The Influence of Regression Models on Genome-Wide Association Studies of Alcohol Dependence: A Comparison of Binary and Quantitative Analyses

Running head: Regression models in alcohol dependence GWAS

[1]Wenqianglong Li, [1]Johan Hilge Thygesen, [1]Niamh Louise O'Brien, [2]Mathis Heydtmann, [3]Iain Smith, [4,5]Franziska Degenhardt, [4]Markus Maria Nöthen, [6]Marsha Yvonne Morgan, [1]Nicholas James Bass, [1]Andrew McQuillin.

[1]Molecular Psychiatry Laboratory, Division of Psychiatry, University College London, UK

[2]Royal Alexandria Hospital, NHS Greater Glasgow and Clyde, Paisley, UK

[3]Glasgow Addiction Services, NHS Greater Glasgow and Clyde, Glasgow, UK

[4]Institute of Human Genetics, University of Bonn, School of Medicine & University Hospital Bonn, Bonn, Germany

[5]Department of Child and Adolescent Psychiatry, University of Duisburg-Essen, Essen, Germany

[6]UCL Institute for Liver & Digestive Health, Division of Medicine, Royal Free Campus, University College London, UK

Corresponding author:

Andrew McQuillin, Molecular Psychiatry Laboratory, Division of Psychiatry, University College London, London, UK. E-mail: a.mcquillin@ucl.ac.uk

Sources of support

The UCL clinical and control samples were collected with support from, the Neuroscience Research Charitable Trust, the Central London NHS (National Health Service) Blood Transfusion Service, the Camden and Islington NHS Foundation Trust, a research lectureship from the Priory Hospitals and the National Institutes for Health Research (NIHR) Mental Health Research Network (MHRN). Genotyping of the control samples was funded by the Stanley Center for Psychiatric Research at the Broad Institute. Professor McQuillin and Dr Bass are supported by the UCLH NIHR BRC. MMN and FD were supported through grants SysMedAlcoholism (01ZX1611B) and SysMedSUDs by the German Federal Ministry of Education and Research (BMBF) within the e: Med programme. Mr Wenqianglong Li is supported by China Scholarship Council (CSC) for his PhD studies.

Disclaimer

The authors have no conflict of interest to report.

Abstract

Introduction Genome-wide association studies (GWAS) of alcohol dependence syndrome (ADS) to date, have utilized a case-control design and with the exception of protective variants within the alcohol metabolizing genes, have failed to identify consistently replicable loci. The variability of the ADS phenotype means that the use of quantitative variables as a proxy for the severity of ADS has the potential to facilitate identification of risk loci by increasing statistical power. The current study aims to examine the influences of using binary and adjusted quantitative measures of ADS on GWAS outcomes and on calculated polygenic risk scores (PRS).

Methods A GWAS was performed in 1251 healthy controls with no history of excess alcohol use and 739 patients with ADS classified using binary DMS-IV criteria. Additional two GWASs were undertaken using a quantitative score based on DSM-IV criteria, which were applied assuming both normal and non-normal distributions of the phenotypic variable. PRS analyses were performed utilizing the data from the binary and the quantitative trait analyses.

Results No associations were identified at genomewide significance in any of the individual GWAS; results were comparable in all three. The top associated single nucleotide polymorphism was located on the alcohol dehydrogenase gene cluster on chromosome 4, consistent with previous ADS GWAS. The quantitative trait analysis adjusted for the distribution of the criterion score and the associated PRS had the smallest standard errors and thus the greatest precision.

Conclusion Further exploitation of the use of qualitative trait analysis in GWAS in ADS is warranted.

Key words: alcohol dependence; DSM-IV criterion count; genome-wide association study; polygenic risk score, quantitative trait; regression models.

Introduction

Alcohol dependence syndrome (ADS) is a common disorder characterized by the excessive and compulsive use of alcohol, often resulting in physical, emotional and social harm; it poses significant problems for both health and social services alike. ADS is a complex disorder which is influenced by both environmental and genetic factors (Stickel, Moreno, Hampe, & Morgan, 2017). Twin, family and adoption studies provide evidence for significant heritability of the alcohol dependence and alcohol misuse phenotypes. However, the magnitude of the effect is still debated. Verhulst, Neale, & Kendler (2015) undertook a meta-analysis of data from 12 twin and five adoption studies and provided an overall estimate of the heritability for alcohol use disorders of 49%. However, in an earlier meta-analysis of over 50 family, twin and adoption studies of alcohol misuse phenotypes Walters (2002) showed that there was significant heterogeneity across studies and provided a mean heritability estimate of 24%. Walters (2002) also showed that the heritability was much stronger in men with severe alcoholism/alcohol dependence with heritability estimates of the order of 30% to 36%.

While it is generally agreed that inheritance is polygenic (Goldman, Oroszi, & Ducci, 2005; Schork & Schork, 1998), identifying the genes involved and their relative contribution is difficult because of the considerable variations observed in the design of studies, the phenotypes of the included cohorts, the general failure to control for potential confounders such as co-morbid psychiatric conditions and co-occurring substance misuse, and because many studies have lacked statistical power (Ali et al., 2015; Hirschhorn, Lohmueller, Byrne, & Hirschhorn, 2002).

The majority of the GWAS in ADS, undertaken to date, have failed to identify consistently replicable novel loci (Walters *et al.*, 2018; Kranzler *et al.*, 2019; Liu *et al.*, 2019). However, meta-analyses and studies in populations with greater phenotypic surety have identified

genome-wide significant associations between genetic risk variants in the alcohol dehydrogenase (*ADH*) gene cluster on chromosome 4, which includes *ADH1B*, *ADH1C*, and aldehyde dehydrogenase 2 (*ALDH2*). These polymorphisms, which confer protection against problematic drinking, are significantly more prevalent in populations with East Asian ancestry (Bierut et al., 2012; Li, Zhao, & Gelernter, 2011, 2012). Other significant associations, appear to be specific to individual studies and have failed to replicate (Gelernter et al., 2014; Kranzler et al., 2019; Liu et al., 2019; R. K. Walters et al., 2018). This lack of consistency and failure to replicate may reflect genetic heterogeneity in the susceptibility to ADS and the fact that most of the published studies were likely underpowered to detect variants with small effect sizes, particularly given the stringent genome-wide significance threshold.

Most of the published ADS GWAS have utilised a case-control design (Hart & Kranzler, 2015; Stickel et al., 2017). However, within the ADS phenotype there is substantial uncertainty about which defining features might be inherited. Thus, for example, the *Diagnostic and Statistical Manual of Mental Disorders of the American Psychiatric Association (DSM) 4th edition* (American Psychiatric Association, 2000) lists seven defining symptoms: tolerance, withdrawal, desire/failed efforts to control drinking, drinking more than intended, reduced/impaired activities, primacy of drinking, and drinking despite problems. Individuals who report three or more of these features within one 12-month period are classified as having ADS. However, scores may range from 3 to 7 and it is clear that the contributing features may carry different weighting in relation to the clinical condition and its severity (Lane & Sher, 2015).

Thus, one approach to improving the identification of risk loci in ADS GWAS would be to create a quantitative variable based on the total number of DSM-IV criteria fulfilled. A score based on the DSM-IV, which is a proxy for the severity of ADS, has been used as a quantitative variable in a number of studies (Gelernter et al., 2014; Heath, Whitfield, Martin, Pergadia,

Goate, Lind, Mcevoy, et al., 2011; Kendler et al., 2011; Lai et al., 2019; Mcgue, Zhang, & Miller, 2013; Wang et al., 2013). However, this approach has not, to date, identified any novel associations at genome-wide significance or else novel findings that have been replicated. However, this may reflect the fact that distribution of the ADS criterion scores within GWAS populations may be skewed towards the more severe end of the spectrum, as they are likely treatment–seeking, and hence will demonstrate a quasi-Poisson distribution. No studies examining the impact of the assumptions made with regards to the distribution of quantitative symptom data and hence no studies which have examined the influence of regression models on the ADS GWAS data.

The aims of the present study were:

1. To compare the results of a classic case-control GWAS in ADS with those obtained using the DSM-IV criterion score as a quantitative trait.

2. To determine the proportion of the variance for ADS explained by ADS risk alleles utilising PRS analyses based on the case-control and quantitative trait analyses.

Methods

Participants

Individuals with ADS (n=741, 67 % men) were recruited from a variety of UK community and hospital-based services providing support and treatment for individuals with alcohol use disorders. The diagnosis of ADS was made by senior clinicians and trained researchers using the Alcohol Dependence Syndrome section of the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA-II) (Bucholz et al., 1994). This instrument incorporates the scoring items for the diagnosis of ADS delineated in the *Diagnostic and Statistical Manual of Mental Disorders 4th Edition* (DSM-IV) (American Psychiatric Association, 2000). All participants were of English, Scottish, Welsh or Irish descent with a maximum of one grandparent of non-British but Western European ancestry; none of these individuals was related.

Ancestrally matched, healthy controls (n = 797) were recruited from London branches of the National Health Service (NHS) blood transfusion service, from family doctor clinics and from among university students. Individuals were excluded if screening with the Schedule for Affective Disorders and Schizophrenia (SAD-L) (Endicott & Spitzer, 1978) revealed a lifetime history of neurosis, depression, bipolar disorder, schizophrenia or alcohol use disorders. DNA from a separate set of healthy controls of British ancestry (n = 454) was purchased from the European Collection of Cell Cultures (ECACC; Health Protection Agency Culture Collections, Salisbury, UK).

All procedures involving human subjects were approved by the NHS Metropolitan Multi-centre Research Ethics Committee (now the South Central - Hampshire A Research Ethics Committee) approval number MREC/03/11/090. All participants provided signed informed consent.

Genotyping, Imputation and Quality Control

Participants provided either blood or saliva samples for genomic DNA extraction. Genotyping of the DNA samples in the ADS cohort was performed at Life and Brain GmbH, Bonn, Germany, using the Illumina PsychArray. The DNA samples from the healthy controls were genotyped at the Broad Institute, MA, USA, using the same array. Quality control of the genotype data was performed in two stages, pre-imputation with more inclusive parameters aimed at retaining a maximal number of subjects, and post-imputation with more stringent parameters aimed at obtaining a high-quality data set. Pre-imputation quality control parameters were: individuals were excluded if they had incorrect gender assignment; excessive heterozygosity (in ADS samples and healthy controls: 3 standard deviation > the mean); more than 10% of missing genotype data and evidence of relatedness. Data on SNPs which had a minor allele frequency (MAF) < 5% and/or deviated substantially from the Hardy-Weinberg equilibrium (HWE) ($p < 10^{-6}$) were excluded.

Imputation was undertaken using the Haplotype Reference Consortium (release 1.1) (McCarthy et al., 2016) reference panel on the Sanger Imputation server (McCarthy et al., 2016). Genotypes were prepared as instructed and checks were performed using the HRC-1000G-check-bim tool Version 4.2.3 (Rayner, 2015) prior to the upload of data. A total of 393,270 SNPs with a MAF > 0.01 were uploaded for imputation(McCarthy et al., 2016). Pre-phasing was undertaken with EAGLE2 (Loh et al., 2016) and imputation was performed using the with Positional Burrows-Wheeler Transform (Durbin, 2014) method.

Post imputation quality control (QC) parameters used on the hard-called best-guess SNP genotypes. SNPs were included if they met the following criteria: INFO scores > 0.9, call rates > 99%, HWE p-value > 1×10^{-5}, and MAF > 5%. All QC steps were performed in PLINK2 (Chang et al., 2015). A total of 2.1 million SNPs were available for analysis following this process.

Statistical Analysis

Data Processing

Participants with a DSM-IV ADS score of ≥ 3 were classified as alcohol dependent. The total number of positive responses to the DSM-IV listed criteria (3 to 7) was used to define the phenotype for the quantitative trait analyses (see figure 1 and table 1). The study population was recruited from treatment centres across the UK so it was anticipated that the majority would have severe ADS according to criteria. In anticipation of a prominent right skew to the criterion scores both Gaussian and quasi-Poisson distributions of the data were assumed.

Association tests

A standardised genetic relationship matrix (GRM) was first estimated from the genotype data using the genome-wide efficient mixed-model association (GEMMA) package, which is considered to be an efficient method to estimate GRM (Zhou & Stephens, 2012). The GRM was then used in a generalized linear mixed model with the *glmmkin* function in the generalized linear mixed model association test (GMMAT) package to fit three models for the binary ADS and quantitative ADS criterion scores.

Model 1 (Binomial): used the binary ADS phenotype data and assumed a Bernoulli distribution.

Model 2 (Gaussian): used a fitted linear mixed model for the quantitative ADS criterion scores.

Model 3 (quasi-Poisson): used the same quantitative ADS criterion scores as Model 2 but to fit the likely right-skewed distribution of the criterion scores used a quasi-Poisson regression.

Individual GWASs were performed using generalized linear mixed model implemented in the GMMAT package with imputed best-guess genotypes and with sex, and the first 10 principal components as covariates (Chen et al., 2016).

Effect sizes and p-values for the top nine independent SNPs from the three individual GWASs, were generated using Wald tests. All analyses were performed using R version 3.5.1 (R Core Team, 2019).

Expression and splicing quantitative trait loci (eQTLs and sQTLs)

The GTEx V8 database was interrogated to identify correlations between genotype and gene expression levels (Carithers et al., 2015) for the top GWAS hits.

Gene-based association and enrichment analyses

The Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) package was used to explore gene prioritization, gene expression, and gene-based analyses (Watanabe et al., 2017).

Polygenic risk scores

Polygenic risk score (PRS) analyses were performed to determine the proportion of the variance for ADS that could be explained by the identified ADS risk alleles from the binary and the quantitative trait analyses. The software package PRSice2 was used to estimate the PRS at a range of nine p-value thresholds (Choi, Mak, & O'Reilly, 2020). For clumping, the linkage disequilibrium (LD) threshold was set to an R^2 of 0.1 and a distance of 250kb. Sex and the first 10 PCAs were used as the covariates. Summary statistics for ADS GWAS were taken from the largest GWAS for ADS (n = 202, 004) available to date (Kranzler et al., 2019).

Results

Data were available for 739 (67% men) participants who fulfilled DSM-IV criteria for alcohol dependence, 2 participants with an ADS score of less than 3, and 1251 (35% men) healthy controls. The ADS criterion scores ranged from 3 to 7 but with a prominent right-skew in the distribution (Figure 1 and Table 1).

GWAS

No associations were identified in the individual GWAS which were significant at genome-wide level (Table 2). The top associated SNP in the GWAS utilising a binary diagnosis for ADS (Model 1) was rs34361428, located in the *ADH* cluster on chromosome 4 ($p = 1.31 \times 10^{-6}$, beta = 0.47). This was also the top associated SNP identified in the GWAS based on the DSM-IV criterion scores although the levels of significance were lower in both (Model 2:$p = 2.1 \times 10^{-6}$, beta = 0.62; Model 3: $p = 1.22 \times 10^{-6}$, beta = 0.24). (For Manhattan and QQ plots see Supplementary Figures 1 - 3). Conditional analyses using directly genotyped data for the functional *ADH1B* SNP rs1229984 (Way et al., 2015) in a subset of the samples with available data, did not lead to substantial changes in the effect sizes for the SNPs reported in Table 2 that are located outside of the ADH gene cluster (t-test $p>0.05$, data not shown).

Expression and splicing quantitative trait loci (eQTLs and sQTLs)

Based on information provided by the GTEx V8 database, rs34361428, the top identified SNP associated with ADS is a significant eQTL for *ADH1B* ($p = 6.5 \times 10^{-17}$, normalised effect size (NES) = -0.27); carriage of the C allele is associated with increased expression of *ADH1B*. rs34361428 is also a significant eQTL for *ADH1C* ($p = 3.6 \times 10^{-12}$, NES = -0.22), *ADH1A* ($p = 1.4 \times 10^{-11}$, NES = -0.33), and *METAP1* (p = 0.0002, NES = -0.065); at both loci the C allele is associated with decreased expression. Additionally, single-tissue sQTLs data from GTEx V8 show that rs34361428 is a significant sQTL for *ADH1C* ($p = 8.6 \times 10^{-7}$, NES = -0.54); splicing is influenced

by the C allele but varies by tissue; carriage of the C allele is associated with decreased splicing in the liver.

Gene-based test, pathway, and enrichment analyses

MAGMA gene-based tests showed no significant gene prioritization and gene expression. Likewise, MAGMA gene-set analysis revealed no significant gene set related to ADS in the data after correcting for multiple testing (Supplementary table 1). MAGMA tissue expression analysis also showed no significant enrichment for any tissue types (Supplementary Figures 4 - 6).

Polygenic risk scores (PRS) analysis

The PRS generated from publicly available ADS GWASs were tested in the sample included in the three individual GWAS. The PRSs explained a proportion of the variance for ADS at all p-value thresholds. At a threshold of 0.5 the binary model explained 3.61% of the variance ($p = 0.023$) for ADS and the quantitative model upwards of 2.64% of the variance ($p = 4.53 \times 10^{-15}$) (Table 3). The PRS based on the binary model seemed to explain more of the variance in ADS than the quantitative models but the difference was not significant ($p = 0.22$).

Discussion

Classical GWAS approaches in ADS have failed to identify consistently replicable loci with the exception of protective variants within the alcohol metabolizing genes, notably *ADH1B*, and to a lesser degree, *ADH1C,* but these explain only a small proportion of the associated risk (Hart & Kranzler, 2015; Stickel et al., 2017). Thus, as with other complex traits, it is likely that a large number of causal risk variants contribute to the development of ADS but that individually they have comparatively small effects sizes (Manolio et al., 2009). This coupled with the stringent correction for the large number of SNPs tested in GWAS, means that very large samples will be necessary to reliably detect the associated loci.

An additional challenge to the identification of novel loci contributing to the risk for developing ADS may be the underlying heterogeneity of the condition. The majority of the GWAS undertaken to date have used a binary case-control definition for the ADS phenotype. Another strategy which might improve the detection of variants would be use the more granular approach of defining quantitative traits. Use of quantitative traits could potentially enhance statistical power, produce unbiased PRSs, and identify quantitative mechanisms. The disconnect between a disorder's dichotic classification and the quantitatively distributed polygenic liabilities was examined by Plomin et al., (2009) who commented that the genetic foundation of common disorders is constituted by 'the quantitative extremes of continuous distributions of genetic risk'.

One approach would be to create a quantitative variable based on the total number of DSM-IV criteria fulfilled in order to reduce clinical and genetic heterogeneity.

Although multiple combinations of criteria and study characteristics may result in a similar criterion count, this proxy for ADS severity has been successfully employed in previous studies although did not result in the identification of any additional findings at genome wide significance (Heath et al., 2011; Kendler et al., 2011; Mcgue et al., 2013; Wang et al., 2013).

Gelernter et al., (2014) included DSM-IV ADS criterion counts in their GWAS of ADS to increase statistical power and to correct for co-occurring dependences. They reported association with the *ADH* cluster (*ADH1B* and *ADH1C*), and other novel loci that included *MTIF2*, *CCDC88A*, *PDLIM5*, and *LOC100507053*. However, the results were not replicated in two much larger meta-analyses (Kranzler et al., 2019; R. K. Walters et al., 2018). Lai et al., (2019) have recently undertaken a GWAS in ADS based on DSM-IV criteria counts in European and African Americans and reported an association with a SNP in *ADH1B* at genome wide significance. In addition, they reported four novel loci associated with individual DSM-IV criteria. However, apart from the *ADH1B* locus, only one SNP on chromosome 8 replicated in an independent dataset and meta-analysis ($P = 3.71 \times 10^{-9}$) and the only gene near this region, *FAM84B,* does not appear to be related to any neuropsychiatric disorder. Nevertheless, while significant increases in sample size can potentially overcome the heterogeneity in ADS, the study of quantitative traits could provide a more detailed picture of how genetic risk variants influence the disorder. Such quantitative approaches have also been applied to the analysis of Alcohol Use Disorders Identification Test (AUDIT) data on alcohol use and these have produced robust findings. Interestingly the findings from the analysis of AUDIT data and data on alcohol dependence syndrome have provided evidence for

some important differences in the liability for the two traits (Sanchez-Roige et al., 2019).

One consideration that needs to be taken into account when adopting a quantitative approach in GWAS is that the phenotypic data used in the analysis may not be normally distributed. Thus, the GWAS employing the DSM-IV criterion scores, as a surrogate for disease severity, was undertaken using both a linear regression model, based on the assumption that the sores would be normally distribute, and a quasi-Poisson regression model based on an assumption that the data would show a prominent right skew, which indeed it did.

All three GWAS produced similar results although those obtained with the quasi-Poisson regression model had the smallest standard errors. No associations were identified which were significant at genome-wide level. The top associated SNP, which was consistent across all three GWAS, was rs34361428, which is located in the *ADH* cluster on chromosome 4. This SNP is an eQTL for *ADH1B*, *ADH1C*, *ADH1A*, and *METAP1* and its association with ADS has been reported previously (Goldman et al., 2005; Li et al., 2011; Luczak, Glatt, & Wall, 2006; R. K. Walters et al., 2018; Zuo et al., 2012). *METAP1* (methionyl aminopeptidase 1) is a protein coding gene and its biological pathways are related to metabolism of fat-soluble vitamins. The C allele of rs34361428 is associated with reduced expression of METAP1. This association has been reported previously but specifically in African Americans classified as having alcohol use disorder (Gelernter et al., 2014; Kranzler et al., 2019).

There were no significant differences in the variance estimates provided by the binary and quantitative models despite the fact that the quantitative PRS offered more precise estimates.

This study had two clear limitations: first the relatively small numbers and second the highly selected nature of both the cases and controls. Thus, the cases were treatment-seeking chronic alcohol misusers, the majority of whom had the most severe ADS phenotype. In contrast, the majority of the controls were healthy blood donors or screened controls with no past or current history of excess alcohol use. Inclusion of a cohort of non-dependent heavy drinkers and a general population sample would have provided a better representation of the alcohol consumption spectrum. The strengths of this study are its novel approach to the analysis of the GWAS data which took into account the distribution of the phenotypic trait of interest and the fact that it can be used as a template for future studies.

References

Ali, M. A., Way, M. J., Marks, M., Guerrini, I., Thomson, A. D., Strang, J., … Morgan, M. Y. (2015). Phenotypic heterogeneity in study populations may significantly confound the results of genetic association studies on alcohol dependence. *Psychiatric Genetics*. https://doi.org/10.1097/YPG.0000000000000105

American Psychiatric Association. (2000). Diagnostic and Statistical Manual of Mental Disorders, 4th Ed. DSM-IV-TR. In *American Journal of Critical Care*. https://doi.org/10.1176/dsm 10.1176/appi.books.9780890420249.dsm-iv-tr

Bierut, L. J., Goate, A. M., Breslau, N., Johnson, E. O., Bertelsen, S., Fox, L., … Edenberg, H. J. (2012). ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. *Molecular Psychiatry*. https://doi.org/10.1038/mp.2011.124

Bucholz, K. K., Cadoret, R., Cloninger, C. R., Dinwiddie, S. H., Hesselbrock, V. M., Nurnberger, J. I., … Schuckit, M. A. (1994). A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *Journal of Studies on Alcohol*, *55*(2), 149–158. https://doi.org/10.15288/jsa.1994.55.149

Carithers, L. J., Ardlie, K., Barcus, M., Branton, P. A., Britton, A., Buia, S. A., … Williams, P. (2015). A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreservation and Biobanking*. https://doi.org/10.1089/bio.2015.0032

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*. https://doi.org/10.1186/s13742-015-0047-8

Chen, H., Wang, C., Conomos, M. P., Stilp, A. M., Li, Z., Sofer, T., … Lin, X. (2016). Control for

Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. *American Journal of Human Genetics*. https://doi.org/10.1016/j.ajhg.2016.02.012

Choi, S. W., Mak, T. S. H., & O'Reilly, P. F. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*. https://doi.org/10.1038/s41596-020-0353-1

Durbin, R. (2014). Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btu014

Endicott, J., & Spitzer, R. L. (1978). A Diagnostic Interview: The Schedule for Affective Disorders and Schizophrenia. *Archives of General Psychiatry*. https://doi.org/10.1001/archpsyc.1978.01770310043002

Gelernter, J., Kranzler, H. R., Sherva, R., Almasy, L., Koesterer, R., Smith, A. H., … Farrer, L. A. (2014). Genome-wide association study of alcohol dependence:significant findings in African- and European-Americans including novel risk loci. *Molecular Psychiatry*, *19*(1), 41–49. https://doi.org/10.1038/mp.2013.145

Goldman, D., Oroszi, G., & Ducci, F. (2005). The genetics of addictions: Uncovering the genes. *Nature Reviews Genetics*, *6*(7), 521–532. https://doi.org/10.1038/nrg1635

Hart, A. B., & Kranzler, H. R. (2015). Alcohol Dependence Genetics: Lessons Learned From Genome-Wide Association Studies (GWAS) and Post-GWAS Analyses. *Alcoholism: Clinical and Experimental Research*, *39*(8), 1312–1327. https://doi.org/10.1111/acer.12792

Heath, A. C., Whitfield, J. B., Martin, N. G., Pergadia, M. L., Goate, A. M., Lind, P. A., … Montgomery, G. W. (2011). A Quantitative-Trait Genome-Wide Association Study of

Alcoholism Risk in the Community : Findings and Implications. *BPS*, *70*(6), 513–518. https://doi.org/10.1016/j.biopsych.2011.02.028

Heath, A. C., Whitfield, J. B., Martin, N. G., Pergadia, M. L., Goate, A. M., Lind, P. A., … Montgomery, G. W. (2011). A quantitative-trait genome-wide association study of alcoholism risk in the community: Findings and implications. *Biological Psychiatry*. https://doi.org/10.1016/j.biopsych.2011.02.028

Hirschhorn, J. N., Lohmueller, K., Byrne, E., & Hirschhorn, K. (2002). A comprehensive review of genetic association studies. *Genetics in Medicine*. https://doi.org/10.1097/00125817-200203000-00002

Kendler, K. S., Kalsi, G., Holmans, P. A., Sanders, A. R., Aggen, S. H., Dick, D. M., … Gejman, P. V. (2011). *Genomewide Association Analysis of Symptoms of Alcohol Dependence in the Molecular Genetics of Schizophrenia (MGS2) Control Sample*. *35*(5), 963–975. https://doi.org/10.1111/j.1530-0277.2010.01427.x

Kranzler, H. R., Zhou, H., Kember, R. L., Smith, R. V., Justice, A. C., Damrauer, S., … Gelernter, J. (2019). *Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations*. 1–11. https://doi.org/10.1038/s41467-019-09480-8

Lai, D., Wetherill, L., Bertelsen, S., Carey, C. E., Kamarajan, C., Kapoor, M., … Foroud, T. (2019). Genome-wide association studies of alcohol dependence, DSM-IV criterion count and individual criteria. *Genes, Brain and Behavior*, *18*(6), 1–14. https://doi.org/10.1111/gbb.12579

Lane, S. P., & Sher, K. J. (2015). Limits of current approaches to diagnosis severity based on criterion counts: An example with DSM-5 alcohol use disorder. *Clinical Psychological Science*. https://doi.org/10.1177/2167702614553026

Li, D., Zhao, H., & Gelernter, J. (2011). Strong association of the alcohol dehydrogenase 1B gene (ADH1B) with alcohol dependence and alcohol-induced medical diseases. *Biological Psychiatry*. https://doi.org/10.1016/j.biopsych.2011.02.024

Li, D., Zhao, H., & Gelernter, J. (2012). Strong protective effect of the aldehyde dehydrogenase gene (ALDH2) 504lys (*2) allele against alcoholism and alcohol-induced medical diseases in Asians. *Human Genetics*. https://doi.org/10.1007/s00439-011-1116-4

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., … Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*. https://doi.org/10.1038/s41588-018-0307-5

Loh, P. R., Danecek, P., Palamara, P. F., Fuchsberger, C., Reshef, Y. A., Finucane, H. K., … Price, A. L. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics*. https://doi.org/10.1038/ng.3679

Luczak, S. E., Glatt, S. J., & Wall, T. J. (2006). Meta-analyses of ALDH2 and ADH1B with alcohol dependence in asians. *Psychological Bulletin*. https://doi.org/10.1037/0033-2909.132.4.607

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., … Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*. https://doi.org/10.1038/nature08494

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., … Marchini, J. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*. https://doi.org/10.1038/ng.3643

Mcgue, M., Zhang, Y., & Miller, M. B. (2013). *ORIGINAL RESEARCH A Genome-Wide*

Association Study of Behavioral Disinhibition. 363–373. https://doi.org/10.1007/s10519-013-9606-x

Plomin, R., Haworth, C. M. A., & Davis, O. S. P. (2009). Common disorders are quantitative traits. *Nature Reviews Genetics*, *10*(12), 872–878. https://doi.org/10.1038/nrg2670

R Core Team 2019. (2019). : A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria. URL Http://Www.R-Project.Org/.*

Rayner, W. (2015). HRC-1000G-check-bim.

Sanchez-Roige, S., Palmer, A. A., Fontanillas, P., Elson, S. L., Adams, M. J., Howard, D. M., … Wilson, C. H. (2019). Genome-wide association study meta-analysis of the alcohol use disorders identification test (AUDIT) in two population-based cohorts. *American Journal of Psychiatry*. https://doi.org/10.1176/appi.ajp.2018.18040369

Schork, N. J., & Schork, C. M. (1998). Issues and strategies in the genetic analysis of alcoholism and related addictive behaviors. *Alcohol*. https://doi.org/10.1016/S0741-8329(97)00179-1

Stickel, F., Moreno, C., Hampe, J., & Morgan, M. Y. (2017). The genetics of alcohol dependence and alcohol-related liver disease. *Journal of Hepatology*. https://doi.org/10.1016/j.jhep.2016.08.011

Verhulst, B., Neale, M. C., & Kendler, K. S. (2015). The heritability of alcohol use disorders: A meta-analysis of twin and adoption studies. *Psychological Medicine*, *45*(5), 1061–1072. https://doi.org/10.1017/S0033291714002165

Walters, G. D. (2002). The heritability of alcohol abuse and dependence: A meta-analysis of behavior genetic research. *American Journal of Drug and Alcohol Abuse*. https://doi.org/10.1081/ADA-120006742

Walters, R. K., Polimanti, R., Johnson, E. C., McClintick, J. N., Adams, M. J., Adkins, A. E., … Agrawal, A. (2018). Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nature Neuroscience*. https://doi.org/10.1038/s41593-018-0275-1

Wang, J., Foroud, T., Hinrichs, A. L., Le, N. X. H., Bertelsen, S., Budde, J. P., … Almasy, L. (2013). *ORIGINAL ARTICLE A genome-wide association study of alcohol-dependence symptom counts in extended pedigrees identifies C15orf53*. (October 2012), 1218–1224. https://doi.org/10.1038/mp.2012.143

Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*. https://doi.org/10.1038/s41467-017-01261-5

Way, M., McQuillin, A., Saini, J., Ruparelia, K., Lydall, G. J., Guerrini, I., … Gurling, H. M. D. (2015). Genetic variants in or near adh1b and adh1c affect susceptibility to alcohol dependence in a british and irish population. *Addiction Biology*. https://doi.org/10.1111/adb.12141

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*. https://doi.org/10.1038/ng.2310

Zuo, L., Gelernter, J., Zhang, C. K., Zhao, H., Lu, L., Kranzler, H. R., … Luo, X. (2012). *Genome-Wide Association Study of Alcohol Dependence Implicates KIAA0040 on Chromosome 1q*. 557–566. https://doi.org/10.1038/npp.2011.229