L., et al., 2010. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat. Genet.* 42, 436–440.
- Loke, A.Y., Mak, Y.W., 2013. Family process and peer influences on substance use by adolescents. *Int. J. Environ. Res. Public Health* 10, 3868–3885.
- Lubke, G.H., Hottenga, J.J., Walters, R., Laurin, C., de Geus, E.J.C., Willemsen, G., Smit, J.H., Middeldorp, C.M., Penninx, B.W.J.H., Vink, J.M., Boomsma, D.I., 2012. Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms. *Biol. Psychiatry* 72, 707–709.
- Macgregor, S., Lind, P.A., Bucholz, K.K., Hansell, N.K., Madden, P.A., Richter, M.M., Montgomery, G.W., Martin, N.G., Heath, A.C., Whitfield, J.B., 2009. Associations of ADH and ALDH2 gene variation with self report alcohol reactions, consumption and dependence: an integrated analysis. *Hum. Mol. Genet.* 18, 580–593.
- Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorf, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F., McCarroll, S.A., et al., 2009. Finding the missing heritability of complex diseases. *Nature* 461, 747–753.
- Marouli, E., Graff, M., Medina-Gomez, C., Lo, K.S., Wood, A.R., Kjaer, T.R., Fine, R.S., Lu, Y., Schurmann, C., Highland, H.M., Rueger, S., Thorleifsson, G., Justice, A.E., Lamparter, D., Stirrups, K.E., Turcot, V., Young, K.L., Winkler, T.W., Esko, T., Karaderi, T., Locke, A.E., Masca, N.G., Ng, M.C., Mudgal, P., Rivas, M.A., et al., 2017. Rare and low-frequency coding variants alter human adult height. *Nature* 542, 186–190.
- Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutayavin, T., Stehling-Sun, S., Johnson, A.K., Canfield, T.K., Giste, E., Diegel, M., Bates, D., Hansen, R.S., Neph, S., Sabo, P.J., Heimfeld, S., Raubitschek, A., et al., 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190–1195.
- Mbarek, H., Milaneschi, Y., Fedko, I.O., Hottenga, J.J., de Moor, M.H., Jansen, R., Gelernter, J., Sherva, R., Willemsen, G., Boomsma, D.I., Penninx, B.W., Vink, J.M., 2015. The genetics of alcohol dependence Twin and SNP-based heritability, and genome-wide association study based on AUDIT scores. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 168, 739–748.
- Okbay, A., Baselmans, B.M.L., De Neve, J.E., Turley, P., Nivard, M.G., Fontana, M.A., Meddens, S.F.W., Linner, R.K., Rietveld, C.A., Derringer, J., Gratten, J., Lee, J.J., Liu, J.Z., de Vlaming, R., Ahluwalia, T.S., Buchwald, J., Cavadi, A., Frazier-Wood, A.C., Furlotte, N.A., Garfield, V., Geisel, M.H., Gonzalez, J.R., Haitjema, S., Karlsson, R.,

- van der Laan, S.W., et al., 2016. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses (vol 48, pg 624, 2016). *Nat. Genet.* 48, 1591.
- Olson, E., Saccone, N.L., Johnson, E.O., Chen, L.S., Culverhouse, R., Doheny, K., Foltz, S.M., Fox, L., Gogarten, S.M., Hartz, S., Hettrick, K., Laurie, C.C., Marosy, B., Amin, N., Arnett, D., Barr, R.G., Bartz, T.M., Bertelsen, S., Borecki, I.B., Brown, M.R., Chasman, D.I., van Duijn, C.M., Feitosa, M.F., Fox, E.R., Franceschini, N., et al., 2016. Rare, low frequency and common coding variants in CHRNA5 and their contribution to nicotine dependence in European and African Americans. *Mol. Psychiatry* 21, 601–607.
- Park, J.H., Gail, M.H., Weinberg, C.R., Carroll, R.J., Chung, C.C., Wang, Z., Chanock, S.J., Fraumeni Jr., J.F., Chatterjee, N., 2011. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. *Proc. Natl. Acad. Sci. U. S. A.* 108, 18026–18031.
- Purcell, S., Cherny, S.S., Sham, P.C., 2003. Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19, 149–150.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575.
- Richards, A.L., Leonenko, G., Walters, J.T., Kavanagh, D.H., Rees, E.G., Evans, A., Chambert, K.D., Moran, J.L., Goldstein, J., Neale, B.M., McCarroll, S.A., Pocklington, A.J., Holmans, P.A., Owen, M.J., O'Donovan, M.C., 2016. Exome arrays capture polygenic rare variant contributions to schizophrenia. *Hum. Mol. Genet.* 25, 1001–1007.
- Roden, D.M., Pulley, J.M., Basford, M.A., Bernard, G.R., Clayton, E.W., Balsler, J.R., Masys, D.R., 2008. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin. Pharmacol. Ther.* 84, 362–369.
- Saccone, N.L., Culverhouse, R.C., Schwantes-An, T.H., Cannon, D.S., Chen, X., Cichon, S., Giegling, I., Han, S., Han, Y., Keskitalo-Vuokko, K., Kong, X., Landi, M.T., Ma, J.Z., Short, S.E., Stephens, S.H., Stevens, V.L., Sun, L., Wang, Y., Wenzlaff, A.S., Aggen, S.H., Breslau, N., Broderick, P., Chatterjee, N., Chen, J., Heath, A.C., et al., 2010a. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet.* 6 pii: e1001053.
- Saccone, N.L., Culverhouse, R.C., Schwantes-An, T.H., Cannon, D.S., Chen, X.N., Cichon, S., Giegling, I., Han, S.Z., Han, Y.H., Keskitalo-Vuokko, K., Kong, X.Y., Landi, M.T., Ma, J.Z., Short, S.E., Stephens, S.H., Stevens, V.L., Sun, L.W., Wang, Y.F., Wenzlaff, A.S., Aggen, S.H., Breslau, N., Broderick, P., Chatterjee, N., Chen, J.C., Heath, A.C., et al., 2010b. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet.* 6 pii: e1001053.
- Schumann, G., Liu, C., O'Reilly, P., Gao, H., Song, P., Xu, B., Ruggeri, B., Amin, N., Jia, T., Preis, S., Segura Lepe, M., Akira, S., Barbieri, C., Baumeister, S., Cauchi, S., Clarke, T.K., Enroth, S., Fischer, K., Hallfors, J., Harris, S.E., Hieber, S., Hofer, E., Hottenga, J.J., Johansson, A., Joshi, P.K., Kaartinen, N., Laitinen, J., Lemaître, R., Loukola, A., et al., 2016. KLB is associated with alcohol drinking, and its gene product beta-Klotho is necessary for FGF21 regulation of alcohol preference. *Proc. Natl. Acad. Sci. U. S. A.* 113, 14372–14377.
- Siedlinski, M., Tingley, D., Lipman, P.J., Cho, M.H., Litonjua, A.A., Sparrow, D., Bakke, P., Gulsvik, A., Lomas, D.A., Anderson, W., Kong, X.Y., Rennard, S.I., Beaty, T.H., Hokanson, J.E., Crapo, J.D., Lange, C., Silverman, E.K., Investigator, C., Investigator, E., 2013. Dissecting direct and indirect genetic effects on chronic obstructive pulmonary disease (COPD) susceptibility. *Hum. Genet.* 132, 431–441.
- Simons, Y.B., Sella, G., 2016. The impact of recent population history on the deleterious mutation load in humans and close evolutionary relatives. *Curr. Opin. Genet. Dev.* 41, 150–158.
- Skol, A.D., Scott, L.J., Abecasis, G.R., Boehnke, M., 2006. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* 38, 209–213.
- Smoller, J.W., 2013. Cross Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis (vol 381, pg 1371–9, 2013). *Lancet* 381, 1360.
- Speed, D., Cai, N., Consortium, U., Johnson, M.R., Nejentsev, S., Balding, D.J., 2017. Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* 49, 986–992.
- Stead, L.F., Lancaster, T., 2012. Combined pharmacotherapy and behavioural interventions for smoking cessation. *The Cochrane database of systematic reviews* 10, CD008286.
- Stringer, S., Kahn, R.S., de Witte, L.D., Ophoff, R.A., Derks, E.M., 2014. Genetic liability for schizophrenia predicts risk of immune disorders. *Schizophr. Res.* 159, 347–352.
- Thomasson, H.R., Edenberg, H.J., Crabb, D.W., Mai, X.L., Jerome, R.E., Li, T.K., Wang, S.P., Lin, Y.T., Lu, R.B., Yin, S.J., 1991. Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. *Am. J. Hum. Genet.* 48, 677–681.
- Thorgerisson, T.E., Steinberg, S., Reginsson, G.W., Bjornsdottir, G., Rafnar, T., Jonsdottir, I., Helgadóttir, A., Gretarsdóttir, S., Helgadóttir, H., Jonsson, S., Matthiasson, S.E., Gislason, T., Tyrffingsson, T., Gudbjartsson, T., Isaksson, H.J., Hardardóttir, H., Sigvaldason, A., Kiemeny, L.A., Haugen, A., Zienolddiny, S., Wolf, H.J., Franklin, W.A., Panadero, A., Mayordomo, J.I., Hall, I.P., et al., 2016. A rare missense mutation in CHRNA4 associates with smoking behavior and its consequences. *Mol. Psychiatry* 21, 594–600.
- Tobacco, Genetics, C., 2010. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* 42, 441–447.
- Van Ryzin, M.J., Fosco, G.M., Dishion, T.J., 2012. Family and peer predictors of substance use from early adolescence to early adulthood: an 11-year prospective analysis. *Addict. Behav.* 37, 1314–1324.
- Verhulst, B., Neale, M.C., Kendler, K.S., 2015. The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. *Psychol. Med.* 45, 1061–1072.
- Vink, J.M., Willemsen, G., Boomsma, D.I., 2005. Heritability of smoking initiation and nicotine dependence. *Behav. Genet.* 35, 397–406.
- Vink, J.M., Hottenga, J.J., de Geus, E.J.C., Willemsen, G., Neale, M.C., Furberg, H., Boomsma, D.I., 2014. Polygenic risk scores for smoking: predictors for alcohol and cannabis use? *Addiction* 109, 1141–1151.
- Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., Yang, J., 2017. 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* 101, 5–22.
- Vrieze, S.I., McGue, M., Miller, M.B., Hicks, B.M., Iacono, W.G., 2013. Three mutually informative ways to understand the genetic relationships among behavioral disinhibition, alcohol use, drug use, nicotine use/dependence, and their co-occurrence. *Twin biometry, GCTA, and genome-wide scoring.* *Behav. Genet.* 43, 97–107.
- Vrieze, S.I., Feng, S., Miller, M.B., Hicks, B.M., Pankratz, N., Abecasis, G.R., Iacono, W.G., McGue, M., 2014. Rare nonsynonymous exonic variants in addiction and behavioral disinhibition. *Biol. Psychiatry* 75, 783–789.
- Willer, C.J., Li, Y., Abecasis, G.R., 2010. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190–2191.
- World Health Organization, 2009. Global health risks: mortality and burden of disease attributable to selected major risks. World Health Organization. http://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf. (Accessed April 13 2018).
- World Health Organization, 2014. Global status report on alcohol and health. http://www.who.int/substance_abuse/publications/global_alcohol_report/en/. (Accessed April 13 2018).
- World Health Organization, 2015. WHO report on the global tobacco epidemic. http://www.who.int/tobacco/global_report/2015/en/ (Accessed April 13, 2018).
- Wray, N.R., Sullivan, P.F., 2017. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *bioRxiv* 167577. <http://dx.doi.org/10.1101/167577>. Accessed April 13, 2018.
- Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., Lin, X., 2011. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* 89, 82–93.
- Xie, P.X., Kranzler, H.R., Krauthammer, M., Cosgrove, K.P., Oslin, D., Anton, R.F., Farrer, L.A., Picciotto, M.R., Krystal, J.H., Zhao, H.Y., Gelernter, J., 2011. Rare nonsynonymous variants in alpha-4 nicotinic acetylcholine receptor gene protect against nicotine dependence. *Biol. Psychiat.* 70, 528–536.
- Yang, J., Wang, S., Yang, Z., Hodgkinson, C.A., Iarikova, P., Ma, J.Z., Payne, T.J., Goldman, D., Li, M.D., 2015. The contribution of rare and common variants in 30 genes to risk nicotine dependence. *Mol. Psychiatry* 20, 1467–1478.
- Yoshida, A., Hsu, L.C., Yasunami, M., 1991. Genetics of human alcohol-metabolizing enzymes. *Prog. Nucleic Acid Res. Mol. Biol.* 40, 255–287.
- Young-Wolff, K.C., Enoch, M.A., Prescott, C.A., 2011. The influence of gene-environment interactions on alcohol consumption and alcohol use disorders: a comprehensive review. *Clin. Psychol. Rev.* 31, 800–816.
- Zayats, T., Jacobsen, K.K., Kleppe, R., Jacob, C.P., Kittel-Schneider, S., Ribases, M., Ramos-Quiroga, J.A., Richarte, V., Casas, M., Mota, N.R., Grevet, E.H., Klein, M., Corominas, J., Bralten, J., Galesloot, T., Vasquez, A.A., Herms, S., Forstner, A.J., Larsson, H., Breen, G., Asherson, P., Gross-Lesch, S., Lesch, K.P., Cichon, S., Gabrielsen, M.B., et al., 2016. Exome chip analyses in adult attention deficit hyperactivity disorder. *Transl. Psychiatry* 6, e923.
- Zhan, X., Hu, Y., Li, B., Abecasis, G.R., Liu, D.J., 2016. RVTESTS: An efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* 32, 1423–1426.
- Zhou, W., Jiang, Y., Zhu, M., Hang, D., Chen, J., Zhou, J., Dai, J., Ma, H., Hu, Z., Jin, G., Sha, J., Shen, H., 2017. Low-frequency nonsynonymous variants in FKBPL and ARPC1 B genes are associated with breast cancer risk in Chinese women. *Mol. Carcinog.* 56, 774–780.