



Full length article

Investigating the genetic and causal relationship between initiation or use of alcohol, caffeine, cannabis and nicotine

Lun-Hsien Chang^{a,*}, Jue-Sheng Ong^{b,*}, Jiyuan An^b, Karin J.H. Verweij^c, Jacqueline M. Vink^d, Joëlle Pasman^d, Mengzhen Liu^e, Stuart MacGregor^b, Marilyn C. Cornelis^f, Nicholas G. Martin^a, Eske M. Derks^{g,*}

^a Genetic Epidemiology, QIMR Berghofer Medical Research Institute, 300 Herston Road, Brisbane, QLD, 4006, Australia

^b Statistical Genetics, QIMR Berghofer Medical Research Institute, 300 Herston Road, Brisbane, QLD, 4006, Australia

^c Department of Psychiatry, Amsterdam UMC Location AMC, University of Amsterdam, Meibergdreef 5, 1105 AZ, Amsterdam, the Netherlands

^d Behavioural Science Institute, Developmental Psychopathology, Radboud University, Postbus 9104 6500 HE Nijmegen, the Netherlands

^e Institute for Behavioural Genetics, University of Colorado, Boulder, CO, 80309-0447, United States

^f Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, 680 N Lake Shore Dr Suite 1400, Chicago, IL, 60611, United States

^g Translational Neurogenomics, QIMR Berghofer Medical Research Institute, 300 Herston Road, Brisbane, QLD 4006, Australia

ARTICLE INFO

Keywords

Substance use

Genetic instruments

Tobacco

Coffee

Tea

Single nucleotide polymorphism

Two-sample Mendelian randomisation

ABSTRACT

Background: Caffeine, alcohol, nicotine and cannabis are commonly used psychoactive substances. While the use of these substances has been previously shown to be genetically correlated, causality between these substance use traits remains unclear. We aimed to revisit the genetic relationships among different measures of SU using genome-wide association study (GWAS) summary statistics from the UK Biobank, International Cannabis Consortium, and GWAS & Sequencing Consortium of Alcohol and Nicotine use.

Methods: We obtained GWAS summary statistics from the aforementioned consortia for ten substance use traits including various measures of alcohol consumption, caffeine consumption, cannabis initiation and smoking behaviours. We then conducted SNP-heritability (h^2) estimation for individual SU traits, followed by genetic correlation analyses and two-sample Mendelian randomisation (MR) studies between substance use trait pairs.

Results: SNP h^2 of the ten traits ranged from 0.03 to 0.11. After multiple testing correction, 29 of the 45 trait pairs showed evidence of being genetically correlated. MR analyses revealed that most SU traits were not causally associated with each other. However, we found evidence for an MR association between regular smoking initiation and caffeine consumption (40.17 mg; 95 % CI: [24.01, 56.33] increase in caffeine intake per doubling of odds in smoking initiation). Our findings were robust against horizontal pleiotropy, SNP-outliers, and the direction of causality was consistent in all MR analyses.

Conclusions: Most of the substance traits were genetically correlated but there is little evidence to establish causality apart from the relationship between smoking initiation and caffeine consumption.

1. Introduction

1.1. Observational associations between use of substances

Caffeine, alcohol, nicotine and cannabis are among the most commonly used psychoactive substances worldwide (Gowing et al., 2015). Observational studies frequently associate the use of two or more of these substances. High phenotypic correlations are found between the use of caffeine, alcohol, nicotine and cannabis (Kendler et al., 2008) during adolescence. Smoking initiation is positively associated with coffee consumption (Swanson et al., 1994; Treur et al., 2016). Cannabis use is associated with increased alcohol consumption (Wardell et al., 2018), but also with less success in quitting tobacco smoking (Vogel et al., 2018).

1.2. Genetic associations between use of substances

Genetic influences play an important role in the use of multiple substances. Twin studies reported a wide range of heritability estimates of substance use phenotypes: 0.39–0.51 for coffee consumption (Luciano et al., 2005; Vink et al., 2009), 0.44–0.48 for the initiation of cannabis use (Verweij et al., 2010; Vink et al., 2010), -0.75–0.44 for the initiation of smoking (Chang et al., 2018; Maes et al., 2004; Vink et al., 2005), and 0.24–0.53 for a range of alcohol consumption and alcohol use disorder phenotypes (Chang et al., 2018; Grant et al., 2009; Sartor et al., 2009; van Beek et al., 2012).

Recent genetic association studies based on the variation in single nucleotide polymorphisms (SNPs) provide the support that the genes underlying

* Corresponding authors.

E-mail addresses: lun-hsien.chang@qimrberghofer.edu.au (L-H Chang); JueSheng.Ong@qimrberghofer.edu.au (J-S Ong); jiyuan.an@qimrberghofer.edu.au (J. An); karin.verweij@amsterdamumc.nl (K.J.H. Verweij); J.Vink@pwo.ru.nl (J.M. Vink); j.pasman@bsi.ru.nl (J. Pasman); mengzhen.liu@colorado.edu (M. Liu); Stuart.MacGregor@qimrberghofer.edu.au (S. MacGregor); marilyn.cornelis@northwestern.edu (M.C. Cornelis); Nick.Martin@qimrberghofer.edu.au (N.G. Martin); Eske.Derks@qimrberghofer.edu.au (E.M. Derks)

ing use of one substance are overlapping with those underlying use of other substances. For instance, coffee consumption (Coffee and Caffeine Genetics Consortium et al., 2015) is genetically correlated with various dimensions of cigarette smoking, including smoking initiation (rG: 0.28), persistence (rG: 0.25) and number of cigarettes per day (rG: 0.44) (Nivard et al., 2016; Treur et al., 2017). Number of drinks consumed per week is correlated with different stages of smoking, such as initiation (rG: 0.40 (Chang et al., 2019)), cigarettes per day (rG ranging from 0.07 (Liu et al., 2019) to 0.44 (Nivard et al., 2016)), and smoking cessation (rG: 0.10 (Chang et al., 2019)). Lifetime cannabis use is also genetically correlated with smoking initiation, alcohol consumption, and alcohol dependence (Pasman et al., 2018). While these studies were able to demonstrate the common genetic influences underlying substance use, it remains unclear whether the co-occurring conditions are driven by pleiotropy (i.e., the same set of genes influences multiple traits) or causality (i.e., one trait is causally associated with the other).

1.3. Causality in the genetic risk of substance use

Mendelian randomisation (MR) studies typically use genetic variants as instrumental variables (IVs) to estimate the causal effect of an exposure on a related outcome (Davey Smith and Hemani, 2014). Given that variation in substance use behaviour is heritable, we can identify suitable genetic instruments through large scale GWAS to perform causal inferences. MR is subject to three assumptions (Burgess et al., 2015b): (1) the genetic instrument should be reliably associated with the exposure, (2) the instrument should not be associated with (un)measured confounders, and (3) the instrument should not influence the outcome other than through the exposure. Among a variety of MR approaches available in the literature, the two-sample MR approach (Burgess et al., 2015a) is particularly efficient when individual-level data on both the exposure and outcome traits are unavailable in the same set of individuals. Two-sample MR allows higher statistical power to be leveraged across larger genome-wide association study (GWAS) datasets for both the exposure and the outcome.

The intertwining causal relationships between the use of different substances have been examined by a limited number of studies (Bjørngaard et al., 2017; Verweij et al., 2018). While MR analysis causally associated genetic liability of heavier smoking with higher coffee consumption in current smokers (N = 10,951–16,243) (Bjørngaard et al., 2017), a more recent and larger MR study found little evidence for causal associations between alcohol (N = 112,117), caffeine (N = 92,501), cannabis (N = 32,330) and nicotine use (N = 38,181 in cigarettes per day; N = 74,035 in smoking initiation) (Verweij et al., 2018). However, the power in this study to detect subtle effect sizes remained limited.

1.4. Aims

In this study, we aim to revisit the genetic relationships among different measures of substance use using GWAS summary statistics from larger samples which recently became available. After removal of sample overlap (e.g., 23andme and UKB), sample sizes range from N = 25,153 to N = 296,735. We investigated genetic correlations between ten phenotypes representing the initiation or use of alcohol, caffeine, cannabis and nicotine through large scale studies conducted in the UK Biobank (UKB) cohort, International Cannabis Consortium (ICC) (Pasman et al., 2018), and GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN) (Liu et al., 2019). We then investigated the causal associations between some of these traits using two-sample MR.

2. Materials and methods

2.1. Measures

GWAS summary statistics were available for ten traits from three samples, including (1) number of drinks consumed per week (GSCAN),

day (GSCAN, UKB), (9) pack-years of smoking (UKB), and (10) smoking cessation (GSCAN). Note that among the ten traits, we included two similar traits for quantitative use of alcohol (i.e., number of drinks consumed per week, and estimated standard drinks per week) and two identical traits for quantitative smoking measured separately in the GSCAN and UKB for internal validation. We provided the definitions of the ten traits in Table 1. See definitions of substance use phenotypes in the supporting information for full details.

2.2. Analyses

In this study, we estimated SNP heritability (SNP h^2), examined pairwise genetic correlations, and then followed up with a two-sample MR analysis to evaluate evidence for causality. We curated each GWAS dataset and re-

Table 1
Sources (Sample), definitions, types and sample sizes of substance use traits used in genetic associations. Genetic analyses, including SNP-based heritability estimation, genetic correlation, and two-sample Mendelian randomisation, were conducted using GWAS summary statistics in the UKB, ICC and GSCAN samples. UKB: UK Biobank. ICC: International Cannabis Consortium. GSCAN: GWAS & Sequencing Consortium of Alcohol and Nicotine use.

Substance	Sample	Traits	Type	Trait definition	Sample size
nicotine	GSCAN	Age of initiation of regular smoking	continuous	Age of initiating regular smoking	119,239
nicotine	GSCAN	Cigarettes smoked per day	continuous	Cigarettes smoked per day using averaged number or 5 response categories: 1–5, 6–15, 16–25, 26–35, or 36 +	122,027
alcohol	GSCAN	Drinks consumed per week	continuous	Number of drinks consumed per week in ever drinkers	185,828
nicotine	GSCAN	Smoking cessation	binary	Current smokers versus former smokers	125,361
nicotine	GSCAN	Regular smoking initiation	binary	Ever versus never smoked regularly. A regular smoker was defined as having smoked > 100 cigarettes during lifetime, ever having smoked every day for at least one month, or simply ever smoking regularly	207,726
cannabis	ICC	Cannabis initiation	binary	Ever versus never taken cannabis	164,741
caffeine	UKB	Caffeine consumed per day	continuous	Caffeine consumed per day through regular coffee and tea (mg/day; UKB Data-Field: 1488, 1498, 1508)	168,919
nicotine	UKB	Cigarettes smoked per day	continuous	Number of cigarettes currently smoked daily (UKB Data-Field: 3456)	25,153
alcohol	UKB	Estimated standard drinks per week	continuous	Number of standard drinks consumed per week based on alcohol intake	296,735

moved all sample overlap by removing overlapping individual studies to avoid biased causal estimates from the MR analysis. The details of these analyses are described below.

2.2.1. Genome-wide association studies (GWAS)

The UKB cohort was the largest contributor to both the GSCAN and ICC consortia. To avoid potential inflation of causal estimates in the two-sample MR framework due to sample overlap, we obtained revised GWAS estimates for substance use phenotypes reported in the GSCAN and ICC studies upon removal of UKB and 23andMe participants. Firstly, the five GSCAN meta-analytic GWAS from previous work (Liu et al., 2019) was repeated after removing all contributing UKB participants, the 23andMe participants, and participants from studies that overlapped the ICC sample. Next, GWAS for cannabis initiation was re-conducted after removing the 23andMe participants in the ICC cohort. Because the ICC contained fewer individuals than GSCAN and UKB, we prioritised to retain as many cases from the ICC. Thirdly, we identified 153,501 UKB participants who responded to the survey concerning their lifetime use of cannabis and excluded these participants from our UKB GWASs on four continuous traits, including cigarettes per day, pack-years of smoking, estimated standard drinks per week, and caffeine consumed per day. The UKB GWAS was conducted using a Bayesian linear mixed model implemented via the BOLT-LMM software (version 2.3) (Loh et al., 2015), which takes cryptic relatedness between participants into account. Covariates included age at recruitment, sex, and the first ten ancestral principal components (PCs). SNPs with ambiguous strands and/or a minor allele frequency (MAF) < 0.01 were excluded. Our leave-one-study-out approach in deriving the genetic summary statistics allowed two-sample MR analyses to be conducted without worrying about potential bias due to sample overlap. All individuals in the three independent samples ICC, GSCAN and UKB were of European ancestry.

2.2.2. Heritability estimation and genetic correlations

SNP- h^2 was estimated for the ten traits using LD-Score Regression (LDSC) (B. K. Bulik-Sullivan et al., 2015), which is based on GWAS summary statistics. Bivariate genetic correlations (Bulik-Sullivan et al., 2015) were then calculated using the same software between all possible pairs of traits, giving a total of 45 combinations of tests. Our pairwise genetic correlation results were adjusted using Bonferroni correction with the threshold 0.001 (0.05/45). More details on the SNP- h^2 estimation and the calculation of genetic correlations are provided in the supporting information (See procedure for heritability estimation and genetic correlation analyses).

2.2.3. Mendelian randomization

Mendelian randomisation (MR) analyses rely on the use of genetic instruments to estimate the causal effect of an exposure on a related outcome. The two-sample MR method (Burgess et al., 2015a) typically uses data from GWASs that are conducted in two non-overlapping samples of the same ancestry. This method is efficient because individual-level data are not required, yet GWAS summary results are increasingly available from larger studies.

We selected genetic instruments from GWAS data on cannabis initiation, regular smoking initiation, estimated standard drinks per week and caffeine consumption (as exposures). Traits that had their GWAS performed on stratified samples (i.e. traits that can only be assessed in a subset of the population, such as number of cigarettes per day in smokers) were not considered in the analysis due to the potential presence of collider bias when interpreting these findings unless the number of non-users is negligibly low. A more detailed explanation was provided in the supporting information (See the section avoiding collider bias in Mendelian randomisation findings).

For the MR analyses, we used linkage disequilibrium (LD) based clumping to obtain a set of independent (genome-wide) significant SNPs in each dataset using the following criteria: (1) LD R-squared between instruments < 0.01 (2) minimal distance of 10,000 kb. The clumping procedure was performed in the software package PLINK (version: 1.90 beta) (Chang et al., 2015). Following the previously reported procedure (Verweij et al., 2018), we defined our MR instruments on each of the four exposures at two association criteria: (1) one set of

8 and (2) the other at a less stringent threshold of p-value < 1e-5. MR analyses were conducted for each exposure-outcome pair using both sets of instruments.

For the first set of the instruments, we calculated the proportion of phenotypic variance explained by the instruments (R^2) for each of the four exposure traits (i.e., regular smoking initiation, cannabis initiation, number of standard drinks per week, and caffeine consumption). The R^2 of our instruments ranged from 0.004 to 0.02 (Regular smoking initiation: 8 SNPs, R^2 = 0.004; Cannabis initiation: 4 SNPs, R^2 = 0.02; Caffeine consumption: 22 SNPs, R^2 = 0.011; Estimated standard drinks per week: 37 SNPs, R^2 = 0.008; See Table S3). See the section “strength of genetic instruments” in the supporting information for the calculation of the R^2 .

We conducted two-sample MR analysis using the R package TwoSampleMR curated from the MR-Base platform (Hemani et al., 2018). First, we harmonised each pair of clumped exposure and raw outcome GWAS, which ensures that the effect of an SNP on the exposure trait and the effect of that SNP on the outcome trait correspond to the same allele. Next, the causal associations were then evaluated using the fixed effect inverse-variance weighted (IVW) estimator (Burgess et al., 2013) from an exposure to an outcome, and in the other direction if reverse causality could possibly exist. We tested the association in a total of 23 pairs of exposure and outcome traits. Multiple testing correction was applied using Bonferroni correction with a threshold of 0.002 (0.05/23). For binary exposures, we presented the causal estimates to reflect an average change in the outcome per doubling of odds in the prevalence of the exposure.

2.2.4. Sensitivity analyses

We reported our main MR findings using estimates derived from the IVW estimator. Subsequently, we conducted different sensitivity analyses, with different underlying assumptions, to examine how robust the IVW estimates are. These alternative MR methods make weaker assumptions on instrument validity and horizontal pleiotropy than the IVW method which in turn allows further triangulation of causality (Burgess and Thompson, 2017; Verbanck et al., 2018). In brief, for each exposure-outcome pair, we conducted heterogeneity tests, MR-Egger regression, weighted median, and weighted mode in addition to the IVW estimator. Cochran's Q statistics were derived from the IVW or MR-Egger estimates to detect the presence of heterogeneous causal effects among the genetic instruments, an indication of potential violations of the third assumption (i.e. horizontal pleiotropy) (Greco et al., 2015). MR-Egger regression (Bowden et al., 2015, 2016b) relaxes the third assumption of the IVW analysis that the average pleiotropic effect is zero by allowing a non-zero intercept. A non-zero intercept means that the IVW estimate is biased (Burgess and Thompson, 2017). The weighted median method (Bowden et al., 2016a) allows up to 50 % of the instruments to be invalid. The weighted mode method relaxes the third assumption in a different manner from the MR-Egger regression and weighted median and is recommended to be used in combination with other sensitivity analyses (Hartwig et al., 2017).

To evaluate bias due to heterogeneity of causal effect among instruments, we further applied two additional analyses: the leave-one-out MR analysis (Burgess et al., 2017), and the Mendelian randomization pleiotropy residual sum and outlier (MR-PRESSO; Verbanck et al., 2018), where the lack of presence of heterogeneity would suggest that MR estimates are unlikely to be driven by SNP outliers. The rationale behind each of these six sensitivity analyses are further elaborated in the supporting information (see the section Procedure for two-sample Mendelian randomisation analyses).

Scaling differences between the exposure and outcomes used in the MR analyses were removed (See the section Effect estimate conversion in the supporting information for details).

3. Results

3.1. SNP heritability and genetic correlations

The ten substance use traits show strong evidence of being heritable, with SNP h^2 estimates ranging from 0.033 to 0.109 (Table

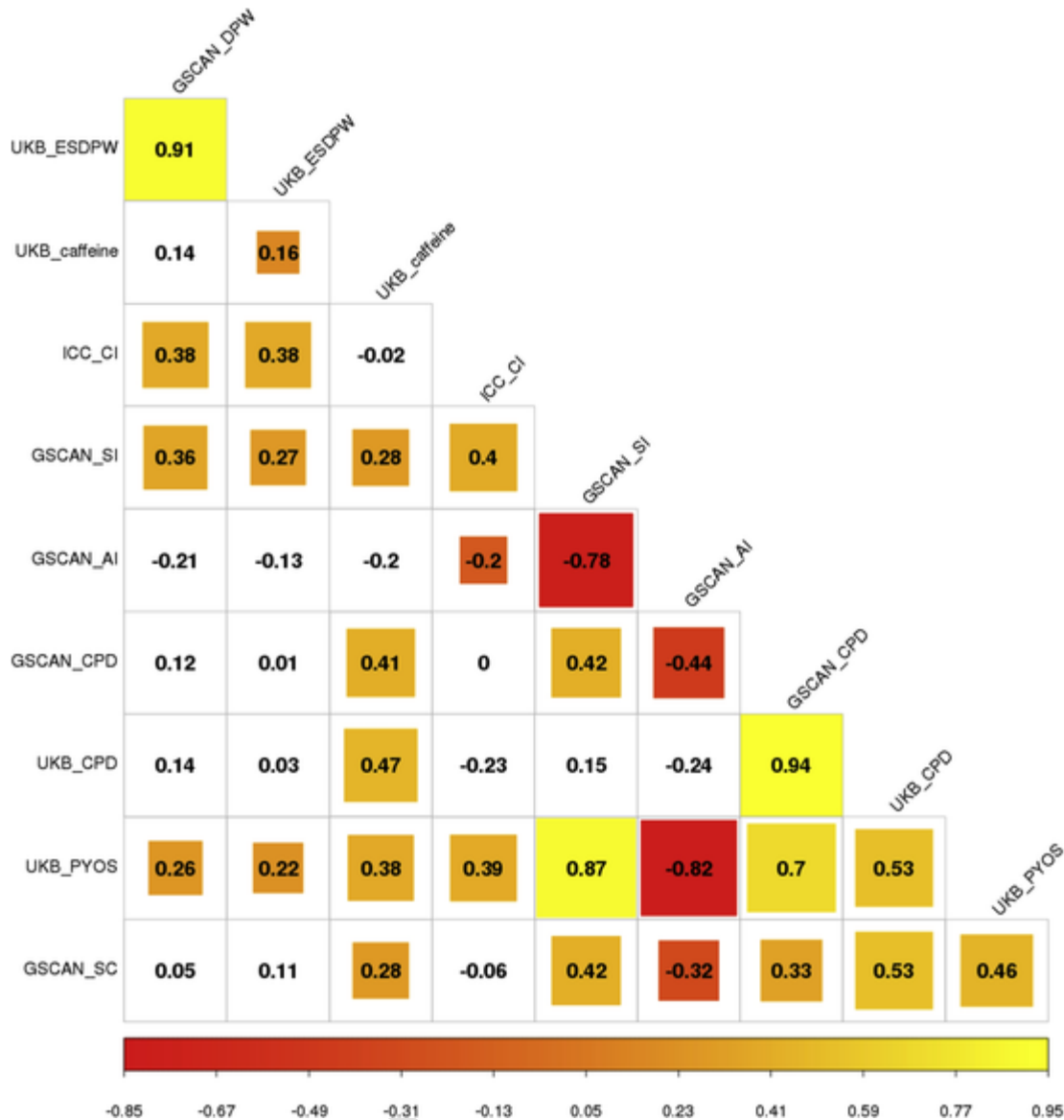


Fig. 1. Pair-wise genetic correlation coefficients between ten GWAS on substance use or initiation from three samples. Sizes of coloured squares are in proportion to the magnitude of the correlations that remained significant after the correction for multiple testing using Bonferroni correction. Correlations that did not survive multiple testing are shown in white background. Abbreviated sample names: (1) GSCAN: GWAS & Sequencing Consortium of Alcohol and Nicotine use, (2) ICC: International Cannabis Consortium, (3) UKB: UK Biobank. Abbreviated trait names: (1) DPW: Drinks consumed per week, (2) ESDPW: Estimated standard drinks per week, (3) caffeine: Caffeine consumed per day, (4) CI: Cannabis initiation, (5) SI: Regular smoking initiation, (6) AI: Age at initiation of regular smoking, (7) CPD: Cigarettes smoked per day, (8) PYOS: Pack years of smoking, and (9) SC: Smoking cessation.

from -0.821 to -0.196. Negative coefficients were only seen in the correlations between age of initiation of regular smoking and the other nine traits (see Fig. 1).

3.2. Two-sample Mendelian randomisation analyses and sensitivity analysis

For each exposure investigated, we adopted two sets of instruments: (i) instruments including SNPs with genome-wide significance (SNP-exposure) to control for weak instrument bias; (ii) a less stringent set of instruments defined at SNP-exposure p-value < 1e-5 to leverage better trait prediction using more SNPs.

Based on the MR associations, regular smoking initiation and caffeine consumption were the only exposure-outcome pair that showed evidence for a causal association across both instrument thresholds (Table S4). The effect estimates were broadly similar across other MR models (Table S4), with minimal evidence of bias due to SNP heterogeneity (MR-PRESSO global test: Table S5; leave-one-out analysis: Table S6) or directional pleiotropy (MR Egger intercept p-value 0.85), ensuring robustness of the IVW finding for regular smoking ini-

tiation on caffeine consumption (Table 2). In our reverse direction of MR analysis, we did not observe evidence for a causal effect of caffeine consumption on regular smoking initiation (Table S4).

To benchmark the MR results, we performed phenotypic associations. See the section Phenotypic associations and definitions of some covariates used in phenotypic associations in the supporting information for details and Table S7 for results.

4. Discussion

In this study, we comprehensively evaluated the genetic relationships between the use or initiation of ten traits measuring different aspects of use for four commonly used substances alcohol, caffeine, cannabis and nicotine. Our pairwise genetic correlation results showed significant genetic overlap between most of the trait pairs (29 out of the 45). We further investigated the causal associations between regular smoking initiation, cannabis initiation, caffeine consumption, or the number of standard drinks per week (as exposures) and the other substance use traits (as outcomes) using various two-sample MR methods and sensitivity analyses. We found evidence for a positive causal effect of regul-

Table 2

Two-sample Mendelian randomisation results on the causal associations between the initiation of regular smoking and caffeine consumption. Independent SNPs associated with exposure trait were clumped with LD window 10,000 kb, R-squared 0.01. n SNPs: Number of independent SNPs used in each exposure-outcome pair. p1: Significance threshold for index SNPs. Heterogeneity tests were available from MR-Egger and inverse variance weighted MR (IVW). Raw estimates (β , SE) are the estimates from the MR estimators. These estimates were converted to interpretable scales as described in the method (Effect size [95 %CI]). Abbreviated consortia: (1) GSCAN: GWAS & Sequencing Consortium of Alcohol and Nicotine use, (2) UKB: UK Biobank. The significance threshold was 0.002173913 (0.05/23). Details of multiple testing correction was provided in the method section. MR analyses that survived heterogeneity tests (i.e. Q p value \geq 0.05) and multiple testing are shown in blue and boldface. Full results for all the exposure and outcome traits are shown in Table S4.

Exposure GWAS				Outcome GWAS		Two-sample MR analysis			Converted Estimates	
No.	Sample	Trait	p1	Sample	Trait	n SNPs	Method	Q p value	Effect size [95 %CI]	p value
1	GSCAN	Regular smoking initiation	1.0E-05	UKB	Caffeine consumed per day	67	IVW	1.1E-01	15.28 [8.389, 22.16]	1.38E-05
2	GSCAN	Regular smoking initiation	1.0E-05	UKB	Caffeine consumed per day	67	Egger	9.5E-02	9.686 [-18.4, 37.74]	5.01E-01
3	GSCAN	Regular smoking initiation	1.0E-05	UKB	Caffeine consumed per day	67	W Median		14.12 [5.076, 23.16]	2.21E-03
4	GSCAN	Regular smoking initiation	1.0E-05	UKB	Caffeine consumed per day	67	W Mode		12.33 [-9.36, 34.02]	2.69E-01
5	GSCAN	Regular smoking initiation	5.0E-08	UKB	Caffeine consumed per day	7	IVW	8.9E-01	40.17 [24.01, 56.33]	1.10E-06
6	GSCAN	Regular smoking initiation	5.0E-08	UKB	Caffeine consumed per day	7	Egger	8.1E-01	57.69 [-111, 226.2]	5.32E-01
7	GSCAN	Regular smoking initiation	5.0E-08	UKB	Caffeine consumed per day	7	W Median		38.67 [18.27, 59.07]	2.03E-04
8	GSCAN	Regular smoking initiation	5.0E-08	UKB	Caffeine consumed per day	7	W Mode		37.92 [7.917, 67.92]	4.80E-02

lar smoking initiation on caffeine consumption, with two-fold increased prevalence of smoking initiation associated with increased consumption of caffeine up to 40.17 mg, an equivalent to the caffeine in about half a cup of coffee. Although this effect size might not reach the magnitude to be of high clinical relevance, nonetheless our findings attempt to provide an explanation to answer whether the observed links are a product of confounding or reflect a true causal nature.

4.1. Genetic correlations

The observed pattern of genetic correlations was consistent with previous studies on smaller GWAS samples (Nivard et al., 2016; Treur et al., 2017). Substantial genetic overlap was detected between pack-years of smoking and each of the other nine traits. To our knowledge, past studies have not reported LD-based genetic correlation between this trait and alcohol, caffeine or cannabis use. Similarly, regular smoking initiation was correlated with almost every other trait, except for cigarettes per day in the UKB sample. However, regular smoking initiation was genetically correlated with cigarettes per day in GSCAN. These inconsistent correlations were likely due to the sample size difference between the UKB (N = 25,153) and GSCAN (N = 122,027).

We found moderate genetic correlations between regular smoking initiation and cannabis initiation (rG: 0.4), and between GSCAN cigarettes per day and caffeine consumed per day (rG: 0.41), but no significant genetic correlation between cannabis initiation and caffeine consumption (rG: -0.02; Table S2). Likewise, Nivard et al. (2016) observed a high genetic correlation between cannabis initiation and smoking initiation (rG: 0.83; SE: 0.15), between cigarettes per day and coffee consumption (rG: 0.44; SE: 0.17), but limited evidence for a genetic overlap between coffee consumption and cannabis initiation (Nivard et al., 2016). Treur et al., 2017 reported an rG of 0.28 between smoking initiation and coffee use (Treur et al., 2017). The pattern in our genetic correlations was similar to those studies, but our correlation coefficients were slightly lower than their estimates. We detected a modest genetic correlation between regular smoking initiation and cannabis initiation; however, this result would require further validation. Moreover, our analyses revealed substantial overlap in the genetic architecture between alcohol consumption and smoking-related traits, such as regular smoking initiation, and pack-years of smoking.

4.2. Genetic evidence for a causal association between liability to smoking initiation and caffeine consumption

Although we observed an abundance of genetic correlations between regular smoking initiation and several other substance use traits, we

with a doubling in odds of regular smoking initiation associated with increased consumption of caffeine up to 40.17 mg. While strong evidence for this causality was only seen in the weighted median model (Table S4) but not in the other two sensitivity analyses MR-Egger regression (Table S4) and weighted mode (Table S4), the direction of causality was similar across all the MR estimators. Our results yield similar conclusions to findings from Bjørngaard et al. (2017) where the authors reported that each additional cigarette consumed per day was associated with 0.1 more cups of coffee intake (Bjørngaard et al., 2017).

It is important to note that our genetic causality analyses do not warrant a definitive conclusion on (non-)causality due to existing shortfalls of our study design. Firstly, genetic correlation analyses conducted using the LD-score regression technique assumes a strong degree of polygenicity (such that LD-scores are informative) for each of the trait evaluated. Whilst our MR-approach complements the LD-score restriction on polygenicity as long as instruments explain sufficient phenotypic variance, MR analyses typically require large sample sizes to enable an adequately powered analysis. Finally, causal estimates derived via MR assumes a lifelong predisposition of the exposure. These findings are not necessarily compatible with results from intervention studies (such as RCTs) that evaluate a temporal change in substance use on the outcome of interest. In lieu of these limitations, the wide CIs around some of the MR findings would remain consistent with a clinically relevant causal effect.

It remains unclear which biological mechanism underlies the positive causal association between regular smoking initiation and caffeine consumption. One possible biological explanation pertains to the cytochrome P450 1A2 (CYP1A2) enzyme that is induced by tobacco smoking (Pelkonen et al., 2008). Caffeine is primarily metabolised by the CYP1A2 enzyme (Pelkonen et al., 2008), which can be induced by cigarette smoking via the aryl hydrocarbon receptor (AHR; Butler et al., 1989). Smoking accelerates caffeine metabolism (Joeres et al., 1988; Langmann et al., 2000) and ultimately increases caffeine consuming behaviour. While the pleiotropic function of CYP1A2 might explain the postulated link between smoking and caffeine intake, none of our instruments for smoking was located in or nearby CYP1A2 nor AHR, making this explanation unlikely. Findings from recent meta-analyses of GWAS have shown that substance use traits are highly polygenic (Minica et al., 2018; Pasman et al., 2018; Prom-Wormley et al., 2017; Walters et al., 2018), which means the variation of the traits is likely to be influenced by many genetic variants that has yet to reach genome-wide significance, which are plausible candidate for MR instruments. The present study cannot reliably distinguish forms of (e.g. vertical and horizontal) pleiotropy (for instance the effect of CYP1A2 variants on coffee and smoking) via statistical techniques given the limited number of instruments. These efforts ought to be revisited when more instruments

5. Conclusion

Our findings showed that the majority of the substance use phenotypes exerted moderate to strong genetic correlations, supporting the notion of a common genetic influence across these behaviours. However, mendelian randomization only revealed evidence for a causal link between regular smoking initiation and caffeine consumption, with two-fold increased prevalence of smoking initiation associated with increased consumption of caffeine up to 40.17 mg, an amount which is equivalent to the caffeine in about half a cup of coffee. Our findings suggest that the association between most addictive substances are likely explained by a common genetic liability but does not warrant strong evidence for a causal relationship.

Uncited reference

Bulik-Sullivan et al. (2020)

Contributors

LHC performed the analysis and wrote the manuscript; KJV, JMV, JAP, and MZL conducted GWAS; JSO, JA, EMD, KJV, JMV, JAP, SM, MCC and NGM conceived, designed and assisted drafting this manuscript.

Declaration of Competing Interest

The authors certify that they have no commercial associations that might pose a conflict of interest in connection with this article.

Acknowledgements

This work was conducted using the UK Biobank Resource (application number 25331). LHC was supported by an international PhD scholarship from QIMR Berghofer Medical Research Institute (QIMR Berghofer). SM is supported by an Australian NHMRC fellowship and an NHRMC project grant (1123248). KJHV is supported by the Foundation Volksbond Rotterdam. We thank Scott Wood and John Pearson from QIMR Berghofer for IT support.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.drugalcdep.2020.107966>.

References

- Björngaard, J H, Nordestgaard, A T, Taylor, A E, Treur, J L, Gabrielsen, M E, Munafò, M R, Nordestgaard, B G, Åsvold, B O, Romundstad, P, Davey Smith, G, 2017. Heavier smoking increases coffee consumption: findings from a Mendelian randomization analysis. *Int. J. Epidemiol.* 46, 1958–1967. doi:10.1093/ije/dyx147.
- Bowden, J, Davey Smith, G, Burgess, S, 2015. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* 44, 512–525. doi:10.1093/ije/dyv080.
- Bowden, J, Davey Smith, G, Haycock, P C, Burgess, S, 2016. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* 40, 304–314. doi:10.1002/gepi.21965.
- Bowden, J, Del Greco, M F, Minelli, C, Davey Smith, G, Sheehan, N A, Thompson, J R, 2016. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I² statistic. *Int. J. Epidemiol.* 45, 1961–1974. doi:10.1093/ije/dyw220.
- Bulik-Sullivan, B K, Loh, P-R, Finucane, H K, Ripke, S, Yang, J, Consortium, S.W.G. of the P.G., Patterson, N, Daly, M J, Price, A L, Neale, B M, 2015. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 47, 291–295. doi:10.1038/ng.3211.
- Bulik-Sullivan, B, Finucane, H K, Anttila, V, Gusev, A, Day, F R, Loh, P-R, 2020. ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan, L, Perry, J.R.B., Patterson, N., Robinson, E.B., Daly, M.J., Price, A.L., Neale, B.M., 2015. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* 47, 1236–1241. doi:10.1038/ng.3406.
- Burgess, S, Thompson, S G, 2017. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* 32, 377–389. doi:10.1007/s10654-017-0255-x.
- Burgess, S, Butterworth, A, Thompson, S G, 2013. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* 37, 658–665. doi:10.1007/s10654-015-0011-z.
- Burgess, S, Small, D S, Thompson, S G, 2015. A review of instrumental variable estimators for Mendelian randomization. *Stat. Methods Med. Res.* 0962280215597579. doi:10.1177/0962280215597579.
- Burgess, S, Bowden, J, Fall, T, Ingelsson, E, Thompson, S G, 2017. Sensitivity analyses for robust causal inference from Mendelian randomization analyses with multiple genetic variants. *Epidemiology* 28, 30–42. doi:10.1097/EDE.0000000000000559.
- Butler, M A, Iwasaki, M, Guengerich, F P, Kadlubar, F F, 1989. Human cytochrome P-450A (P-450IA2), the phenacetin O-deethylase, is primarily responsible for the hepatic 3-demethylation of caffeine and N-oxidation of carcinogenic arylamines. *Proc. Natl. Acad. Sci. U. S. A.* 86, 7696–7700.
- Chang, C C, Chow, C C, Tellier, L C, Vattikuti, S, Purcell, S M, Lee, J J, 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4. doi:10.1186/s13742-015-0047-8.
- Chang, L-H, Couvy-Duchesne, B, Medland, S E, Gillespie, N A, Hickie, I B, Parker, R, Martin, N G, 2018. The genetic relationship between psychological distress, somatic distress, affective disorders, and substance use in young Australian adults: a multivariate twin study. *Twin Res. Hum. Genet.* 1–14. doi:10.1017/thg.2018.33.
- Chang, L-H, Couvy-Duchesne, B, Liu, M, Medland, S E, Verhulst, B, Benotsch, E G, Hickie, I B, Martin, N G, Gillespie, N A, GSCAN Consortium, 2019. Association between polygenic risk for tobacco or alcohol consumption and liability to licit and illicit substance use in young Australian adults. *Drug Alcohol Depend.* 197, 271–279. doi:10.1016/j.drugalcdep.2019.01.015.
- Coffee and Caffeine Genetics Consortium, Cornelis, M C, Byrne, E M, Esko, T, Nalls, M A, et al., 2015. Genome-wide meta-analysis identifies six novel loci associated with habitual coffee consumption. *Mol. Psychiatry* 20, 647–656. doi:10.1038/mp.2014.107.
- Davey Smith, G, Hemani, G, 2014. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* 23, R89–R98. doi:10.1093/hmg/ddu328.
- Gowing, L R, Ali, R L, Allsop, S, Marsden, J, Turf, E E, West, R, Witton, J, 2015. Global statistics on addictive behaviours: 2014 status report. *Addiction* 110, 904–919. doi:10.1111/add.12899.
- Grant, J D, Agrawal, A, Bucholz, K K, Madden, P A F, Pergadia, M L, Nelson, E C, Lynskey, M T, Todd, R D, Todorov, A A, Hansell, N K, Whitfield, J B, Martin, N G, Heath, A C, 2009. Alcohol consumption indices of genetic risk for alcohol dependence. *Biol. Psychiatry* 66, 795–800. doi:10.1016/j.biopsych.2009.05.018.
- Greco, M F D, Minelli, C, Sheehan, N A, Thompson, J R, 2015. Detecting Pleiotropy in Mendelian Randomisation Studies with Summary Data and a Continuous Outcome. p. 2940. doi:10.1002/sim.6522.
- Hartwig, F P, Davey Smith, G, Bowden, J, 2017. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* 46, 1985–1998. doi:10.1093/ije/dyx102.
- Hemani, G, Zheng, J, Elsworth, B, Wade, K H, Haberland, V, Baird, D, Laurin, C, Burgess, S, Bowden, J, Langdon, R, Tan, V Y, Yarmolinsky, J, Shihab, H A, Timpson, N J, Evans, D M, Relton, C, Martin, R M, Davey Smith, G, Gaunt, T R, Haycock, P C, 2018. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 7. doi:10.7554/eLife.34408.
- Joeres, R, Klinker, H, Heusler, H, Epping, J, Zilly, W, Richter, E, 1988. Influence of smoking on caffeine elimination in healthy volunteers and in patients with alcoholic liver cirrhosis. *Hepatology* 8, 575–579. doi:10.1002/hep.1840080323.
- Kendler, K S, Schmitt, E, Aggen, S H, Prescott, C A, 2008. Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. *Arch. Gen. Psychiatry* 65, 674–682. doi:10.1001/archpsyc.65.6.674.
- Langmann, P, Bienert, A, Zilly, M, Väh, T, Richter, E, Klinker, H, 2000. Influence of smoking on cotinine and caffeine plasma levels in patients with alcoholic liver cirrhosis. *Eur. J. Med. Res.* 5, 217–221.
- Liu, M, Jiang, Y, Wedow, R, Li, Y, Brazel, D M, Chen, F, Datta, G, Davila-Velderrain, J, McGuire, D, Tian, C, Zhan, X, Choquet, H, Docherty, A R, Faul, J D, Forster, J R, Fritsche, L G, Gabrielsen, M E, Gordon, S D, Haessler, J, Hottenga, J-J, Huang, H, Jang, S-K, Jansen, P R, Ling, Y, Mägi, R, Matoba, N, McMahon, G, Mulas, A, Orrù, V, Palviainen, T, Pandit, A, Reginsson, G W, Skogholt, A H, Smith, J A, Taylor, A E, Turman, C, Willemsen, G, Young, H, Young, K A, Zajac, G J M, Zhao, W, Zhou, W, Björnsdóttir, G, Boardman, J D, Boehnke, M, Boomsma, D I, Chen, C, Cucca, F, Davies, G E, Eaton, C B, Ehringer, M A, Esko, T, Fiorillo, E, Gillespie, N A, Gudbjartsson, D F, Haller, T, Harris, K M, Heath, A C, Hewitt, J K, Hickie, I B, Hokanson, J E, Hopfer, C J, Hunter, D J, Iacono, W G, Johnson, E O, Kamatani, Y, Kardia, S L R, Keller, M C, Kellis, M, Kooperberg, C, Kraft, P, Krauter, K S, Laakso, M, Lind, P A, Loukola, A, Lutz, S M, Madden, P A F, Martin, N G, McGue, M, McQueen, M B, Medland, S E, Metspalu, A, Mohlke, K L, Nielsen, J B, Okada, Y, Peters, U, Polderman, T J C, Posthuma, D, Reiner, A P, Rice, J P, Rimm, E, Rose, R J, Runarsdóttir, V, Stallings, M C, Stancáková, A, Stefansson, H, Thai, K K, Tindle, H A, Tyringsson, T, Wall, T L, Weir, D R, Weisner, C, Whitfield, J B, Winsvold, B S, Yin, J, Zuccolo, L, Bierut, L J, Hveem, K, Lee, J J, Munafò, M R, Saccone, N L, Willer, C J, Cornelis, M C, David, S P, Hinds, D A, Jorgenson, E, Kaprio, J, Stitzel, J A, Stefansson, K, Thorgerirsson, T E, Abecasis, G, Liu, D J, Vrieze, S, 2019. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* 1. doi:10.1038/s41588-018-0307-5.
- Loh, P-R, Tucker, G, Bulik-Sullivan, B K, Vilhjálmsson, B J, Finucane, H K, Salem, R M, Chasman, D I, Ridker, P M, Neale, B M, Berger, B, Patterson, N, Price, A L, 2015. Efficient Bayesian mixed model analysis increases association power in large cohorts. *Nat. Genet.* 47, 284–290. doi:10.1038/ng.3190.
- Luciano, M, Kirk, K M, Heath, A C, Martin, N G, 2005. The genetics of tea and coffee drinking and preference for source of caffeine in a large community sample of Australian twins. *Addiction* 100, 1510–1517. doi:10.1111/j.1360-0443.2005.01223.x.
- Maes, H H, Sullivan, P F, Bulik, C M, Neale, M C, Prescott, C A, Eaves, L J, Kendler, K S, 2004. A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use and nicotine dependence. *Psychol. Med.* 34, 1251–1261. doi:10.1017/s0033291704002405.
- Minica, C C, Verweij, K J H, van der Most, P J, Mbarek, H, Bernard, M, van Eijk, K R, Lind, P A, Liu, M Z, Maciejewski, D F, Palviainen, T, Sanchez-Mora, C, Sherva, R, Taylor, M, W

- R K, Abdellaoui, A, Bigdeli, T B, Branje, S J T, Brown, S A, Casas, M, Corley, R P, Smith, G D, Davies, G E, Ehli, E A, Farrer, L, Fedko, I O, Garcia-Martinez, I, Gordon, S D, Hartman, C A, Heath, A C, Hickie, I B, Hickman, M, Hopfer, C J, Hottenga, J J, Kahn, R S, Kaprio, J, Korhonen, T, Kranzler, H R, Krauter, K, van Lier, P A C, Madden, P A F, Medland, S E, Neale, M C, Meeus, W H J, Montgomery, G W, Nolte, I M, Oldehinkel, A J, Pausova, Z, Ramos-Quiroga, J A, Richarte, V, Rose, R J, Shin, J, Stallings, M C, Wall, T L, Ware, J J, Wright, M J, Zhao, H, Koot, H M, Paus, T, Hewitt, J K, Ribases, M, Loukola, A, Boks, M P, Snieder, H, Munafò, M R, Gelernter, J, Boomsma, D I, Martin, N G, Gillespie, N A, Vink, J M, Derks, E M, 2018. Genome-wide association meta-analysis of age at first cannabis use. *Addiction* (Abingdon, England). doi:10.1111/add.14368.
- Nivard, M G, Verweij, K J H, Minică, C C, Treur, J L, International Cannabis Consortium, Vink, J M, Boomsma, D I, 2016. Connecting the dots, genome-wide association studies in substance use. *Mol. Psychiatry* 21, 733–735. doi:10.1038/mp.2016.14.
- Pasman, J A, Verweij, K J H, Gerring, Z, Stringer, S, Sanchez-Roige, S, Treur, J L, Abdellaoui, A, Nivard, M G, Baselmans, B M L, Ong, J-S, Ip, H F, Zee, M D, van der Bartels, M, Day, F R, Fontanillas, P, Elson, S L, Wit, H de, Davis, L K, MacKillop, J, Derringer, J L, Branje, S J T, Hartman, C A, Heath, A C, van Lier, P A C, Madden, P A F, Mägi, R, Meeus, W, Montgomery, G W, Oldehinkel, A J, Pausova, Z, Ramos-Quiroga, J A, Paus, T, Ribases, M, Kaprio, J, Boks, M P M, Bell, J T, Spector, T D, Gelernter, J, Boomsma, D I, Martin, N G, MacGregor, S, Perry, J R B, Palmer, A A, Posthuma, D, Munafò, M R, Gillespie, N A, Derks, E M, Vink, J M, 2018. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat. Neurosci.* 21, 1161–1170. doi:10.1038/s41593-018-0206-1.
- Pelkonen, O, Turpeinen, M, Hakkola, J, Honkakoski, P, Hukkanen, J, Raunio, H, 2008. Inhibition and induction of human cytochrome P450 enzymes: current status. *Arch. Toxicol.* 82, 667–715. doi:10.1007/s00204-008-0332-8.
- Prom-Wormley, E C, Ebejer, J, Dick, D M, Bowers, M S, 2017. The genetic epidemiology of substance use disorder: a review. *Drug Alcohol Depend.* 180, 241–259. doi:10.1016/j.drugalcdep.2017.06.040.
- Sartor, C E, Lynskey, M T, Bucholz, K K, Madden, P A F, Martin, N G, Heath, A C, 2009. Timing of first alcohol use and alcohol dependence: evidence of common genetic influences. *Addiction* 104, 1512–1518. doi:10.1111/j.1360-0443.2009.02648.x.
- Swanson, J A, Lee, J W, Hopp, J W, 1994. Caffeine and nicotine: a review of their joint use and possible interactive effects in tobacco withdrawal. *Addict. Behav.* 19, 229–256.
- Treur, J L, Taylor, A E, Ware, J J, McMahon, G, Hottenga, J, Baselmans, B M L, Willemsen, G, Boomsma, D I, Munafò, M R, Vink, J M, 2016. Associations between smoking and caffeine consumption in two European cohorts. *Addiction* 111, 1059–1068. doi:10.1111/add.13298.
- Treur, J L, Taylor, A E, Ware, J J, Nivard, M G, Neale, M C, McMahon, G, Hottenga, J-J, Baselmans, B M L, Boomsma, D I, Munafò, M R, Vink, J M, 2017. Smoking and caffeine consumption: a genetic analysis of their association. *Addict. Biol.* 22, 1090–1102. doi:10.1111/adb.12391.
- van Beek, J H D A, Kendler, K S, de Moor, M H M, Geels, L M, Bartels, M, Vink, J M, van den Berg, S M, Willemsen, G, Boomsma, D I, 2012. Stable genetic effects on symptoms of alcohol abuse and dependence from adolescence into early adulthood. *Behav. Genet.* 42, 40–56. doi:10.1007/s10519-011-9488-8.
- Verbanck, M, Chen, C-Y, Neale, B, Do, R, 2018. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* 50, 693. doi:10.1038/s41588-018-0099-7.
- Verweij, K J, Zietsch, B P, Lynskey, M T, Medland, S E, Neale, M C, Martin, N G, Boomsma, D I, Vink, J M, 2010. Genetic and environmental influences on cannabis use initiation and problematic use: a meta-analysis of twin studies. *Addiction* 105, 417–430. doi:10.1111/j.1360-0443.2009.02831.x.
- Verweij, K J H, Treur, J L, Vink, J M, 2018. Investigating causal associations between use of nicotine, alcohol, caffeine and cannabis: a two-sample bidirectional Mendelian randomization study. *Addiction*. doi:10.1111/add.14154.
- Vink, J M, Willemsen, G, Boomsma, D I, 2005. Heritability of smoking initiation and nicotine dependence. *Behav. Genet.* 35, 397–406. doi:10.1007/s10519-004-1327-8.
- Vink, J M, Staphorsius, A S, Boomsma, D I, 2009. A genetic analysis of coffee consumption in a sample of Dutch twins. *Twin Res. Hum. Genet.* 12, 127–131. doi:10.1375/twin.12.2.127.
- Vink, J M, Wolters, L M C, Neale, M C, Boomsma, D I, 2010. Heritability of cannabis initiation in Dutch adult twins. *Addict. Behav.* 35, 172–174. doi:10.1016/j.addbeh.2009.09.015.
- Vogel, E A, Rubinstein, M L, Prochaska, J J, Ramo, D E, 2018. Associations between marijuana use and tobacco cessation outcomes in young adults. *J. Subst. Abuse Treat.* 94, 69–73. doi:10.1016/j.jsat.2018.08.010.
- Walters, R K, Polimanti, R, Johnson, E C, McClintick, J N, Adams, M J, Adkins, A E, Aliev, F, Bacanu, S-A, et al., 2018. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* 21, 1656–1669. doi:10.1038/s41593-018-0275-1.
- Wardell, J D, Shuper, P A, Hendershot, C S, 2018. A longitudinal investigation of the association between cannabis use and alcohol use among people living with HIV. *Drug Alcohol Depend.* 193, 7–13. doi:10.1016/j.drugalcdep.2018.08.026.