

Published in final edited form as:

*Biol Psychiatry*. 2011 September 15; 70(6): 513–518. doi:10.1016/j.biopsych.2011.02.028.

## A quantitative-trait genome-wide association study of alcoholism risk in the community: findings and implications

Andrew C. Heath<sup>1,4,3</sup>, John B. Whitfield<sup>2,3</sup>, Nicholas G. Martin<sup>2,3</sup>, Michele L. Pergadia<sup>1</sup>, Alison M. Goate<sup>1</sup>, Penelope A. Lind<sup>2</sup>, Brian P. McEvoy<sup>2</sup>, Andrew J. Schrage<sup>1</sup>, Julia D. Grant<sup>1</sup>, Yi-Ling Chou<sup>1</sup>, Rachel Zhu<sup>1</sup>, Anjali K. Henders<sup>2</sup>, Sarah E. Medland<sup>2</sup>, Scott D. Gordon<sup>2</sup>, Elliot C. Nelson<sup>1</sup>, Arpana Agrawal<sup>1</sup>, Dale R. Nyholt<sup>2</sup>, Kathleen K. Bucholz<sup>1</sup>, Pamela A.F. Madden<sup>1,5</sup>, and Grant W. Montgomery<sup>2,5</sup>

<sup>1</sup>Department of Psychiatry, Washington University School of Medicine

<sup>2</sup>Queensland Institute of Medical Research, Brisbane, Australia

### Abstract

**Background**—Given moderately strong genetic contributions to variation in alcoholism and heaviness of drinking (50–60% heritability), with high correlation of genetic influences, we have conducted a quantitative trait genomewide association study for phenotypes related to alcohol use and dependence.

**Methods**—Diagnostic interview and blood/buccal samples were obtained from sibships ascertained through the Australian Twin Registry. Genomewide SNP genotyping was performed with 8754 individuals [2062 alcohol dependent cases] selected for informativeness for alcohol use disorder and associated quantitative traits. Family-based association tests were performed for alcohol dependence, dependence factor score and heaviness of drinking factor score, with confirmatory case–population control comparisons using an unassessed population control series of 3393 Australians with genomewide SNP data.

**Results**—No findings reached genomewide significance ( $p=8.4\times 10^{-8}$  for this study), with lowest  $p$ -value for primary phenotypes of  $1.2\times 10^{-7}$ . Convergent findings for quantitative consumption and diagnostic and quantitative dependence measures suggest possible roles for a transmembrane protein gene (*TMEM108*) and for *ANKK1*. The major finding, however, was small effect sizes estimated for individual SNPs, suggesting that hundreds of genetic variants make modest contributions (1/4% of variance or less) to alcohol dependence risk.

**Conclusions**—We conclude that (i) meta-analyses of consumption data may contribute usefully to gene-discovery; (ii) translation of human alcoholism GWAS results to drug discovery or clinically useful prediction of risk will be challenging; (iii) through accumulation across studies, GWAS data may become valuable for improved genetic risk differentiation in research in biological psychiatry (e.g. prospective high-risk or resilience studies).

© 2011 Society of Biological Psychiatry. Published by Elsevier Inc. All rights reserved.

<sup>4</sup>Correspondence: Andrew C. Heath, DPhil, Washington University School of Medicine, Department of Psychiatry, 660 S. Euclid, Campus Box 8134, St. Louis, MO 63110, aheath@wustl.edu.

<sup>3</sup>Contributed equally to this work

<sup>5</sup>Contributed equally as senior authors to this work

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

All authors report no biomedical financial interest or potential conflicts of interest.

## Keywords

Alcoholism; genome-wide association; quantitative-trait; non-replication

---

## INTRODUCTION

Evidence for important genetic contributions to the intergenerational transmission of alcohol use disorder has accumulated over many years. This began with the observation of pedigrees with multiple alcoholic family members (1), continuing with adoption study data suggesting that adopted-away offspring risk correlates with biological parent alcoholism (2,3) and twin studies suggesting that, across a range of phenotype definitions ranging from severe to broad, genetic factors may explain as much as 60% of the variance in risk of alcoholism (4,5). Genetic linkage studies (6,7,8) and, most recently, the first generation of GWAS studies of alcoholic case-control series(9,10,11), have sought to identify genes contributing to risk of alcohol use disorder. In parallel, a second literature has developed, mostly based on twin studies, showing comparably high genetic variance in alcohol consumption patterns (12) and noting genetic overlap between alcohol dependence and heaviness of consumption (13). This overlap persists even when consumption data from alcohol dependent individuals are omitted from the analysis, to avoid an artefactual correlation due to escalating consumption secondary to the onset of alcoholism (14). Studies of alcohol metabolism genes (*ALDH2*, *ADH* gene family), initially in Asian samples, and then in those of Jewish, European or African ancestry, have shown how genetic variants can make important contributions to differences in risk of alcohol dependence, consumption, or other aspects of response to alcohol (15,16,17,18,19). Finally, a third approach using various forms of latent structure modeling has emphasized the quasi-continuous nature of alcoholism (20,21,22), drawing attention to the undesirability, for some research purposes, of dichotomizing into affected and unaffected.

Published first-generation case-control genomewide association studies of alcoholism have yielded generally disappointing findings. They have not achieved the robustly replicated associations seen using quantitative smoking phenotypes, for example, between the *CHRNA3/CHRNA5/CHRNA4* gene complex and heaviness of smoking (23,24,25). Here we report results of a GWA study using a different research strategy, that seeks to take advantage of the quantitative information that can be obtained for heaviness of alcohol consumption (as operationalized in (14)) and for alcoholism symptom severity. This is the first study to apply a genome-wide association approach to a quantitative measure of alcoholism risk in the general community, and potentially complements the clinical case-control studies.

## METHODS

### Samples

Samples were ascertained from a pool of approximately 11,700 Australian families identified through diagnostic interview surveys of two cohorts of like-sex and unlike-sex twin pairs from a volunteer Australian twin panel (cohort 1, born 1895–1964 [N=5995 interviewed twins (4)], but for the purposes of this study mostly born 1940–1964; and cohort 2, born 1964–1971 [N=6257 twins;(26)], as well as through an interview survey of the spouses/partners of the former cohort (N=3846,(27)). Index cases from these families, their full siblings and parents were recruited for three coordinated studies: (i) the NAG (Nicotine Addiction Genetics) Study (28) which ascertained heavy smoking index cases; (ii) the OZALC-EDAC study, which ascertained index cases with a history of alcohol dependence or scoring above the 85<sup>th</sup> centile for heaviness of drinking factor score (operationalized as

in(14)); (iii) the OZALC-BIGSIB study, which ascertained large sibships (4–14 full siblings), regardless of sibling phenotypic values. Additional cases and controls were recruited from Cohort 1 participants, and additional Cohort 2 participants, who did not complete the new interview protocol but had comparable alcohol use/dependence assessments. Finally an unassessed population control series is included in some analyses, comprising twins and their families with GWAS data who had participated in an adolescent twin study (29). Further details of the sampling design are given in the Supplemental Methods in Supplement 1. All projects underwent IRB review at the participating institutions. GWAS genotyping, using the Illumina 370K array, was performed on a total of 6852 individuals selected from the BIGSIB and EDAC series (including N=336 parents); from a subsample of the NAG families that had previously been selected for 10cM microsatellite scans (28), and a smaller number of additional alcohol dependent cases and controls from Cohorts 1 and 2. Additional GWAS data are included here for sample members who had been genotyped for other projects (29). Tables S1 and S2 in the Supplement summarize numbers of participants and distribution of sibship size for individuals with GWAS genotyping.

### Assessment

The diagnostic interview assessment was modified from the SSAGA (30,31) for telephone administration, with deletion of certain diagnostic sections and elimination of non-DSM items. Assessment of history of alcohol abuse and dependence was supplemented with detailed questions about alcohol consumption (frequency of use, frequency of heavy drinking [using 5 or more drinks in a day], frequency of drinking to intoxication, drinks per typical drinking day). These were coded categorically, with 10 response categories and a wide range of values (e.g. 1–2 to 31 or more drinks in a typical drinking day) used to encourage more accurate reports; and were asked of both heaviest drinking period of at least 12 months duration, and of past 12 months (if not included as the heaviest period). Two additional open-ended items coded maximum drinks (“Max Drinks”) in a 24-hour period, lifetime and in the past 12 months. All questions used standard Australian drinks [=10g of alcohol]. Diagnostic sections on smoking, anxiety, depression and conduct disorder were included based on their relevance for understanding alcoholism genetics. Given the potential for inaccurate reporting by interviewed parents (mean age 66, range 49–91), parent reports were limited to smoking history. In some cases interview data were only available from previous diagnostic assessments, with slight variations in assessment protocol, so that not all consumption measures were available for all individuals in the GWAS sample (Table S3 in the Supplement).

### Genotyping and Quality Control

Genotyping and the standard quality control filters that were applied are described in greater detail in Medland et al. (29) (see especially Table 1 in that publication). All genotyping was conducted on Illumina platforms, with genotypes called using Illumina BeadStudio software. Quality-control excluded SNPs with mean GenCall score less than 70%, with call rate less than 95%, with deviation from Hardy-Weinberg significant at  $p < 10^{-6}$ , or Minor Allele Frequency less than 1%. For the present study, Illumina CNV370-Quadv3 GWAS data were available on 4241 individuals (including most alcohol dependent cases) genotyped at CIDR and an additional 2611 individuals genotyped by deCODE for the OZALC project; Illumina 317K data were available for 53 individuals genotyped at the University of Helsinki Genome Center; and Illumina 610 Quad data were available for the remaining individuals genotyped by deCODE. Duplicate samples allowed comparison of genotyping across platforms/locations: a single SNP was identified, genotyped using the CNV370-Quadv3, that was called very differently at CIDR versus deCODE, and therefore deleted

from the data-set. Checks were run on genetic relatedness, with misspecified relationships corrected prior to analysis.

Cohorts 1 and 2 are almost entirely of European ancestry, reflecting restrictive Australian immigration policies through 1972; however, Eigenstrat analyses (32), which included data from other Australian GWAS series, identified as outliers (operationalized by  $\pm 6$  standard deviations) a small number of families of mixed European and Asian ancestry (principally Chinese, Burmese, Indian or Malaysian), of middle eastern (Lebanese) ancestry; with one or more grandparents of Aboriginal, Torres Strait Islander or Maori ancestry; or with some African heritage (including individuals of self-report Maltese ancestry, consistent with the known population genetics of the Maltese population) (see (29)). A total of 153 individuals from 60 families were thus identified, and excluded from further analyses: this included 34 alcohol dependent and 119 unaffected individuals. In the analyses presented here, we use only data from the approximately 300,000 SNPs that passed quality-control and were available on all platforms. We do not attempt imputation, given the potential for statistical problems in the analysis of imputed data (e.g. biased estimates of effect size) in family-structured data-sets.

## Analyses

Consistent with our previous work (14), a Heaviness of Drinking [HOD] factor score was derived, using four items: lifetime Max Drinks, and 3 heaviest period measures of frequency of heavy drinking, frequency of drinking to intoxication, and log-transformed average weekly consumption (derived as the product of frequency of use and drinks per drinking day measures). MaxDrinks was log-transformed and, for males only, winsorized in the left tail at 3 standard deviations below the mean, based on mean and standard deviation for the BIGSIB series. For the alcohol use disorder factor score, we included both DSM-IV dependence and DSM-IV abuse items (excluding recurrent legal problems) (AUD-FS), consistent with the perfect genetic correlation between Abuse and Dependence previously reported (14), and the anticipated operationalization of alcohol dependence in DSM-V. For comparison, an alcohol dependence factor (AD-FS) limited to dependence items was also estimated. Factor analyses were conducted separately for the OZALC-NAG and Cohorts 1 and 2 datasets; for the former, factor scoring coefficients were generated using the BIGSIB sample, by gender, and then applied to the combined OZALC-NAG data-set. Factor loadings were in the range 0.69–0.86 for women, 0.63 to 0.92 for men, for the HOD factor, 0.41 to 0.64 and 0.43 to 0.66 for the AUD factor. Quantitative measures were adjusted for gender, age and its quadratic and cubic terms, gender, and age\*gender interaction using linear regression. Adjusted scores were then rank-normalized, separately for the BIGSIB, OZALC-NAG EDAC and NAG, Cohort 1 only and Cohort 2 only subsamples and combined in a single genetic association analysis. Analysis by subsample, with results combined in a meta-analysis, yielded similar results, so only the former results are reported.

Preliminary linkage analyses were conducted with MERLIN-REGRESS(33), using a panel of SNPs selected for high minor allele frequency and low linkage disequilibrium ( $r^2 < 0.02$ ). Family-based association analyses of quantitative phenotypes were conducted using the FASTASSOC option in MERLIN (34); family-based analyses of the categorical DSM-IV alcohol dependence diagnosis were conducted using MQLS (35) as implemented in GDT (36). Results for our primary phenotypes were compared to those from the COGA (10), SAGE (9) and German (11) alcoholism GWAS studies. Finally, case-population control analyses were implemented using Huber-White adjustment for familial clustering in STATA (37). We use  $\alpha = 1.67E-7$  (Bonferroni correction for 300,000 SNPs) as the threshold for genome-wide significance for a single phenotype, thus  $\alpha = 8.35E-8$  for our primary HOD and AUD-FS outcome measures, allowing for testing of 2 phenotypes. We also report some key findings for other consumption phenotypes that would be more readily available in other

data-sets. We looked for consistency of evidence of genetic effects across phenotypes – between HOD and AUD-FS; for each with dependence diagnosis, and with other consumption measures. Finally, we used case-population control comparisons as a further check for consistency of evidence. To provide context for our findings, and guidelines for replication, we conducted power calculations for power to detect genetic association, under an additive genetic model, at  $\alpha=5E-8$ , for a SNP in complete linkage disequilibrium, or with specified  $D'$ , with a variant, for a range of assumed effect sizes, using the Genetic Power Calculator (38).

## RESULTS

### Sample characteristics

The component subsamples shared several characteristics typical of a general community sample (Table S4 in the Supplement): (i) most alcohol dependent cases were mild, with 70% of those meeting alcohol dependence criteria reporting only 3 or 4 dependence symptoms, and fewer than 5% reporting 7 dependence symptoms (not shown); (ii) a moderately high percentage of these affected individuals denied weekly drinking to intoxication, and a minority denied even weekly drinking of 5 or more standard drinks in a single day, during their 12-month period of heaviest drinking, implying that a not insignificant number experienced an episode of less than 12 months duration. Alcohol consumption histories, stratified by alcohol dependence history and gender, were comparable across subsamples, supporting the decision to combine the subsamples in a single analysis.

**Alcohol factor measures**—Neither of our primary measures gave genomewide significant evidence for linkage (Table 1). Table 2 summarizes genetic association results, for primary HOD and AUD-FS measures and for AD-FS and individual consumption measures, tabulating lowest observed p-values, and effect sizes (genetic variance explained by each SNP under an additive model, for SNPs with nominal associations at  $p<.0001$  or less). No SNPs reached genomewide significance, with lowest p-values for our primary measures being  $1.2E-7$  for HOD, and  $7.2E-7$  for AUD. Effect sizes were consistently small, half a percent or less. Out of the top 400 SNPs, ranked by p-value for association with HOD, 65% had effect sizes of less than 0.25% of the variance (but greater or equal to 0.20%), and only 7.5% had effect sizes greater than 0.3%, with a median effect size of 0.23% (not shown); while for AUD-FS the median effect size among the top 400 SNPs was 0.18% (range 0.16–0.35%), with 94% of these top SNPs having effect sizes less than 0.25%.

SNPs showing strongest association with HOD and AUD-FS measures are summarized in Tables S5 and S6 in the Supplement. Additional Tables S7 and S8 in the Supplement show AD-FS and 12-month weekly consumption results. The small number of SNPs showing convergent evidence across phenotypes (at nominal  $p<.0001$ , for the primary phenotype and  $p<.005$  for the confirmatory phenotype) are shown in Table S9 in the Supplement. A chromosome 3 SNP, rs2369955, that was the 3<sup>rd</sup> most highly associated SNP for HOD ( $p=1.6E-6$ ) also showed associations with AUD-FS ( $p=1.3E-4$ ), AD-FS ( $p=7.3E-5$ ) heaviest period Frequency of Heavy Drinking ( $p=4.7E-6$ ), Frequency Drunk ( $p=3.9E-5$ ), Frequency of Any Use ( $p=4.9E-5$ ) and with Weekly Alcohol Consumption ( $p=2.7E-5$ ) measures; and was modestly associated with DSM-IV alcohol dependence diagnosis ( $p=1.5E-3$ ). This SNP is intergenic but in moderate linkage disequilibrium with an intronic SNP in *TMEM108* (rs10935045, bp 134282836:  $r^2=0.49$ ,  $D'=0.73$ ) that is the 4<sup>th</sup> most highly associated SNP for HOD factor score ( $p=1.7E-6$ ) and also weakly associated with AUD-FS ( $p=6.4E-4$ ). Both SNPs were only weakly associated with AD diagnosis in case-population control (CPC) comparisons (OR=1.18, 95% CI 1.02–1.37,  $p=.02$ ; OR=1.17, 95% CI 1.02–1.31,  $p=0.03$ ).



Of the remaining SNPs nominally associated with HOD at  $p < .0001$ , two intronic SNPs in the *ANKK1* gene were associated with HOD, AUD-FS, AD-FS and with AD diagnosis in CPC comparisons (rs1737727:  $p = 5.5E-5$ ,  $7.4E-6$ ,  $6.6E-6$  and  $3.4E-4$  respectively; rs2140418:  $p = 4.4E-5$ ,  $4.3E-6$ ,  $4.1E-6$  and  $1.6E-4$ ) (see Table S9 in the Supplement); with the former also associated with heaviest period weekly consumption ( $p = 6.3E-4$ ), frequency of heavy drinking ( $p = 3.2E-4$ ) and MaxDrinks ( $p = 4.6E-3$ ). Neither were associated with AD diagnosis in family-based analyses. Additional *ANKK1* SNPs show association with AUD-FS at  $p < 0.0001$  as well as with HOD at  $p < .005$  (rs847851, rs847848, rs2273006 – the latter a *TAF11* SNP but in complete LD with an *ANKK1* SNP). An intronic SNP in *CNGB3* (rs4961216) was associated with HOD ( $p = 2.8E-5$ ), AUD-FS ( $p = 2.4E-4$ ), binary AD in family-based analyses ( $p = 3.0E-4$ ) and AD in CPC comparisons ( $p = 1.3E-4$ ). Additional SNPs in *USH2A*, *ITIH5*, *SHANK2* and *C15orf32* showed nominal association with HOD, modest association with AUD-FS, but without confirmatory association with AD diagnosis (see Table S9 in the Supplement), as did a number of non-genic SNPs. Overall, 22 of 64 SNPs nominally associated with HOD also showed association with AUD-FS at  $p < .005$ , compared with only 4 of 51 SNPs associated with 12-month weekly consumption that were also associated with AUD-FS (not shown).

**Alcohol dependence diagnosis**—Of the SNPs nominally associated with alcohol dependence diagnosis at  $p < .0001$  in family based analyses (Table S10 in the Supplement), four also showed association in case-population control comparisons (rs1541918, Chr5,  $p = 7.4E-6$ ; CPC OR=1.23 (95% CI 1.06–1.42),  $p = .007$ ; rs10089021, Chr8,  $p = 4.3E-5$ , CPC OR=0.82 (0.77–0.94),  $p = .003$ ; rs786870, Chr10,  $p = 8.4E-5$ , CPC OR =0.81 (0.71–0.02),  $p < .001$ ; rs2789686, Chr10,  $p = 3.8E-5$ , CPC OR=0.76 (0.64–0.90),  $p < .001$ ). Only one of the 29 SNPs associated with alcohol dependence at  $p < .0001$  also shows association with HOD, the intergenic chromosome 12 SNP rs2463107 SNP that is in moderate LD with intronic SNPs in the *PAWR* and *SYT1* genes. This SNP also shows predicted case-population control differences (OR=0.62, 95%CI 0.49–0.79,  $p = 7E-5$ ).

**Replication failures**—For associations identified by the COGA project, after correction for multiple testing, we found no SNPs or tagging SNPs that confirmed associations reported for alcohol dependence diagnosis. For BBX, we found 9 SNPs that show association with one or more alcohol consumption measures, particularly rs1403774 (heaviest period drinks per drinking day,  $p = 9.5E-5$ ; current drinks per drinking day,  $p = 1.2E-4$ ); rs2049339 (HOD,  $p = 4.7E-3$ ; heaviest period drink per week,  $p = 3.6E-4$ ); and rs9875732 (heaviest period drinks per week,  $p = 6.6E-4$ ; drinks per drinking day,  $p = 4.6E-4$ ); with only this last SNP associated with AUD-FS ( $p = 7.9E-4$ ) and with AD-FS ( $p = 4.5E-3$ ). One of these, rs2049339, was in weak LD with the COGA SNP (rs10511260:  $r^2 = .20$ ,  $D' = .57$ ), with the two others having  $r^2 < .05$ . No SNPs or tagging SNPs confirmed in the German Alcoholism GWAS study were replicated in our analyses, nor were any SNPs or tagging SNPs identified as the most strongly associated SNPs in the SAGE study.

**Sample size projections**—The power of our study, because of its family-structured sampling, approximates that of a study using a random sample of 7600 individuals (a conservative figure, since it ignores the oversampling of individuals from the upper tail of the distribution of alcohol consumption). The range of effect sizes that we actually observed, of the order of 0.15–0.25% of the variance, implies that at least several hundred genetic variants are contributing to variation; but in a discovery sample, selecting top hits will overestimate true effect sizes. Thus in hindsight the achieved power of our study to detect the effect of a specific variant will be low (Table S11 in the Supplement). Given a median effect size of 0.0017 for SNPs associated with AUD-FS at  $p < .0001$ , replication of a true association with 90% probability at  $\alpha = 5E-8$ , in the ideal case of complete linkage

disequilibrium between marker and trait variant, would require approximately 27000 unrelated individuals, with more realistic effect sizes of 0.001 or 0.00075 requiring samples of 45,000 or 61,000. Imperfect linkage disequilibrium ( $D'=.9$ ) with a modest mismatch between genetic marker allele frequency and trait allele frequency (0.15 versus 0.05) would increase required sample sizes approximately 4-fold (e.g. 188,000 for a variant accounting for 0.001% of the variance).

## DISCUSSION

The primary conclusion from these analyses is that, as for many other complex phenotypes (e.g. body-mass index: (39)), effect sizes for the contribution of individual genetic variants to differences in heaviness of alcohol consumption and alcoholism risk are small, perhaps accounting for as little as one-tenth of one percent of the variance. The approximately log-normal distribution of alcohol consumption in the general population is consistent with the hypothesis that this variation is being explained by small effects of many variants acting additively, rather than a few rare family-specific variants (40). For traits such as height (41) cumulative results do appear to support an important polygenic contribution to variation; and it seems plausible, given the many central and peripheral effects of alcohol, that this will also be true for variation in alcohol consumption and alcoholism.

Given the low power of genetic linkage analyses in general community samples, relative to finding genetic association (42), and the accumulating evidence for small effect sizes, our failure to find genomewide significant linkage signals is unsurprising. However, whereas findings for heaviness of drinking were uniformly negative (LOD scores  $<1.5$ ), for AUD-FS our second highest peak coincided with a location identified in previous alcoholism linkage studies (Chr2, LOD 2.22 at 112cM (8,43)) while a second peak (Chr10, LOD 2.02 at 152cM) occurred within approximately 20cM of the peak reported in (44). We did not find evidence for clustering of SNP associations in these regions. In association analyses, while in a few cases we found suggestive convergence for consumption versus dependence phenotypes, or between inferences from family-based versus case-population control comparisons, in no case did results reach genomewide significance. Of the genes noted as of interest, *TMEM108* codes for a transmembrane protein of unknown function, but has previously been reported as associated with smoking cessation in a pooling GWAS study (45). *SHANK2* is a scaffolding gene implicated in the formation of the postsynaptic density at glutamergic synapses, and there have been reports of rare *SHANK2* variants overrepresented in cases of Autism Spectrum Disorder (46). For the remainder no obvious link with alcoholism risk can be identified. In no case do we have confidence that a true positive association has been identified; and because of small effect sizes, confirmation by the first generation of alcohol GWAS studies is not to be expected.

For alcoholism, published GWAS studies have been seriously underpowered so that accumulation of many more alcohol dependent cases with GWAS will be necessary. Our secondary analyses of current (12-month) alcohol consumption measures, which had effect sizes and p-values comparable to those for heaviest drinking period, also provide some limited grounds for hope that large scale cross-study analyses will ultimately be successful in identifying some of the variants that contribute to consumption differences, and thus indirectly to differences in dependence risk. Current though not heaviest period alcohol consumption measures will have been obtained in studies of many medical phenotypes as part of a dietary assessment (e.g. food frequency questionnaires, (47)), albeit typically using truncated scales that do not well characterize individuals with highly elevated consumption levels, and in older age-groups whose consumption may have declined substantially from their heaviest drinking period. Still, accumulating such data on one hundred thousand or more individuals (necessary to detect effect sizes of the order of 0.001% [one tenth of one

percent] of the variance with reasonable power) would be feasible. The immediate clinical value of identifying a number of very weakly but significantly associated variants by such an approach may be low, but identification of genes and pathways involved in individual variation in liability to alcohol use disorders could be great. Perhaps more narrowly defined consequences of alcohol effects (alcoholic liver disease; severe alcohol withdrawal) will give more hopeful outcomes. It is also possible however that work on genomic profiling, using random effects modeling of genomewide SNP data (48) will point to new directions in biological psychiatry, yielding greater understanding of the genetic contributions to individual phenotypes and enabling better quantification of genetic risk, thereby overcoming one of the primary challenges in prospective research on high-risk groups, and in resilience research, namely the inability to achieve sharp differentiation of genetic risk between groups.

There are several limitations of this research. The low density coverage of the 370K array may have contributed to negative findings. While the family-structured sampling design that we have used remains powerful for quantitative phenotypes, this is achieved at the cost of a loss of power for alcohol dependence diagnosis: in a general community sample the majority of such cases will be mild, and cases and their unaffected siblings may differ only modestly, in terms of symptom count. Second, while the study at the time it was implemented was powerful, considering anticipated effect sizes, subsequent findings across many complex phenotypes suggests that it is seriously underpowered given effect sizes that are likely to emerge for alcohol consumption and dependence outcomes. Third, our strategy of relying upon convergence of findings across consumption and dependence phenotypes (14) could cause us to miss associations that are specific to dependence. Finally, the cohorts that we have used in this research mostly were raised at a time of restrictive Australian divorce practices, so that even in families with parental alcoholism, it was usual for that parent to remain in the family. While this might increase effect size, through GxE interaction effects, it also requires confirmation of generalizability to more contemporary cohorts.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Supported by NIH grants AA07535, AA07728, AA13320, AA13321, AA14041, AA11998, AA17688, DA012854, DA019951; by grants from the Australian National Health and Medical Research Council (241944, 339462, 389927, 389875, 389891, 389892, 389938, 442915, 442981, 496739, 552485, 552498); by grants from the Australian Research Council (A7960034, A79906588, A79801419, DP0770096, DP0212016, DP0343921); and by the FP-5 GenomEUtwin Project (QLG2-CT-2002-01254). GWAS genotyping at CIDR was supported by a grant to the late Richard Todd, PhD, MD, former PI of grant AA13320 and a key contributor to research described in this manuscript. S.E.M., D.R.N., A.F.M., M.A.R.F., S.M., D.L.D., and G.W.M. are supported by the National Health and Medical Research Council (NHMRC) Fellowship Scheme. We acknowledge the contributions of project investigator Alexandre Todorov, PhD at Washington University. We also thank Dixie Statham, Ann Eldridge, Marlene Grace, Kerrie McAloney (sample collection); Lisa Bowdler, Steven Crooks (DNA processing); David Smyth, Harry Beeby, and Daniel Park (IT support) at Queensland Institute of Medical Research, Brisbane Australia. Last, but not least, we thank the twins and their families for their participation.

## Reference List

1. Cotton NS. The familial incidence of alcoholism: a review. *J Stud Alcohol*. 1979; 40:89–116. [PubMed: 376949]
2. Cloninger CR, Bohman M, Sigvardsson S. Inheritance of alcohol abuse. Cross-fostering analysis of adopted men. *Arch Gen Psychiatry*. 1981; 38:861–868. [PubMed: 7259422]



3. Goodwin DW, Schulsinger F, Moller N, Hermansen L, Winokur G, Guze SB. Drinking problems in adopted and nonadopted sons of alcoholics. *Arch Gen Psychiatry*. 1974; 31:164–169. [PubMed: 4851437]
4. Heath AC, Bucholz KK, Madden PA, Dinwiddie SH, Slutske WS, Statham DJ, et al. Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men. *Psychol Med*. 1997; 27:1381–1396. [PubMed: 9403910]
5. Kendler, KS.; Prescott, CA. *Genes, Environment, and Psychopathology: Understanding the Causes of Psychiatric and Substance Use Disorders*. Guilford Press; 2006.
6. Hill SY, Shen S, Zezza N, Hoffman EK, Perlin M, Allan W. A genome wide search for alcoholism susceptibility genes. *Am J Med Genet B Neuropsychiatr Genet*. 2004; 128B:102–113. [PubMed: 15211641]
7. Kuo PH, Neale MC, Riley BP, Webb BT, Sullivan PF, Vittum J, et al. Identification of susceptibility loci for alcohol-related traits in the Irish Affected Sib Pair Study of Alcohol Dependence. *Alcohol Clin Exp Res*. 2006; 30:1807–1816. [PubMed: 17067344]
8. Reich T, Edenberg HJ, Goate A, Williams JT, Rice JP, Van Eerdewegh P, et al. Genome-wide search for genes affecting the risk for alcohol dependence. *Am J Med Genet*. 1998; 81:207–215. [PubMed: 9603606]
9. Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E, et al. A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci USA*. 2010; 107:5082–5087. [PubMed: 20202923]
10. Edenberg HJ, Koller DL, Xuei X, Wetherill L, McClintick JN, Almasy L, et al. Genome-wide association study of alcohol dependence implicates a region on chromosome 11. *Alcohol Clin Exp Res*. 2010; 34:840–852. [PubMed: 20201924]
11. Treutlein J, Cichon S, Ridinger M, Wodarz N, Soyka M, Zill P, et al. Genome-wide association study of alcohol dependence. *Arch Gen Psychiatry*. 2009; 66:773–784. [PubMed: 19581569]
12. Heath AC, Meyer J, Jardine R, Martin NG. The inheritance of alcohol consumption patterns in a general population twin sample: II. Determinants of consumption frequency and quantity consumed. *J Stud Alcohol*. 1991; 52:425–433. [PubMed: 1943097]
13. Whitfield JB, Zhu G, Madden PA, Neale MC, Heath AC, Martin NG. The genetics of alcohol intake and of alcohol dependence. *Alcohol Clin Exp Res*. 2004; 28:1153–1160. [PubMed: 15318113]
14. Grant JD, Agrawal A, Bucholz KK, Madden PA, Pergadia ML, Nelson EC, et al. Alcohol consumption indices of genetic risk for alcohol dependence. *Biol Psychiatry*. 2009; 66:795–800. [PubMed: 19576574]
15. Birley AJ, James MR, Dickson PA, Montgomery GW, Heath AC, Whitfield JB, et al. Association of the gastric alcohol dehydrogenase gene ADH7 with variation in alcohol metabolism. *Hum Mol Genet*. 2008; 17:179–189. [PubMed: 17921519]
16. Hasin D, Aharonovich E, Liu X, Mamman Z, Matseoane K, Carr L, et al. Alcohol and ADH2 in Israel: Ashkenazis, Sephardics, and recent Russian immigrants. *Am J Psychiatry*. 2002; 159:1432–1434. [PubMed: 12153842]
17. Higuchi S. Polymorphisms of ethanol metabolizing enzyme genes and alcoholism. *Alcohol Alcohol Suppl*. 1994; 2:29–34. [PubMed: 8974313]
18. Higuchi S, Matsushita S, Muramatsu T, Murayama M, Hayashida M. Alcohol and aldehyde dehydrogenase genotypes and drinking behavior in Japanese. *Alcohol Clin Exp Res*. 1996; 20:493–497. [PubMed: 8727242]
19. Macgregor S, Lind PA, Bucholz KK, Hansell NK, Madden PA, Richter MM, et al. Associations of ADH and ALDH2 gene variation with self report alcohol reactions, consumption and dependence: an integrated analysis. *Hum Mol Genet*. 2009; 18:580–593. [PubMed: 18996923]
20. Bucholz KK, Heath AC, Reich T, Hesselbrock VM, Kramer JR, Nurnberger JI Jr, et al. Can we subtype alcoholism? A latent class analysis of data from relatives of alcoholics in a multicenter family study of alcoholism. *Alcohol Clin Exp Res*. 1996; 20:1462–1471. [PubMed: 8947326]
21. Heath AC, Bucholz KK, Slutske WS, Madden PAF, Dinwiddie SH, Dunne MP, et al. The assessment of alcoholism in surveys of the general community: What are we measuring? Some insights from the Australian twin panel interview study. *Int Rev Psychiatry*. 1994; 6:295–307.

22. Saha TD, Chou SP, Grant BF. Toward an alcohol use disorder continuum using item response theory: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Psychol Med.* 2006; 36:931–941. [PubMed: 16563205]
23. Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature.* 2008 Apr 3; 452:638–642. [PubMed: 18385739]
24. Saccone NL, Culverhouse RC, Schwantes-An TH, Cannon DS, Chen X, Cichon S, et al. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet.* 2010; 6:e1001053.
25. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet.* 2008; 40:616–622. [PubMed: 18385676]
26. Knopik VS, Heath AC, Madden PA, Bucholz KK, Slutske WS, Nelson EC, et al. Genetic effects on alcohol dependence risk: re-evaluating the importance of psychiatric and other heritable risk factors. *Psychol Med.* 2004; 34:1519–1530. [PubMed: 15724882]
27. Grant JD, Heath AC, Bucholz KK, Madden PA, Agrawal A, Statham DJ, et al. Spousal concordance for alcohol dependence: evidence for assortative mating or spousal interaction effects? *Alcohol Clin Exp Res.* 2007; 31:717–728. [PubMed: 17378921]
28. Saccone SF, Pergadia ML, Loukola A, Broms U, Montgomery GW, Wang JC, et al. Genetic linkage to chromosome 22q12 for a heavy-smoking quantitative trait in two independent samples. *Am J Hum Genet.* 2007; 80:856–866. [PubMed: 17436240]
29. Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G, et al. Common variants in the trichohyalin gene are associated with straight hair in Europeans. *Am J Hum Genet.* 2009; 85:750–755. [PubMed: 19896111]
30. Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI Jr, et al. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *J Stud Alcohol.* 1994; 55:149–158. [PubMed: 8189735]
31. Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V. A validity study of the SSAGA—a comparison with the SCAN. *Addiction.* 1999; 94:1361–1370. [PubMed: 10615721]
32. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38:904–909. [PubMed: 16862161]
33. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet.* 2002; 30:97–101. [PubMed: 11731797]
34. Chen WM, Abecasis GR. Family-based association tests for genomewide association scans. *Am J Hum Genet.* 2007; 81:913–926. [PubMed: 17924335]
35. Thornton T, McPeck MS. Case-control association testing with related individuals: a more powerful quasi-likelihood score test. *Am J Hum Genet.* 2007; 81:321–337. [PubMed: 17668381]
36. Chen WM, Manichaikul A, Rich SS. A generalized family-based association test for dichotomous traits. *Am J Hum Genet.* 2009; 85:364–376. [PubMed: 19732865]
37. StataCorp. *Statistical Software: Release 7.0.* College Station, TX: Stata Corporation; 2001.
38. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics.* 2003; 19:149–150. [PubMed: 12499305]
39. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet.* 2010; 42:937–948. [PubMed: 20935630]
40. Goldstein DB. Common genetic variation and human traits. *N Engl J Med.* 2009; 360:1696–1698. [PubMed: 19369660]
41. Yang J, Benyamin G, McEvoy BP, Gordon S, Henders AK, Nyholt D, et al. Common SNPs explain a large proportion of heritability for human height. *Nat Genet.* 2010; 42:565–569. [PubMed: 20562875]

42. Sham PC, Cherny SS, Purcell S, Hewitt JK. Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *Am J Hum Genet.* 2000; 66:1616–1630. [PubMed: 10762547]
43. Dick DM, Meyers J, Aliev F, Nurnberger J Jr, Kramer J, Kuperman S, et al. Evidence for genes on chromosome 2 contributing to alcohol dependence with conduct disorder and suicide attempts. *Am J Med Genet B Neuropsychiatr Genet.* 2010; 153B:1179–1188. [PubMed: 20468071]
44. Panhuysen CI, Kranzler HR, Yu Y, Weiss RD, Brady K, Poling J, et al. Confirmation and generalization of an alcohol-dependence locus on chromosome 10q. *Neuropsychopharmacology.* 2010; 35:1325–1332. [PubMed: 20147890]
45. Uhl GR, Liu QR, Drgon T, Johnson C, Walther D, Rose JE, et al. Molecular genetics of successful smoking cessation: convergent genome-wide association study results. *Arch Gen Psychiatry.* 2008; 65:683–693. [PubMed: 18519826]
46. Berkel S, Marshall CR, Weiss B, Howe J, Roeth R, Moog U, et al. Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nat Genet.* 2010; 42:489–491. [PubMed: 20473310]
47. Block G, Subar AF. Estimates of nutrient intake from a food frequency questionnaire: the 1987 National Health Interview Survey. *J Am Diet Assoc.* 1992; 92:969–977. [PubMed: 1640041]
48. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.* 2007; 17:1520–1528. [PubMed: 17785532]

**Table 1**

Summary of regression-based multipoint linkage results for primary quantitative phenotype AUD Factor score. Only LOD scores > 1.5 are shown. No LOD scores  $\geq$  1.5 were observed for HOD Factor score.

AUD-Factor Score		
Chromosome	Location (CM)	LOD
1	153.06	2.30
2	112.31	2.22
10	152.45	2.02

**Table 2**

Effect size ( $\%h^2$  = QTL heritability) range for most strongly associated SNPs ( $p < .0001$ ) with heaviness of drinking (HOD) and alcohol use disorder (AUD) factor score measures, and for secondary consumption and alcohol dependence (AD) factor score measures. Also given is the most extreme p-value observed for each measure. Items marked with an asterisk are components of the heaviness of drinking (HOD) factor score. Primary phenotype results are bolded.

	N	Effect sizes ( $\%h^2$ )	Lowest p-value
<b>AUD Factor Score</b>	<b>7490–8209</b>	<b>0.22–0.35</b>	<b>7.2E-7</b>
AD Factor Score	7490–8209	0.21–0.33	8.2E-7
Heaviest drinking period			
<b>HOD factor score</b>	<b>6194–6300</b>	<b>0.27–0.50</b>	<b>1.2E-7</b>
Frequency of any use	6411–6481	0.26–0.39	2.4E-6
* Frequency of heavy drinking <sup>a</sup>	6198–6481	0.26–0.47	3.7E-7
* Frequency drunk <sup>b</sup>	6017–6098	0.27–0.40	4.7E-6
Drinks per drinking day	6198–6481	0.26–0.40	1.5E-6
* Drinks per week	6204–6481	0.26–0.39	3.7E-6
* Max drinks	8218–8305	0.21–0.36	5.4E-7
Past 12-month consumption			
Frequency of any use	7665–7947	0.22–0.42	1.0E-7
Drinks per drinking day	7798–7910	0.22–0.34	1.1E-6
Drinks per week	7819–7907	0.22–0.32	3.1E-6
Max drinks	7453–7529	0.23–0.32	4.6E-6

Notes:

<sup>a</sup>Frequency of drinks 5 or more standard drinks in a day.

<sup>b</sup>Excludes individuals who had never been drunk; these individuals were re-scored as zero on this item and included in the HOD-FS analyses.