

# Examining the role of common genetic variants on alcohol, tobacco, cannabis and illicit drug dependence: genetics of vulnerability to drug dependence

Rohan H. C. Palmer<sup>1,2</sup>, Leslie Brick<sup>1</sup>, Nicole R. Nugent<sup>1,2</sup>, L. Cinnamon Bidwell<sup>1,2,3</sup>, John E. McGeary<sup>1,2,4</sup>, Valerie S. Knopik<sup>1,2</sup> & Matthew C. Keller<sup>5,6</sup>

Division of Behavioral Genetics, Department of Psychiatry, Rhode Island Hospital, Providence, RI, USA,<sup>1</sup> Department of Psychiatry and Human Behavior, Alpert Medical School of Brown University, Providence, RI, USA,<sup>2</sup> Center for Alcohol and Addiction Studies, Department of Behavioral and Social Science, Brown University, Providence, RI, USA,<sup>3</sup> Department of Veterans Affairs, Providence VA Medical Center, Providence, RI, USA,<sup>4</sup> Institute for Behavioral Genetics, University of Colorado at Boulder, Boulder, CO, USA<sup>5</sup> and Department of Psychology and Neuroscience, University of Colorado at Boulder, Boulder, CO, USA<sup>6</sup>

## ABSTRACT

**Background and Aims** Twin and family studies suggest that genetic influences are shared across substances of abuse. However, despite evidence of heritability, genome-wide association and candidate gene studies have indicated numerous markers of limited effects, suggesting that much of the heritability remains missing. We estimated (1) the aggregate effect of common single nucleotide polymorphisms (SNPs) on multiple indicators of comorbid drug problems that are typically employed across community and population-based samples, and (2) the genetic covariance across these measures. **Participants** A total of 2596 unrelated subjects from the Study of Addiction: Genetics and Environment provided information on alcohol, tobacco, cocaine, cannabis and other illicit substance dependence. Phenotypic measures included: (1) a factor score based on DSM-IV drug dependence diagnoses (DD), (2) a factor score based on problem use (PU; i.e. 1+ DSM-IV symptoms) and (3) dependence vulnerability (DV; a ratio of DSM-IV symptoms to the number of substances used). **Findings** Univariate and bivariate genome-wide complex trait analyses of this selected sample indicated that common SNPs explained 25–36% of the variance across measures, with DD and DV having the largest effects [ $h^2_{\text{SNP}}$  (standard error) = 0.36 (0.13) and 0.33 (0.13), respectively; PU = 0.25 (0.13)]. Genetic effects were shared across the three phenotypic measures of comorbid drug problems [ $r_{\text{DD-PU}} = 0.92$  (0.08),  $r_{\text{DD-DV}} = 0.97$  (0.08) and  $r_{\text{PU-DV}} = 0.96$  (0.07)]. **Conclusion** At least 20% of the variance in the generalized vulnerability to substance dependence is attributable to common single nucleotide polymorphisms. The additive effect of common single nucleotide polymorphisms is shared across important indicators of comorbid drug problems.

**Keywords** Addiction, dependence vulnerability, drug dependence, genetics, genome-wide association studies (GWAS), genome-wide complex trait analysis

*Correspondence to:* Rohan H. C. Palmer, Division of Behavioral Genetics at Rhode Island Hospital, Department of Psychiatry and Human Behavior at Brown University, Bradley Hasbro Children's Research Center, 1 Hoppin Street, Suite 204, Providence, RI 02903, USA. E-mail: Rohan\_Palmer@Brown.edu  
Submitted 20 August 2014; initial review completed 9 October 2014; final version accepted 14 November 2014

## INTRODUCTION

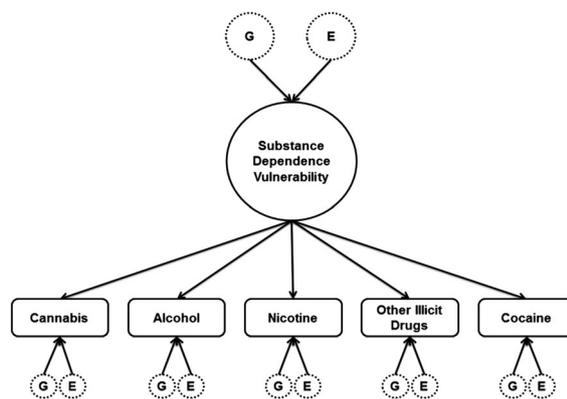
Substance dependence is a complex behavior influenced by both genetic and environmental factors. The role of genetics in substance dependence is well established in twin and family studies. Unfortunately, molecular genetic studies have had limited success identifying individual genetic variants that are common across multiple substances of abuse [1–10]. First, the primary conclusions from molecular genetic studies are that common single nucleotide polymorphisms (SNPs) contribute modestly to substance

dependence phenotypes. For example, rs1614972 in the alcohol dehydrogenase gene (*ADH1C*) was the only replicable SNP in a recent alcoholism genome-wide association study (GWAS) [11], and variants in the *CHRNA5-A3-B4* gene cluster have been linked repeatedly to tobacco addiction/dependence [12,13]. Secondly, multiple genetic polymorphisms influence substance dependence. In part to address the possibility that most variants truly associated with complex traits have effect sizes too small to detect individually using GWAS studies, Yang *et al.* [14] developed a new method, Genome-wide Complex Trait Analysis

(GCTA), that focuses on the estimation of the phenotypic variance explained by genome-wide similarity at genotyped SNPs. Rather than testing each SNP individually, GCTA decomposes the phenotypic variance into two components: (1) effects due to the additive influences of all measured SNPs ( $h_{\text{SNP}}^2$ ) and (2) the effects due to unmeasured environmental influences, random noise or the effects of genetic variants that were not measured by the genotyping array. This approach allows for an estimate of phenotypic variability explained by genome-wide SNP data.

In part, some of the variability in findings across studies can also be attributed to differences in how drug dependence phenotypes are scored for analysis. Prior studies have primarily utilized clinically defined phenotypes [based on the Diagnostic and Statistical Manual of Mental Disorders dependence symptoms (version four; DSM-IV [15,16]), such as dependence diagnosis (i.e. 3+ DSM-IV dependence symptoms all occurring in a 12-month period). However, alternative and dimensional summary scores, such as problem usage (i.e. 1+ DSM-IV dependence symptoms) and symptom counts have also been utilized to help overcome low-level levels of diagnosis often observed in community and population-based samples. Epidemiological studies [17,18] show that individuals who meet the clinical criteria for dependence diagnosis for one substance are at greatly increased risk of using or becoming dependent/addicted to other substances, suggesting a generalized pattern of problematic drug usage. Furthermore, studies suggest that the 'common liability' (i.e. each substance has its own set of genetic and environmental liabilities that are shared with other substances) and 'alternative forms' (i.e. comorbidity across substances arises because each substance is an alternate manifestation of a common underlying liability for deviant behaviors) models best explain the observed comorbidity for dependence across different substances [19–22]. Like Fig. 1, these models assume that dependence upon multiple substances is attributable to correlated latent liabilities or a single latent continuous liability, respectively. Recently, a comparison of three multivariate genetic models indicated that a model which attributes the covariance among different substances to a single latent trait parsimoniously describes alcohol, tobacco and cannabis dependence in a community-based sample [21]. The identified factor, which was referred to as 'substance dependence vulnerability', was highly heritable (64%) across genders and has been demonstrated to be stable over time [23]. Similarly, evidence for a general tendency to misuse substances has also given rise to dimensional measures of comorbid drug problems; in particular, dependence vulnerability (DV), which is a heritable ( $h^2 = 0.40$ ) summary measure that reflects the average number of DSM-IV dependence symptoms across substances used [24].

The current study aimed to identify genetic effects on the vulnerability to substance dependence. We hypothe-



**Figure 1** The Common Pathway Model. The observed phenotypic variance/covariance of the measured dependence phenotypes (rectangles) represent alternate manifestations of a generalized vulnerability to dependence (indicated by the latent trait (solid circle), substance dependence vulnerability). Variation in the latent trait and the variance in the observed phenotypes are decomposable into genetic (G) and environmental/residual (E) influences (dashed circles)

sized that common genetic variants account for at least half of the genetic variance observed in twin/adoption studies. Furthermore, to demonstrate the validity of these findings, we utilized three DSM-IV based definitions of comorbid drug dependence and hypothesized that a common set of genetic factors would account for any and all genetic variances identified across the definitions. To test these hypotheses, we first used factor analysis to replicate prior work [21] and determine the existence of a common factor indicated by either drug dependence diagnosis (DD; i.e. 3+ DSM-IV criteria in a 12-month period) or problem use [PU; i.e. 1+ DSM-IV criteria (life-time)]. Secondly, we examined genome-wide additive genetic influences on the observed factors, as well as the summary measure, DV. Clinical definitions (DD), subclinical thresholds (PU), and the DV summary score were analyzed separately to explore how the measurement approach used to reflect substance-related problems affects the magnitude of identified additive genetic effects. Finally, we used a bivariate GCTA model to test whether DD, PU and DV index the same genetic liability.

## METHODS

### Sample

Data were from the Study of Addiction: Genetics and Environment (SAGE), which is part of the National Human Genome Research Institute's Gene Environment Association Study Initiative [Database for Genotypes and Phenotypes (dbGaP) study accession phs000092.v1.p1]. SAGE is a multi-ethnic sample of 4121 unrelated individuals from three large, complementary data sets designed to study drug addiction: the Collaborative Study on the Genetics of Alcoholism (COGA), the Family Study of Cocaine

Dependence (FSCD) and the Collaborative Genetic Study of Nicotine Dependence (COGENN). Subjects in COGA were recruited from several sites throughout the United States by identifying alcoholic probands from treatment facilities and recruiting other family members, as well as comparison families from the same communities. COGA subjects utilized in the current study are a case–control subset of independent individuals. The FSCD is a case–control family study that ascertained cocaine-dependent individuals from chemical dependency treatment units in the greater St Louis metropolitan area, along with community-based control subjects matched on several criteria such as age and race. The COGENN is a community-based case–control family study that recruited nicotine-dependent cases and non-dependent smoking controls from the cities of Detroit and St Louis. The current analyses focused on a subset of 2596 (COGA  $n = 957$ , FSCD  $n = 541$ , COGENN  $n = 1098$ ) unrelated participants (44% male, mean age = 38.58 years, standard deviation = 9.80) of European descent [confirmed via principal component analysis (PCA), including HAPMAP CEPH, Yoruban, Han Chinese and Japanese samples as ancestral reference groups] drawn from all three studies that comprise SAGE. SAGE was selected for this study because it is the largest cohort of complementary data sets that is publicly available (additional description of SAGE is available at [http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000092.v1.p1](http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1)). While the ascertainment strategy of each study focuses on a particular substance, the fact that substance users have a tendency to become involved with multiple substances allows for sampling of both single and multiple drug users, thereby increasing our power to examine the latent trait that is vulnerability to substance addiction (Supporting information, Table S1 provides a description of rates of drug dependence across studies in SAGE).

## Measures

Self-report data on the endorsement of DSM-IV symptoms for dependence on alcohol, nicotine, cocaine, cannabis and other illicit drugs [i.e. dependence on drugs other than cannabis or cocaine (e.g. opiates, phencyclidine, hallucinogens, sedatives)] were gathered using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) [25] and its modified versions (the Semi-Structured Assessment of Nicotine Dependence used in COGENN and the Semi-Structured Assessment for Cocaine Dependence used in FSCD).

## Derivation of phenotypes

Several studies strongly support a general tendency to use and become dependent on multiple substances

[18,26–28]. As such, we defined substance addiction phenotypically by combining dependence data across various classes of drugs. Further, because phenotypic definitions affect heritability estimates, we created three DSM-IV-based phenotypes intended to approximate those that have been used commonly in the literature. Consistent with the extant literature, we defined dependence vulnerability (DV) as a summary score based on the ratio of the total dependence symptoms endorsed over the total number of substances used [mean = 2.13, standard deviation (SD) = 1.72] [24,29]. Exploratory and confirmatory analysis was used to for the remaining measures.

To understand the factor structure of DD and PU items, we first used exploratory factor analysis (EFA) on a random half of the subjects and then conducted confirmatory factor analysis (CFA) on the remaining half of the sample. For all factor analyses, we used MPlus version 7 [30] using weighted least-squares mean variance estimation. Parallel analysis [31] and scree plots were used to determine the number of factors, and standard EFA/CFA fit indices [e.g. root mean square error of approximation (RMSEA), comparable fit index (CFI) and Tucker–Lewis index (TLI)] to compare nested models [32,33]. EFA and CFA models indicated a single factor when using either the substance dependence diagnosis items or problem use items (see Supporting information, Table S2 and Fig. S1 for detailed results). Based on the consensus between the EFA and CFA models, genetic analyses utilized factor scores (mean = 0, SD = 1) extracted from CFA models using all individuals of European descent. Genetic analyses of DV utilized rank normalized scores (using the BLOM approximation method<sup>1</sup>; mean = 0, SD = 1). These analyses yielded a factor score based on the dichotomous DSM-IV diagnoses of DD and a factor score of PU that was based on items measuring the presence or absence of at least one of the seven total substance dependence symptoms for each of the five diagnostic categories. Unlike recent GWAS studies using factor scores based on dependence diagnoses, DV has been shown previously to localize regions of interest using linkage analysis [29].

## Genetic analyses

### Genotyping

Genotyping of SAGE was performed using the Illumina 1M platform using blood samples deposited at the Rutgers University Cell and DNA Repository (<http://www.rucdr.org>), and was carried out at the Johns Hopkins Center for Inherited Disease Research (CIDR) using Illumina

<sup>1</sup>SAS Institute, Inc. The analyses for this paper were generated using SAS software version 9.3 of the SAS System for Windows, copyright© 2000–2002 SAS Institute, Inc. SAS and all other SAS Institute Inc. products or service names are registered trademarks or trademarks of SAS Institute, Inc., Cary, NC, USA; 2000.

Human1Mv1\_C BeadChips and the Illumina Infinium II assay protocol. SNP calls were made using Illumina BeadStudio Genotyping Module version 3.1.14. Strict quality control (QC) standards were implemented, and genotypes were released by CIDR for 1 040 106 SNPs (99.15% of attempted). Further details are provided in the comprehensive data cleaning report posted at dbGaP [http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/document.cgi?study\\_id=phs000092.v1.p1&phv=22928&phd=2274&pha=&pht=116&phvf=&phdf=20&phaf=&phtf=&dssp=1&consent=&temp=1](http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/document.cgi?study_id=phs000092.v1.p1&phv=22928&phd=2274&pha=&pht=116&phvf=&phdf=20&phaf=&phtf=&dssp=1&consent=&temp=1).

#### Population stratification

We performed PCA using a random sample of 30 000 independent SNPs to control for any residual population stratification effects among the individuals of European descent. Principal components have been demonstrated to capture subtle ancestral differences even within ethnicity, as well certain technical artifacts (e.g. batch effects) [34]. Accordingly, all mixed effects analyses included the first five ancestral principal components as covariates.

#### QC

For all genetic analyses, we retained 796 125 autosomal markers with an allele frequency >1%, a call rate  $\geq 99\%$  and a Hardy–Weinberg equilibrium (HWE) *P*-value greater than 0.0001 among subjects of European descent.

#### Estimation of variance/covariance explained by the SNPs

We used GCTA to decompose the phenotypic variance in each measure into variance components due to the additive effects of all genotyped SNPs and residual effects.

GCTA consists of two steps in which the genetic similarity between all pairs of individuals is obtained via a pairwise genetic relationship matrix (GRM), followed by construction of a mixed-effects model using genetic similarity as a random effect to predict each phenotype. In our identification of subjects of European descent from SAGE, we used GCTA to systematically remove one of any pair of individuals who were more related than second cousins in order to control for cryptic relatedness, which could artificially inflate SNP heritability estimates. The GRM used in all analyses comprised the 2596 unrelated individuals. Univariate and bivariate models were fitted to the phenotypic data while controlling for age, gender, study origin (to account for mean differences/batch effects between the different samples within SAGE) and the first five ancestral principal components to account for stratification effects within individuals of European descent [11].

## RESULTS

### Prevalence and comorbidity of drug dependence and problem use

Table 1 shows the prevalence rates and correlations between the variables used to construct the DD and PU factor scores. Alcohol and nicotine dependence represented the most common substance dependence diagnoses (~45% prevalence). Rates for the problem use items were higher, and followed the same pattern. Phenotypic tetrachoric correlations among all items were generally high, and are shown in Table 1. For example, correlations among the problem use items ranged from 0.47 (between other drug and nicotine problem use) to ~0.84 (between other drug and cocaine problem use).

**Table 1** Prevalence, sample size, and correlations among drug dependence (DD)/problem use (PU) variables.

Phenotype	% (n)	Tetrachoric correlations (ASE)				
<i>Drug dependence diagnosis (3+ symptoms)</i>		<i>Alcohol</i>	<i>Cocaine</i>	<i>Cannabis</i>	<i>Nicotine</i>	<i>Other drugs</i>
Alcohol	45.65 (1185)	1.00				
Cocaine	19.29 (500)	0.78 (0.02)	1.00			
Cannabis	16.76 (434)	0.79 (0.02)	0.82 (0.02)	1.00		
Nicotine	45.53 (1146)	0.66 (0.02)	0.55 (0.03)	0.60 (0.03)	1.00	
Other drugs	15.63 (405)	0.80 (0.02)	0.83 (0.02)	0.77 (0.02)	0.54 (0.03)	1.00
<i>Problem use (1+ symptoms)</i>		<i>Alcohol</i>	<i>Cocaine</i>	<i>Cannabis</i>	<i>Nicotine</i>	<i>Other drugs</i>
Alcohol	70.80 (1838)	1.00				
Cocaine	22.71 (589)	0.73 (0.03)	1.00			
Cannabis	30.35 (788)	0.67 (0.02)	0.80 (0.02)	1.00		
Nicotine	70.69 (1835)	0.59 (0.02)	0.47 (0.03)	0.50 (0.03)	1.00	
Other drug	22.13 (574)	0.74 (0.03)	0.84 (0.01)	0.78 (0.02)	0.47 (0.03)	1.00

Table showing the level of dependence diagnoses (3+ symptoms in a year) and problem use (1+ symptoms) for each substance in the sample, as well as the tetrachoric correlation between substances. ASE = asymptotic standard error.

### Phenotypic variance/covariance attributable to common SNPs

The additive genetic variance due to all SNPs ( $h_{\text{SNP}}^2$ ) was similar across DD, PU and DV. The univariate  $h_{\text{SNP}}^2$  estimate of DD was 0.36 [standard error (SE) = 0.13,  $P$ -value ( $P$ ) = 2.30E-3]. The observed univariate  $h_{\text{SNP}}^2$  estimates for PU ( $h_{\text{SNP}}^2 = 0.25$ , SE = 0.13,  $P = 0.03$ ) and DV ( $h_{\text{SNP}}^2 = 0.32$ , SE = 0.13,  $P = 8.91\text{E-}3$ ) were slightly smaller. Consistent with the hypothesis that a large number of alleles throughout the genome contribute to substance dependence, longer chromosomes accounted for more phenotypic variation than smaller ones (Fig. 2), but not for all phenotypes. For PU,  $h_{\text{SNP}}^2$  increased with chromosome length (Fig. 2;  $R^2 = 0.27$ ,  $\beta = 1.87 \times 10^{-4}$ ,  $t_{(20)} = 2.75$ ,  $P = 0.01$ ). This was not the case with DD ( $R^2 = 0.12$ ,  $\beta = 1.47 \times 10^{-4}$ ,  $t_{(20)} = 1.64$ ,  $P = 0.12$ ) and DV ( $R^2 = 0.00$ ,  $\beta = 1.60 \times 10^{-5}$ ,  $t_{(20)} = 0.20$ ,  $P = 0.84$ ) phenotypes. *Post-hoc* analyses of the  $h_{\text{SNP}}^2$  estimates within each cohort of SAGE (see Supporting information, Table S3) showed internal consistency in the GCTA estimates for each of the phenotypes, although the small sample sizes and differences in ascertainment strategies across studies complicate direct comparison of these estimates.

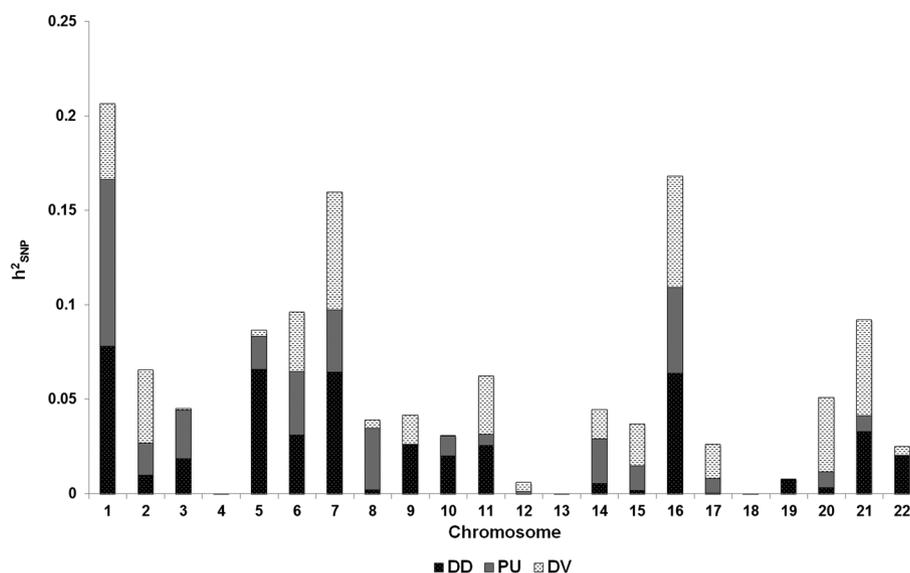
Table 2 presents SNP heritability and correlation estimates from the bivariate analyses. As expected, the three phenotypes had high phenotypic correlations, ranging from 0.84 to 0.86. A strong significant SNP correlation was observed from bivariate analyses between DD and DV ( $r_{\text{SNP}} = 0.97$ , SE = 0.08) and strong SNP correlations were found between DD and PU ( $r_{\text{SNP}} = 0.92$ , SE = 0.08) and between PU and DV ( $r_G = 0.96$ , SE = 0.07). This suggested that the different

ways of scoring substance dependence or vulnerability to substance dependence are associated with largely overlapping SNPs.

### DISCUSSION

This is the first study of its kind to estimate the SNP heritability of generalized vulnerability to substance dependence and compare the genetic liability across multiple phenotypic definitions. The results provide an indication of the additive effect of common SNPs on nicotine, cannabis, cocaine and alcohol, in particular. Given the high loadings of these substances on the DD and PU factors, the additive effect of the common SNPs on individual substance dependence/problems ranges from ~25 to 36%. Although DD evidenced the highest heritability, the effects attributed to common SNPs are correlated highly across the three phenotypes. Overall, the evidence supports the utility of common SNPs to index the genetic liability to substance addiction.

Despite the paucity of findings from GWAS on substance addiction, the present analyses indicate that common SNPs on existing GWA platforms capture substantial genetic information regarding the vulnerability to substance dependence in subjects of European ancestry. The bivariate findings also confirm the expectation that different phenotypic representations of generalized substance dependence are influenced by the same SNPs. Our ability to find largely overlapping SNP effects across substances of abuse may be attributable to our analytic approach. As described and demonstrated elsewhere [35–38], relative to GWAS or candidate studies that look for the independent effects of individual SNPs, GCTA aggregates variance



**Figure 2** Proportion of variance explained (VG/VP) of individual chromosomes on each phenotype while imposing strict quality control (QC) and adjusting for cryptic relatedness. Additive effects attributable to variants across chromosome were estimated at  $1\text{E-}6$  across all phenotypes. DD = dependence diagnosis factor; PU = problem use factor; DV = dependence vulnerability score; SNP = single nucleotide polymorphisms

**Table 2** Single nucleotide polymorphism (SNP) heritability and SNP correlation estimates (standard error).

Trait 1	Trait 2	$V_{\text{SNP-trait 1}}$	$V_{\text{SNP-trait 2}}$	$C_{\text{SNP}}$	$h^2_{\text{SNP-trait 1}}$	$h^2_{\text{SNP-trait 2}}$	$r_{\text{SNP}}$	$P (r_{\text{SNP}} = 0)$
DD	PU	0.16 (0.06)	0.13 (0.07)	0.13 (0.06)	0.36 (0.13)	0.24 (0.13)	0.92 (0.08)	0.0096
DD	DV	0.16 (0.06)	0.30 (0.12)	0.21 (0.08)	0.36 (0.13)	0.33 (0.13)	0.97 (0.05)	0.0029
PU	DV	0.13 (0.07)	0.31 (0.12)	0.20 (0.09)	0.25 (0.13)	0.35 (0.13)	0.96 (0.07)	0.0002

DD = dependence diagnosis factor; PU = problem use factor; DV = dependence vulnerability score;  $V_{\text{SNP-Trait}}$  = genetic variance of each trait;  $C_{\text{SNP}}$  = genetic covariance between trait-1 and trait-2;  $h^2_{\text{SNP}}$  = SNP heritability;  $r_{\text{SNP}}$  = genetic correlation.

across all SNPs and typically explains a greater proportion of variance because the effects of individually non-significant SNPs are included in the estimate. Our results suggest that GWA studies on these phenotypes that have larger samples will discover additional significant genetic associations, as has occurred for other phenotypes [39,40]. It should be noted that additional research is needed to confirm these effects in African populations and to determine whether the observed effects can be attributed to shared genetic factors across different ethnic backgrounds.

Our findings also suggest that the subtle differences in substance vulnerability definitions may not be of major concern in genetic analyses. All three alternative measures of substance vulnerability showed broadly consistent estimates, and the high SNP correlations among them suggest that the same genetic influences affect the three phenotypes. Thus, it is likely that results of genetic analyses across these phenotypes can be compared meaningfully, and meta- or mega-analyses of data sets using alternative phenotypic definitions of substance vulnerability will be informative.

An additional implication of these findings is that we should expect to have a better chance of finding and replicating genetic association analyses (using common SNPs) across phenotypes indicative of a known underlying dimension of risk. The current findings suggest that the same SNPs affect each of these three phenotypes, and therefore genetic associations from studies across these three phenotypes can be compared meaningfully. While the current study is limited to DSM criteria, the evidence suggests the possibility of replication success across studies using non-DSM phenotypes, but that likelihood is dependent upon the extent to which the selected phenotypes reflect common biological variation. Non-DSM phenotypes were beyond the scope of the current paper, and additional research comparing additive genetic effects across DSM and non-DSM-based phenotypes is necessary in order to obtain a definitive conclusion on expectations of replication using broad phenotypes and endophenotypes that may not necessarily harmonize. Findings of common underlying genetic contributions across different phenotypic representations of the vulnerability to substance dependence tested here may also be extended to non-genetic studies to inform

models of substance etiology and treatment. This is possible because the phenotypes used here summarize involvement with disparate substances of abuse (such as alcohol versus heroin).

### Limitations

The current study utilized subjects ascertained for studying alcohol (COGA), nicotine (COGEND) and cocaine (FSCD) dependence, and none of the studies excluded individuals because of their involvement with other substances. Our analyses accounted for mean differences/batch effects between the different samples within SAGE by controlling for study origin in all analyses. However, our estimates are uncorrected for ascertainment, because there is currently no way in GCTA to correct for ascertainment on continuous phenotypes, such as liability for substance dependence, which we focused on here. Our estimates of SNP heritability are therefore relevant to a population with the same distribution of liability as that observed in our sample; we expect that the SNP heritability is lower in the general (non-ascertained) population [41]. Future research, using larger epidemiologic samples recruited for population representativeness rather than substance abuse is needed to replicate the present findings. However, given the low rates of comorbid drug involvement in community- and general-population samples, samples in the tens of thousands will be necessary to obtain sufficient variability across alcohol, tobacco and cannabis, and in particular illicit substances. It should also be noted that the strength of the associations among substances and the evidence for the common phenotypic factors defined by drug dependence and problem use items, may depend upon the prevalence of symptoms in our sample. Larger samples of substance users might show different patterns of comorbidity. However, large studies among drug users in community- and general-population samples have provided substantial evidence in support of a general vulnerability to drug dependence [18,42]. As such, one would expect to identify at least a common factor [under similar genetic influence (as evidenced in the current study)] that accounts for the majority of the phenotypic covariance and possibly one or more additional factors that reflect unique variance shared by other substances [22].

## CONCLUSIONS

The present findings provide important support for a significant contribution of common SNPs in the prediction of vulnerability to substance dependence. Although our findings provided some support for highest heritability observed in DD relative to other ways of defining vulnerability to substance dependence, the associations among the definitions point to shared underlying genetic variation across all three different ways of defining vulnerability to substance dependence. Taken together, these important findings reinforce the utility of examining common variation in SNPs as well as some ability to generalize across studies using different definitions of vulnerability to substance dependence.

### Declaration of interests

None.

### Acknowledgements

Funding for this study was provided by AA021113 (to R.H. C.P.), DA023134 (to V.S.K.), DA033302 (to L.C.B.), MH087240 (to N.R.N.) and MH100141 (to M.C.K.). NIAAA, NIMH and NIDA had no role in the study design, collection, analysis or interpretation of the data, writing the manuscript or the decision to submit the paper for publication. Funding support for the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIH Genes, Environment and Health Initiative [GEI] (U01HG004422). SAGE is one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of data sets and samples was provided by the Collaborative Study on the Genetics of Alcoholism (COGA; U10AA008401), the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01CA089392) and the Family Study of Cocaine Dependence (FSCD; R01DA013423). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and the NIH contract 'High throughput genotyping for studying the genetic contributions to human disease' (HHSN268200782096C). The data sets used for the analyses described in this manuscript were obtained from dbGaP at: [http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000092.v1.p1](http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1) through dbGaP accession number phs000092.v1.p1. The views expressed in this paper are those of the authors

and do not necessarily reflect the position or policy of the Department of Veterans Affairs.

## References

1. Kendler K. S., Kalsi G., Holmans P. A., Sanders A. R., Aggen S. H., Dick D. M. *et al.* Genomewide association analysis of symptoms of alcohol dependence in the molecular genetics of schizophrenia (MGS2) control sample. *Alcohol Clin Exp Res* 2011; **35**: 963–75.
2. Yang C., Li C., Kranzler H. R., Farrer L. A., Zhao H., Gelernter J. Exploring the genetic architecture of alcohol dependence in African-Americans via analysis of a genomewide set of common variants. *Hum Genet* 2013; **133**(5): 617–24.
3. Vrieze S. I., McGue M., Iacono W. G. The interplay of genes and adolescent development in substance use disorders: leveraging findings from GWAS meta-analyses to test developmental hypotheses about nicotine consumption. *Hum Genet* 2012; **131**: 791–801.
4. Yang B. Z., Han S., Kranzler H. R., Farrer L. A., Gelernter J. A genomewide linkage scan of cocaine dependence and major depressive episode in two populations. *Neuropsychopharmacology* 2011; **36**: 2422–30.
5. Gelernter J., Panhuysen C., Weiss R., Brady K., Hesselbrock V., Rounsaville B. *et al.* Genomewide linkage scan for cocaine dependence and related traits: significant linkages for a cocaine-related trait and cocaine-induced paranoia. *Am J Med Genet B Neuropsychiatric Gen* 2005; **136B**: 45–52.
6. Cornelis M. C., Monda K. L., Yu K., Paynter N., Azzato E. M., Bennett S. N. *et al.* Genome-wide meta-analysis identifies regions on 7p21 (AHR) and 15q24 (CYP1A2) as determinants of habitual caffeine consumption. *PLoS Genet* 2011; **7**: e1002033.
7. Gelernter J., Kranzler H. R., Sherva R., Koesterer R., Almasy L., Zhao H. *et al.* Genome-wide association study of opioid dependence: multiple associations mapped to calcium and potassium pathways. *Biol Psychiatry* 2014; **76**: 66–74.
8. Nishizawa D., Fukuda K., Kasai S., Hasegawa J., Aoki Y., Nishi A. *et al.* Genome-wide association study identifies a potent locus associated with human opioid sensitivity. *Mol Psychiatry* 2014; **19**: 55–62.
9. Drgon T., Zhang P. W., Johnson C., Walther D., Hess J., Nino M. *et al.* Genome wide association for addiction: replicated results and comparisons of two analytic approaches. *PLoS One* 2010; **5**: e8832.
10. Yu Y., Kranzler H. R., Panhuysen C., Weiss R. D., Poling J., Farrer L. A. *et al.* Substance dependence low-density whole genome association study in two distinct American populations. *Hum Genet* 2008; **123**: 495–506.
11. Biernacka J. M., Geske J. R., Schneekloth T. D., Frye M. A., Cunningham J. M., Choi D. S. *et al.* Replication of genome wide association studies of alcohol dependence: support for association with variation in ADH1C. *PLoS One* 2013; **8**: e58798.
12. Bierut L. J. Convergence of genetic findings for nicotine dependence and smoking related diseases with chromosome 15q24–25. *Trends Pharmacol Sci* 2010; **31**: 46–51.
13. Berrettini W. H., Doyle G. A. The CHRNA5-A3-B4 gene cluster in nicotine addiction. *Mol Psychiatry* 2012; **17**: 856–66.
14. Yang J., Lee S. H., Goddard M. E., Visscher P. M. Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. *Methods Mol Biol* 2013; **1019**: 215–36.
15. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (4th ed.). Washington, DC: 1994.

16. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (4th ed., text rev.). Washington, DC; 2000.
17. Substance Abuse and Mental Health Services Administration (SAMHSA). Results from the 2010 National Survey on Drug Use and Health. Summary of National Findings. HHS Publication no. (SMA) 11-4658. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2011.
18. Palmer R. H., Young S. E., Hopfer C. J., Corley R. P., Stallings M. C., Crowley T. J. *et al.* Developmental epidemiology of drug use and abuse in adolescence and young adulthood: evidence of generalized risk. *Drug Alcohol Depend* 2009; **102**: 78–87.
19. Kendler K. S., Jacobson K. C., Prescott C. A., Neale M. C. Specificity of genetic and environmental risk factors for use and abuse/dependence of cannabis, cocaine, hallucinogens, sedatives, stimulants, and opiates in male twins. *Am J Psychiatry* 2003; **160**: 687–95.
20. Kendler K. S., Myers J., Prescott C. A. Specificity of genetic and environmental risk factors for symptoms of cannabis, cocaine, alcohol, caffeine, and nicotine dependence. *Arch Gen Psychiatry* 2007; **64**: 1313–20.
21. Palmer R. H., Button T. M., Rhee S. H., Corley R. P., Young S. E., Stallings M. C. *et al.* Genetic etiology of the common liability to drug dependence: evidence of common and specific mechanisms for DSM-IV dependence symptoms. *Drug Alcohol Depend* 2012; **123**(Suppl 1): S24–32.
22. Rhee S. H., Hewitt J. K., Young S. E., Corley R. P., Crowley T. J., Neale M. C. *et al.* Comorbidity between alcohol dependence and illicit drug dependence in adolescents with antisocial behavior and matched controls. *Drug Alcohol Depend* 2006; **84**: 85–92.
23. Palmer R. H., Young S. E., Corley R. P., Hopfer C. J., Stallings M. C., Hewitt J. K. *et al.* Stability and change of genetic and environmental effects on the common liability to alcohol, tobacco, and cannabis DSM-IV dependence symptoms. *Behav Genet* 2013; **43**: 374–85.
24. Button T. M., Hewitt J. K., Rhee S. H., Young S. E., Corley R. P., Stallings M. C. *et al.* Examination of the causes of covariation between conduct disorder symptoms and vulnerability to drug dependence. *Twin Res Hum Genet* 2006; **9**: 38–45.
25. Bucholz K. K., Cadoret R., Cloninger C. R., Dinwiddie S. H., Hesselbrock V. M., Nurnberger J. L., Jr *et al.* A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *J Stud Alcohol* 1994; **55**: 149–58.
26. Kapusta N. D., Plener P. L., Schmid R., Thau K., Walter H., Lesch O. M. Multiple substance use among young males. *Pharmacol Biochem Behav* 2007; **86**: 306–11.
27. Sartor C. E., Grant J. D., Bucholz K. K., Madden P. A., Heath A. C., Agrawal A. *et al.* Common genetic contributions to alcohol and cannabis use and dependence symptomatology. *Alcohol Clin Exp Res* 2010; **34**: 545–54.
28. Iacono W. G., Malone S. M., McGue M. Behavioral disinhibition and the development of early-onset addiction: common and specific influences. *Annu Rev Clin Psychol* 2008; **4**: 325–48.
29. Stallings M. C., Corley R. P., Hewitt J. K., Krauter K. S., Lessem J. M., Mikulich S. K. *et al.* A genome-wide search for quantitative trait loci influencing substance dependence vulnerability in adolescence. *Drug Alcohol Depend* 2003; **70**: 295–307.
30. Muthen, L. K. Muthen, B. O. Mplus User's Guide, 7th edn. Los Angeles, CA: Muthen & Muthen; 1998–2012.
31. Hayton J. C., Allen D. G., Scarpello V. Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. *Organ Res Meth* 2004; **7**: 191–205.
32. Hu L. T., Bentler P. M. Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equat Model* 1999; **6**: 1–55.
33. Yu C. Y. Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. Unpublished dissertation, 2002. Available at Mplus website: <http://www.statmodel.com/download/Yudissertation.pdf> (accessed 02 February 2014).
34. Price A. L., Patterson N. J., Plenge R. M., Weinblatt M. E., Shadick N. A., Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–9.
35. Yang J., Ferreira T., Morris A. P., Medland S. E., Madden P. A. E., Heath A. C. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 2012; **44**: 369–75.
36. Galarneau G., Palmer C. D., Sankaran V. G., Orkin S. H., Hirschhorn J. N., Lettre G. Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. *Nat Genet* 2010; **42**: 1049–51.
37. Sanna S., Li B., Mulas A., Sidore C., Kang H. M., Jackson A. U. *et al.* Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. *PLoS Genet* 2011; **7**: e1002198.
38. Vrieze S. I., McGue M., Miller M. B., Hicks B. M., Iacono W. G. Three mutually informative ways to understand the genetic relationships among behavioral disinhibition, alcohol use, drug use, nicotine use/dependence, and their co-occurrence: twin biometry GCTA, and genome-wide scoring. *Behav Genet* 2013; **43**: 97–107.
39. Rahmioglu N., Macgregor S., Drong A. W., Hedman A. K., Harris H. R., Randall J. C. *et al.* Genome-wide enrichment analysis between endometriosis and obesity-related traits reveals novel susceptibility loci. *Hum Mol Genet* 2014 Oct 8. pii: ddu516. [Epub ahead of print] PubMed PMID: 25296917.
40. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**(7510): 421–7.
41. Lee S. H., Yang J., Goddard M. E., Visscher P. M., Wray N. R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* 2012; **28**: 2540–2.
42. Derringer J., Krueger R. F., Dick D. M., Agrawal A., Bucholz K. K., Foroud T. *et al.* Measurement invariance of DSM-IV alcohol, marijuana and cocaine dependence between community-sampled and clinically overselected studies. *Addiction* 2013; **108**: 1767–76.

### Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

**Figure S1.** Scree plots from EFA of Drug Dependence (A) and Problem Use (B) items.

**Table S1.** Prevalence (N) of DSM-IV diagnosis of drug dependence by sub-sample of SAGE

**Table S2.** Model fit and loadings for EFA/CEFA models

**Table S3:** Univariate  $h^2_{\text{SNP}}$  (SE) for DD, PU, and DV by subsamples of SAGE