Alcohol consumption and mate choice in UK Biobank: comparing observational and Mendelian randomization estimates

Authors: Laurence J Howe*, [1, 3] Daniel J Lawson, [1] Neil M Davies, [1] Beate St. Pourcain, [2] Sarah J Lewis, [1] George Davey Smith, [1, #] and Gibran Hemani [1, #]

[1] Medical Research Council Integrative Epidemiology Unit, Population Health Sciences, University of Bristol, United Kingdom

[2] Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

[3] Institute of Cardiovascular Science, University College London, United Kingdom

* Correspondence to lh14833@bristol.ac.uk

# These authors contributed equally

Abstract

Alcohol use is correlated within spouse-pairs, but it is difficult to disentangle the effects of alcohol consumption on mate-selection from social factors or cohabitation leading to spouses becoming more similar over time. We hypothesised that genetic variants related to alcohol consumption may, via their effect on alcohol behaviour, influence mate selection.

Therefore, in a sample of over 47,000 spouse-pairs in the UK Biobank we utilised a well-characterised alcohol related variant, rs1229984 in *ADH1B*, as a genetic proxy for alcohol use. We compared the phenotypic concordance between spouses for self-reported alcohol use with the association between an individual's self-reported alcohol use and their partner's rs1229984 genotype using Mendelian randomization. This was followed up by an exploration of the spousal genotypic concordance for the variant and an analysis determining if relationship length may be related to spousal alcohol behaviour similarities.

We found strong evidence that both an individual's self-reported alcohol consumption and rs1229984 genotype are associated with their partner's self-reported alcohol use. The Mendelian randomization analysis found that each unit increase in an individual's weekly alcohol consumption increased their partner's alcohol consumption by 0.26 units (95% C.I. 0.15, 0.38; $P=1.10 \times 10^{-5}$). Furthermore, the rs1229984 genotype was concordant within spouse-pairs, suggesting that some spousal concordance for alcohol consumption existed prior to cohabitation. Although the SNP is strongly associated with ancestry, our results suggest that this concordance is unlikely to be explained by population stratification. Overall, our findings suggest that alcohol behaviour directly influences mate selection.

Introduction

Human mate choice is highly non-random; spouse-pairs are generally more phenotypically similar than would be expected by chance [1-6]. Previous studies suggest that alcohol related phenotypes, ranging from consumption to alcohol dependence, are highly correlated within spouse-pairs [7-13]. However, the extent to which the spousal concordance is due to the effect of alcohol behaviour on mate selection (assortative mating) is currently unclear. Indeed, the spousal concordance may be related to assortment on other social and environmental factors (social homogamy) or be a consequence of an individual's partner influencing their alcohol behaviour after the individuals have paired up (partner interaction effects) or even relate to spousal similarities influencing relationship length (relationship dissolution) [11-13]. The mechanism explaining spousal concordance for alcohol consumption could have important implications relating to human social and reproductive behaviour. **Figure 1** illustrates possible explanations for spousal concordance on alcohol consumption.

One biological mechanism that partially explains the phenotypic concordance between spouse-pairs is that they are on average more genetically similar across the genome than non-spouse-pairs [14]. Genotypes implicated in the aetiology of height, education, blood pressure and several chronic diseases have been shown to be correlated within spouse-pairs [15-18]. It is not known whether genetic variants implicated in alcohol metabolism, via their effect on alcohol behaviour, contribute to mate selection.

Alcohol behaviour has been shown to be highly heritable with estimates of 30-50% for alcohol use disorders [19 20] and a common variant heritability of 13% for

self-reported alcohol consumption [21]; Genome-wide Association Studies (GWAS) have identified more than 15 loci implicated in either the aetiology of alcohol dependence [22-26] or alcohol consumption volume [21 24 27-29]. Notably, genetic variants in the Alcohol Dehydrogenase (*ADH*) and Aldehyde Dehydrogenase (*ALDH*) gene families are associated with differences in alcohol consumption [30]. For example, *ADH1B* is involved in the production of enzymes that oxidise alcohol and so individuals with certain alleles may find alcohol consumption unpleasant, resulting in lower intake. Similarly, a genetic variant in *ALDH2*, rare in non-east Asian populations, is associated with a "flush reaction" to alcohol [31 32].

Alcohol consumption-related genetic variants can be useful to determine the most likely explanation for the spousal phenotypic concordance for alcohol use, by analogy with Mendelian randomization studies [33 34]. Genetic variants for alcohol consumption are in theory less susceptible to confounding from socioeconomic and behavioural factors than measured alcohol consumption so can be used to rule out the possibility that social homogamy is driving the spousal phenotypic concordance [33 35]. The timing of the effects of alcohol consumption can be discerned by evaluating the spousal genotypic concordance for alcohol use-related variants. Genotypic concordance would imply that an effect exists prior to pairing, suggesting that some degree of the spousal phenotypic concordance is attributable to assortative mating (**Figure 2**).

In this study we aimed to explore spousal similarities for alcohol consumption using observational and genetic data. First, we estimated the association of an individual's self-reported alcohol use with the self-reported alcohol use of their partner. Second, we used a Mendelian randomization framework to estimate the effect of an individual's alcohol use on their spouse's alcohol use. Here, we used

their partner's rs1229984 genotype, a missense mutation in *ADH1B* strongly associated with alcohol consumption as an instrumental variable for self-reported alcohol consumption. Third, we estimated the association of rs1229984 genotype between spouses, to evaluate the timing of possible causal effects, and investigate the possibility of bias from population stratification. Fourth, using the mean age of each couple as a proxy for relationship duration, we determined if there was an association between longer relationships and more similar spousal alcohol behaviour. As a positive control, to demonstrate the validity of derived spouse pairs and the usage of a Mendelian randomization framework, we also analysed height, known to be correlated between spouses, using similar methods.

Materials and Methods

Study participants

UK Biobank

UK Biobank is a large-scale cohort study, including 502,655 participants aged between 40-69 years. Study participants were recruited from 22 recruitment centres across the United Kingdom between 2006 and 2010 [36 37]. For the purposes of our analyses, we restricted the dataset to a subset of 463,827 individuals of recent European descent with available genotype data, with individuals of non-European descent removed based on a k-means cluster analysis on the first 4 genetic principal components [38]. The different subsets of UK Biobank utilised in our analyses are illustrated in **Supplementary Figure 1**.

Spouse-pair subsample

Spouse information is not explicitly available, therefore we used similar methods to previous studies [15-17] to identify spouse-pairs in the UK Biobank. Starting

with the European subsample described above, household sharing information was used to extract pairs of individuals who (a) report living with their spouse (6141-0.0), (b) report the same length of time living in the house (699-0.0), (c) report the same number of occupants in the household (709-0.0), (d) report the same number of vehicles (728-0.0), (e) report the same accommodation type and rental status (670-0.0, 680-0.0), (f) have identical home coordinates (rounded to the nearest km) (20074-0.0, 20075-0.0), (g) are registered with the same UK Biobank recruitment centre (54-0.0) and (h) both have available genotype data. If more than two individuals shared identical information across all variables, these individuals were excluded from analysis. At this stage, we identified 52,471 potential spouse-pairs.

We excluded 4,866 potential couples who were the same sex (9.3% of the sample), as unconfirmed same sex pairs may be more likely to be false positives. Although sexual orientation data was collected in UK Biobank, access is restricted for privacy/ethical reasons. To reduce the possibility that identified spouse-pairs are in fact related or non-related familial, non-spouse pairs; we removed 3 pairs reporting the same age of death for both parents (1807-0.0, 3526-0.0). Then we constructed a genetic relationship matrix (GRM) amongst derived pairs and removed 53 pairs with estimated relatedness (IBD >0.1). To construct the GRM; we used a pool of 78,341 markers which were derived by LD pruning (50KB, steps of 5 KB, r2<0.1) 1,440,616 SNPs from the HapMap3 reference panel [39] using the 1000 Genomes CEU genotype data [40] as a reference panel. The final-sample included 47,549 spouse-pairs.

Non-spouse-pair samples

For secondary analyses requiring data from unrelated individuals, we derived a sample of individuals of European descent and a more restrictive sample believed

146 to be of white British descent. Starting with the UK Biobank subset of 463,827
147 individuals of recent European descent, we removed 78,540 related individuals
148 (relevant methodology has been described previously [38]) to generate the European
149 sample and using lists provided by UK Biobank, further restricted this sample to
150 337,114 individuals identifying as being of “white British” descent.

151

152 **Height and educational attainment**

153 At baseline, the height (cm) of UK Biobank participants was measured using a
154 Seca 202 device at the assessment centre (ID: 50-0.0). Measured height was used
155 as a positive control for the application of a Mendelian randomization framework in
156 the context of assortative mating.

157 Educational attainment as characterised by years in full-time education was
158 defined as in a previous publication [41]. Individuals born outside England, Scotland or
159 Wales were removed because of schooling system differences, participants with a
160 college or university degree were classified with a leaving age of 21 years and
161 participants who self-reported leaving school when younger than 15 years were
162 classified with a leaving age of 15. Educational attainment was included as a
163 covariate in phenotypic analyses of spousal alcohol behaviour similarities as a
164 possible confounder.

165

166 **Self-reported alcohol variables**

167 At baseline, study participants completed a questionnaire. Participants were
168 asked to describe their current drinking status (never, previous, current, prefer not to

say) (ID: 20117-0.0) and estimate their current alcohol intake frequency (daily or almost daily, three or four times a week, once or twice a week, one to three times a month, special occasions only, never, prefer not to say) (ID: 1558-0.0). Individuals reporting a current intake frequency of at least “once or twice a week” were asked to estimate their average weekly intake of a range of different alcoholic beverages (red wine, white wine, champagne, beer, cider, spirits, fortified wine) (ID: 1568-0.0, 1578-0.0, 1588-0.0, 1598-0.0, 1608-0.0).

From these variables, we derived three measures: ever or never consumed alcohol (current or former against never), a binary measure of current drinking for self-reported current drinkers (three or more times a week against less than three times a week) and an average intake of alcoholic units per week, derived by combining the self-reported estimated intakes of the different alcoholic beverages consumptions across the five drink types, as in a previous study [21]. The questionnaire used the following measurement units for each of the five alcoholic drink types: measures for spirits, glasses for wines and pints for beer/cider which were estimated to be equivalent to 1, 2 and 2.5 units respectively. Individuals reporting current intake frequency of “one to three times a month”, “special occasions only” or “never” (for whom this phenotype was not collected), were assumed to have a weekly alcohol consumption volume of 0. More information on alcohol variables used in this study is contained in **Supplementary Table 1**.

Genotyping

488,377 UK Biobank study participants were assayed using two similar genotyping arrays, the UK BiLEVE Axiom™ Array by Affymetrix1 (N= 49,950) and the closely-related UK Biobank Axiom™ Array (N= 438,427). Directly genotyped variants were pre-phased using SHAPEIT3 [42] and then imputed using Impute4 using

194 the UK10K [43], Haplotype Reference Consortium [44] and 1000 Genomes Phase 3 [40]
195 reference panels. Post-imputation, data were available on approximately ~96 million
196 genetic variants.

197 **Statistical analysis**

198 *Utilising genetic variation to disentangle spousal correlations*

199 In general, the effects of genetic variation on a phenotype can be assumed to
200 be via the variant's effect on intermediary observable or unobservable phenotypes.
201 In the context of assortative mating, it is unlikely that individuals would assort based
202 directly on genotype but rather on an observed phenotype influenced by genetic
203 factors. Assuming that a phenotype is influenced by genetic factors G and individuals
204 assort on the phenotype such that the phenotypic correlation between spouses is
205 equal to C, then expected correlations between an index individual's G and their
206 partner's phenotype and G induced by assortment can be shown to be a function of
207 the heritability of the phenotype and the spousal phenotypic correlation C
208 (**Supplementary Methods**). This implies that estimates of assortative mating
209 utilising genetic data are likely to be attenuated compared to the true value of
210 phenotypic assortment, unless genetic factors completely explain variation in the
211 phenotype of interest.

212 However, there are notable advantages of applying genetic approaches such
213 as Mendelian randomization and genetic correlation analyses to the context of
214 assortative mating for mechanistic understanding. In conventional Mendelian
215 randomization studies [33 34], genetic variants are used as proxies for a measured
216 exposure to evaluate potential causal relationships between an exposure and an
217 outcome (e.g. LDL cholesterol and coronary heart disease [45]). Genetic proxies may

be more reliable than the measured exposure because of the reduced potential for confounding and reverse causation.

In the context of Mendelian randomization across spouses, the premise is largely similar; the exposure is an individual's phenotype (e.g. alcohol consumption), proxied by a genetic instrument, and the outcome is their partner's phenotype (e.g. alcohol consumption). A Mendelian randomization approach can evaluate a direct effect of an individual's alcohol consumption on the alcohol consumption of their partner as opposed to effects of social homogamy. A direct effect captured by a Mendelian randomization framework could capture; individuals being likely to select a mate with similar behaviour (assortative mating), an individual's alcohol consumption influencing their partner's during the relationship (partner interaction effects) or more similar couples staying together for longer (relationship dissolution). Interpretation can be nuanced, as for example, it seems unlikely that an individual's height could influence the height of their partner, but partner interaction effects are highly plausible for alcohol behaviour.

Similarly, estimating the genotypic concordance between-spouses for variants relating to a trait of interest can be used to improve mechanistic understanding. The interpretation of genotypic concordance is comparable to that of Mendelian randomization across spouses with two important distinctions. First, genotypic concordance will not capture partner interaction effects as germline DNA is fixed for both spouses prior to assortment. Second, concordance induced by assortment will be further attenuated compared to a Mendelian randomization approach.

Phenotypic spousal concordance for height

To verify the validity of the derived spouse-pair sample, we evaluated the spousal phenotypic concordance for height. Previous studies have found strong evidence of spousal concordance for height, so comparable results would be consistent with derived spouses being genuine. The spousal phenotypic concordance was estimated using a linear regression of an individual's height against the height of their partner, adjusting for sex. With one unique phenotype pairing within couples (male spouse height/ female spouse height), each individual in the data-set was included only once as either the reference individual or their partner.

Mendelian randomization: Genetically influenced height and measured height of partner

We validated the application of a Mendelian randomization approach to assortative mating using height as a positive control; genotypes influencing height have previously demonstrated to be highly correlated between spouse-pairs [15]. As a measure of genetically influenced height, we started with 382 independent SNPs, generated using LD clumping (r^2<0.001) in MR-Base [46], from a recent Genome-wide Association Study (GWAS) of adult height in Europeans [47].

For the purposes of the Mendelian randomization analysis, we restricted analyses to spouse-pairs with complete measured height data and genotype data. First, we estimated the association between 378 SNPs (4 SNPs were unavailable in the QC version of the data-set) and height in the same individual, using the spouse-pair sample with sex included as a covariate. Second, we estimated the association between the 378 SNPs and spousal height. PLINK [48] was used to estimate the SNP-

phenotype associations also including sex as a covariate. We then estimated the effect of a 1 cm increase in an individual's height on their partner's height using the TwoSampleMR R package [46] and the internally derived weights described above. The fixed-effects Inverse-Variance Weighted (IVW) method was used as the primary analysis. Cochran's Q test and the I^2 statistic were used to test for heterogeneity in the fixed-effects IVW [49]. MR Egger [50] was used to test for directional pleiotropy. The weighted median [51] and mode [52] were used to test the consistency of the effect estimate. With two unique pairings between genotype and phenotype in each couple (male spouse genotype/ female spouse height and the converse), each individual in the data-set was included twice as both the reference individual and as the partner.

Spousal genetic concordance for height

To evaluate spousal genotypic concordance for height, we evaluated the association between height genetic risk scores (GRS) across spouse-pairs. Height GRS were constructed using previously described height loci in PLINK [48]. The cross-spouse association was estimated using linear regression of an individual's GRS against the GRS of their partner. With one unique genotype pairing within couples (male spouse genotype/female spouse genotype), each individual in the dataset was included only once as either the reference individual or their partner.

Phenotypic spousal concordance for self-reported alcohol behaviour

To evaluate the phenotypic concordance on alcohol use we compared self-reported alcohol behaviour between spouses. We estimated the spousal concordance for the two binary measures (ever or never consumed alcohol, three or more times a week) using a logistic regression of the relevant variable for an individual against the relevant variable for their partner, adjusting for sex, age and

partner's age. In addition, we included recruitment centre, height and education (of both spouses) in the model as potential confounders. Similarly, linear regression was used to estimate the spousal-concordance for continuous weekly alcohol consumption volume, adjusting for the same covariates. Spouse-pairs with any missing phenotype data, or where one or more spouses reported their weekly alcohol consumption volume to be more than five standard deviations away from the mean (calculated using the sample of individuals with non-zero weekly drinking) were removed from relevant analyses. With one unique phenotype pairing within couples (male alcohol variable/ female alcohol variable), each individual in the data-set was included only once as either the reference individual or their partner.

Mendelian randomization: Genetically influenced alcohol consumption volume and self-reported alcohol consumption of partner

We then applied the Mendelian randomization framework to investigate if an individual's genotype at rs1229984 in *ADH1B* affects the self-reported alcohol consumption volume of their partner. Given the rarity of individuals homozygous for the minor allele in European populations, the MAF is 2.9% in the 1000 Genomes CEU population [40], we first determined whether an additive or a dominant model (as used in previous studies [45 53]) was most appropriate for the SNP by comparing the association of genotype at rs1229984 with self-reported weekly alcohol consumption in the European and British samples. We found strong evidence to suggest that the SNP has an additive effect on alcohol consumption (**Supplementary Table 2**) and assumed this model in all relevant analyses.

For the Mendelian randomization analysis, we restricted analysis to spouse-pairs where both members had genotype data, and one or more members had self-reported alcohol consumption volume. First, we estimated the association of the

313 rs1229984 genotype with alcohol consumption in the same individual after adjusting
314 for sex, age, centre and the first 10 principal components of the reference individual.
315 Second, we estimated the association between rs1229984 and spousal alcohol
316 consumption after adjusting for sex, age (of both spouses), centre and the first 10
317 principal components of both spouses. PLINK [48] was used to estimate the SNP-
318 phenotype associations. We then estimated the effect of a 1 unit increase in an
319 individual's weekly alcohol consumption volume on the same variable in their
320 partner. The Wald ratio estimate was obtained using mr_wald_ratio function in the
321 TwoSample MR R package [46] using internally derived weights. Sensitivity analyses
322 were limited due to the use of a single genetic instrument. With two unique pairings
323 between genotype and phenotype in each couple (male alcohol variable/ female
324 genotype and the converse), each individual in the data-set was included twice as
325 both the reference individual and as the partner.

326 *Spousal genotypic concordance for rs1229984 genotype*

327 We then investigated properties of the rs122984 variant in the UK Biobank
328 that may be relevant to assortative mating. Starting with the UK Biobank subset of
329 463,827 individuals of recent European descent, we removed 78,540 related
330 individuals (relevant methodology has been described previously [38]) and tested
331 Hardy-Weinberg Equilibrium (HWE) in the resulting sample of 385,287 individuals.
332 To evaluate the possibility of population stratification, we investigated the association
333 of both the SNP and self-reported alcohol consumption with genetic principal
334 components and birth coordinates. As a sensitivity analysis, we also restricted the
335 sample to a more homogeneous sample of British individuals, provided by the UK
336 Biobank, and repeated analyses.

We then estimated the genotypic concordance between derived spouse-pairs for rs1229984 genotype using linear regression. As a sensitivity analysis, we then investigated the possibility that spousal-concordance for rs1229984 was driven by fine-scale assortative mating due to geography, which is itself associated with genetic variation within the UK [54 55]. For this, we restricted the sample to include only 28,653 spouse-pairs born within 100 miles of each other. To test the validity of this sensitivity analysis, we explored whether birth or genetic differences (as determined by principal components) between spouses are associated with alcohol behaviour or rs122984 genotype differences in the restricted and full spouse-pair samples. The spouse-pairs were then stratified into the 22 different UK Biobank recruitment centres and logistic regression analyses were re-run to estimate the spousal-concordance of the *ADH1B* genotype by centre. With one unique genotype pairing within couples (male genotype/female genotype), each individual in the dataset was included only once as either the reference individual or their partner. Geographical patterns of heterogeneity across the different UK Biobank recruitment centres would provide evidence of population stratification.

Relationship duration and spousal alcohol behaviour

Relationship length may influence spousal similarities for alcohol behaviour because spouses become more similar over time or because pairs with similar alcohol behaviour tend to have longer relationships. To explore these possibilities, we investigated the association between relationship length and alcohol behaviour and rs122984 genotype similarities. Without available data on relationship length, we used the mean age of each couple as a proxy and evaluated associations using a linear regression of mean couple age against spousal difference in weekly alcohol

consumption and rs1229984 genotype. Analyses were adjusted for the sex of reference individual.

A list of derived spouse-pairs has been returned to UK Biobank. For details please contact access@ukbiobank.ac.uk.

Results

Spousal concordance for height

Phenotypic concordance for height

Measured height was strongly concordant between spouse-pairs. In a sample of 47,377 spouse-pairs, a 1 unit increase in an individual's height was associated with a 0.24-unit increase (95% C.I. 0.23, 0.25, $P<10^{-16}$) in their partner's height. This result is consistent with previous findings [56 57], validating the derived spouse pairs.

Mendelian randomization framework: Genetically influenced height and height of partner

The application of Mendelian randomization to spousal height was consistent with the previous evidence for assortative mating on height. Across 47,377 spouse-pairs, a 1 cm increase in an individual's height was associated with a 0.19 cm increase in their partner's height (95% C.I. 0.18, 0.21; $P<10^{-16}$), distinctly smaller than the phenotype estimate (Z-test for difference of means: $P=8.3\times10^{-8}$). The I^2 statistic (2.9%) and Cochran's Q test (P=0.64) suggested consistent effects across SNPs, and estimates were consistent across the weighted median, weighted modal

and MR-Egger estimators with the MR-Egger intercept test finding no strong evidence for directional pleiotropy (**Table 1)**.

Table 1: Mendelian randomization estimates for the effect of a 1 cm increase in height on partner's height

Test	Interpretation	Estimate (95% C.I.)	P-value
Phenotypic association for comparison	N/A	0.24 (0.23, 0.25)	$<10^{-16}$
Inverse variance weighted	Primary causal estimate[1]	0.19 (0.18, 0.21)	$<10^{-16}$
Heterogeneity of Inverse variance weighted	Balanced pleiotropy	I^2=3.6%	0.68
MR-Egger	Intercept test for directional pleiotropy[2]	0.001 (-0.006, 0.008)	0.75
	Regression estimate[1]	0.19 (0.15, 0.21)	$<10^{-16}$
Weighted median	Consistency[1]	0.18 (0.15, 0.21)	$<10^{-16}$
Weighted mode	Consistency[1]	0.17 (-0.23, 0.57)	0.41

1 Units: mm change in partner's height per 1-unit increase in individual's height

2 Units: Average pleiotropic effect of a height genetic variant on partner's height

Genotypic concordance for height

Similarly, the genotypic concordance analysis for height was strongly concordant with previous findings; we found strong evidence that spouses have similar genotypes at height influencing loci. Each 1 S.D. increase in an individual's height GRS was associated with a 0.024 S.D. increase in their partner's GRS (95% C.I. 0.015, 0.033; $P=1.96\times10^{-7}$).

Spousal concordance for self-reported alcohol behaviour

Phenotypic spousal concordance

The majority of derived spouse-pairs had complete data for relevant self-reported alcohol behaviour phenotypes. Strong evidence was found for phenotypic concordance between spouse-pairs for all self-reported alcohol variables. Amongst 45,066 spouse-pairs, an individual self-reporting as a never-drinker was associated with increased odds (OR 13.03, 95% C.I., 10.98, 15.44 $P<10^{-16}$) of their partner self-reporting as a never-drinker. Similarly, when restricting to 40,723 pairs who both reported being current-drinkers, an individual drinking three or more times a week had increased odds (OR 6.24, 95% C.I., 5.95, 6.54 $P<10^{-16}$) of their partner also drinking three or more times a week.

For self-reported alcohol consumption volume; 44,886 spouse-pairs had either complete phenotype data or reported their consumption frequency as less than weekly (in which case their weekly volume was assumed to be 0). After removing 189 pairs with outlying values (>5 S.D from the mean) from one or more members, the final sample included 47,321 spouse-pairs. In this sample, each unit increase in an individual's weekly alcohol consumption volume was associated with a 0.37-unit increase (95% C.I. 0.36, 0.38 $P<10^{-16}$) in the same variable in their partner.

Mendelian randomization: Genetically influenced alcohol consumption and self-reported alcohol behaviour of partner

To evaluate the degree to which an individual's alcohol consumption is affected by their partner's genetically influenced alcohol consumption, we used a sample of 47,321 spouse-pairs with available data on weekly alcohol consumption. In this sample, each additional copy of the *ADH1B* major allele was associated with

an increased weekly alcohol consumption of 3.98 units a week (95% C.I. 3.51, 4.43; $P<10^{-16}$) in the same individual. Each additional copy of the major allele was associated with an increased weekly alcohol consumption of 1.04 units a week (95% C.I. 0.58, 1.51; $P=1.09x10^{-5}$) in the reference individual's partner. After scaling the estimate using a Wald estimator; a 1 unit increase in an individual's alcohol consumption led to having partner's with alcohol consumption 0.26 units higher than baseline (95% C.I. 0.15, 0.38; $P=1.10x10^{-5}$). This effect is slightly lower than the phenotypic estimate of 0.37 units (95% C.I. 0.36, 0.38) although confidence intervals overlap (Z-test for difference of means: P=0.064).

Characteristics of rs1229984 in the UK Biobank

In the sample of 385,287 individuals of recent European descent, the MAF of rs1229984 was 2.8% and very strong evidence was found for the SNP violating HWE (Chi^2 = 275, $P <10^{-16}$) due to fewer heterozygotes compared to expectation (expected=20,972, observed=20,194). However, when restricting to the sample of 337,114 individuals of British descent, the MAF of rs1229984 was 2.2% and there was little evidence of the SNP violating HWE (Chi^2 = 2.0, P=0.16) and there were more heterozygotes compared to expected (expected= 14,506 observed=14,743) (**Supplementary Table 3**). Evidence was found of allele frequency differences for rs1229984 between the two samples (Chi^2=445, $P<10^{-16}$) suggesting that population substructure differences may explain the HWE results.

The SNP was found to be strongly associated with both genetic principal components and birth coordinates in both samples. In the less restrictive European sample, each additional major allele of rs1229984 was associated with being born

24.6 miles farther north (95% C.I. 22.2, 27.0) and 13.3 miles farther west (95% C.I. 12.1, 14.5). The SNP was similarly associated with principal components and birth coordinates in the sample of British descent although there were differences in effect estimates between the two samples (**Supplementary Table 4**). We also found strong evidence that self-reported alcohol consumption is strongly associated with birth coordinates and principal components in both samples concordant directionally with the SNP associations (**Supplementary Table 5**).

Genotypic concordance

Amongst 47,549 spouse-pairs, strong concordance was observed for the genotype of rs1229984. Each additional copy of the major rs1229984 allele was associated with an increased number of major alleles in their partner (Beta 0.019; 95% C.I. 0.010, 0.028; $P=5.0x10^{-5}$).

As a sensitivity analysis, we restricted the sample to 28,653 spouse-pairs born within 100 miles of each other and stratified spouse-pairs by the 22 different UK Biobank recruitment centres. In this sample, we did not find strong evidence that birth location differences were associated with similarities in alcohol behaviour or rs1229984 genotype, contrasting with clear evidence of associations in the full spouse-sample. However, we did find evidence that genomic principal component differences were associated with spousal similarities for these variables, likely reflecting the fine-scale population structure of UK Biobank (**Supplementary Table 6**). Of the 22 centres, 2 centres were omitted from the meta-analysis because the limited sample sizes led to convergence issues in regression. A fixed-effects meta-analysis was then used to estimate the spousal-concordance across the remaining 20 centres and 28,615 spouse-pairs. Evidence was found of spousal concordance

for rs1229984 (Beta 0.016; 95% C.I. 0.004, 0.028; P=0.011), consistent with the previous analysis. Cochran's Q test for heterogeneity across the betas suggested no strong evidence for heterogeneity (P= 0.34) across the different centres (**Table 2**).

Table 2: Meta-analysis of spousal-concordance for rs1229984 across the UK Biobank recruitment centres

Recruitment Centre	Number of spouse-pairs born within 100 km of each other	Beta (95% C.I.)
Stockport	9	N/A[1]
Manchester	662	0.024 (-0.088, 0.0675)
Oxford	669	-0.010 (-0.088, 0.067)
Cardiff	930	0.022 (-0.043, 0.088)
Glasgow	1046	0.072 (0.019, 0.125)
Edinburgh	611	-0.047 (-0.166, 0.070)
Stoke	1215	-0.012 (-0.075, 0.051)
Reading	1352	0.003 (-0.055, 0.060)
Bury	2244	0.012 (-0.031, 0.055)
Newcastle	2976	-0.025 (-0.064, 0.013)
Leeds	2563	0.041 (0.001, 0.081)
Bristol	2117	0.015 (-0.030, 0.060)
St Bartholomew's Hospital	122	-0.073 (-0.220, 0.074)
Nottingham	2342	0.025 (-0.017, 0.066)
Sheffield	2260	0.037 (-0.009, 0.082)
Liverpool	2632	0.023 (-0.020, 0.066)
Middlesbrough	1477	0.002 (-0.050, 0.053)
Hounslow	838	0.073 (-0.000, 0.147)
Croydon	1034	0.044 (-0.027, 0.115)
Birmingham	1440	-0.019 (-0.068, 0.031)
Swansea	85	-0.068 (-0.283, 0.146)
Wrexham	29	N/A[1]
Combined (Fixed effects)	**28,615**	**0.016 (0.004, 0.028) P=0.011**

1 Linear regression estimates did not converge due to limited sample sizes, these studies were excluded from the meta-analysis.

Relationship length and spousal alcohol behaviour similarities

We did not find strong evidence that increased mean couple age, used as a proxy for relationship length, was associated with more concordant spousal alcohol behaviour. Per 1-year increase in couple mean age, spousal differences in terms of weekly alcohol units consumed were 0.017 smaller (95% C.I. -0.040, 0.007, P=0.16). In terms of genotypic differences at rs1229984, we found weak evidence that older

483 couples were more dissimilar at the locus. Per 1-year increase in couple mean age,
484 spousal allelic differences at rs1229984 were 0.0004 larger (95% C.I. 0.0000,
485 0.0009; P=0.035).

486

487 ## Discussion

488 In this study, we used a large sample of derived spouse-pairs in a UK-based
489 cohort to demonstrate that an individual's self-reported alcohol use and their
490 genotype for an alcohol implicated variant, rs1229984 in *ADH1B*, are associated with
491 their partner's self-reported alcohol use. Furthermore, we showed that the genotype
492 of the variant is concordant within spouse-pairs. There are several possible
493 explanations for our findings. First, that rs1229984 influences alcohol behaviour,
494 which has a downstream effect on mate selection. Second, that a participant's
495 alcohol use is influenced by their partner's alcohol use. Third, spouse-pairs with
496 more similar alcohol behaviour were more likely to remain in a relationship, and so
497 be present in our study sample. Fourth, that given the strong association of the SNP
498 with both genetic principal components and birth coordinates, the spousal
499 concordance is related to factors influencing social homogamy, independent of
500 alcohol behaviour, such as place of birth, ancestry or socio-economic status. Indeed,
501 the allele frequency of rs1229984 was found to deviate between European and
502 British subsets of the UK Biobank.

503 However, we presented evidence suggesting that a substantial proportion of
504 the spousal concordance is likely to be explained by the biological effects of the
505 variant on alcohol consumption in the index individual. Firstly, we have tested the
506 association between a causal SNP for alcohol consumption, and not the measured

consumption itself, thereby avoiding any post-birth confounding factors suggesting that alcohol use has a direct effect on spousal alcohol use. Secondly, because rs1229984 is concordant between spouses, there must be some degree of assortment on alcohol consumption prior to cohabitation. Furthermore, we found little evidence to suggest that the mean age of each spouse-pair, used as a proxy for relationship length, was associated with alcohol behaviour similarities. These findings suggest that the spousal concordance is unlikely to be due to relationship dissolution after the age of 40. Thirdly, we accounted for possible effects of ancestral factors, which could have induced confounding, by including principal components as covariates in the Mendelian randomization analysis. Additionally, as a sensitivity analysis, we conducted a within centre sensitivity analysis excluding spouse-pairs born more than 100 miles apart, finding a consistent effect estimate.

The strong evidence for spousal-concordance on the variant has implications for conventional Mendelian randomization studies (i.e. estimating the causal effect of an exposure on an outcome) [33] which use the SNP as a genetic proxy for alcohol intake [45]. Assortative mating could lead to a violation of the Mendelian randomization assumption, that the genetic instrument for the exposure is not strongly associated with confounders of the exposure-outcome relationship. If both genetic and environmental factors affect alcohol consumption, then assortative mating on alcohol consumption could contribute to associations between genetic and environmental factors in the offspring, with the strength of association dependent on the degree of assortative mating [58].

Interestingly, the minor allele of rs1229984 (i.e. associated with lower alcohol consumption) has been previously found to be positively associated with years in

education [45] and socio-economic related variables, such as the Townsend deprivation index and number of vehicles in household [59 60]. Each copy of the minor allele was associated with an additional 0.023 (95% C.I. 0.012 to 0.034, P=0.00005) years of education and a 0.016 S.D. (95% C.I. -0.001 to 0.033, P=0.06) increase in intelligence [61 62]. These associations may be down-stream causal effects of alcohol consumption, which implies that some of the spousal concordance for alcohol consumption could be explained by assortative mating on educational attainment [15] or alternatively these associations may reflect maternal genotype and intrauterine effects [63]. Over time, assortative mating on alcohol consumption may further strengthen the associations between rs1229984 and socio-economic related variables [58]. Of further interest is that the variant has previously been shown to be under selection [64] suggesting that the variant has historically had a substantial effect on reproductive fitness and may partially explain the violation of HWE observed across Europeans in our analyses.

The analyses in this study extended previous work on the concordance between spouse-pairs for alcohol behaviour [7-12] by comparing the phenotypic concordance with analyses utilising a genetic variant strongly associated with alcohol consumption. A major strength of this study is the use of distinct methods with different non-overlapping limitations, allowing for improved inference by triangulating the results from the different methods [65]. First, we evaluated the spousal phenotypic concordance for self-reported alcohol consumption, second we investigated the effect of an individual's rs1229984 genotype on the alcohol consumption of their spouse using Mendelian randomization, third we demonstrated spousal genotypic concordance for rs1229984 and fourth we explored whether older couples have more similar alcohol behaviour. The use of the UK Biobank data-set was a

556 considerable strength for these analyses because of the low frequency of the
557 rs1229984 minor allele; the large scale of the UK Biobank allowed for the
558 identification of thousands of genotyped spouse-pairs. A further strength of these
559 analyses is that we have demonstrated the utility of a Mendelian randomization
560 framework for application to assortative mating by applying it to height and alcohol
561 use. Indeed, the evidence for differences between the observational and Mendelian
562 randomization estimates for spousal height suggest that the observational estimate
563 may be inflated by confounding factors although differences could also be related to
564 the attenuated effects of phenotypic assortment on genetic associations. A similar
565 approach using polygenic risk scores has previously demonstrated assortative
566 mating on educational attainment [18]. However, the use of Mendelian randomization
567 has a notable advantage over polygenic approaches because of the possibility of
568 using various sensitivity analyses to test for heterogeneity and consistency of the
569 effect estimate [50-52].

570 There are several limitations of this study. First, although spouse-pairs were
571 identified using similar methods to previous studies [15-17], the identified spouse-pairs
572 have not been confirmed. However, the phenotypic spousal concordance estimate
573 for height found in this study is highly concordant with previous estimates [56],
574 consistent with derived couples being genuine. Second, despite follow-up analyses,
575 it is difficult to definitively prove that the spousal concordance is a direct result of
576 assortative mating on alcohol consumption. Assortment independent of alcohol use,
577 potentially relating to ancestral or geographical factors, cannot be completely ruled
578 out and down-stream pleiotropic effects of the variant may influence mate selection.
579 Third, the use of a single genetic instrument in the Mendelian randomization
580 analysis, limited the use of sensitivity analyses [50-52] and meant it is not possible to

infer similar associations for other alcohol-implicated variants. Fourth, selection into the UK Biobank, particularly with regards to participation of spouse-pairs is a potential source of bias [66]. Fifth, it is unclear whether the mean age of each couple is a suitable proxy for relationship length, which limits conclusions regarding the possibilities of partner interactions and relationship dissolution. Indeed, patterns of assortment on alcohol behaviour changing over time would confound the use of this proxy. Finally, it is difficult to extrapolate the results of this study in the UK Biobank to non-European populations. This is because of potential contextual influences; for example, in some East Asian populations, males are much more likely to consume alcohol than females[67 68]. Indeed, even within the UK, there may be regional variation that we were unable to detect in this study. Additionally, there is some evidence that the effect of genetic contributors to alcohol varies across different populations [29].

To conclude, our results suggest that there is non-random mating on rs1229984 in *ADH1B*, likely related to the effect of the variant on alcohol behaviour. These results suggest that alcohol use influences mate selection and argue for a more nuanced approach to considering social and cultural factors when examining causality in epidemiological studies. Further research investigating other alcohol-implicated variants, and other societies and ethnicities, and assortment on other phenotypes, would strengthen these conclusions.

Figure titles and descriptions

Figure 1 Possible explanations for spousal concordance on alcohol use.

(A) Assortative mating. Alcohol behaviour influences mate selection; individuals are more likely to select a mate with similar alcohol consumption.

(B) Social homogamy or confounding. An unknown confounder influence mate selection independent of alcohol behaviour. For example, ancestry or socio-economic status may influence both alcohol use and mate choice.

(C) Partner interaction effects. As spouse-pairs cohabitate their alcohol behaviour becomes more similar over time.

(D) Relationship dissolution. Spouse-pairs with more similar alcohol behaviour are more likely to remain in a relationship and be recruited into UK Biobank or similarly, are more likely to participate in the study together.

Figure 2 Interpretations of phenotypic concordance, Mendelian randomization and genotypic concordance analyses between-spouses.

(A) Phenotypic concordance. Spousal concordance for alcohol use could be explained by a direct effect of an individual's alcohol consumption on their partner's alcohol consumption (assortative mating, partner interaction effects or relationship

626 dissolution) or confounding factors such as assortment on social factors (social
627 homogamy) leading to spousal correlation for alcohol use.

628

629 (B) Mendelian randomization framework. An association between an individual's
630 alcohol influencing genotype and their spouse's alcohol use would suggest that the
631 spousal concordance is explained by a direct effect of alcohol consumption. Genetic
632 variants are unlikely to be associated with socio-economic confounders suggesting
633 that social homogamy is unlikely. Spousal phenotype/genotype associations induced
634 by assortment are dependent on the heritability of the trait (see Supplementary
635 Methods).

636

637 (C) Genotypic concordance. Genotypic concordance for alcohol related genetic
638 variants would suggest that some degree of the spousal concordance is explained
639 by assortative mating or relationship dissolution. Partner interaction effects cannot
640 lead to genotypic concordance because genotypes are fixed from birth. Spousal
641 genotypic concordance induced by assortment is dependent on the trait heritability
642 (see Supplementary Methods).

643

644

645 ## Acknowledgements

646 LJH was a Medical Research Council funded PhD student at the University of Bristol
647 and is now funded by the British Heart Foundation and University College London.
648 NMD, SJL and GDS work in the Medical Research Council Integrative Epidemiology

Unit at the University of Bristol (MC_UU_00011/1) which is supported by the Medical Research Council and the University of Bristol. NMD is supported by the Economics and Social Research Council (ESRC) via a Future Research Leaders grant [ES/N000757/1]. DJL [WT104125MA] and GH [208806/|/17/|] are both supported by the Wellcome Trust. UK Biobank received ethical approval from the Research Ethics Committee (11/NW/0383). This research was approved as part of application 15825 (PI: Dr Philip Haycock).

Conflicts of interest

Neil Davies reports a grant for research unrelated to this work from the Global Research Awards for Nicotine Dependence (GRAND), an independent grant making body funded by Pfizer.

Contributions

LJH, GDS, GH and DJL formulated the project outline and analysis plan. LJH performed all statistical analyses and drafted the first manuscript draft under supervision from GDS, GH, SJL, BSP and NMD. All authors contributed to interpretation of results and writing of the final manuscript.

References

1. Vandenberg SG. Assortative mating, or who marries whom? *Behavior Genetics* 1972;2(2-3):127-57.
2. Buss DM. Human mate selection: Opposites are sometimes said to attract, but in fact we are likely to marry someone who is similar to us in almost every variable. *American Scientist* 1985;73(1):47-51.
3. Mare RD. Five decades of educational assortative mating. *American Sociological Review* 1991:15-32.
4. Silventoinen K, Kaprio J, Lahelma E, et al. Assortative mating by body height and BMI: Finnish twins and their spouses. *American Journal of Human Biology* 2003;15(5):620-27.

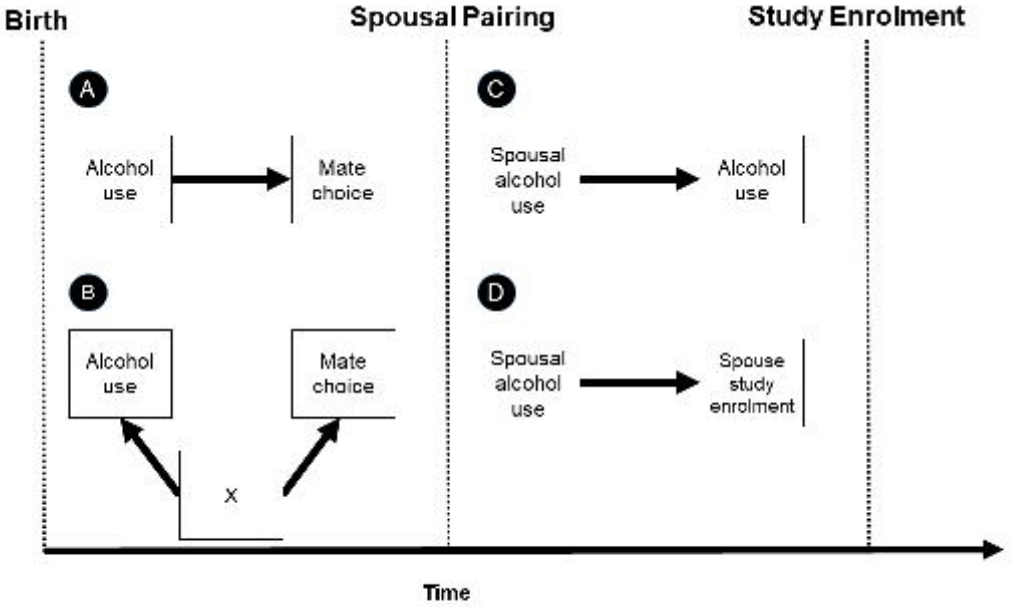
5. Krueger RF, Moffitt TE, Caspi A, et al. Assortative mating for antisocial behavior: Developmental and methodological implications. *Behavior Genetics* 1998;28(3):173-86.
6. Watson D, Klohnen EC, Casillas A, et al. Match makers and deal breakers: Analyses of assortative mating in newlywed couples. *Journal of Personality* 2004;72(5):1029-68.
7. Agrawal A, Heath AC, Grant JD, et al. Assortative mating for cigarette smoking and for alcohol consumption in female Australian twins and their spouses. *Behavior Genetics* 2006;36(4):553-66.
8. Hall RL, Hesselbrock VM, Stabenau JR. Familial distribution of alcohol use: II. Assortative mating of alcoholic probands. *Behavior Genetics* 1983;13(4):373-82.
9. Hall RL, Hesselbrock VM, Stabenau JR. Familial distribution of alcohol use: I. Assortative mating in the parents of alcoholics. *Behavior Genetics* 1983;13(4):361-72.
10. McLeod JD. Spouse concordance for alcohol dependence and heavy drinking: Evidence from a community sample. *Alcoholism: Clinical and Experimental Research* 1993;17(6):1146-55.
11. Reynolds CA, Barlow T, Pedersen NL. Alcohol, tobacco and caffeine use: spouse similarity processes. *Behavior Genetics* 2006;36(2):201.
12. Grant JD, Heath AC, Bucholz KK, et al. Spousal concordance for alcohol dependence: evidence for assortative mating or spousal interaction effects? *Alcoholism: Clinical and Experimental Research* 2007;31(5):717-28.
13. Ask H, Rognmo K, Torvik FA, et al. Non-random mating and convergence over time for alcohol consumption, smoking, and exercise: the Nord-Trøndelag Health Study. *Behavior Genetics* 2012;42(3):354-65.
14. Domingue BW, Fletcher J, Conley D, et al. Genetic and educational assortative mating among US adults. *Proceedings of the National Academy of Sciences* 2014;111(22):7996-8000.
15. Robinson MR, Kleinman A, Graff M, et al. Genetic evidence of assortative mating in humans. *Nature Human Behaviour* 2017;1:0016.
16. Rawlik K, Canela-Xandri O, Tenesa A. Indirect assortative mating for human disease and longevity. *bioRxiv* 2017:185207.
17. Tenesa A, Rawlik K, Navarro P, et al. Genetic determination of height-mediated mate choice. *Genome Biology* 2016;16(1):269.
18. Hugh-Jones D, Verweij KJ, Pourcain BS, et al. Assortative mating on educational attainment leads to genetic spousal resemblance for polygenic scores. *Intelligence* 2016;59:103-08.
19. Verhulst B, Neale MC, Kendler KS. The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. *Psychological Medicine* 2015;45(5):1061-72.
20. Walters GD. The heritability of alcohol abuse and dependence: a meta-analysis of behavior genetic research. *The American Journal of Drug and Alcohol Abuse* 2002;28(3):557-84.
21. Clarke T-K, Adams MJ, Davies G, et al. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N= 112 117). *Molecular Psychiatry* 2017;22(10):1376.
22. Gelernter J, Kranzler H, Sherva R, et al. Genome-wide association study of alcohol dependence: significant findings in African-and European-Americans including novel risk loci. *Molecular Psychiatry* 2014;19(1):41.

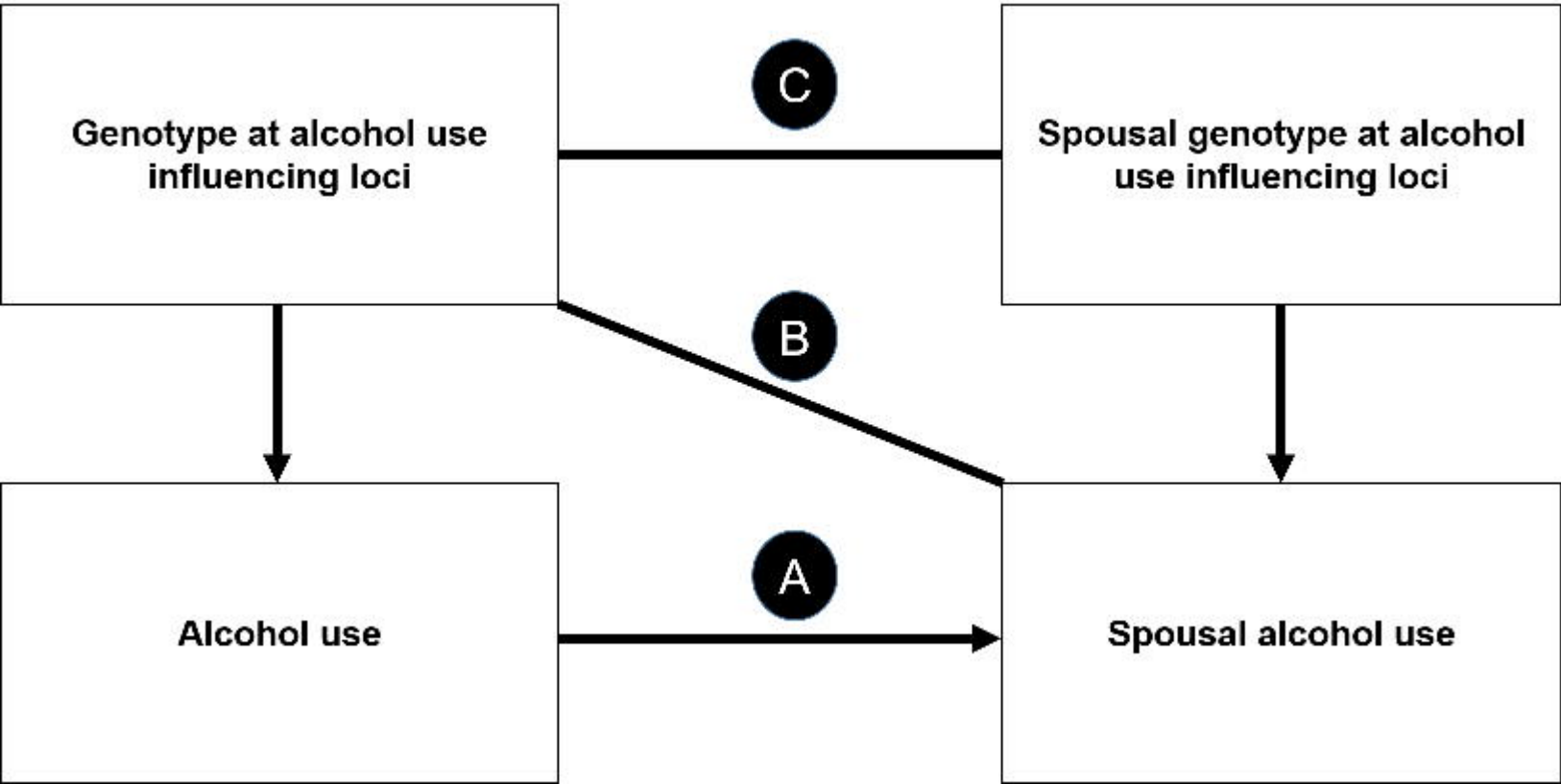
23. Park BL, Kim JW, Cheong HS, et al. Extended genetic effects of ADH cluster genes on the risk of alcohol dependence: from GWAS to replication. *Human Genetics* 2013;132(6):657-68.
24. Bierut LJ, Goate AM, Breslau N, et al. ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. *Molecular Psychiatry* 2012;17(4):445.
25. Bierut LJ, Agrawal A, Bucholz KK, et al. A genome-wide association study of alcohol dependence. *Proceedings of the National Academy of Sciences* 2010;107(11):5082-87.
26. Treutlein J, Cichon S, Ridinger M, et al. Genome-wide association study of alcohol dependence. *Archives of General Psychiatry* 2009;66(7):773-84.
27. Schumann G, Coin LJ, Lourdusamy A, et al. Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. *Proceedings of the National Academy of Sciences* 2011;108(17):7119-24.
28. Schumann G, Liu C, O'Reilly P, et al. KLB is associated with alcohol drinking, and its gene product β-Klotho is necessary for FGF21 regulation of alcohol preference. *Proceedings of the National Academy of Sciences* 2016;113(50):14372-77.
29. Jorgenson E, Thai KK, Hoffmann TJ, et al. Genetic contributors to variation in alcohol consumption vary by race/ethnicity in a large multi-ethnic genome-wide association study. *Molecular Psychiatry* 2017;22(9):1359.
30. Edenberg HJ, McClintick JN. Alcohol Dehydrogenases, Aldehyde Dehydrogenases, and Alcohol Use Disorders: A Critical Review. *Alcoholism: Clinical and Experimental Research* 2018;42(12):2281-97.
31. Thomasson HR, Edenberg HJ, Crabb DW, et al. Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. *American Journal of Human Genetics* 1991;48(4):677.
32. Luczak SE, Glatt SJ, Wall TJ. Meta-analyses of ALDH2 and ADH1B with alcohol dependence in Asians: American Psychological Association, 2006.
33. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* 2003;32(1):1-22.
34. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human molecular genetics* 2014;23(R1):R89-R98.
35. Davey Smith G, Lawlor DA, Harbord R, et al. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Medicine* 2007;4(12):e352.
36. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine* 2015;12(3):e1001779.
37. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;562(7726):203.
38. Mitchell RE, Hemani G, Dudding T, et al. UK Biobank Genetic Data: MRC-IEU Quality Control, version 1, 13/11/2017 2017 [
39. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010;467(7311):52-58.
40. Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015;526(7571):68-74.

41. Mountjoy E, Davies NM, Plotnikov D, et al. Education and myopia: assessing the direction of causality by mendelian randomisation. *BMJ* 2018;361:k2022.
42. O'Connell J, Sharp K, Shrine N, et al. Haplotype estimation for biobank-scale data sets. *Nature Genetics* 2016;48(7):817-20.
43. Consortium UK. The UK10K project identifies rare variants in health and disease. *Nature* 2015;526(7571):82-90.
44. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics* 2016;48(10):1279.
45. Holmes MV, Dale CE, Zuccolo L, et al. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *BMJ* 2014;349:g4164.
46. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* 2018;7:e34408.
47. Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* 2014;46(11):1173.
48. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 2007;81(3):559-75.
49. Greco M, Del F, Minelli C, et al. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Statistics in Medicine* 2015;34(21):2926-40.
50. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* 2015;44(2):512-25.
51. Bowden J, Davey Smith G, Haycock PC, et al. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic Epidemiology* 2016;40(4):304-14.
52. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *International Journal of Epidemiology* 2017;46(6):1985-98.
53. Zuccolo L, Lewis SJ, Davey Smith G, et al. Prenatal alcohol exposure and offspring cognition and school performance. A 'Mendelian randomization'natural experiment. *International Journal of Epidemiology* 2013;42(5):1358-70.
54. Leslie S, Winney B, Hellenthal G, et al. The fine-scale genetic structure of the British population. *Nature* 2015;519(7543):309.
55. Haworth S, Mitchell R, Corbin L, et al. Common genetic variants and health outcomes appear geographically structured in the UK Biobank sample: Old concerns returning and their implications. *bioRxiv* 2018:294876.
56. Price RA, Vandenberg SG. Spouse similarity in American and Swedish couples. *Behavior Genetics* 1980;10(1):59-71.
57. Pearson K, Lee A. On the laws of inheritance in man: I. Inheritance of physical characters. *Biometrika* 1903;2(4):357-462.
58. Hartwig FP, Davies NM, Davey Smith G. Bias in Mendelian randomization due to assortative mating. *Genetic epidemiology* 2018
59. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nature Genetics* 2018;50(11):1593.

827 60. Howe LJ, Sharp GC, Hemani G, et al. Prenatal alcohol exposure and facial
828 morphology in a UK cohort. *Drug and Alcohol Dependence* 2019
829 61. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from
830 a genome-wide association study of educational attainment in 1.1 million
831 individuals. *Nature Genetics* 2018;50(8):1112.
832 62. Hill W, Marioni R, Maghzian O, et al. A combined analysis of genetically
833 correlated traits identifies 187 loci and a role for neurogenesis and myelination
834 in intelligence. *Molecular Psychiatry* 2018:1.
835 63. Hemani G, Bowden J, Haycock PC, et al. Automating Mendelian randomization
836 through machine learning to construct a putative causal map of the human
837 phenome. *bioRxiv* 2017:173682.
838 64. Galinsky KJ, Bhatia G, Loh P-R, et al. Fast principal-component analysis reveals
839 convergent evolution of ADH1B in Europe and East Asia. *The American*
840 *Journal of Human Genetics* 2016;98(3):456-72.
841 65. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology.
842 *International Journal of Epidemiology* 2016;45(6):1866-86.
843 66. Munafò MR, Tilling K, Taylor AE, et al. Collider scope: when selection bias can
844 substantially influence observed associations. *International Journal of*
845 *Epidemiology* 2017;47(1):226-35.
846 67. Spiller W, Slichter D, Bowden J, et al. Detecting and correcting for bias in
847 Mendelian randomization analyses using Gene-by-Environment interactions.
848 *International Journal of Epidemiology* 2018:187849.
849 68. Chen L, Davey Smith G, Harbord RM, et al. Alcohol intake and blood pressure: a
850 systematic review implementing a Mendelian randomization approach. *PLoS*
851 *Medicine* 2008;5(3):e52.

852


Birth
Spousal Pairing
Study Enrolment
A
Alcohol use
Mate choice
B
Alcohol use
Mate choice
X
C
Spousal alcohol use
Alcohol use
D
Spousal alcohol use
Spouse study enrolment
Time


Genotype at alcohol use influencing loci
C
Spousal genotype at alcohol use influencing loci
B
Alcohol use
A
Spousal alcohol use