Genome-wide association meta-analysis of cocaine dependence: Shared genetics with comorbid conditions

Judit Cabana-Domínguez[1,2,3,4], Anu Shivalikanjli[1,2,3,4], Noèlia Fernàndez-Castillo[1,2,3,4]*, Bru Cormand[1,2,3,4]*

[1]Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain

[2]Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, Spain

[3]Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Catalonia, Spain

[4]Institut de Recerca Sant Joan de Déu (IR-SJD), Esplugues de Llobregat, Catalonia, Spain

*Equally contributed

CORRESPONDING AUTHORS:

Bru Cormand. Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Avinguda Diagonal 643, edifici Prevosti, 3ª planta, 08028, Barcelona, Catalonia, Spain. Tel.: (+34) 93 4021013. Fax: (+34) 93 4034420; Email: bcormand@ub.edu

Noèlia Fernàndez Castillo. Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Avinguda Diagonal 643, edifici Prevosti, 3ª planta, 08028, Barcelona, Catalonia, Spain. Tel.: (+34) 93 4037082. Fax: (+34) 93 4034420; Email: noefernandez@ub.edu

Abstract

Cocaine dependence is a complex psychiatric disorder that is highly comorbid with other psychiatric traits. Twin and adoption studies suggest that genetic variants contribute substantially to cocaine dependence susceptibility, which has an estimated heritability of 65-79%. Here we performed a meta-analysis of genome-wide association studies of cocaine dependence using four datasets from the dbGaP repository (2,085 cases and 4,293 controls, all of them selected by their European ancestry). Although no genome-wide significant hits were found in the SNP-based analysis, the gene-based analysis identified *HIST1H2BD* as associated with cocaine-dependence (10% FDR). This gene is located in a region on chromosome 6 enriched in histone-related genes, previously associated with schizophrenia (SCZ). Furthermore, we performed LD Score regression analysis with comorbid conditions and found significant genetic correlations between cocaine dependence and SCZ, ADHD, major depressive disorder (MDD) and risk taking. We also found, through polygenic risk score analysis, that all tested phenotypes can significantly predict cocaine dependence status: SCZ (R^2=2.28%; P=1.21e-26), ADHD (R^2=1.39%; P=4.5e-17), risk taking (R^2=0.60%; P=2.7e-08), MDD (R^2=1.21%; P=4.35e-15), children's aggressiveness (R^2=0.3%; P=8.8e-05) and antisocial behavior (R^2=1.33%; P=2.2e-16). To our knowledge, this is the largest reported cocaine dependence GWAS meta-analysis in European-ancestry individuals. We identified suggestive associations in regions that may be related to cocaine dependence and found evidence for shared genetic risk factors between cocaine dependence and several comorbid psychiatric traits. However, the sample size is limited and further studies are needed to confirm these results.

Keywords: Cocaine addiction, cocaine dependence, genetic correlation, GWAS, poligenic risk score, meta-analysis.

1. Introduction

Cocaine is one of the most used illicit drugs worldwide and its abuse produces serious health problems. In Europe, around 5.2% of adults (from 15 to 64 years old) have tried cocaine (EMCDDA, 2017), but at most 20% will develop addiction (Wagner and Anthony, 2002). This information allows us to estimate the prevalence of cocaine dependence in the European population around 1.1%, similar to the prevalence observed in American populations (Compton et al., 2007).

Cocaine dependence is a complex psychiatric disorder that results from the interaction of environmental and genetic risk factors. It is one of the most heritable psychiatric conditions, with an estimated heritability of 65-79% (Ducci and Goldman, 2012). Although many case-control association studies in candidate genes have been performed, only a few risk variants for cocaine dependence have been identified and replicated so far, such as rs16969968 in the *CHRNA5* gene, encoding the cholinergic receptor nicotinic alpha 5 subunit, and rs806368 in *CNR1*, coding for the cannabinoid receptor 1 (Bühler et al., 2015). To date, only one genome-wide association study (GWAS) on cocaine dependence has been performed in European- and African-American individuals (Gelernter et al., 2014). When combining the two populations, one genome-wide finding was identified in the *FAM53B* gene, using a symptom count approach, but this hit could not be replicated in a subsequent study (Pineda-Cirera et al., 2018).

Several studies have shown that substance use disorders (SUD), and especially cocaine dependence, is highly comorbid with other psychiatric disorders and related phenotypes like aggressive, antisocial or risk-taking behaviors (Bezinović and Malatestinić, 2009; Hasin and Kilcoyne, 2012). For example, the occurrence of lifetime SUD in patients with schizophrenia (SCZ) is 70-80%, in attention deficit/hyperactivity disorder (ADHD) it is 39.2% and in major depressive disorder (MDD) it is 16.1% (Currie et al., 2005; Piñeiro-Dieguez et al., 2016;

Westermeyer, 2006). Conversely, about 81% of substance abuse/dependence patients have at least one comorbid mental disorder: 33% MDD, 11% SCZ and 9% personality disorders (Shantna et al., 2009). Such comorbidity is associated with an increase of severity for all disorders, although it is unclear whether this relationship is causal or the result of shared genetic and/or environmental risk factors. Some studies have started to inspect these relationships using both genetic correlation and polygenic risk score approaches, supporting the hypothesized role of shared genetic risk factors in the lifetime co-occurrence of several psychiatric disorders and SUD (Carey et al., 2016; Du Rietz et al., 2017; Hartz et al., 2017; Reginsson et al., 2018).

Here we performed a GWAS meta-analysis of cocaine dependence in samples with European ancestry using datasets from the dbGaP repository. Then we investigated the shared genetics between cocaine dependence and other psychiatric traits.

2. Experimental procedures

Detailed description of the materials and methods used, as well as supplementary figures, can be found in the Supplementary Information.

2.1 Subjects

The case-control GWAS meta-analysis was performed using four datasets from the dbGaP repository (https://www.ncbi.nlm.nih.gov/gap) under the project 17170 (Table 1). All cases used met DSM-IV criteria for cocaine dependence, although most of them are also dependent to other drugs of abuse. Drug abuse or dependence were discarded only in controls of the SAGE sample, whereas in the other studies general population individuals were used as controls.

Since samples 2-4 did not have enough controls to perform the association studies, we used controls from other datasets but genotyped with the same array and built using the same genome assembly. In all cases, patients and controls were from the same geographic area. In order to reduce bias, we merged controls with cases prior to quality control (QC) and imputation (Mitchell et al., 2014) (see Supplementary Information for detailed description of subjects used in this study).

The study was approved by the ethics committee of our institution, in accordance with the Helsinki Declaration and with the dbGaP protocols.

2.2 Quality control and association analyses

Prior to analysis, stringent QC was performed on the genotyped markers and individuals in each sample using the standardized pipeline "Ricopili" (Ripke, 2014). European-ancestry individuals were selected based on principal component analysis (PCA): PC1 and PC2 were used to define a genetic homogenous population, excluding individuals with PC values greater than three standard deviations from the reference population (European individuals from 1000 Genomes Project Phase 3 (1KGP3)). Related individuals and genetic outliers were excluded. A permutation test for between-group IBS differences with fixed 10,000 permutations was performed to discard population stratification between cases and controls.

After QC, non-genotyped markers were imputed using the European individuals from the 1KGP3 reference panel in MINIMAC3 (https://genome.sph.umich.edu/wiki/Minimac3).

For each sample, case-control GWAS was conducted using logistic regression under the additive model in PLINK v1.9 (http://pngu.mgh.harvard.edu/purcell/plink/). The 10 firsts PCs were included as covariates to correct for population stratification, and only variants with imputation INFO score >0.8 and minor allele frequency (MAF) >0.01 were considered.

2.3 GWAS meta-analysis

In total, 2,085 cases and 4,293 controls were meta-analyzed using an inverse-variance weighted fixed effects model in METAL (http://csg.sph.umich.edu//abecasis/Metal/)(Willer et al., 2010). Association results were considered only for variants with an effective sample size [N=2/((1/Ncases)+(1/Ncontrols))] greater than 70% of the full meta-analysis. Heterogeneity across studies was tested with the Cochran's Q test and quantified with the I^2 heterogeneity index, implemented in METAL.

Manhattan plot and Q-Q plot from each sample and the meta-analysis results were generated with the library "qqman" implemented in R.

2.4 LD Score intercept evaluation

LD score (LDSC) regression analysis was used to calculate LDSC intercept by regressing the chi-square statistics from GWAS against LD scores for each SNP (downloaded from GitHub website, https://github.com/bulik/ldsc)(Bulik-Sullivan et al., 2015b).

2.5 SNP heritability

LDSC 1.0.0 (https://github.com/bulik/ldsc/) was used to assess the proportion of phenotypic variance explained by common SNPs in the liability-scale, using a population prevalence for cocaine dependence of 1.1% (Compton et al., 2007).

Partitioned heritability analysis was performed based on 24 functional overlapping annotations described previously (Finucane et al., 2015). Enrichment in the heritability of a functional category was defined as the proportion of SNP heritability explained divided by the proportion of SNPs. The threshold for significance was calculated using the Bonferroni correction for multiple testing (P<2e-03).

2.6 Functional annotation of risk SNPs

SNPs were functionally annotated using FUMA (http://fuma.ctglab.nl/)(Watanabe et al., 2017). FUMA define lead SNPs as signals that are significantly associated with the phenotype of interest (we considered suggestive associations (*P*<1e-05)) and independent to each other (r^2<0.1)). For each lead SNP, a "Genomic risk locus" is defined, including all independent signals that are physically close or overlapping in a single locus. The variants located in a "Genomic risk locus" were explored considering the following functional annotations: eQTL (from GTEx v6/v7 and BRAINEAC), CADD_v1.3, ANNOVAR, RegulomeDB_v1.1, 15-core chromatin state and GWAS-catalog e91.

2.7 Gene-based and gene-set association analyses

Gene-based and gene-set association analyses were performed with MAGMA 1.05b (Willer et al., 2010) using the summary statistics from the cocaine dependence GWAs meta-analysis. For gene-based analysis, the SNP-wise mean model was used as the statistic test, considering p-values for SNPs located within the transcribed region. For multiple testing corrections, 10% FDR was considered.

Gene-set analysis was used to test for enrichment in association signals in genes belonging to specific biological pathways or processes. We performed a competitive test using: "All Canonical Pathways" (1329 gene sets), "GO" (4436 gene sets) and "BioCarta" (217 gene sets) provided by MsigDB 5.1 (https://software.broadinstitute.org/gsea/msigdb/)(Subramanian et al., 2005). Multiple testing corrections were applied to each gene set separately. When gene sets are strongly overlapping, the Bonferroni correction can be quite conservative, and for that reason, we used an empirical multiple testing correction implemented in MAGMA, based on a permutation procedure.

2.8 Shared genetic factors between cocaine dependence and comorbid conditions

2.8.1 Subjects

We studied shared genetic factors between cocaine dependence and six previously described comorbid conditions using publicly available GWAS summary statistics of SCZ, ADHD, MDD, children's aggressive behavior, antisocial behavior and risk-taking behavior (Table 2). Summary statistics from the vitamin D levels GWAS of the UK Biobank was used as a negative control.

2.8.2 LDSC Genetic correlation

Genetic correlation between cocaine dependence and the six selected comorbid disorders/traits was calculated using LDSC 1.0.0 (Bulik-Sullivan et al., 2015a). Summary statistics from all samples and pre-computed LD scores from HapMap3_SNPs calculated on 378 phased European-ancestry individuals from 1KGP3 were used (available at https://github.com/buligk/ldsc). After Bonferroni correction, the significance threshold was adjusted to P<7.1e-03.

Furthermore, the genetic correlation of cocaine dependence with other traits available at LD Hub (http://ldsc.broadinstitute.org/ldhub/)(Zheng et al., 2017) was evaluated (Bonferroni correction threshold P<7.2e-05).

2.8.3 Polygenic risk scores for cocaine dependence

Poligenic risk score (PRS) analysis was performed using the PRSice 2.1.0 software (https://github.com/choishingwan/PRSice)(Euesden et al., 2015).

The four cocaine dependence datasets were merged using PLINK v1.9 and used as a target sample. After merging, strict QC was applied resulting in 5,957,307 SNPs in 2,083 cases and 4,287 controls. To assess population stratification PCA was performed, and the 10 first PCs and a dummy variable indicating genotyping-study was included in the PRS analysis as a covariate.

Summary statistics of the seven phenotypes described above were used as discovery samples, which were clumped (r^2<0.1 in a 250-kb window) to remove SNPs in linkage disequilibrium

(LD). Then, PRSs were estimated for each discovery sample using a wild range of meta-analysis p-value thresholds (P_T) between P_T=1e-04 and P_T=1 at increments of 5e-05. For each P_T, the proportion of variance explained (R^2) by each discovery sample was computed by comparing the full model (PRS + covariates) score to a reduced model (covariates only). The reported R^2 value is the difference between R^2 from the two models. For quantitative traits we performed linear regression analysis, and for qualitative traits we used logistic regression and *Nagelkerke's pseudo-R^2* values are shown.

As recommended, we used the significance threshold of *P*=0.004 (Euesden et al., 2015). Bonferroni correction was applied considering the seven tested phenotypes (P<5.7e-04).

3. Results

3.1 GWAS results

We performed a GWAS meta-analysis of cocaine dependence using four datasets from the dbGaP repository. In total, we meta-analized 9,290,362 common genetic variants in 2,085 cases and 4,293 controls of European ancestry. No marker demonstrated significant heterogeneity between studies (Figure S1). The Q-Q plot (Figure 1A) displayed a λ of 1.06, comparable to other GWAS. The LDSC analyses estimated an intercept of 1.01 (SE=0.006; *P*=0.1), not significantly greater than 1, discarding residual population stratification or cryptic relatedness (Bulik-Sullivan et al., 2015b).

None of the analyzed variants reached the threshold for genome-wide significance (*P*<5e-08) in the SNP-based analysis, although we identified several suggestive associations (*P*<1e-05) (Table S1; Figure 1B).

3.2 Polygenic architecture of cocaine dependence

We applied LDSC analysis to assess the proportion of phenotypic variance explained by common SNPs. The estimated SNP heritability in liability-scale was h^2_{snp}=0.30 (SE=0.06; *P*=2.4e-07). Then we performed partitioned heritability analysis based on functional genomic categories and found significant enrichment in the heritability by SNPs located in intronic regions (enrichment=2.17; SE=0.45; *P*=1.2e-03), and a nominal result for conserved genomic regions (enrichment=23.63; SE=8.57; *P*=4e-03) (Figure S2).

3.3 Functional annotation of risk SNPs

To identify potentially interesting regions with FUMA we considered the SNPs showing suggestive associations (*P*<1e-05), as the SNP-based analysis did not reveal genome-wide significant hits (*P*<5e-08). We identified 23 lead SNPs which correspond to 22 genomic risk loci including 112 genes (Table 3). Interestingly, the risk locus located on chromosome 6 overlaps with a genomic region previously associated with schizophrenia. This region is defined by two lead SNPs (rs806973 and rs56401801, GWAs *P*=3.1e-06 and 3.4e-06, respectively) and it includes 77 genes and 458 nominally associated SNPs. Moreover, most of the SNPs in this region (447) are brain eQTLs for at least one member of a small group of 12 genes, including *BTN3A2*, *HIST1H2AK*, *ZSCAN31*, *PRSS16* and *ZNF184* (Figure 2 and S3).

3.4 Gene-based and gene-set analyses

The gene-based analysis mapped 3,418,270 SNPs from the GWAS meta-analysis to 18,069 protein-coding genes (Figure S4 and Table S2). The *HIST1H2BD* gene, located in a genomic region on chromosome 6 that showed a suggestive association in the SNP-based analysis, displayed a significant gene-wise association with cocaine dependence (10% FDR), although it did not overcome the Bonferroni correction for multiple testing. Then we performed competitive gene-set tests for all BioCarta, GO and Canonical Pathways. No gene sets attained a significant association with cocaine dependence after correction for multiple testing (Table

S3-5). However, interestingly, from the 10 GO gene sets with lower p-values, seven were related to synapse organization, glutamatergic neurotransmission and brain functions.

3.5 Cocaine dependence and shared genetic factors with comorbid conditions

Cocaine dependence is highly comorbid with other psychiatric disorders like SCZ, ADHD and MDD, and also with other phenotypes like aggressive behavior in children, antisocial behavior or risk taking. In order to investigate whether these phenotypic correlations are genetically mirrored, we performed a genetic correlation analysis using LDSC analysis and found significant genetic correlations (P<7.1e-03) between cocaine dependence and SCZ (rg=0.2; SE=0.05; P=1e-04), ADHD (rg=0.5; SE=0.08; P=1.6e-09), MDD (rg=0.4; SE=0.08; P=6.6e-07) and risk taking (rg=0.35; SE=0.06; P=9.1e-08) (Figure 3A). No significant correlations were found with children's aggression or antisocial behavior, nor with a negative control (Vitamin D levels).

On the other hand, we tested genetic correlations between cocaine dependence and all the GWAS summary statistics publicly available in the LD Hub. From the 690 tested traits, 109 demonstrated significant correlations after applying the Bonferroni correction, including a negative correlation with educational achievement (e.g. college completion) or with reproductive traits (e.g. age at first child) and a positive correlation with familiar situation (e.g. tobacco smoke exposure at home) or with several psychological or psychiatric traits like neuroticism, depressive symptoms or loneliness (Figure S5 and Table S6). The high number of significant associations may be explained by the high redundancy of the phenotypes of the LD Hub.

We also investigated the shared genetic etiology between cocaine dependence and comorbid phenotypes through PRS analysis, and tested whether these phenotypes can predict cocaine dependence status. For all the discovery samples tested, PRS significantly predicted cocaine dependence: SCZ (pseudo-R^2=2.28%, P_T=0.4911, P=1.21e-26), ADHD (pseudo-R^2=1.39%,

P_T=0.04275, *P*=4.5e-17), antisocial behavior (pseudo-R^2=1.33%, P_T=0.4055, *P*=2.2e-16), MDD (pseudo-R^2=1.21%, P_T=0.0129, *P*=4.35e-15), risk taking (R^2=0.60%, P_T=0.00135, *P*=2.7e-08) and children's aggressive behavior (R^2=0.30%, P_T=0.3552, *P*=8.8e-05) (Figure 3B and S6). In all cases, the quantile plot shows the positive nature of this relationship as cocaine dependence increases with greater polygenic load at each discovery dataset (Figure S7). As expected, the negative control based on vitamin D levels did not predict cocaine dependence (R^2=0.07%, P_T=0.03075, *P*=0.06).

4. Discussion

To our knowledge, this is the largest GWAS meta-analysis on cocaine dependence performed so far in individuals with European ancestry, although the sample size is still limited. No genome-wide significant (GWS) signals emerged from the SNP-based analysis, but the gene-based study identified *HIST1H2BD* as significantly associated with cocaine dependence. This gene is located in a region of chromosome 6 enriched in immune system and histone-related genes. These pathways have previously been associated with other psychiatric disorders like schizophrenia, major depressive disorder and bipolar disorder (O'Dushlaine, 2015). Despite the lack of GWS findings in this region, we identified many subthreshold variants. Based on the omnigenic model that has recently been proposed for complex disorders, these variants could point at regulatory elements of core genes (Boyle et al., 2017; Wray et al., 2018) and, therefore, contribute to the susceptibility to cocaine dependence, as most of them are brain eQTLs. Interestingly, this region overlaps with the genomic region that has been most consistently associated with schizophrenia. Indeed, it contains five SNPs (rs16897515, rs17693963, rs34706883, rs41266839 and rs55834529) nominally associated with cocaine dependence (*P*<1e-04) that had previously been associated with schizophrenia and with bipolar disorder, being the risk allele the same in all studies. The genetic variant most consistently associated with schizophrenia is rs17693963, reported in five different studies

(Bergen et al., 2012; Ripke et al., 2011, 2014; Ruderfer et al., 2014; Sleiman et al., 2013), which is a brain eQTL for *PRSS16*, *ZSCAN9, ZNF184* and *ZSCAN31*. Furthermore, a transcriptomic study performed in lymphoblastoid cell lines of 413 schizophrenic cases and 446 controls found that the top differentially expressed genes are located in this region (e.g. *HIST1H2BD*, *HIST1H2BC*, *HIST1H2BH*, *HIST1H2BG* and *HIST1H4K*) (Sanders et al., 2013).

Cocaine dependence is a highly heritable disorder (around 65-79% (Ducci and Goldman, 2012)). The results from LDSC analysis showed that common SNPs significantly capture almost a half of the cocaine dependence heritability (h^2_{SNP}=0.3; *P*=2.4e-07). Similar results were obtained previously for alcohol dependence (Mbarek et al., 2015) and other psychiatric disorders (Lee, 2013). However, a recent study estimated a lower SNP-heritability for alcohol dependence (h^2_{SNP}=0.09 for European-American subjects) using a larger GWAS meta-analysis on alcohol dependence (Walters et al., 2018).

It is well known that most psychiatric disorders are highly comorbid. About 73.4% of cocaine abuse/dependence patients have comorbid mental disorders: 49.7% have personality disorders (e.g. 5.3% schizoid and 17% antisocial personality) and 61.5% other mental disorders (e.g. 23.4% MDD and 20.5% anxiety) (Arias et al., 2013). However, the reasons for these covariations remain largely unknown. We investigated whether the phenotypic correlations between cocaine dependence and six comorbid psychiatric traits are genetically mirrored by performing genetic correlation analyses using LDSC. For the first time we found significant genetic correlation with ADHD, SCZ, MDD and risk-taking behavior, although these results should be taken with caution and need to be followed-up in a larger sample of cocaine-dependent individuals. Furthermore, we used the PRS method that, in contrast to LDSC, uses individual-level SNP data, resulting in higher statistical power and allowing for direct testing of interaction effects. According to our results, all the tested comorbid conditions can predict cocaine dependence status, suggesting that cocaine dependence is more likely in individuals

with many risk alleles for the tested conditions than in those with fewer risk alleles. To our knowledge, this is the first report of a shared genetic etiology between cocaine dependence and ADHD, antisocial behavior, risk-taking behavior and children's aggressive behavior based on genome-wide data. Previous studies have reported significant PRS associations between cocaine dependence and SCZ or MDD (Carey et al., 2016; Hartz et al., 2017; Reginsson et al., 2018), and also between SUD and other psychiatric disorders (Du Rietz et al., 2017; Gurriarán et al., 2018), although our study used the largest sample of cocaine dependence for this type of analysis so far. This correlation can reflect biological pleiotropy, where similar genetic mechanisms influence more than one trait, or mediated pleiotropy, where one phenotype is causally related to a second phenotype, so that the variants associated with this phenotype are indirectly associated with the second one.

This study has strengths and limitations that need some discussion. We performed a GWAS meta-analysis using all the cocaine dependence datasets available at the dbGaP repository, but we could not find any GWS association at the SNP-based level, as expected given the limited sample size, with a total of around 6,000 subjects, one third of them cases. However, these data allowed us to detect genetic correlations between cocaine dependence and several co-occurring conditions. Also, we calculated polygenic risk scores that explain a small fraction of the variance in the target phenotype, with figures that are similar to those obtained for other pairs of psychiatric conditions. To obtain a more comprehensive picture of the etiological overlap between cocaine dependence and comorbid conditions, larger studies will be needed, and other genetic factors should be included in the analyses (e.g. CNVs and rare variants). On the other hand, some of the dbGaP datasets used included only cases but not control individuals. For this reason, we used controls from other datasets that can introduce potential biases into the experimental design. Nevertheless, we performed very strict quality controls to avoid population stratification: the paired case and control samples were genotyped with the same platform and come from the same geographical area, the merging of the different

datasets was performed prior to quality control and imputation, and after that a permutation test was performed to discard population stratification (Mitchell et al., 2014). Moreover, the LDSC analyses confirmed that most of the observed inflation (λ =1.06) can be attributed to polygenicity rather than to residual population stratification or cryptic relatedness (Bulik-Sullivan et al., 2015b). Finally, the disease phenotype has not been excluded in most of the control samples, which may potentially dilute positive findings in the association study (but not lead to false positive results). However, based on the prevalence of cocaine dependence (about 1.1%), the probability of false negative results due to this effect is low, and similar controls were used in other GWAS of drug addiction (Ikeda et al., 2013; Johnson et al., 2016).

In conclusion, we reported the largest cocaine dependence GWAS meta-analysis on individuals of European ancestry, even though no GWS hits could be identified. Enlarging the sample size of this study would increase the chances to detect significant associations. However, the fact that our analyses highlighted a region on chromosome 6 that also pops-up in several schizophrenia GWAS supports the idea of shared genetic risk factors in these two comorbid disorders. This is in line with the significant results derived from the genetic correlation and PRS analyses in our study and in others. Finally, it would also be interesting to investigate the genetic pathways and neurobiological mechanisms that underlie the genetic overlap between cocaine dependence and comorbid traits.

Author disclosures

Contributors

JC-D and AS contributed to the acquisition of genotype data; JC-D performed the GWAS meta-analysis and secondary analyses with the assistance of AS; JC-D prepared the first draft of the manuscript and all figures and tables; BC and NF-C coordinated the study and supervised the

manuscript preparation. All authors contributed to the design of the study and approved the final manuscript.

Role of Funding Source

Major financial support for this research was received by BC from the Spanish 'Ministerio de Economía y Competitividad' (SAF2015-68341-R) and AGAUR, 'Generalitat de Catalunya' (2017-SGR-738). The research leading to these results has also received funding from the European Union Seventh Framework Program [FP7/2007-2013] under grant agreement n° 602805 (Aggressotype) and from the European Union H2020 Program [H2020/2014-2020] under grant agreements n° 667302 (CoCA) and 643051 (MiND). JC-D was supported by 'Generalitat de Catalunya' (2015 FI_B 00448), AS by the H2020 MiND project and NF-C by 'Centro de Investigación Biomédica en Red de Enfermedades Raras' (CIBERER) and by an EMBO short-term fellowship (ASTF 573-2016).

Conflict of interests

The authors declare no conflicts of interest.

Acknowledgments

We are really thankful to Paula Rovira, Marta Ribasés, Nina Roth Mota, Concepció Arenas and Oscar Lao for their support with the GWAS analysis. We are also very grateful to the developers of the software tools used, especially Stephan Ripke and Raymond Walters.

The authors acknowledge the contribution of data from dbGAP accessed through project number 17170: the Study of Addiction: Genetics and Environment (SAGE), one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI (supported by U01 HG004422, U01 HG004446, U01HG004438, HHSN268200782096C; and support for collection of datasets and samples was provided by the

COGA; U10 AA008401, COGEND; P01 CA089392 and FSCD; R01 DA013423); GWAS of Heroin Dependence (supported by R01DA17305); GWAS of Cocaine Dependence in Two Populations (supported by RC2 DA028909, R01 DA12690, R01 DA12849, R01 DA18432, R01 AA11330, and R01 AA017535); CRIC Study (supported by U01DK060990, U01DK060984, U01DK061022, U01DK061021, U01DK061028, U01DK60980, U01DK060963 and U01DK060902); CIDR-Gelernter Study, a genome-wide association study funded as part of the Genetics of Alcohol Dependence in American Populations (supported by U01HG004438 and HHSN268200782096C); the COPDGene study (supported by U01HL089856 and U01HL089897, and the COPD Foundation through contributions made by an Industry Advisory Board comprised of Pfizer, AstraZeneca, Boehringer Ingelheim, Novartis, and Sunovion); and Personalized Medicine Research Project (PMRP) (supported by U01HG004608, U01HG004438 and U01HG004603).

References

Arias, F., Szerman, N., Vega, P., Mesias, B., Basurte, I., Morant, C., Ochoa, E., Poyo, F., Babin, F., 2013. Abuso o dependencia a la cocaína y otros trastornos psiquiátricos. Estudio Madrid sobre la prevalencia de la patología dual. Rev. Psiquiatr. Salud Ment. 6, 121–128. https://doi.org/10.1016/j.rpsm.2012.09.002

Bergen, S.E., O'Dushlaine, C.T., Ripke, S., Lee, P.H., Ruderfer, D.M., Akterin, S., Moran, J.L., Chambert, K.D., Handsaker, R.E., Backlund, L., Ösby, U., McCarroll, S., Landen, M., Scolnick, E.M., Magnusson, P.K.E., Lichtenstein, P., Hultman, C.M., Purcell, S.M., Sklar, P., Sullivan, P.F., 2012. Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. Mol. Psychiatry 17, 880–886. https://doi.org/10.1038/mp.2012.73

Bezinović, P., Malatestinić, D., 2009. Perceived exposure to substance use and risk-taking behavior in early adolescence: cross-sectional study. Croat. Med. J. 50, 157–64. https://doi.org/10.3325/CMJ.2009.50.157

Boyle, E.A., Li, Y.I., Pritchard, J.K., 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169, 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Bühler, K.M., Giné, E., Echeverry-Alzate, V., Calleja-Conde, J., De Fonseca, F.R., López-Moreno, J.A., 2015. Common single nucleotide variants underlying drug addiction: More than a decade of research. Addict. Biol. 20, 845–871. https://doi.org/10.1111/adb.12204

Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., Duncan, L., Perry, J.R.B., Patterson, N., Robinson, E.B., Daly, M.J., Price, A.L., Neale, B.M., 2015a. An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47, 1236–1241. https://doi.org/10.1038/ng.3406

Bulik-Sullivan, B., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M., Neale, B.M., 2015b. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47, 291–295. https://doi.org/10.1038/ng.3211

Carey, C.E., Agrawal, A., Bucholz, K.K., Hartz, S.M., Lynskey, M.T., Nelson, E.C., Bierut, L.J., Bogdan, R., 2016. Associations between Polygenic Risk for Psychiatric Disorders and Substance Involvement. Front. Genet. 7, 149. https://doi.org/10.3389/fgene.2016.00149

Compton, W.M., Thomas, Y.F., Stinson, F.S., Grant, B.F., 2007. Prevalence, correlates, disability, and comorbidity of DSM-IV drug abuse and dependence in the United States: Results from the national epidemiologic survey on alcohol and related conditions. Arch. Gen. Psychiatry 64, 566–576. https://doi.org/10.1001/archpsyc.64.5.566

Currie, S.R., Patten, S.B., Williams, J.V.A., Wang, J., Beck, C.A., El-Guebaly, N., Maxwell, C., 2005. Comorbidity of major depression with substance use disorders. Can. J. Psychiatry 50, 660–666. https://doi.org/10.1177/070674370505001013

Du Rietz, E., Coleman, J., Glanville, K., Choi, S.W., O'Reilly, P.F., Kuntsi, J., 2017. Association of Polygenic Risk for Attention-Deficit/Hyperactivity Disorder With Co-occurring Traits and Disorders. Biol. Psychiatry Cogn. Neurosci. Neuroimaging. https://doi.org/10.1016/J.BPSC.2017.11.013

Ducci, F., Goldman, D., 2012. The Genetic Basis of Addictive Disorders. Psychiatr. Clin. North Am. https://doi.org/10.1016/j.psc.2012.03.010

Euesden, J., Lewis, C.M., O'Reilly, P.F., 2015. PRSice: Polygenic Risk Score software. Bioinformatics 31, 1466–1468. https://doi.org/10.1093/bioinformatics/btu848

European Monitoring Centre for Drugs and Drug Addiction, 2017. European Drug Report 2017: Trends and Developments, Publications Office of the European Union, Luxembourg. https://doi.org/10.2810/610791

Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., Ripke, S., Day, F.R., Purcell, S., Stahl, E., Lindstrom, S., Perry, J.R.B., Okada, Y., Raychaudhuri, S., Daly, M.J., Patterson, N., Neale, B.M., Price, A.L., Patterson, N., Neale, B.M., Price, A.L., 2015. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235. https://doi.org/10.1038/ng.3404

Gelernter, J., Sherva, R., Koesterer, R., Almasy, L., Zhao, H., Kranzler, H.R., Farrer, L., 2014. Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. Mol. Psychiatry 19, 717–723. https://doi.org/10.1038/mp.2013.99

Gurriarán, X., Rodríguez-López, J., Flórez, G., Pereiro, C., Fernández, J.M., Fariñas, E., Estévez, V., Arrojo, M., Costas, J., GenPol Study Group, 2018. Relationships between substance abuse/dependence and psychiatric disorders based on polygenic scores. Genes, Brain Behav. e12504. https://doi.org/10.1111/gbb.12504

Hartz, S.M., Horton, A.C., Oehlert, M., Carey, C.E., Agrawal, A., Bogdan, R., Chen, L.-S., Hancock, D.B., Johnson, E.O., Pato, C.N., Pato, M.T., Rice, J.P., Bierut, L.J., 2017. Association Between Substance Use Disorder and Polygenic Liability to Schizophrenia. Biol. Psychiatry 82, 709–715. https://doi.org/10.1016/j.biopsych.2017.04.020

Hasin, D., Kilcoyne, B., 2012. Comorbidity of psychiatric and substance use disorders in the United States: current issues and findings from the NESARC. Curr. Opin. Psychiatry 25, 165–71. https://doi.org/10.1097/YCO.0b013e3283523dcc

Ikeda, M., Okahisa, Y., Aleksic, B., Won, M., Kondo, N., Naruse, N., Aoyama-Uehara, K., Sora, I., Iyo, M., Hashimoto, R., Kawamura, Y., Nishida, N., Miyagawa, T., Takeda, M., Sasaki, T., Tokunaga, K., Ozaki, N., Ujike, H., Iwata, N., 2013. Evidence for Shared Genetic Risk Between Methamphetamine-Induced Psychosis and Schizophrenia. Neuropsychopharmacology 38, 1864–1870. https://doi.org/10.1038/npp.2013.94

Johnson, E.O., Hancock, D.B., Levy, J.L., Gaddis, N.C., Page, G.P., Glasheen, C., Saccone, N.L., Bierut, L.J., Kral, A.H., 2016. KAT2B polymorphism identified for drug abuse in African Americans with regulatory links to drug abuse pathways in human prefrontal cortex. Addict. Biol. 21, 1217–1232. https://doi.org/10.1111/adb.12286

Lee, S.H., Cross-Disorder Group of the Psychiatric GenomicsConsortium, 2013. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat. Genet. 45, 984–994. https://doi.org/10.1038/ng.2711

Mbarek, H., Milaneschi, Y., Fedko, I.O., Hottenga, J.-J., de Moor, M.H.M., Jansen, R., Gelernter,

J., Sherva, R., Willemsen, G., Boomsma, D.I., Penninx, B.W., Vink, J.M., 2015. The genetics of alcohol dependence: Twin and SNP-based heritability, and genome-wide association study based on AUDIT scores. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. 168, 739–748. https://doi.org/10.1002/ajmg.b.32379

Mitchell, B.D., Fornage, M., McArdle, P.F., Cheng, Y.C., Pulit, S.L., Wong, Q., Dave, T., Williams, S.R., Corriveau, R., Gwinn, K., Doheny, K., Laurie, C.C., Rich, S.S., de Bakker, P.I.W., 2014. Using previously genotyped controls in genome-wide association studies (GWAS): Application to the Stroke Genetics Network (SiGN). Front. Genet. 5, 95. https://doi.org/10.3389/fgene.2014.00095

O'Dushlaine, C., The Network and Pathway Analysis Subgroup of the Psychiatric GenomicsConsortium, 2015. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. Nat. Neurosci. 18, 199–209. https://doi.org/10.1038/nn.3922

Pappa, I., St Pourcain, B., Benke, K., Cavadino, A., Hakulinen, C., Nivard, M.G., Nolte, I.M., Tiesler, C.M.T., Bakermans-Kranenburg, M.J., Davies, G.E., Evans, D.M., Geoffroy, M.C., Grallert, H., Groen-Blokhuis, M.M., Hudziak, J.J., Kemp, J.P., Keltikangas-Järvinen, L., McMahon, G., Mileva-Seitz, V.R., Motazedi, E., Power, C., Raitakari, O.T., Ring, S.M., Rivadeneira, F., Rodriguez, A., Scheet, P.A., Seppälä, I., Snieder, H., Standl, M., Thiering, E., Timpson, N.J., Veenstra, R., Velders, F.P., Whitehouse, A.J.O., Smith, G.D., Heinrich, J., Hypponen, E., Lehtimäki, T., Middeldorp, C.M., Oldehinkel, A.J., Pennell, C.E., Boomsma, D.I., Tiemeier, H., 2016. A genome-wide approach to children's aggressive behavior: The EAGLE consortium. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. 171, 562–572. https://doi.org/10.1002/ajmg.b.32333

Pineda-Cirera, L., Cabana-Domínguez, J., Roncero, C., Cozar, M., Grau-López, L., Abad, A.C., Martínez-Luna, N., Robles-Martínez, M., Sánchez-Mora, C., Ramos-Quiroga, J.A., Casas,

M., Ribasés, M., Fernàndez-Castillo, N., Cormand, B., 2018. Evaluation of previous substance dependence genome-wide significant findings in a Spanish sample. Drug Alcohol Depend. 187, 358–362. https://doi.org/10.1016/j.drugalcdep.2018.03.013

Piñeiro-Dieguez, B., Balanzá-Martínez, V., García-García, P., Soler-López, B., Domingo, M.A., Labarra, J.D.A., Lobato, P.A., Salamanca, A.A., Bes, J.A., Fernández, F.J.B., Moraga, R.B., Blanco, J.B., Perona, A.B., Boniatti, T.C., Gras, J.M.C., Martínez, J.M.C., Machado, I.C., Pena, M.C.C., Flores, C.C., Dono, C.C., Taracena, M.T.D.L., García, B.D.P., Martín, P.F., Del Valle, E.F.G., Fontova, E., Eroles, F.F., Chapa, R.F., Tejedor, I.G., Palomares, A.G., Campayo, J.G., Moreno, J.M.G., Pérez, X.G., Escrig, M.D.C.G., Vives, S.G., Queralt, E.G., Gonochategui, A.G., Escajedo, A.G., Labrador, R.G., Mendoza, A.H., Lorenzo, F.G.I., Iribarren, M.M., Yepez, G.J.D.F., Vazquez, M.A.L., Borrajo, J.M.L., Arribas, S.L., García, J.M.M., Flores, A.M., Fernández-Mayoralas, D.M., Navarro, N.M., Recuero, L.M., Gras, I.M., Raga, J.M., López, P.A.M., Horche, J.L.M., Flores, P.L.M., Diago, E.O., Panicali, F., Pastor, F.P., Alfaro, G.P., Rosati, S.P., Sanz, M.P., Presa, A.M.I.Q., Alandes, C.R., Solano, J.J.R., Buguña, J.R.S., Izquierdo, G.S.N., García, M.S., De La Garza, C.L.S., Granado, O.S., Murla, F.S., Ceballos, P.S., Cabra, O.S., Rodríguez, P.A.S., Sender, V.T., Moreira, R.V., Abarquero, M.A.V., Peri, R.V., Mateo, L.V., Zambrano, D., Ramos, J.M.Z., 2016. Psychiatric Comorbidity at the Time of Diagnosis in Adults With ADHD: The CAT Study. J. Atten. Disord. 20, 1066–1075. https://doi.org/10.1177/1087054713518240

Reginsson, G.W., Ingason, A., Euesden, J., Bjornsdottir, G., Olafsson, S., Sigurdsson, E., Oskarsson, H., Tyrfingsson, T., Runarsdottir, V., Hansdottir, I., Steinberg, S., Stefansson, H., Gudbjartsson, D.F., Thorgeirsson, T.E., Stefansson, K., 2018. Polygenic risk scores for schizophrenia and bipolar disorder associate with addiction. Addict. Biol. 23, 485. https://doi.org/10.1111/ADB.12496

Ripke, S., Schizophrenia Psychiatric GWAS Consortium, 2011. Genome-wide association study

identifies five new schizophrenia loci. Nat. Genet. 43, 969–76. https://doi.org/10.1038/ng.940

Ripke, S., Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421–427. https://doi.org/10.1038/nature13595

Ruderfer, D.M., Fanous, A.H., Ripke, S., McQuillin, A., Amdur, R.L., Gejman, P. V, O'Donovan, M.C., Andreassen, O.A., Djurovic, S., Hultman, C.M., Kelsoe, J.R., Jamain, S., Landén, M., Leboyer, M., Nimgaonkar, V., Nurnberger, J., Smoller, J.W., Craddock, N., Corvin, A., Sullivan, P.F., Holmans, P., Sklar, P., Kendler, K.S., 2014. Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. Mol. Psychiatry 19, 1017–1024. https://doi.org/10.1038/mp.2013.138

Sanders, A.R., Göring, H.H.H., Duan, J., Drigalenko, E.I., Moy, W., Freda, J., He, D., Shi, J., MGS, Gejman, P. V, 2013. Transcriptome study of differential expression in schizophrenia. Hum. Mol. Genet. 22, 5001–14. https://doi.org/10.1093/hmg/ddt350

Shantna, K., Chaudhury, S., Verma, A., Singh, A., 2009. Comorbid psychiatric disorders in substance dependence patients: A control study. Ind. Psychiatry J. 18, 84–87. https://doi.org/10.4103/0972-6748.62265

Sleiman, P., Wang, D., Glessner, J., Hadley, D., Gur, R.E., Cohen, N., Li, Q., Hakonarson, H., Sleiman, P., Glessner, J., Hadley, D., Kao, C., Wu, Z., Kim, C., Thomas, K., Hakonarson, H., Wang, D., Favis, R., Fu, D.-J., Chung, H., Savitz, A., Gopal, S., Cohen, N., Li, Q., 2013. GWAS meta analysis identifies TSNARE1 as a novel Schizophrenia / Bipolar susceptibility locus. Sci. Rep. 3, 3075. https://doi.org/10.1038/srep03075

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment

analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. 102, 15545–15550. https://doi.org/10.1073/pnas.0506580102

Tielbeek, J.J., Johansson, A., Polderman, T.J.C., Rautiainen, M.-R., Jansen, P., Taylor, M., Tong, X., Lu, Q., Burt, A.S., Tiemeier, H., Viding, E., Plomin, R., Martin, N.G., Heath, A.C., Madden, P.A.F., Montgomery, G., Beaver, K.M., Waldman, I., Gelernter, J., Kranzler, H.R., Farrer, L.A., Perry, J.R.B., Munafò, M., LoParo, D., Paunio, T., Tiihonen, J., Mous, S.E., Pappa, I., de Leeuw, C., Watanabe, K., Hammerschlag, A.R., Salvatore, J.E., Aliev, F., Bigdeli, T.B., Dick, D., Faraone, S. V., Popma, A., Medland, S.E., Posthuma, D., 2017. Genome-Wide Association Studies of a Broad Spectrum of Antisocial Behavior. JAMA Psychiatry. https://doi.org/10.1001/jamapsychiatry.2017.3069

Wagner, F.A., Anthony, J.C., 2002. From first drug use to drug dependence: Developmental periods of risk for dependence upon marijuana, cocaine, and alcohol. Neuropsychopharmacology 26, 479–488. https://doi.org/10.1016/S0893-133X(01)00367-0

Walters, R.K., Adams, M.J., Adkins, A.E., Aliev, F., Bacanu, S.-A., Batzler, A., Bertelsen, S., Biernacka, J., Bigdeli, T.B., Chen, L.-S., Clarke, T.-K., Chou, Y.-L., Degenhardt, F., Docherty, A.R., Fontanillas, P., Foo, J., Fox, L., Frank, J., Giegling, I., Gordon, S., Hack, L., -, Hartz, S.M., Heilmann-Heimbach, S., Herms, S., Hodgkinson, C., Hoffmann, P., Hottenga, J.-J., Kennedy, M.A., Alanne-Kinnunen, M., Konte, B., Lahti, J., Lahti-Pulkkinen, M., Ligthart, L., Loukola, A.-M., Maher, B.S., Mbarek, H., McIntosh, A.M., McQueen, M.B., Milaneschi, Y., Palviainen, T., Pearson, J.F., Peterson, R.E., Polimanti, R., Ripatti, S., Ryu, E., Saccone, N.L., Salvatore, J.E., Sanchez-Roige, S., Schwandt, M., Sherva, R., Streit, F., Strohmaier, J., Thomas, N., Wang, J.-C., Webb, B.T., Wedow, R., Wetherill, L., Wills, A.G., Team, 23andMe Research, Boardman, J.D., Chen, D., Choi, D.-S., Copeland, W.E., Culverhouse, R.C., Dahmen, N., Degenhardt, L., Domingue, B.W., Elson, S.L., Frye, M., Gäbel, W., Ising,

M., Johnson, E.C., Keyes, M., Kiefer, F., Kramer, J., Kuperman, S., Lucae, S., Lynskey, M.T., Maier, W., Mann, K., Männistö, S., McClintick, J.N., Meyers, J.L., Müller-Myhsok, B., Nurnberger, J.I., Palotie, A., Preuss, U., Räikkönen, K., Reynolds, M.D., Ridinger, M., Scherbaum, N., Shuckit, M., Soyka, M., Treutlein, J., Witt, S., Wodarz, N., Zill, P., Adkins, D.E., Boden, J.M., Boomsma, D., Bierut, L.J., Brown, S.A., Bucholz, K.K., Cichon, S., Costello, E.J., Wit, H. de, Diazgranados, N., Dick, D.M., Eriksson, J.G., Farrer, L.A., Foroud, T.M., Gillespie, N.A., Goate, A.A., Goldman, D., Grucza, R.A., Hancock, D.B., Harris, K.M., Heath, A.C., Hesselbrock, V., Hewitt, J.K., Hopfer, C., Horwood, J., Iacono, W., Johnson, E.O., Kaprio, J.A., Karpyak, V., Kendler, K.S., Kranzler, H.R., Krauter, K., Lichtenstein, P., Lind, P.A., McGue, M., MacKillop, J., Madden, P.A.F., Maes, H., Magnusson, P., Martin, N.G., Medland, S.E., Montgomery, G.W., Nelson, E.C., Nöthen, M., Palmer, A.A., Pedersen, N.L., Penninx, B.W.J.H., Porjesz, B., Rice, J.P., Rietschel, M., Riley, B.P., Rose, R., Rujescu, D., Shen, P.-H., Silberg, J., Stallings, M.C., Tarter, R.E., Vanyukov, M.M., Vrieze, S., Wall, T.L., Whitfield, J.B., Zhao, H., Neale, B.M., Gelernter, J., Edenberg, H.J., Agrawal, A., 2018. Trans-ancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. bioRxiv 257311. https://doi.org/10.1101/257311

Watanabe, K., Taskesen, E., Van Bochoven, A., Posthuma, D., 2017. Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. 8, 1826. https://doi.org/10.1038/s41467-017-01261-5

Westermeyer, J., 2006. Comorbid schizophrenia and substance abuse: A review of epidemiology and course. Am. J. Addict. https://doi.org/10.1080/10550490600860114

Willer, C.J., Li, Y., Abecasis, G.R., 2010. METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191. https://doi.org/10.1093/bioinformatics/btq340

Wray, N.R., The Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, 2018. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat. Genet. 50, 668–681. https://doi.org/10.1038/s41588-018-0090-3

Wray, N.R., Wijmenga, C., Sullivan, P.F., Yang, J., Visscher, P.M., 2018. Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. Cell 173, 1573–1580. https://doi.org/10.1016/j.cell.2018.05.051

Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., Pourcain, B. St., Warrington, N.M., Finucane, H.K., Price, A.L., Bulik-Sullivan, B.K., Anttila, V., Paternoster, L., Gaunt, T.R., Evans, D.M., Neale, B.M., 2017. LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics 33, 272–279. https://doi.org/10.1093/bioinformatics/btw613

Figures

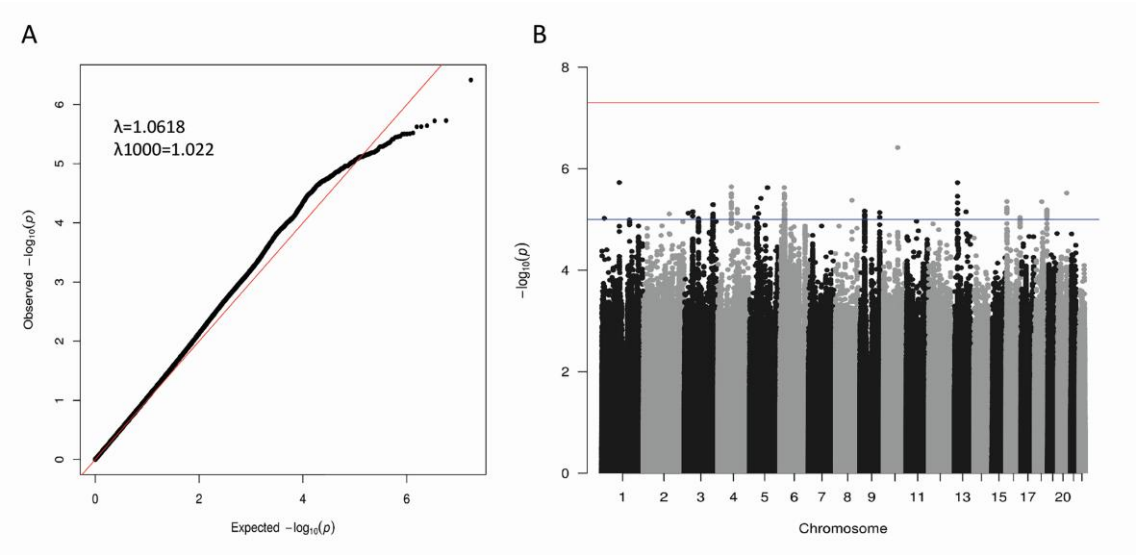

Figure 1. Results from the GWAS meta-analysis on cocaine dependence. A) Q-Q plot and B) Manhattan plot. Red line: threshold for genome-wide significance (*P* < 5e-08). Blue line: threshold for suggestive associations (*P* < 1e-05).

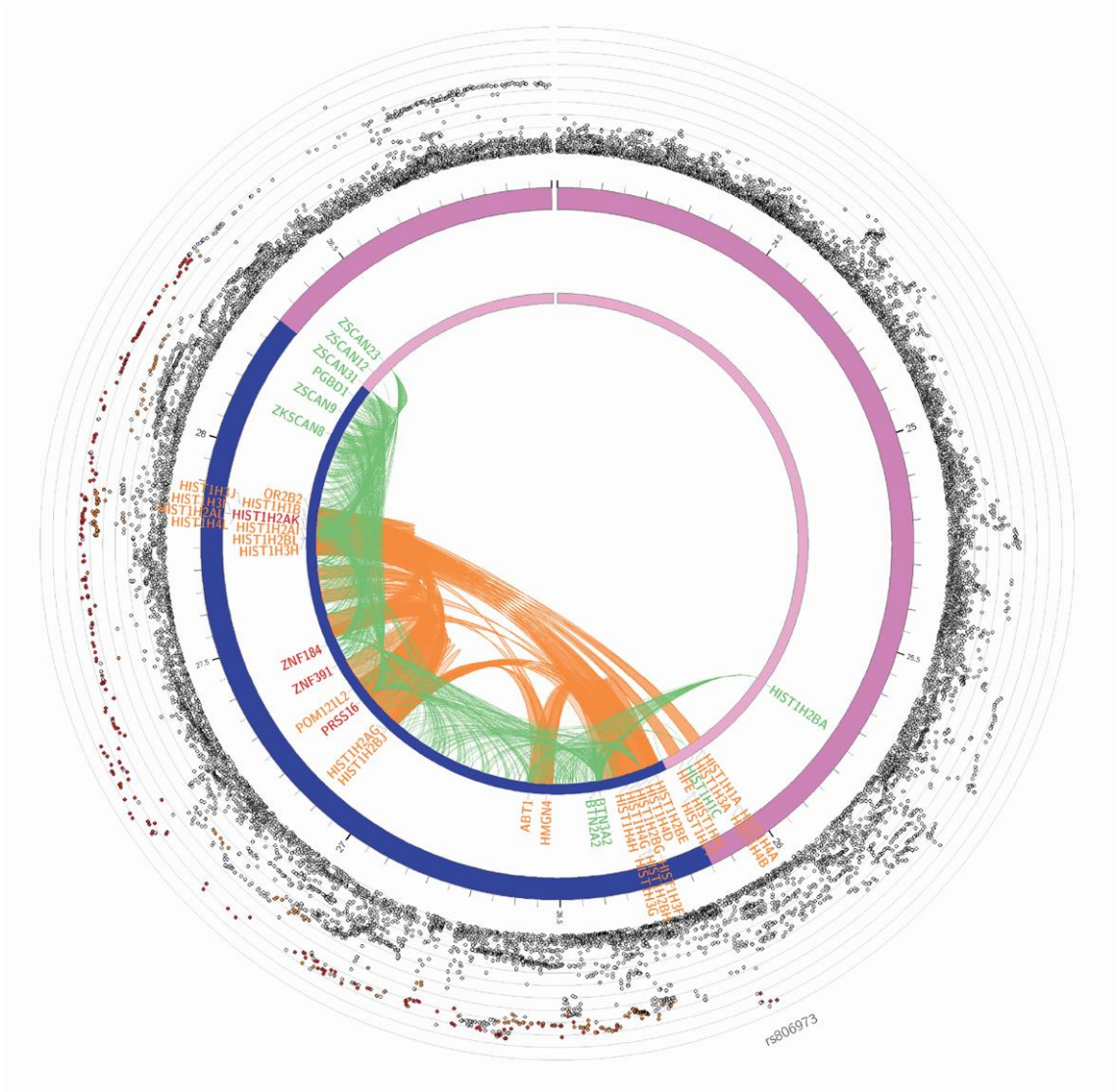

Figure 2. Circo-plot from genomic risk loci on chromosome 6 by FUMA. The most outer layer is the Manhattan plot (only SNPs with $P < 0.05$ are displayed). SNPs in genomic risk loci are color-coded as a function of their maximum r^2 to the lead SNPs in the locus, as follows: red ($r^2 > 0.8$), orange ($r^2 > 0.6$), green ($r^2 > 0.4$), blue ($r^2 > 0.2$) and grey ($r^2 \leq 0.2$). The rs ID of the top SNPs in the risk locus is displayed in the most outer layer. Y-axis is ranged between 0 to the maximum $-\log_{10}$(p-value) of the SNPs. The second layer is the chromosome ring, with the genomic risk locus highlighted in blue. Here genes are mapped by chromatin interactions (orange) or eQTLs (green). When the gene is mapped by both, it is colored in red.

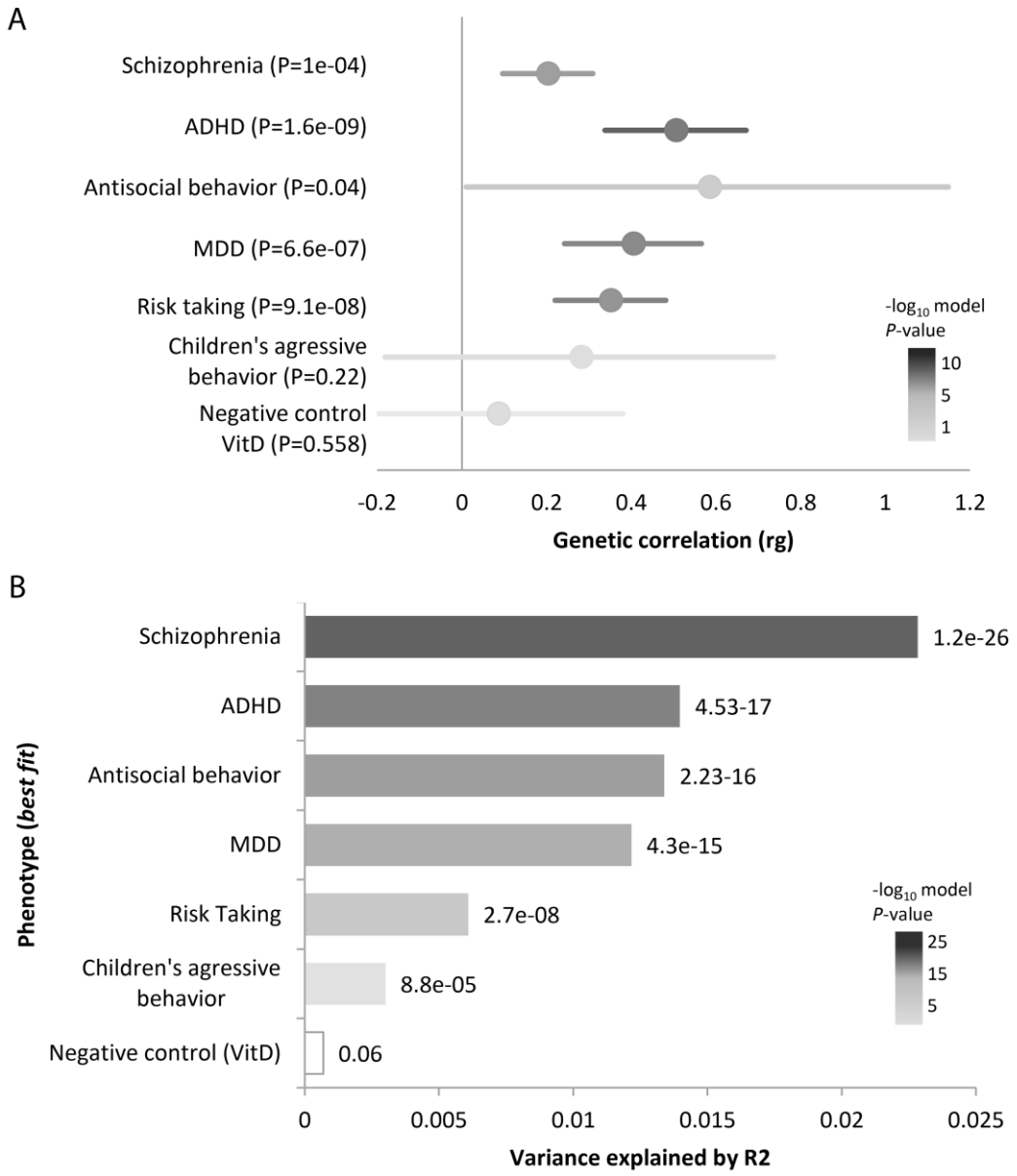

Figure 3. Shared genetic factors between cocaine dependence and comorbid conditions. A) Results from LD Score (LDSC) regression analysis showing genetic correlation (rg) between cocaine dependence and several traits. Error bars indicate 95% confidence limits. The significance threshold was set to $P < 7.1e-03$. B) Best fit results from Polygenic Risk Score (PRS) analysis for each tested phenotype. Values displayed next to each bar represent the p-value for significance for the most predictive models. The significance threshold was set at $P < 5.7e-04$.

Table 1. Description of dbGaP samples used in the cocaine dependence GWAS meta-analysis

Sample	N cases	%F cases	N controls	%F controls	Genotyping chip	dbGaP code
Sample 1 (SAGE)	468	39.3%	1284	69.9%	Illumina ILMN_Human-1	phs000092.v1.p1 (cases/controls)
Sample 2	609	45.9%	410	50.2%	Illumina HumanOmni1-Quad_v1-0_B	phs000952.v1.p1 (cases/controls) + phs000179.v5.p2 (controls)
Sample 3	504	44.3%	1190	62%	Illumina Human660W-Quad_v1_A	phs000277.v1.p1 (cases/controls) + phs000170.v2.p1 (controls)
Sample 4	504	36.7%	1409	41.9%	Illumina HumanOmni1-Quad_v1-0_B	phs000425.v1.p1 (cases/controls) + phs000524.v1.p1 (controls)
TOTAL	**2,085**		**4,293**			

%F= percentage of females

Table 2. Description of samples used to inspect shared genetic factors between cocaine dependence and comorbid conditions

GWAS	Abbreviation	N individuals	Reference	Website
Schizophrenia European meta-analysis	SCZ	34,241 cases and 45,604 controls 1,235 parent affected-offspring trios	Ripke *et al*, 2014	Psychiatric Genomics Consortium (PGC): https://www.med.unc.edu/pgc/results-and-downloads
Attention deficit/hyperactivity disorder European meta-analysis	ADHD	19,099 cases and 34,194 controls	Demontis *et al*, 2017	
Major depressive disorder meta-analysis	MDD	59,851 cases and 113,154 controls	Wray *et al*, 2018	
Children's aggressive behavior GWAS from EAGLE	Child-Aggre	18,988 individuals	Pappa *et al*, 2016	
Antisocial behavior meta-analysis	ASB	16,400 individuals	Tielbeek *et al*, 2017	BroadABC: http://broadabc.ctglab.nl/summary_statistics
Risk-taking behavior dataset	RT	325,821 individuals		UK Biobank: https://sites.google.com/broadinstitute.org/ukbbgwasresults/home?authuser=0

Table 3. FUMA analysis of genetic risk loci for cocaine dependence as identified from the GWAS meta-analysis

Genomic Locus	Position*	Lead SNP	SNP position	GWAS P-value	nSNPs	nGenes	Genes
1	chr1:15916714-16049893	rs148069235	15950064	9.52e-06	236	6	*DNAJC16, AGMAT, DDI2, RSC1A1, PLEKHM2, SLC25A34*
2	chr1:104966872-104966872	rs72685414	104966872	1.87e-06	1	0	
3	chr3:50615472-51356217	rs148179194	50798652	7.03e-06	152	3	*C3orf18, HEMK1, DOCK3*
4	chr3:84862871-84961810	rs6767407	84955841	9.61e-06	27	0	
5	chr3:172170104-172214001	rs57361543	172212144	5.09e-06	28	2	*GHSR, TNFSF10*
6	chr4:82943149-83005137	rs7675557	82970816	2.28e-06	16	1	*RASGEF1B*
7	chr4:117560979-117607439	rs67769911	117607070	6.33e-06	18	1	*MIR1973*
8	chr5:30884915-31001774	rs62357000	30884915	9.20e-06	4	1	*CDH6*
9	chr5:44051305-44130776	rs4410642	44129423	9.42e-06	61	0	
10	chr5:54436897-54754893	rs334878	54519878	7.67e-06	90	9	*ESM1, CDC20B, GPX8, MCIDAS, CCNO, DHX29, SKIV2L2, PLPP1, DDX4*
11	chr5:106788817-106796026	rs71575441	106788817	2.37e-06	2	1	*EFNA5*
12	chr6:26148326-28301195	rs806973; rs56401801	26148326; 27301762	3.14e-06; 3.44e-06	458	77	Locus too broad
13	chr8:99193765-99226821	rs4734387	99193765	4.20e-06	35	1	*NIPAL2*
14	chr9:28890331-28993271	rs35735220	28963289	6.86e-06	113	0	
15	chr9:118176789-118273407	rs10121366	118244022	7.33e-06	54	0	
16	chr13:36947826-37019186	rs79309473	36972202	1.89e-06	26	5	*SOHLH2, SPG20, SPG20-AS1, CCNA1, SERTM1*
17	chr13:88150884-88150884	rs7332726	88150884	7.10e-06	1	0	
18	chr16:6654017-6695032	rs112252907	6675141	4.43e-06	42	1	*RBFOX1*
19	chr16:84581684-84590225	rs247831	84581893	9.17e-06	8	2	*TLDC1, COTL1*
20	chr18:43206985-43231622	rs1484873	43206985	4.45e-06	14	1	*SLC14A2*
21	chr18:73743937-73775398	rs73973283	73757906	6.46e-06	88	0	
22	chr20:54516338-54516338	rs11086525	54516338	3.01e-06	1	0	

*Gene coordinates based on GRCh37/hg19. NSNPs: Number of nominally associated SNPs per genomic locus.