# 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], Dr. Noèlia Fernàndez-Castillo[1,2,3,4]*, Dr. 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

**Running title:** GWAS meta-analysis of cocaine dependence

## Abstract

Cocaine dependence is a complex neuropsychiatric disorder that is highly comorbid with other psychiatric traits. Association studies suggest that common genetic variants contribute substantially to cocaine dependence susceptibility. Also, increasing evidence supports the role of shared genetic risk factors in the lifetime co-occurrence of psychiatric traits and cocaine dependence. Here we performed a genome-wide association study (GWAS) meta-analysis of cocaine dependence using four different dbGaP datasets (2,085 cases and 4,293 controls). Although no genome-wide significant hits were found in the SNP-based analysis, the gene-based analysis identified *HIST1H2BD* as significantly 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. The top SNPs of this region, rs806973 and rs56401801 ($P$=3.14e-06 and 3.44e-06, respectively), are eQTLs for different genes in multiple brain areas. Furthermore, we performed LD Score regression (LDSC) analysis with comorbid conditions and found significant genetic correlations between cocaine dependence and ADHD, SCZ, MDD and risk-taking behavior. We also found, through polygenic risk score (PRS) 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 behavior ($R^2$=0.60%; $P$=2.7e-08), MDD ($R^2$=1.21%; $P$=4.35e-15), children's aggressive behavior ($R^2$=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 on European ancestry individuals. Despite the small sample size, we identified suggestive associations in regions that may be related to cocaine dependence. Furthermore, we found evidence for shared genetic risk factors between cocaine dependence and several comorbid psychiatric traits.

**Keywords:** Cocaine addiction, genetic correlation, GWAS

## 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 & Anthony 2002). Cocaine dependence prevalence in the European population can be estimated 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 around 65-79% (Ducci & 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 and rs806368 in *CNR1* (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 genomewide 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 & Malatestinić 2009; Hasin & 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; Westermeyer 2006; Piñeiro-Dieguez *et al.* 2016). 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; Hartz *et al.* 2017; Du Rietz *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.

## Materials and methods

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

### 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.

## 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 *et al.* 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.

### 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, Li, & Abecasis 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.

**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).

**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. Additionally, we performed a heritability analysis based on the cell type-specific regulatory element H3K4me3, highly associated to promoters (Howe *et al.* 2017; Woo *et al.* 2017), from the Roadmap Epigenomic Mapping Consortium (Kundaje *et al.* 2015; Ernst & Kellis 2015). The threshold for significance was calculated using the Bonferroni correction for multiple testing ($P$<2e-03 for partitioned heritability and $P$<5.6-e-04 for cell type-specific analysis).

## 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.

## 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.

## Shared genetic factors between cocaine dependence and comorbid conditions

### 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.

**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).

**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, Lewis, & O'Reilly 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 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).

**Gene-based meta-analysis with comorbid conditions**

Six independent gene-based meta-analyses were performed between cocaine dependence and the six phenotypes previously described using MAGMA 1.05b. Gene-based analysis of each independent dataset was performed as previously described. Then, the weighted Stouffer's Z method was used to combine the Z-scores for each gene across cohorts, with weights set to the square root of the sample size. Signification threshold was calculated using the Bonferroni correction to control for multiple testing (*P*<2.8e-06).

## Results

**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, indicating that most of the observed inflation can be attributed to polygenicity rather than to 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).

**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). Furthermore, the analysis of the promoter-associated mark H3K4me3 in individual cells/tissues, revealed an enrichment in ten brain tissues and neuronal cell lines, although none of them overcame the Bonferroni correction for multiple testing (Figure S3).

**Functional annotation of risk SNPs**

In order to identify potentially interesting regions, we performed functional annotation analyses with FUMA considering only SNPs showing suggestive associations. We identified 23 lead SNPs which correspond to 22 genomic risk loci including 112 genes (Table 3). The most interesting risk locus, located on chromosome 6, 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. Interestingly, 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). Also, it is important to notice that many SNPs in this region have a CADD score greater than 12.37, highlighting a deleterious effect, and a low RegulomeDB score, suggesting regulatory functionality (Figures 2 and S4). Another interesting region was identified on chromosome 1 defined by the lead SNP rs148069235 (GWAS *P*=9.5e-06). This region includes 236 SNPs that

cover six genes, and three of them (*PLEKHM2*, *DDI2* and *SLC25A34*) have several brain eQTLs in this region (Figure S5). Finally, another region on chromosome 13, defined by the lead SNP rs79309473 (GWAs *P*=1.8e-06), includes five genes and 26 nominally associated SNPs, many of them brain eQTLs for *SOHLH2* (Figure S6).

**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 S7 and Table S2). The *HIST1H2BD* gene, located in the genomic region on chr6 showing a suggestive association in the SNP-based analysis, displayed a significant gene-wise association with cocaine dependence (surpassing 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 associated with cocaine dependence after correction for multiple testing (Table S3-5), although the BioCarta gene set "BIOCARTA TNFR2 PATHWAY" showed a trend (uncorrected *P*=5.38e-04, corrected *P*=0.09), being also the most significantly associated canonical pathway gene set. On the other hand, from the 10 GO gene sets with lower p-values, seven were related to synapse organization, glutamatergic neurotransmission and brain functions.

**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 (Child-Aggre), antisocial behavior (ASB) or risk taking (RT). 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

RT (rg=0.35; SE=0.06; *P*=9.1e-08) (Figure 3A). However, no significant genetic correlations were found with Child-Aggre or ASB, 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 and intelligence) or with reproductive traits (e.g. age of the first birth and age starting oral contraceptive pills) and a positive correlation with familiar situation (e.g. being adopted as a child or tobacco smoke exposure at home) or with several psychological or psychiatric traits like neuroticism, depressive symptoms or loneliness (Figure S8 and Table S6).

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), ASB (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), RT ($R^2$=0.60%, $P_T$=0.00135, *P*=2.7e-08) and Child-Aggre ($R^2$=0.30%, $P_T$=0.3552, *P*=8.8e-05) (Figure 3B and S9). 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 S10). 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).

**Gene-based meta-analysis with comorbid conditions**

Finally, we performed a gene-based meta-analysis between cocaine dependence and each of the six comorbid phenotypes from the previous section in order to identify novel genes involved in each pair of disorders. However, the differences in the sample size of the different dataset

pairs hampered this approach. The most interesting results were obtained when we combined cocaine dependence with ADHD. We found significant association with two new genes: *AMIGO3* and *BSN* (combined *P*=4.24e-07 and 5.13e-07, respectively), and also with 19 of the 20 genes previously identified in a gene-based analysis performed as part of the ADHD meta-analysis by Demontis *et al*. (2017). Four of these 19 associations increased significance by at least one order of magnitude in the combined analysis (*TIE1*, *CDH8*, *ELOVL1* and *CCDC24*) (Table S7).

The gene-based meta-analysis between cocaine dependence and SCZ revealed 360 associations: 28 new genes and 332 that were already highlighted in the SCZ study alone. However, all of the new findings were driven by the SCZ dataset. Three of the 332 genes increased the significance of the association by three orders of magnitude (*BTN3A2, HIST1H2BD* and *SLC17A1*), 15 genes by two orders of magnitude and 73 by one (Table S8).

No relevant results were obtained in the gene-based meta-analysis of cocaine dependence with MDD, RT, ASB or Child-Aggre.

## Discussion

To our knowledge, this is the largest GWAS meta-analysis on cocaine dependence performed to date in individuals with European ancestry. Although no genome-wide significant (GWS) hits were found, gene-based analysis 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. Interestingly, immune and histone pathways have previously been associated with other psychiatric disorders like schizophrenia, major depressive disorder and bipolar disorder (O'Dushlaine *et al.* 2015). Moreover, this region overlaps with the genomic region that has been most consistently associated with SCZ. Indeed, we found five SNPs in this region (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. The most relevant is rs17693963, reported in three different studies (Bergen *et al.* 2012; Sleiman *et al.* 2013; Ruderfer *et al.* 2014) and being 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 encoded 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 & 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 *et al.* 2013). Furthermore, the analysis of the promoter-associated mark H3K4me3 in individual cells/tissues (Howe *et al.* 2017; Woo *et al.* 2017) highlighted a nominal enrichment of SNP heritability in brain tissues and neuronal cell lines. The highest enrichment was observed in substantia nigra, which plays an important role in reward and thus it may be related to the development of the dependence (Beaudoin *et al.* 2017).

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 analysis using LDSC and, for the first time, we found significant genetic correlation with ADHD, SCZ, MDD and risk-taking behavior. Furthermore, we used the PRS method that, in contrast to LDSC, uses individual-level SNP data, resulting in greater 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 probable 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. Further studies are needed to identify the genetic pathways and neurobiological mechanisms underlying this genetic overlap.

Gene-based meta-analysis between different disorders allows identification of new genes involved at the same time in the different disorders considered. We performed this analysis for cocaine dependence with each of the comorbid conditions where genetic correlations were detected. The most interesting results were obtained when combining cocaine dependence and ADHD, identifying two new genes: *AMIGO3* and *BSN,* both involved in brain function. *AMIGO3* inhibits maturation of oligodendrocyte precursor cells and myelin production (Foale *et al.* 2017). *BSN* is a brain specific gene involved in the organization of the cytomatrix at the presynaptic active zone which regulates neurotransmitter release (Südhof 2004). Previously, we reported an association between another component of the neurotransmitter release machinery, NSF, and cocaine dependence (Fernàndez-Castillo *et al.* 2012; Cabana-Domínguez *et al.* 2016).

This study has strengths and limitations that need some discussion. We performed a GWAS meta-analysis using all the cocaine dependence datasets available in the dbGaP repository, but

we could not find any GWS association at the SNP-based level, as expected given the limited sample size. 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. To obtain a more comprehensive picture of the etiological overlap between cocaine dependence and comorbid conditions, a larger sample will be needed, and other genetic factors should be included in the analysis (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 took 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). 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 (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. Increasing the number of individuals of this study would increase the chances to detect significant associations. Furthermore, we found evidence for a shared genetic etiology between cocaine dependence and several comorbid conditions. Further studies are needed to identify the genetic pathways and neurobiological mechanisms that underlie this genetic overlap.

## Authors' contributions

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.

## Acknowledgments

The authors declare no conflicts of interest.

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

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).

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.

Beaudoin GMJ, Gomez JA, Perkins J, Bland JL, Petko AK & Paladini CA (2017) Cocaine selectively reorganizes excitatory inputs to Substantia Nigra pars compacta dopamine neurons. *J Neurosci* 38:1975–17.

Bergen SE, O'Dushlaine CT, Ripke S, Lee PH, Ruderfer DM, Akterin S, Moran JL, Chambert KD, Handsaker RE, Backlund L, et al. (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.

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.

Bühler KM, Giné E, Echeverry-Alzate V, Calleja-Conde J, De Fonseca FR & López-Moreno JA

(2015) Common single nucleotide variants underlying drug addiction: More than a decade of research. *Addict Biol* 20:845–871.

Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, Duncan L, Perry JRB, Patterson N, Robinson EB, et al. (2015a) An atlas of genetic correlations across human diseases and traits. *Nat Genet* 47:1236–1241.

Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL, Neale BM & Neale BM (2015b) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 47:291–295.

Cabana-Domínguez J, Roncero C, Grau-López L, Rodríguez-Cintas L, Barral C, Abad AC, Erikson G, Wineinger NE, Torrico B, Arenas C, et al. (2016) A Highly Polymorphic Copy Number Variant in the NSF Gene is Associated with Cocaine Dependence. *Sci Rep* 6:31033.

Carey CE, Agrawal A, Bucholz KK, Hartz SM, Lynskey MT, Nelson EC, Bierut LJ & Bogdan R (2016) Associations between Polygenic Risk for Psychiatric Disorders and Substance Involvement. *Front Genet* 7:149.

Compton WM, Thomas YF, Stinson FS & Grant BF (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.

Currie SR, Patten SB, Williams JVA, Wang J, Beck CA, El-Guebaly N & Maxwell C (2005) Comorbidity of major depression with substance use disorders. *Can J Psychiatry* 50:660–666.

Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, Belliveau R, Bybjerg-Grauholm J, Bækved-Hansen M, Cerrato F, et al. (2017) *Discovery Of The First Genome-Wide Significant Risk Loci For ADHD*. Cold Spring Harbor Laboratory.

Ducci F & Goldman D (2012) The Genetic Basis of Addictive Disorders. *Psychiatr Clin North Am* 35:495–519.

EMCDDA (2017) *Informe Europeo sobre Drogas 2017: Tendencias y novedades*.

Ernst J & Kellis M (2015) Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat Biotechnol* 33:364–376.

Euesden J, Lewis CM & O'Reilly PF (2015) PRSice: Polygenic Risk Score software. *Bioinformatics* 31:1466–1468.

Fernàndez-Castillo N, Cormand B, Roncero C, Sánchez-Mora C, Grau-Lopez L, Gonzalvo B, Miquel L, Corominas R, Ramos-Quiroga JA, Casas M, et al. (2012) Candidate pathway association study in cocaine dependence: the control of neurotransmitter release. *World J Biol Psychiatry* 13:126–134.

Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, Anttila V, Xu H, Zang C, Farh K, et al. (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* 47:1228–1235.

Foale S, Berry M, Logan A, Fulton D & Ahmed Z (2017) LINGO-1 and AMIGO3, potential therapeutic targets for neurological and dysmyelinating disorders? *Neural Regen Res* 12:1247–1251.

Gelernter J, Sherva R, Koesterer R, Almasy L, Zhao H, Kranzler HR & Farrer L (2014) Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. *Mol Psychiatry* 19:717–723.

Gurriarán X, Rodríguez-López J, Flórez G, Pereiro C, Fernández JM, 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.

Hartz SM, Horton AC, Oehlert M, Carey CE, Agrawal A, Bogdan R, Chen L-S, Hancock DB, Johnson EO, Pato CN, et al. (2017) Association Between Substance Use Disorder and Polygenic Liability to Schizophrenia. *Biol Psychiatry* 82:709–715.

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.

Howe FS, Fischl H, Murray SC & Mellor J (2017) Is H3K4me3 instructive for transcription activation? *BioEssays* 39:1–12.

Johnson EO, Hancock DB, Levy JL, Gaddis NC, Page GP, Glasheen C, Saccone NL, Bierut LJ & Kral AH (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.

Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, et al. (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518:317–330.

Lee SH, Ripke S, Neale BM, Faraone S V, Purcell SM, Perlis RH, Mowry BJ, Thapar A, Goddard ME, Witte JS, et al. (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 45:984–994.

Mbarek H, Milaneschi Y, Fedko IO, Hottenga J-J, de Moor MHM, Jansen R, Gelernter J, Sherva R, Willemsen G, Boomsma DI, et al. (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.

Mitchell BD, Fornage M, McArdle PF, Cheng YC, Pulit SL, Wong Q, Dave T, Williams SR, Corriveau R, Gwinn K, et al. (2014) Using previously genotyped controls in genome-wide association studies (GWAS): Application to the Stroke Genetics Network (SiGN). *Front Genet* 5:95.

O'Dushlaine C, Rossin L, Lee PH, Duncan L, Parikshak NN, Newhouse S, Ripke S, Neale BM, Purcell SM, Posthuma D, et al. (2015) Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci* 18:199–209.

Pappa I, St Pourcain B, Benke K, Cavadino A, Hakulinen C, Nivard MG, Nolte IM, Tiesler CMT, Bakermans-Kranenburg MJ, Davies GE, et al. (2016) A genome-wide approach to children's aggressive behavior: The EAGLE consortium. *Am J Med Genet Part B Neuropsychiatr Genet* 171:562–572.

Pineda-Cirera L, Cabana-Domínguez J, Roncero C, Cozar M, Grau-López L, Abad AC, Martínez-Luna N, Robles-Martínez M, Sánchez-Mora C, Ramos-Quiroga JA, et al. (2018) Evaluation of previous substance dependence genome-wide significant findings in a Spanish sample.

*Drug Alcohol Depend* 187:358–362.

Piñeiro-Dieguez B, Balanzá-Martínez V, García-García P, Soler-López B, Domingo MA, Labarra JDA, Lobato PA, Salamanca AA, Bes JA, Fernández FJB, et al. (2016) Psychiatric Comorbidity at the Time of Diagnosis in Adults With ADHD: The CAT Study. *J Atten Disord* 20:1066–1075.

Reginsson GW, Ingason A, Euesden J, Bjornsdottir G, Olafsson S, Sigurdsson E, Oskarsson H, Tyrfingsson T, Runarsdottir V, Hansdottir I, et al. (2018) Polygenic risk scores for schizophrenia and bipolar disorder associate with addiction. *Addict Biol* 23:485.

Du Rietz E, Coleman J, Glanville K, Choi SW, O'Reilly PF & Kuntsi J (2017) Association of Polygenic Risk for Attention-Deficit/Hyperactivity Disorder With Co-occurring Traits and Disorders. *Biol Psychiatry Cogn Neurosci Neuroimaging*.

Ripke S, Neale BM, Corvin A, Walters JTR, Farh KH, Holmans PA, Lee P, Bulik-Sullivan B, Collier DA, Huang H, et al. (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511:421–427.

Ruderfer DM, Fanous AH, Ripke S, McQuillin A, Amdur RL, Gejman P V, O'Donovan MC, Andreassen OA, Djurovic S, Hultman CM, et al. (2014) Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol Psychiatry* 19:1017–1024.

Sanders AR, Göring HHH, Duan J, Drigalenko EI, 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.

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.

Sleiman P, Wang D, Glessner J, Hadley D, Gur RE, Cohen N, Li Q, Hakonarson H, Sleiman P, Glessner J, et al. (2013) GWAS meta analysis identifies TSNARE1 as a novel Schizophrenia / Bipolar susceptibility locus. *Sci Rep* 3:3075.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* 102:15545–15550.

Südhof TC (2004) The synaptic vesicle cycle: Neurotransmitter release and the synaptic vesicle cycle. *Annu Rev Neurosci* 27:509–47.

Tielbeek JJ, Johansson A, Polderman TJC, Rautiainen M-R, Jansen P, Taylor M, Tong X, Lu Q, Burt AS, Tiemeier H, et al. (2017) Genome-Wide Association Studies of a Broad Spectrum of Antisocial Behavior. *JAMA Psychiatry*.

Wagner FA & Anthony JC (2002) From first drug use to drug dependence: Developmental periods of risk for dependence upon marijuana, cocaine, and alcohol. *Neuropsychopharmacology* 26:479–488.

Watanabe K, Taskesen E, Van Bochoven A & Posthuma D (2017) Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 8:1826.

Westermeyer J (2006) Comorbid schizophrenia and substance abuse: A review of epidemiology

and course. *Am J Addict* 15:345–355.

Willer CJ, Li Y & Abecasis GR (2010) METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26:2190–2191.

Woo H, Ha SD, Lee SB, Buratowski S & Kim TS (2017) Modulation of gene expression dynamics by co-Transcriptional histone methylations. *Exp Mol Med* 49:e326–e326.

Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, Adams MJ, Agerbo E, Air TM, Andlauer TMF, et al. (2018) Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 50:668–681.

Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, Hemani G, Tansey K, Laurin C, Pourcain B St., et al. (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.

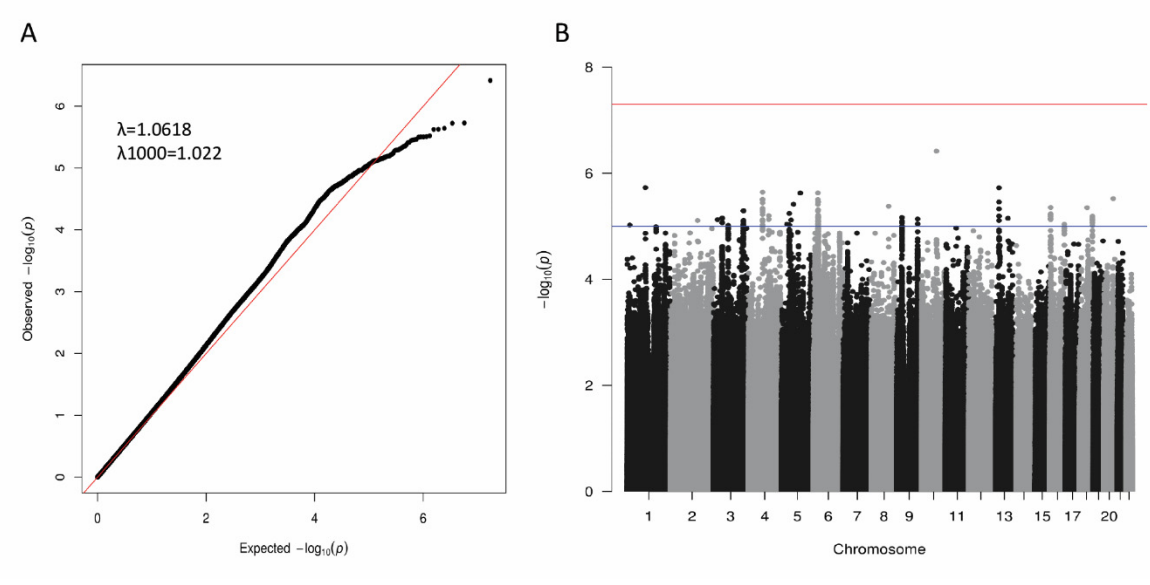


**Figure 1. Results from the GWAS meta-analysis on cocaine dependence.** A) Q-Q plot and B) Manhattan plot. Red line: threshold for genome-wide significance (*P* < 5e-08). Blue line: threshold for suggestive associations (*P* < 1e-05).

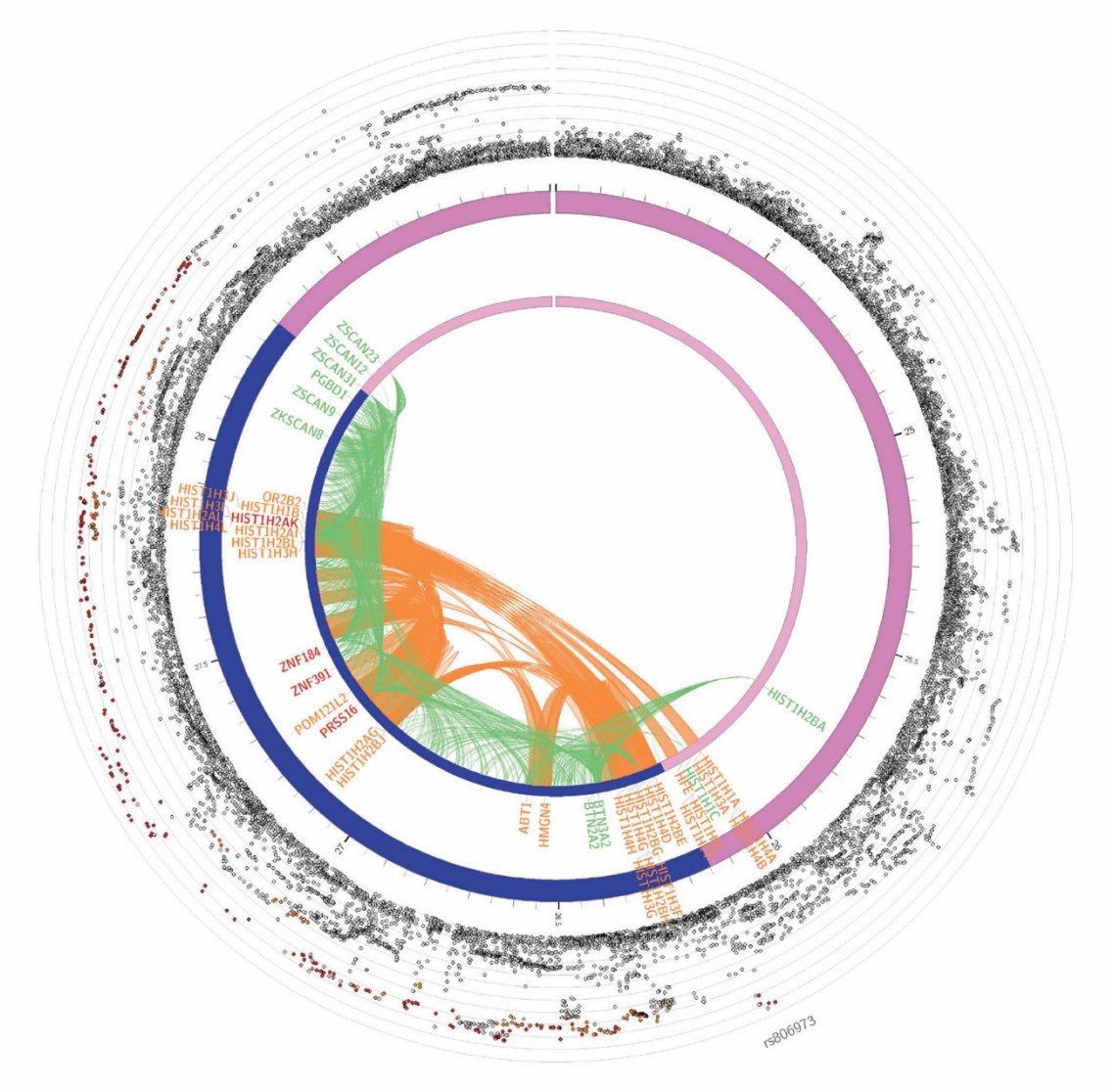


**Figure 2. Circo-plot from genomic risk loci on chromosome 6 by FUMA.** The most outer layer is the Manhattan plot (only SNPs with $P < 0.05$ are displayed). SNPs in genomic risk loci are color-coded as a function of their maximum $r^2$ to the lead SNPs in the locus, as follows: red ($r^2 > 0.8$), orange ($r^2 > 0.6$), 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.

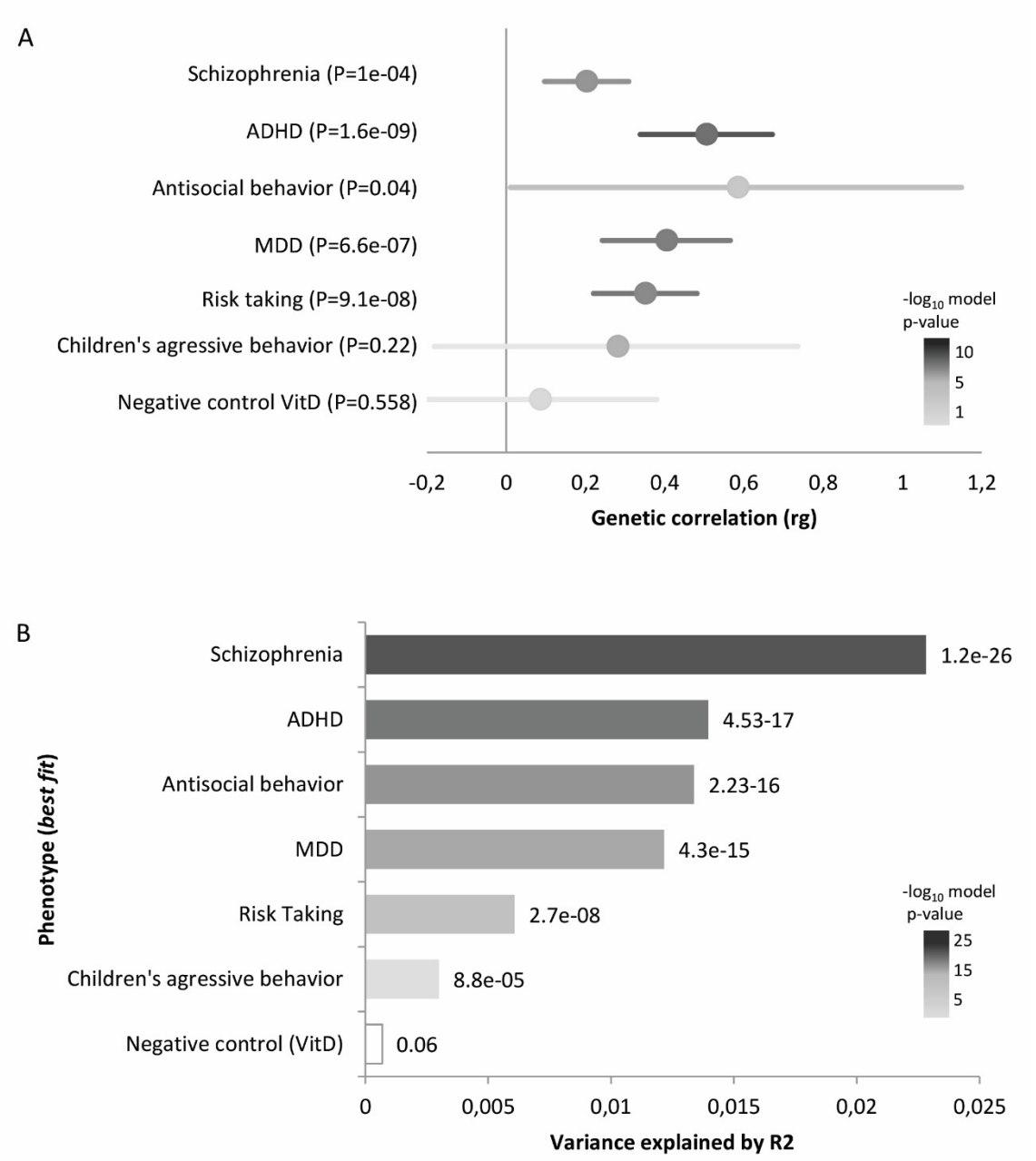


**Figure 3. Shared genetic factors between cocaine dependence and comorbid conditions.** A) Results from LD Score (LDSC) regression analysis showing genetic correlation (rg) between cocaine dependence and several traits. Error bars indicate 95% confidence limits. The significance threshold was set to $P < 7.1e-03$. 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<br>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 SNPs | 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.