Behav Genet (2012) 42:626–635
DOI 10.1007/s10519-012-9531-4

ORIGINAL RESEARCH

The Aggregate Effect of Dopamine Genes on Dependence Symptoms Among Cocaine Users: Cross-Validation of a Candidate System Scoring Approach

Jaime Derringer · Robert F. Krueger · Danielle M. Dick · Fazil Aliev · Richard A. Grucza · Scott Saccone · Arpana Agrawal · Howard J. Edenberg · Alison M. Goate · Victor M. Hesselbrock · John R. Kramer · Peng Lin · Rosalind J. Neuman · John I. Nurnberger Jr. · John P. Rice · Jay A. Tischfield · GENEVA consortium · Laura J. Bierut

Received: 14 April 2011 / Accepted: 6 February 2012 / Published online: 23 February 2012

Abstract Genome-wide studies of psychiatric conditions frequently fail to explain a substantial proportion of variance, and replication of individual SNP effects is rare. We demonstrate a selective scoring approach, in which variants from several genes known to directly affect the dopamine system are considered concurrently to explain individual differences in cocaine dependence symptoms. 273 SNPs from eight dopamine-related genes were tested for association with cocaine dependence symptoms in an initial training sample. We identified a four-SNP score that accounted for 0.55% of the variance in a separate testing sample ($p = 0.037$). These findings suggest that (1) limiting investigated SNPs to those located in genes of theoretical importance improves the chances of identifying replicable effects by reducing statistical penalties for multiple testing, and (2) considering top-associated SNPs in the aggregate can reveal replicable effects that are too small to be identified at the level of individual SNPs.

Keywords Candidate gene · Cocaine dependence · Dopamine

Introduction

Cocaine use is not uncommon, with 14.7% of adult participants in a large, US-representative sample reporting lifetime use (SAMHSA 2009). Cocaine is one of the most addictive recreationally used substances (Nutt et al. 2007). Even occasional users are likely to develop symptoms of problematic use (Gillespie et al. 2007), with tolerance or withdrawal being especially highly associated with the development of cocaine dependence (Shaffer and Eber 2002). The medical consequences of cocaine use are severe

Edited by Valerie Knopik

Electronic supplementary material The online version of this article (doi:10.1007/s10519-012-9531-4) contains supplementary material, which is available to authorized users.

J. Derringer (✉)
Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO 80309-0447, USA
e-mail: jaime.derringer@colorado.edu

R. F. Krueger
University of Minnesota, Minneapolis, USA

D. M. Dick · F. Aliev
Virginia Commonwealth University, Richmond, USA

F. Aliev
Ankara University, Ankara, Turkey

R. A. Grucza · S. Saccone · A. Agrawal · A. M. Goate · P. Lin · R. J. Neuman · J. P. Rice · L. J. Bierut
Washington University in St. Louis, St. Louis, USA

H. J. Edenberg · J. I. Nurnberger Jr.
Indiana University, Bloomington, USA

V. M. Hesselbrock
University of Connecticut, Farmington, USA

J. R. Kramer
University of Iowa, Iowa City, USA

J. A. Tischfield
Rutgers University, Piscataway, USA

and widespread, and include such conditions as arrhythmia, cerebral hemorrhage, and pulmonary edema (Gray 1993). Cocaine dependence is also highly co-morbid with psychiatric and other substance use disorders, with severity of cocaine dependence being associated with increased comorbidity of other substance use disorders (Ford et al. 2009). Approximately 40–80% of the variance in cocaine dependence is attributable to additive genetic factors (Kendler et al. 2000; Kendler and Prescott 1998; van den Bree et al. 1998).

Dopamine and cocaine

Substantial evidence implicates dopamine as the primary neurotransmitter system involved in the rewarding effects of cocaine exposure (Dackis and Gold 1985; Haile et al. 2007; Kuhar et al. 1991). Cocaine competitively inhibits dopamine internalization by binding to the dopamine transporter (Beuming et al. 2008). Administration of a typical dose (i.e., what a user might be expected to take) in humans blocks a majority of dopamine transporter sites (Volkow et al. 1997). The blocking of these sites results in increased synaptic dopamine, which contributes to the reinforcing and addictive properties of cocaine (Haile et al. 2007). Cocaine exposure induces long-term potentiation of dopamine neurons (Liu et al. 2005; Ungless et al. 2001). Extended use results in persisting effects on nuclear accumbens D1 receptors (Henry and White 1991) and prefrontal dopamine D2 receptors (Briand et al. 2008).

Candidate gene studies further support the association between dopamine and response to cocaine, by demonstrating specific variants located in dopaminergic genes that affect risk for cocaine dependence. The dopamine transporter gene solute carrier family 6 (neurotransmitter transporter, dopamine), member 3 (*SLC6A3*, previously known as *DAT* or *DAT1*) has been the most frequently examined gene with respect to cocaine response. *SLC6A3* knockout mice show reduced alteration of behavioral response (Giros et al. 1996) and brain glucose metabolism (Thanos et al. 2008) when exposed to cocaine. However, *SLC6A3* knockout mice still self-administer cocaine (Rocha et al. 1998) and can still develop conditioned place preference in response to cocaine (Sora et al. 1998), demonstrating that *SLC6A3* is not solely responsible for cocaine's reinforcing effects.

Substantial candidate gene research has also examined the dopamine receptor D1 gene (*DRD1*) for association with cocaine response. *DRD1* knockout mice do not self-administer cocaine (Caine et al. 2007). $DRD1^{-/-}$ mice show decreased locomotion with cocaine exposure, compared to the increase in locomotion observed when *DRD1* wild-type mice are exposed to cocaine (Xu et al. 1994). The relationship between dopamine receptors and cocaine response has also been demonstrated experimentally in humans. Evidence from agonist trials suggests that activation of D1-like (i.e., *DRD1* and *DRD5*) receptors has opposite effects to those from activating D2-like (i.e., *DRD2*, *DRD3* and *DRD4*) receptors, with D1-like agonists reducing cocaine-seeking behavior and D2-like agonists increasing this behavior (Self et al. 1996).

While individual differences in the dopamine system are not sufficient to solely account for individual differences in cocaine dependence (e.g., Hnasko et al. 2007), dopamine has been repeatedly shown in both human and animal models to strongly influence the propensity to develop such dependence (see Haile et al. 2007 for a review). Further molecular genetic investigation may be a promising research direction to more fully elucidate the relationship between dopamine and cocaine dependence.

The genetic scoring method

Given the wide range in severity of observed phenotypes in substance use and dependence, it is likely that genetic influences on substance problems are highly polygenic, that is, comprised of numerous small effects (e.g., Goldman et al., 2005). This is similar to other complex psychiatric and behavioral phenotypes (e.g., Plomin et al. 2009). A recent genome-wide study of schizophrenia identified a genetic score consisting of tens of thousands of SNPs that, in the aggregate, explained $\sim$3% of the variance in schizophrenia case–control status (International Schizophrenia Consortium 2009). While such genome-wide scores have the potential to explain significant variance, they do not necessarily provide greater information or discriminative ability than do established variants of known effect, in cases where such "known" variants exist (Evans et al. 2009). This may be in part due to the expected inclusion of "noise" variants (e.g., false-positive SNPs meeting selected significance cut-offs due to chance alone) when casting such a (genome) wide net as part of the scoring approach.

One method for reducing the inclusion of the noise variants would be to increase the size of the score "discovery" sample, similar to the rapidly increasing sample sizes utilized in genome-wide association studies (GWAS; Amin et al. 2009). However, increasing sample size may not be practical for certain phenotypes that are rare, or for those that require thorough assessment or selected samples. In such cases, alternatives to atheoretical genome-wide approaches are needed, in order to take into account the reduced statistical power inherent in genetic analyses of smaller samples. With this in mind, we suggest one possible approach to improving the accuracy of the genetic risk score: reducing the "noise" variants available for inclusion by only examining genes located within a candidate system (i.e., genomic regions specifically theorized to be involved with one or more processes that influence the phenotype of interest).

The current study

Similar to prior scoring endeavors (e.g., International Schizophrenia Consortium 2009; Evans et al. 2009), we made use of a large sample enriched for our phenotypes of interest to conduct intra-sample cross-validation. Because of the amount of evidence implicating the dopamine system in substance use generally, and cocaine dependence specifically, we focused our analyses on this system. The intention behind focusing on a specific neurotransmitter system was to reduce the likelihood of including "noise" SNPs, increasing our ability to identify a valid scoring set of SNPs. Similarly, the construction of a poly-SNP genetic "risk" score is well-aligned with current theories of genetic contributions to complex phenotypes, such as substance dependence, in which numerous genetic variants are likely to each have a relatively small effect on the observed phenotype (e.g., Plomin et al. 2009).

The current study sought to identify SNP variation that was associated with individual differences in cocaine dependence symptoms. It was hypothesized that in studies with modest sample sizes, or small effect sizes, these effects could be most reliably detected and replicated in the aggregate. Given that there is strong evidence implicating the dopamine system in cocaine dependence, we sought to improve our power to detect what were expected to be small effects of individual SNPs on cocaine dependence by limiting our analyses to SNPs in a selected, core set of autosomal genes that are definitely and directly involved in the dopamine system and by examining the joint effect of these SNPs. We therefore constructed a genetic risk score from a subset of a genome-wide data set, including only SNPs from selected dopaminergic genes. The specific SNPs most strongly associated with cocaine dependence symptoms were first identified in an initial discovery sample. The replication of these top SNPs' aggregate effect on cocaine dependence was then examined in an independent testing sample. Because substantial heritable variance is common across multiple substances of dependence, we then examined the potential specificity of this cocaine-derived genetic risk score by testing its correlation with the frequently comorbid phenotypes of alcohol, tobacco, and marijuana dependence symptoms.

Methods

Participants

In order to examine the genetic contribution to problems among cocaine users, only individuals who reported having tried cocaine were included in this study, because experimentation is a necessary precursor to problem use. That is, the inclusion of non-users would make results unclear in terms of whether they identify influences on propensity to use, or influences on problem use severity, or both (Buckland 2008). One consideration in measuring cocaine dependence is phenotypic noise due to environmental constraints on access to cocaine. The current sample was drawn from project centers in densely populated US cities (such as Saint Louis and Detroit) and was specifically designed to over-sample individuals with substance-related problems. This suggests that availability of cocaine is a reasonable assumption for individuals included in the current sample.

The present sample included 1,591 unrelated individuals from the Study of Addiction: Genetics and Environment (SAGE; Bierut et al. 2010) who reported having ever used cocaine; the mean age of these cocaine users was 38.25 years (SD = 7.6, range = 18–63), 42.6% were female, and 37.5 and 4.05% self-reported African or Hispanic ancestry (vs. European or non-Hispanic ancestry), respectively. SAGE was designed to examine the genetic underpinnings of alcohol dependence, with participants drawn from three primary studies: FSCD, the Family Study of Cocaine Dependence (Bierut et al. 2008); COGA, the Collaborative Study on the Genetics of Alcoholism (Begleiter et al. 1995); and COGEND, the Collaborative Genetic Study of Nicotine Dependence (Bierut 2007). All data used in the present study are available via the database of Genotypes and Phenotypes (dbGaP; phs000092.v1.p1).

Measure

The current study utilized DSM-IV-TR cocaine, alcohol, nicotine, and marijuana dependence symptom counts as measures of substance use severity. These counts were derived from the Semi-Structured Assessment for the Genetics of Alcoholism interview (SSAGA-II), which has demonstrated validity and reliability (Bucholz et al. 1994; Hesselbrock et al. 1999). To obtain approximately normally distributed regression residuals, the symptom counts for marijuana dependence were log-transformed [i.e., ln(symptoms + 1)]. Symptom counts for all other substances (i.e., cocaine, alcohol, and nicotine) remained untransformed.

Genotyping

DNA was obtained from blood or lymphoblastoid cell lines, and genotyping was carried out on the Illumina Human IM Bead Chip by the Johns Hopkins University Center for Inherited Disease Research (CIDR). SNPs had a median missing call rate of less than 0.05, and 95% of SNPs resulted in less than 1.4% missingness. Strict quality control procedures were implemented, including assessment

of batch effects, duplication errors, gender and chromosomal anomalies, Hardy–Weinberg disequilibrium, hidden relatedness, Mendelian errors, missing call rates, and population structure (Laurie et al. 2010). Duplicates, outliers, and related subjects were removed. In total, 948,142 SNPs passed this thorough cleaning procedure (Bierut et al. 2010).

A literature search was used to identify relevant autosomal dopamine-related genes. Genes were selected for inclusion if they were known to have definite and direct effects on the dopaminergic system; these were *COMT*, *DBH*, *DDC*, *DRD1*, *DRD2*, *DRD3*, *DRD4*, and *SLC6A3*. SNPs located within each of these genes that are available on the Illumina Human 1M Bead Chip were then identified for analysis and annotated for exact location and function using WGAViewer (Ge and Goldstein 2007). A total of 273 SNPs were selected for inclusion in our analyses (identical to those identified in Derringer et al. 2010). Due to our modest sample size, we chose to be conservative in our inclusion of SNPs. With the inclusion of an increasing number of genomic regions, multiple testing becomes more of a concern and the likelihood of false positives increases as increasingly system-distal gene and gene products are included. Because we used narrow inclusion criteria for "dopaminergic" SNPs, this limited the extent to which spurious SNPs could be included in the risk score.

Statistical analyses

The current analysis is similar to that described by the International Schizophrenia Consortium (2009), with two primary differences. First, our total sample was split into training and testing samples based on random assignment, rather than by sex. Second, due to the limited number of SNPs tested, SNPs were included in the "risk" score one at a time, rather than in clusters defined by *p*-value cut-offs. All analyses were conducted in the open-source statistical package R (R Development Core Team 2009).

Our sample was split randomly into training and testing samples of equal size. Although the psychometric properties of and endorsement rates for substance problems may vary substantially between sexes (Nichol et al. 2007) and races (Harford et al. 2009), the random splitting of our sample resulted in demographically equivalent sub-samples (see Table 1). Given the heterogeneity within the sample, all substance dependence symptom counts were residualized over sex, age (coded in quartiles as three dummy codes, as has been done in previous SAGE analyses, corresponding to ≤34, 35–39, and 40–44, with 45+ as the reference group; Bierut et al. 2010), primary study source (dummy codes, corresponding to COGA and COGEND, with FSCD as the reference group), and ancestry. To account for sample ethnic heterogeneity, a procedure described by Price et al. (2006) was used to estimate ancestry in the form of principal components derived from the entire SAGE sample genome-wide data. This resulted in two major principal components, corresponding to European versus African ancestry (PC1) and Hispanic versus non-Hispanic ancestry (PC2) (Bierut et al. 2010). Both PC1 and PC2 were included as covariates over which symptom counts were residualized.

Initially, the standardized residuals of cocaine dependence symptom counts were regressed separately on each SNP within the training sample. SNPs were coded as 0/1/2 to indicate the number of minor (i.e., less frequent) alleles present for a given individual. This coding scheme assumes a priori that allelic effects will be additive in nature. Although this may decrease our ability to detect alleles with dominant effects that are only weakly tagged by our

Table 1 Demographic characteristics of the two samples

	Sample	
	Training	Testing
N	796	795
Female (%)	41.5	43.8
African American (%)	37.1	38.0
Hispanic (%)	4.5	3.6
Age <35 (%)	25.1	26.4
Age 35–39 (%)	27.0	27.2
Age 40–44 (%)	31.3	28.6
Age >45 (%)	16.6	17.9
Original study (%)		
FSCD	41.5	40.5
COGA	34.3	34.8
COGEND	24.2	24.7
Cocaine symptoms		
N	796	795
Mean	3.92	3.70
SD	3.03	3.04
Alcohol symptoms		
N	794	793
Mean	4.33	4.48
SD	2.27	2.21
Nicotine symptoms		
N	768	762
Mean	3.53	3.51
SD	1.87	1.80
Marijuana symptoms		
N	779	778
Mean	2.33	2.20
SD	2.44	2.37

Original study codes are as follows: *FSCD* the Family Study of Cocaine Dependence, *COGA* the Collaborative Study on the Genetics of Alcoholism, *COGEND* the Collaborative Genetic Study of Nicotine Dependence

genotyped SNPs (Vukcevic et al. 2011), additive allelic effects likely account for the majority of genetic variance within a complex phenotype such as cocaine dependence (Hill et al. 2008). It has been suggested that interactions among SNPs could also account for substantial heritable variance. However, such epistatic models may often be statistically indistinguishable from models in which genetic effects are assumed to be primarily additive, and substantially greater sample sizes are required to achieve adequate power to estimate interactive effects among SNPs, over and above any main effects of those same SNPs (Zuk et al. 2012). Given the additional multiple testing burden of testing more than one effect model for each SNP, we chose to assume an additive model for all tests.

Following the individual SNP association tests within the testing sample, "risk scores" for each SNP were then calculated for each individual in the testing sample by multiplying the number of minor alleles at that SNP by its training-sample-estimated regression coefficient. For the purpose of estimating risk scores, if an individual was missing a genotype for a scoring SNP, they were allocated the mean number of minor alleles (although the missingness rates among the selected SNPs were low, with a median rate of 0.1% and all SNPs having missingness rates less than 2%). To identify the best aggregate SNP score in the testing sample, SNPs were incorporated one at a time to the calculation of a total risk score, in order of ascending *p*-value resulting from the initial association tests run in the training sample, until the variance in the testing sample explained by the included SNPs began decreasing. That is, individual SNP scores were summed across all SNPs to be included in the score, i.e. score $= \Sigma_{1\ -\ i}$ ($N_{\text{Minor Alleles for SNP i}}$*$B_{\text{SNP i}}$), where B is the regression weight for the SNP predicting the standardized residuals of cocaine dependence symptom counts in the training sample.

Risk scores were therefore linear combinations of the included SNPs, weighted by their training-sample-derived regression coefficients. The ability of this score to cross-predict testing sample variance in cocaine dependence scores was estimated as the correlation between the summed risk score and cocaine dependence symptoms (which had been residualized over covariates, i.e., sex, age, study source, and ancestry, as described for the training set). Specificity of the aggregate cocaine SNP score was investigated by correlating the cocaine-derived SNP score with substance dependence symptoms for alcohol, nicotine, and marijuana, each of which were residualized over the same covariates.

Results

Results from the top-ranked SNPs (by *p*-value) associated with cocaine dependence symptoms in the training sample are shown in Table 2. For comparison, results for association tests between these SNPs and cocaine dependence symptoms in the testing sample are also provided. The results of all SNP association tests are provided in online supplementary Table S1.

Maximum variance explained in the testing sample was reached when the top four SNPs from the training sample association tests (along with their training-sample-derived regression weights) were aggregated into a genetic risk score. These top four SNPs (rs1611131, rs5326, rs9817063, and rs1079597) are located in four different genes (dopamine beta-hydroxylase, *DBH*; dopamine receptor D1, *DRD1*; dopamine receptor D3, *DRD3*; and dopamine receptor D2, *DRD2*; respectively). It should be noted that the negative regression weight on the *DBH* SNP shown in Table 2 was treated as such (when it was used to weight the *DBH* SNP for inclusion in the risk score), meaning that a greater number of the coded (in this case, protective) alleles would result in a lower risk score. This four-SNP score explained 2.76% of the sample variance in number of cocaine dependence symptoms in the training set from which the SNP weights were derived ($p < 2.5 \times 10^{-6}$).

Table 2 Training sample top results included in estimation of testing sample genetic risk score, predicting cocaine dependence symptom count

SNP	Gene	Chr	Function	Allele	MAF	Training sample			Testing sample		
						B	Z	P	B	Z	p
rs1611131	*DBH*	9	Intron	G	0.23	−0.40	−2.51	0.012	−0.15	−0.94	0.35
rs5326	*DRD1*	5	UTR-5	A	0.14	0.45	2.42	0.015	0.10	0.54	0.59
rs9817063	*DRD3*	3	NearGene-3	C	0.45	0.30	2.23	0.026	0.20	1.47	0.14
rs1079597	*DRD2*	11	Intron	A	0.16	0.36	2.15	0.032	0.25	1.41	0.16
Aggregate genetic risk score						$R^2 = 2.756\%$		2.5E−6	$R^2 = 0.546\%$		0.037

Note Chr is chromosome on which the gene is located, Function is locational function within the gene; Allele is minor (i.e. less frequent) allele, for which the effect is reported; MAF is minor allele frequency; B is regression weight for the SNP predicting the standardized residuals of cocaine dependence symptom counts; Aggregate genetic risk score is sum of each individual's minor alleles, weighted by the training-sample-derived regression weights

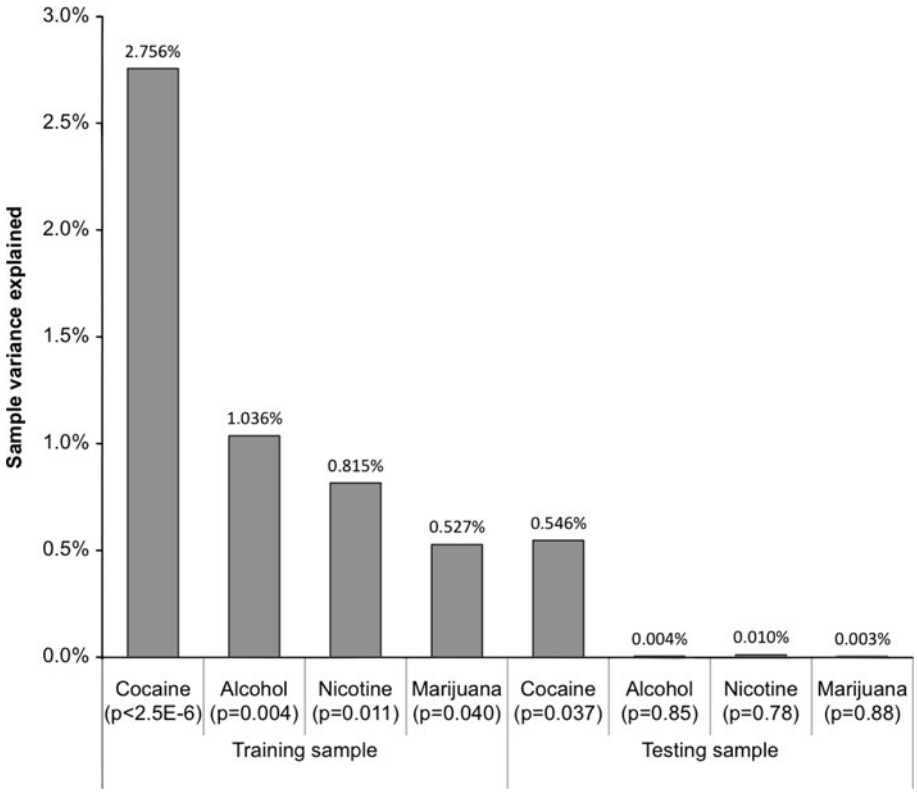

Fig. 1 Ability of cocaine-derived genetic risk score to predict sample variance in cocaine, alcohol, nicotine, and marijuana dependence symptoms

This same genetic risk score explained 0.55% ($p = 0.037$) of the variance in cocaine dependence symptoms in the independent testing sample. This suggests that the effects of these four SNPs replicate, in aggregate, and explain variance in cocaine symptoms in both the original training sample and the independent testing sample.

The specificity of the genetic risk score was examined by correlating it with other substance dependence symptom counts (i.e., alcohol, nicotine, or marijuana, again controlling for demographic covariates). Within the training sample, this cocaine-based score explained 0.53–1.04% of the variance in alcohol, nicotine, or marijuana dependence symptoms ($p = 0.004$–0.04). However, in the testing sample the cocaine-derived genetic risk score explained only 0.004, 0.010, and 0.003% of the sample variance in alcohol, nicotine, and marijuana symptoms, respectively, (all with p-values greater than 0.78, Fig. 1). This demonstrates that this cocaine-identified dopaminergic genetic risk score accounted for variance in cocaine dependence symptoms within the testing sample that was largely independent of variance associated with other substance problems.

Discussion

The current study sought to demonstrate a method of candidate system scoring in a sample that would be underpowered for genome-wide analytic approaches. To that end, we aimed to identify dopaminergic SNP variation that was associated with individual differences in cocaine dependence symptoms. Given the strong evidence implicating the dopamine system in cocaine dependence, we attempted to improve our power to detect what were expected to be small effects of individual SNPs on cocaine dependence symptoms by limiting our analyses to SNPs in a selected, core set of autosomal genes that are definitely and directly involved in the dopamine system and by examining the effect of these SNPs in the aggregate. We found evidence for an association between cocaine dependence symptoms and dopamine-related candidate genes at the biological system level. The optimal risk score, utilizing regression weights derived in the training sample by independently testing each SNP's association with cocaine dependence symptoms, incorporated four SNPs (one each from four separate genes: *DBH*, *DRD1*, *DRD3*, and *DRD4*) and explained 0.55% of the sample variance in the "testing" sample ($p = 0.037$). This reduction in variance explained by the four-SNP score in the training sample (2.76%, $p < 2.5 \times 10^{-6}$) is a common phenomenon even when replications agree with original findings; that is, effects are often smaller in size in replication samples compared to those observed in the original (the "winner's curse" effect).

Several of the four SNPs included in our genetic risk score have been previously investigated for association

with substance use or other psychiatric phenotypes. While a recent report found no association between rs9817063 (in DRD3) and case–control cocaine dependence status in an African American sample (Bloch et al. 2009), rs5326 (in *DRD1*) has been associated with heroin addiction (Levran et al. 2009), as well as with tardive dyskinesia among patients with schizophrenia on long-term antipsychotic treatment (Lai et al. 2011).

Also notable is the inclusion of rs1079597 (formerly known as *DRD2* Taq1B) in the risk score. Rs1079597 is in linkage disequilibrium (i.e., correlated) with rs1800497 (formerly known as *DRD2* Taq1A), a missense mutation located 10 kb downstream of *DRD2* in the ankyrin repeat and kinase domain containing 1(*ANKK1*) gene. rs1800497 has been previously suggested as a candidate polymorphism for substance use phenotypes [see meta-analyses by Munafò et al. (2007, 2009)]. rs1079597 and rs1800497 alleles have R-squared values with each other of 69.6 and 28.1% in European American and African samples, respectively, (as estimated in SNAP, Johnson et al. 2008). While a recent report found that rs1079597 was not associated with cocaine dependence case–control status in a Spanish sample (Fernàndez-Castillo et al. 2010), it has been previously associated with nicotine withdrawal symptoms (Robinson et al. 2007), borderline personality traits (Nemoda et al. 2010), and Tourette syndrome (Herzberg et al. 2010).

Our cocaine-derived genetic risk score demonstrated specificity for cocaine dependence severity in the present analyses. It did not predict significant variance in alcohol, tobacco, or marijuana dependence severity (0.003–0.010%) in the testing sample ($N = 795$), which had at least 80% power to detect effects less than half the size (i.e., 0.25%) observed for cocaine dependence at a very liberal alpha level of 0.50. Given the high p-values ($p > 0.78$) for all substances other than cocaine, it is likely that even in a large replication sample the cocaine genetic risk score would not account for more than 0.25% of the variance in alcohol, tobacco, or marijuana dependence symptoms.

Limitations

The individuals included in the present analyses were not representative of the general population; rather, they were selected on the basis of reporting cocaine use from a larger genome-wide association study (SAGE; Bierut et al. 2010) that was over-sampled for alcohol dependence cases. The total SAGE sample was drawn from three primary studies designed to study alcohol, nicotine, and cocaine dependence. The SAGE study design results in greater levels of comorbidity among cocaine, alcohol, tobacco, and marijuana dependence, which could be expected to result in a more substance-general genetic risk score. Despite the clinical heterogeneity within our sample due to three different ascertainment schemes and despite a bias towards general substance risk factors, we still observed specificity of the genetic risk score to cocaine dependence symptoms within the testing sample.

Findings from genetic association studies such as the current analyses may be sensitive to the specific operationalization of the phenotype. While our measure of dependence symptom count provides greater phenotypic information than would be available from a diagnostic dichotomization of case versus control, future studies of measured genetic influences on substance dependence could potentially benefit by applying phenotype refinement strategies, such as aggregating across informant- and self-report measures (e.g. Achenbach et al. 2005) or including measures of frequency or quantity of use (e.g. Chen and Kandel 2002). By reducing the extent of measurement error in our phenotype of interest, such refinements may provide a way to improve our power to identify genetic associations.

Our approach of selecting only SNPs located in genes known to be directly involved in a neurotransmitter system of interest aimed to reduce the extent to which "noise" SNPs could be included in our genetic risk score. Such an approach also substantially limited the amount of variance in cocaine dependence that we would expect to explain. Although other non-dopamine systems, such as those involved in cocaine metabolism, likely contribute to individual differences in liability to cocaine addiction, the current total sample size of 1,591 individuals who reported having ever used cocaine required that a trade-off be made in terms of specificity of our target system (to reduce the chance of false-positives as a result of multiple testing) versus generality at the level of either the inclusion of multiple systems or even taking a genome-wide scoring approach. As a result, we were unable to compare the results of the current limited-scope investigation to the results of a potential multiple system or genome-wide scoring approach with regard to genetic influences on cocaine dependence. (To our knowledge there is currently no published GWAS of cocaine dependence in the existing literature on which we might be able to base a genome-wide scoring model.) Nevertheless, the substantial heritability of cocaine dependence suggests that there likely exist numerous influential genetic variants whose effects on cocaine dependence could be discovered using genome-wide approaches, given the availability of an adequately powered sample or meta-analysis. Further, the reliance of the current study on genotypes selected from existing genome-wide data necessarily limited our ability to include non-SNP variation (e.g., the frequently-investigated 48 bp VNTR in *DRD4*, Munafò et al. 2008). Thus, there may remain phenotypic variance attributable to non-SNP

polymorphisms within these dopaminergic genes that was unmodeled in our current analysis.

It is clear that our four-SNP score does not fully account for dopaminergic effects on cocaine dependence, nor are genetic influences on cocaine dependence likely to be explained by effects of the dopamine system alone. Future applications of a candidate system scoring approach may seek to include a broader selection of additional candidate systems, and may consider potential SNP–SNP interaction effects (either within genes, such as consideration of haplotype blocks, or between SNPs located in separate genes), although these would necessarily increase the complexity of the model. Of course, even the inclusion of genes from numerous systems a priori hypothesized to be involved in the target phenotype will not allow for this candidate system scoring approach to identify new variants or systems of interest, a goal for which the GWAS and genome-wide scoring approaches are clearly well-suited, given adequate sample sizes.

Conclusions

A risk score based upon four SNPs selected from a set of eight dopaminergic genes in an initial training sample significantly predicted cocaine dependence symptoms in a replication testing sample. We would expect a large number of genetic variants, each with relatively small effect sizes, to substantially impact complex, continuous phenotypes only when considered in the aggregate (e.g., Plomin et al. 2009). Because there is potential for a large number of variants each with small effects, it could greatly benefit researchers to attempt aggregate replication of promising individual variants, especially in cases where the recent trend of rapidly increasing sample sizes may be impractical (e.g., for phenotypes requiring time- or cost-intensive measurement or selected samples).

Given our modest sample size, and substantial evidence implicating the dopamine system in cocaine dependence, we sought to improve our power to detect SNP effects on cocaine dependence by limiting our analyses to SNPs located within a small set of dopamine-related genes and by examining the effects of these SNPs in aggregate. Although such a specific (e.g. candidate system) approach limits the potential for novel findings, such a limitation in scope is informed by the theoretical effect size of the influence of individual SNPs on complex phenotypes, such as substance dependence. The identification of four SNPs explaining 0.55% of the testing sample variance may appear both limited in scope and effect size, but this puts the individual SNPs at effect sizes of just over 0.1% each, which is well-aligned with current expectations for effect sizes of common variants.

The cocaine-derived score identified within the current training sample predicted only cocaine dependence severity in the testing sample; it did not account for genetic variance common across substances. This could be due to the specific relationship between dopamine and cocaine. There is extensive genetic, pharmacologic, and functional evidence, in both human and non-human animal models, explicating the direct relationship between cocaine and the dopaminergic system (e.g., Dackis and Gold 1985; Haile et al. 2007; Kuhar et al. 1991). While a substantial research literature supports an association between dopamine and other substances (e.g., alcohol, tobacco, and marijuana; Blum et al. 2000; Derauf et al. 2009), the effects are generally not as strong and the posited mechanism(s) of action not as direct as between dopamine and cocaine.

We detected four SNPs that in aggregate accounted for 0.55% of testing sample variance in cocaine dependence symptoms. These four SNPs were individually nominally significant in the training sample, but none were significant in the testing sample ($p = 0.14$–0.59). The effects of these SNPs in the testing sample were nominally significant only when considered in the aggregate, and replication of this effect is necessary before it may be considered reliable. Rather than imply that the effects of these SNPs are of primary importance in identifying individuals who are at increased risk of cocaine dependence, we endeavored to illustrate the potential utility of such an aggregate SNP approach to studying the influence of measured genotypes on behavioral or psychiatric phenotypes, for which large datasets are unavailable (making genome-wide association and scoring approaches likely substantially underpowered) and for which a reasonable body of research and/or theoretical literature exists to allow for the a priori selection of a pathway for inclusion in the construction of a targeted scoring approach. The current findings demonstrate that such a candidate system scoring approach may be potentially useful to the investigation of genetic effects on complex phenotypes. In such situations it is likely that numerous genetic variants within a system of biologically plausible genes each exert a relatively small effect, and the influence of these genes may only be reliably detected when they are considered in the aggregate. Therefore, we suggest that using strict criteria for individual candidate SNP replication across samples may discard SNPs that, if their effects were considered in aggregate, may have a significant impact on the phenotype of interest.

Acknowledgments *Financial Disclosures* Drs. LJ Bierut, AM Goate, JR Rice, and S Saccone are listed as inventors on the patent "Markers for Addiction" (US 20070258898) covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction. Dr. Bierut has acted as a consultant for Pfizer, Inc. in 2008. All other authors report no competing interests. *Study of Addiction: Genetics and Environment (SAGE)*. Funding support for

the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIH Genes, Environment and Health Initiative [GEI] (U01 HG004422). SAGE is one of the GWAS 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 (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of datasets and samples was provided by the Collaborative Study on the Genetics of Alcoholism (COGA; U10 AA008401), the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01 CA089392), and the Family Study of Cocaine Dependence (FSCD; R01 DA013423). 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). Dr. A Agrawal receives support from NIH DA23668. Dr. LJ Bierut receives support from NIH DA21237. Dr. J Derringer receives support from NIH DA029377 and GM081739. Dr. RA Grucza receives support from NIH AA017444 and DA026612. Dr. S Saccone receives support from NIH DA024722.

References

Achenbach TM, Krukowski RA, Dumenci L, Ivanova MY (2005) Assessment of adult psychopathology: meta-analyses and implications of cross-informant correlations. Psychol Bull 131: 361–382

Amin N, van Dujin CM, Janssens ACJW (2009) Genetic scoring analysis: a way forward in genome wide association studies? Eur J Epidemiol 24:585–587

Begleiter H, Reich T, Hesselbrock V, Porjesz B, Li T-K, Schuckit M, Edenberg H, Rice J (1995) The collaborative study on the genetics of alcoholism. Alcohol Health Res World 19:228–236

Beuming T, Kniazeff J, Bergmann M et al (2008) The binding sites for cocaine and dopamine in the dopamine transporter overlap. Nat Neurosci 11:780–789

Bierut LJ (2007) Genetic variation that contributes to nicotine dependence. Pharmacogenomics 8:881–883

Bierut LJ, Strickland JR, Thompson JR, Afful SE, Cottler LB (2008) Drug use and dependence in cocaine dependent subjects, community-based individuals, and their siblings. Drug Alcohol Depend 95:14–22

Bierut LJ, Agrawal A, Bucholz KK et al (2010) A genome wide association study of alcohol dependence. Proc Natl Acad Sci USA 107:5082–5087

Bloch PJ, Nall AH, Weller AE, Ferraro TN, Berrettini WH, Kamman KM, Pettinati HM, Dackis CA, O'Brien CP, Oslin DW, Lohoff FW (2009) Association analysis between polymorphisms in the dopamine D3 receptor (DRD3) gene and cocaine dependence. Psychiatr Genet 19:275–276

Blum K, Braverman ER, Holder JM, Lubar JF, Monastra VJ, Miller D, Lubar JO, Chen TJ, Coming DE (2000) Reward deficiency syndrome: a biogenetic model for the diagnosis and treatment of impulsive, addictive, and compulsive behaviors. J Psychoact Drugs 32:1–112

Briand LA, Flagel SB, Garcia-Fuster MJ et al (2008) Persistent alterations in cognitive function and prefrontal dopamine D2 receptors following extended, but not limited, access to self-administered cocaine. Neuropsychopharmacology 33:2969–2980

Bucholz KK, Cadoret R, Cloninger CR et al (1994) A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. J Study Alcohol 55:149–158

Buckland PR (2008) Will we ever find the genes for addiction? Addiction 103:1768–1776

Caine SB, Thomsen M, Gabriel KI et al (2007) Lack of self-administration of cocaine in dopamine D1 receptor knock-out mice. J Neurosci 27:13140–13150

Chen K, Kandel D (2002) Relationship between extent of cocaine use and dependence among adolescents and adults in the United States. Drug Alcohol Depend 68:65–85

Dackis CA, Gold MS (1985) New concepts in cocaine addiction: the dopamine depletion hypothesis. Neurosci Biobehav Rev 9:469–477

Derauf C, Kekatpure M, Neyzi N, Lester B, Kosofsky B (2009) Neuroimaging of children following prenatal drug exposure. Semin Cell Dev Biol 20:441–454

Derringer J, Krueger RF, Dick DM et al (2010) Predicting sensation seeking from dopamine genes: a candidate system approach. Psychol Sci 21:1282–1290

Evans DM, Visscher PM, Wray NR (2009) Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. Hum Mol Genet 18:3525–3531

Fernàndez-Castillo N, Ribasés M, Roncero C, Casas M, Gonzalvo B, Cormand B (2010) Association study between the DAT1, DBH and DRD2 genes and cocaine dependence in a Spanish sample. Psychiatr Genet 20:317–320

Ford JD, Gelernter J, DeVoe JS et al (2009) Association of psychiatric and substance use disorder comorbidity with cocaine dependence severity and treatment utilization in cocaine-dependent individuals. Drug Alcohol Depend 99:193–203

Ge D, Goldstein DB (2007) WGAViewer: package of whole genome association annotation. Version 1.10. Durham, NC

Gillespie N, Neale M, Prescott C, Aggen S, Kendler K (2007) Factor and item-response analysis DSM-IV criteria for abuse of and dependence on cannabis, cocaine, hallucinogens, sedatives, stimulants and opioids. Addiction 102:920–930

Giros B, Jaber M, Jones SR, Wightman RM, Caron MG (1996) Hyperlocomotion and indifference to cocaine and amphetamine in mice lacking the dopamine transporter. Nature 379: 606–612

Goldman D, Oroszi G, Ducci F (2005) The genetics of addictions: uncovering the genes. Nat Rev Genet 6:521–532

Gray JD (1993) Medical consequences of cocaine. Can Fam Physician 39(1975–1976):1979–1981

Haile CN, Kosten TR, Kosten TA (2007) Genetics of dopamine and its contribution to cocaine addiction. Behav Genet 37:119–145

Harford TC, Yi HY, Faden VB, Chen CM (2009) The dimensionality of DSM-IV alcohol used disorders among adolescent and adult drinkers and symptom patterns by age, gender, and race/ethnicity. Alcohol Clin Exp Res 33:868–878

Henry DJ, White FJ (1991) Repeated cocaine administration causes persistent enhancement of D1 dopamine receptor sensitivity within the rat nucleus accumbens. J Pharmacol Exp Ther 258: 882–890

Herzberg I, Valencia-Duarte AV, Kay VA, White DJ, Müller H, Rivas IC, Mesa SC, Cuartas M, Garcia J, Bedova G, Cornejo W, Ruiz-Linares A, Kremeyer B (2010) Association of DRD2 variants and Gilles de la Tourette syndrome in a family-based sample from a South American population isolate. Psychiatr Genet 20:179–183

Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V (1999) A validity study of the SSAGA—a comparison with the SCAN. Addiction 94:1361–1370

Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. PloS Genet 4:e1000008

Hnasko TS, Sotak BN, Palmiter RD (2007) Cocaine-conditioned place preference by dopamine-deficient mice is mediated by serotonin. J Neurosci 27:12484–12488

Johnson AD, Handsaker RE, Pulit S, Nizzari MM, O'Donnell CJ, de Bakker PIW (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24:2938–2939

Kendler KS, Prescott CA (1998) Cocaine use, abuse and dependence in a population-based sample of female twins. Br J Psychiatry 173:345–350

Kendler KS, Karkowski LM, Neale MC, Prescott CA (2000) Illicit psychoactive substance use, heavy use, abuse, and dependence in a US population-based sample of male twins. Arch Gen Psychiatry 57:261–269

Kuhar MJ, Ritz MC, Boja JW (1991) The dopamine hypothesis of the reinforcing properties of cocaine. Trends Neurosci 14:299–302

Lai IC, Mo GH, Chen ML, Wang YC, Chen JY, Liao DL, Bai YM, Lin CC, Chen TT, Liou YJ (2011) Analysis of genetic variations in the dopamine D1 receptor (DRD1) gene and antipsychotic-induced tardive dyskinesia in schizophrenia. Eur J Clin Pharmacol 67:383–388

Laurie CC, Doheny KF, Mirel DB et al (2010) Quality control and quality assurance in genotypic data for genome-wide association studies. Genet Epidemiol 34(6):591–602

Levran O, Londono D, O'Hara K, Randesi M, Rotrosen J, Casadonte P, Linzy S, Ott J, Adelson M, Kreek MJ (2009) Heroin addiction in African Americans: a hypothesis-driven association study. Genes Brain Behav 8:531–540

Liu QS, Pu L, Poo MM (2005) Repeated cocaine exposure in vivo facilitates LTP induction in midbrain dopamine neurons. Nature 437:1027–1031

Munafò MR, Matheson IJ, Flint J (2007) Association of the DRD2 gene Taq1A polymorphism and alcoholism: a meta-analysis of case-control studies and evidence of publication bias. Mol Psychiatry 12:454–461

Munafò MR, Yalcin B, Willis-Owen SA, Flint J (2008) Association of the dopamine D4 Receptor (DRD4) gene and approach-related personality traits: meta-analysis and new data. Biol Psychiatry 63:197–206

Munafò MR, Timpson NJ, David SP, Ebrahim S, Lawlow DA (2009) Association of the DRD2 gene Taq1A polymorphism and smoking behavior: a meta-analysis and new data. Nicotine Tob Res 11:64–76

Nemoda Z, Lyons-Ruth K, Szekely A, Bertha E, Faludi G, Sasvari-Szekely M (2010) Association between dopaminergic polymorphisms and borderline personality traits among at-risk adults and psychiatric inpatients. Behav Brain Funct 6:4

Nichol PE, Krueger RF, Iacono WG (2007) Investigating gender differences in alcohol problems: a latent trait modeling approach. Alcohol Clin Exp Res 31:783–794

Nutt D, King LA, Saulsbury W, Blakemore C (2007) Development of a rational scale to assess the harm of drugs of potential misuse. Lancet 369:1047–1053

Plomin R, Haowrth CM, Davis OS (2009) Common disorders are quantitative traits. Nat Rev Genet 10:872–878

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick Nam Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909

Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P, International Schizophrenia Consortium (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460:748–752

R Development Core Team (2009) R: a language and environment for statistical computing. Version 2.10.1. R foundation for statistical computing. Vienna, Austria. http://www.R-project.org

Robinson JD, Lam CY, Minnix JA, Wetter DW, Tomlinson GE, Minna JD, Chen TT, Cinciripini PM (2007) The DRD2 TaqI-B polymorphism and its relationship to smoking abstinence and withdrawal symptoms. Pharmacogenomics J 7:266–274

Rocha BA, Fumagalli F, Gainetdinov RR et al (1998) Cocaine self-administration in dopamine-transporter knockout mice. Nat Neurosci 1:132–137

Self DW, Barnhart WJ, Lehman DA, Nestler EJ (1996) Opposite modulation of cocaine-seeking behavior by D1- and D2-like dopamine receptor agonists. Science 271:1586–1589

Shaffer HJ, Eber GB (2002) Temporal progression of cocaine dependence symptoms in the US National Comorbidity Survey. Addiction 97:543–554

Sora I, Wichems C, Takahashi N et al (1998) Cocaine reward models: conditioned place preference can be established in dopamine- and in serotonin-transporter knockout mice. Proc Natl Acad Sci USA 95:7699–7704

Substance abuse and mental health services administration (SAMHSA) (2009) Results from the 2008 National Survey on Drug Use and Health: National Findings. NSDUH Series H-36. Office of spplied studies. Rockville, MD. www.samhsa.gov/

Thanos PK, Michaelides M, Benveniste H, Wang GJ, Volkow ND (2008) The effects of cocaine on regional brain glucose metabolism is attenuated in dopamine transporter knockout mice. Synapse 62:319–324

Ungless MA, Whistler JL, Malenka RC, Bonci A (2001) Single cocaine exposure in vivo induces long-term potentiation in dopamine neurons. Nature 411:583–587

van den Bree MB, Johnson EO, Neale MC, Pickens RW (1998) Genetic and environmental influences on drug use and abuse/dependence in male and female twins. Drug Alcohol Depend 52:231–241

Volkow ND, Wang GJ, Fischman MW et al (1997) Relationship between subjective effects of cocaine and dopamine transporter occupancy. Nature 386:827–830

Vukcevic D, Hechter E, Spencer C, Donnelly P (2011) Disease model distortion in association studies. Genet Epidemiol 35:278–290

Xu M, Hu XT, Cooper DC et al (1994) Elimination of cocaine-induced hyperactivity and dopamine-mediated neurophysiological effects in dopamine D1 receptor mutant mice. Cell 79: 945–955

Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: genetic interactions create phantom heritability. Proc Natl Acad Sci USA 109(4):1193–1198