

## **Title**

*Full title:* Combining multivariate genomic approaches to elucidate the comorbidity between ASD and ADHD.

*Short title:* Genetic comorbidity between ASD and ADHD.

## **Authors**

Hugo Peyre<sup>1,2,3,4\*</sup>, Tabea Schoeler<sup>1</sup>, Chaoyu Liu<sup>1</sup>, Camille Michèle Williams<sup>2</sup>, Nicolas Hoertel<sup>5,6</sup>, Alexandra Havdahl<sup>7</sup>, Jean-Baptiste Pingault<sup>1,8</sup>

## **Affiliations**

1. Department of Clinical, Educational and Health Psychology, University College London, London, UK.

2. Laboratoire de Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS), Ecole Normale Supérieure, PSL University, Paris, France.

3. Neurodiderot, INSERM UMR 1141, Paris Diderot University, Paris, France.

4. Department of Child and Adolescent Psychiatry, Robert Debré Hospital, APHP, Paris, France.

5. INSERM UMR 894, Psychiatry and Neurosciences Center; Paris Descartes University, PRES Sorbonne Paris Cité, Paris, France.

6. Department of Psychiatry, Corentin Celton Hospital, APHP, Issy-les-Moulineaux; Paris Descartes University, PRES Sorbonne Paris Cité, Paris, France.

7. Nic Waals Institute, Lovisenberg Diaconal Hospital, 0853 Oslo, Norway; Department of Mental Disorders, Norwegian Institute of Public Health, Oslo, Norway; MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.

8. Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK.

***Full title: Combining multivariate genomic approaches to elucidate the comorbidity between ASD and ADHD.***

**Abstract**

**Background**

Attention Deficit Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD) are two highly heritable neurodevelopmental disorders. Several lines of evidence point toward the presence of shared genetic factors underlying ASD and ADHD. We conducted genomic analyses of common risk variants (i.e. Single-Nucleotide Polymorphisms, SNPs) shared by ASD and ADHD, and those specific to each disorder.

**Methods**

With the summary data from two GWAS, one on ASD (N=46,350) and another on ADHD (N=55,374) individuals, we used genomic structural equation modelling and colocalization analysis to identify SNPs shared by ASD and ADHD and SNPs specific to each disorder. Functional genomic analyses were then conducted on shared and specific common genetic variants. Finally, we performed a bidirectional Mendelian randomization analysis to test whether the shared genetic risk between ASD and ADHD was interpretable in terms of reciprocal relationships between ASD and ADHD.

**Results**

We found that 37.5% of the SNPs associated with ASD (at  $p < 1e-6$ ) colocalized with ADHD SNPs and that 19.6% of the SNPs associated with ADHD colocalized with ASD SNPs. We identified genes mapped to SNPs that are specific to ASD or ADHD and that are shared by ASD and ADHD, including two novel genes *INSM1* and *PAX1*. Our bidirectional Mendelian randomization analyses indicated that the risk of ASD was associated with an increased risk of ADHD and vice versa.

**Conclusions**

Using multivariate genomic analyses, the present study uncovers shared and specific genetic variants associated to ASD and ADHD. Further functional investigation of genes mapped to those shared variants may help identify pathophysiological pathways and new targets for treatment.

### **Keywords**

Autism Spectrum Disorder; Attention Deficit Hyperactivity Disorder; comorbidity; common genetic variants; SNP; colocalization; genomic structural equation modelling; GWAS.

### **Abbreviations**

ADHD: Attention Deficit Hyperactivity Disorder.

ASD: Autism Spectrum Disorder.

CNVs: Copy Number Variants.

SNPs: Single-Nucleotide Polymorphisms.

Genomic SEM: Genomic Structural Equation Modelling.

GWAS: Genome-Wide Association Studies.

LD: Linkage disequilibrium.

MR: Mendelian randomization.

LDSC: Linkage disequilibrium score regression.

IVW: Inverse weighted variance.

MR-RAPS: Mendelian randomization robust adjusted profile score.

### **Tables & Figures**

Number of Figures: 4

Number of Tables: 1

Number of Supplementary Figures: 14

Number of Supplementary Tables: 5

Number of supplementary Data: 4

**Word count**

Abstract: 259

Manuscript: 5962

Attention Deficit Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD) are two common neurodevelopmental disorders (American Psychiatric Association, 2013) with a prevalence in children from 5 to 7% (Willcutt, 2012) and from 1 to 2% (Baio, 2018), respectively. ADHD is characterized by symptoms of inattention, impulsivity, and hyperactivity, and ASD by a deficit in social communication as well as restricted and repetitive patterns of interests and behaviors. Although both disorders differ in terms of diagnostic criteria, there is a considerable overlap in symptomatology. Individuals with ASD commonly display inattention, impulsivity, and hyperactivity symptoms (Lord et al., 2020) and likewise, individuals with ADHD often have impaired social and communication skills (Hollingdale et al., 2019). In addition, ASD and ADHD are frequently comorbid; approximately one third of children with ASD meet diagnostic criteria for ADHD (Lai et al., 2019) and 15% of children with ADHD meet diagnostic criteria for ASD (Grzadzinski et al., 2016).

Little is known about what causes the association between ASD and ADHD. Both disorders are thought to be caused by a complex interplay of environmental and genetic risk factors (Cross-Disorder Group of the Psychiatric Genomics Consortium et al., 2013). Shared environmental risk factors, such as preterm birth (Bhutta et al., 2002) or prenatal exposure to valproate (Christensen et al., 2019) might partially explain the association between the disorders. Several lines of evidence also point toward the presence of shared genetic factors in ASD and ADHD. In family-based studies, relatives of children with ASD are at higher risk for ADHD than relatives of children without ASD (e.g. Ghirardi et al., 2018; Septier et al., 2019) and in twin studies, researchers report strong genetic correlations between ADHD and ASD traits (e.g. Constantino et al., 2003; Ronald et al., 2008). A similar burden of rare protein-truncating variants has been found in individuals with ASD and those with ADHD

(Satterstrom et al., 2019). Moreover, most of the rare Copy Number Variants (CNVs) that are linked to ASD are also associated with ADHD (Gudmundsson et al., 2019). And yet, less than 10% of ASD and ADHD liabilities can be accounted for by rare genetic variants (Gudmundsson et al., 2019).

Recent genomic Structural Equation Modelling (genomic SEM) of Genome-Wide Association Studies (GWAS) identified shared common risk variants (i.e. Single-Nucleotide Polymorphisms, SNPs) underlying several psychiatric disorders (including ASD and ADHD, but also anorexia nervosa, bipolar disorder, major depression, obsessive-compulsive disorder, schizophrenia, and Tourette syndrome) (Lee et al., 2019; Schork et al., 2019). However, by simultaneously including all of these disorders in one multivariate analysis, these studies may have missed SNPs that specifically explain the comorbidity between ASD and ADHD. Moreover, little research has focused on uncovering the SNPs that are uniquely associated to ASD or ADHD. Finally, no study has used a robust tool, such as the colocalization approach, to identify specific and shared SNPs of ASD and ADHD after taking linkage disequilibrium (LD) into account (Giambartolomei et al., 2018). So far, colocalization has been successfully applied to elucidate the specific and shared genetic risk between autoimmune diseases (Fortune et al., 2015), lipid levels and cardiovascular outcomes (Siewert & Voight, 2018), as well as schizophrenia and gene expression within human brain tissue (Dobbyn et al., 2018). By combining colocalization and genomic SEM, the present study offers a novel approach to identify the specific and shared genetic risk between ASD and ADHD.

Alternatively, the association between ASD and ADHD might be explained by direct causal relationships between the two phenotypes, unidirectionally or bidirectionally. ADHD might lead to secondary impairments in social interaction and behavioral flexibility (Taylor et al.,

2013), while ASD might contribute to secondary attention problems, hyperactive and impulsive behaviours. These directional relationships could change our interpretation of the colocalization analysis considering that a given SNP may colocalize between ASD and ADHD not because of variants independently affecting both ASD and ADHD, but because the increase in symptoms of one disorder may lead to symptom increase of the other. To remedy the issue of directionality, the present study will also be the first to conduct a bidirectional Mendelian randomization (MR) analysis on ASD and ADHD GWAS to determine the effect of ADHD on ASD and the effect of ASD on ADHD (Pingault et al., 2018).

This study used genomic SEM and colocalization to identify SNPs shared by ASD and ADHD and SNPs specific to each disorder. Functional genomic analyses were performed on shared and specific common genetic variants. Our analysis plan to identify SNPs shared by ASD and ADHD was (i) to use genomic SEM to determine the association of each SNP with a general factor (corresponding to the shared variance between the ASD and ADHD GWAS) and then (ii) to use a Bayesian colocalization method to independently verify whether those SNPs actually colocalize based on the original ASD and ADHD GWAS. (iii) Finally, a bidirectional Mendelian randomization analysis was performed to determine whether the shared genetic risk between ASD and ADHD was interpretable in terms of (reciprocal) causal relationships. The SNPs that did not colocalize between ASD and ADHD were considered to be specific to each disorder.

## **METHODS**

### **GWAS Summary Statistics**

Summary statistics for ASD and ADHD were obtained from the European ancestry subgroup of the Psychiatric Genomics Consortium and iPSYCH (PGC + iPSYCH). We used the most recent GWAS data on ASD (18,381 diagnosed ASD cases and 27,969 controls) (Grove et al., 2019) and ADHD individuals (20,183 diagnosed ADHD cases and 35,191 controls) (Demontis et al., 2019). There was some overlap in the controls included in these 2 GWAS (iPSYCH).

### **Genomic SEM**

Genomic SEM (GenomicSEM R package) models the genetic covariance structure of GWAS summary statistics using LD score regression (LDSC) (Nivard, 2019) to estimate the association of each SNP with the general factor (a latent variable corresponding to the shared variance between the ASD and ADHD GWAS). In genomic SEM, factor loadings were fixed to be equal between ASD and ADHD GWAS. For liability-scale estimates, we used a population prevalence of 1.12% for ASD and 5% for ADHD (Demontis et al., 2019; Grove et al., 2019).

### **Colocalization Analyses**

Colocalization analyses were performed on the two original ASD and ADHD GWAS. These analyses were conducted on each genome-wide significant ( $p < 5e-8$ ) LD-independent ( $r^2 > 0.2$ ; window of 500kb) SNP associated with either ASD, ADHD, or the general factor of ASD and ADHD in genomic SEM analyses to account for two possible types of false results that occur when analysing individual SNPs (*Supp. Fig. 1*): (i) a SNP associated with a trait A may be falsely associated with the general factor of Trait A and B because a causal SNP of Trait B is in LD with the SNP of Trait A; (ii) the general factor might miss some SNPs that are in fact associated with both traits. For example, a shared causal SNP strongly associated with Trait



A but less strongly associated with Trait B may erroneously not be linked to the general factor.

We used the R packages COLOC (Wallace et al., 2019) and Hyprcoloc (C. Foley & Staley, 2019). Both methods implement a Bayesian test for the colocalization of two association signals in a selected region (250kb around a SNP under study), using the summary statistics of the two traits, to test five hypotheses: H0 (no association between the SNP and either trait), H1 (SNP association with trait 1 only), H2 (SNP association with trait 2 only), H3 (SNP association with both traits but not colocalizing, i.e. two distinct SNPs in LD); and H4 (SNP association with both traits and colocalizing). We set the prior probability that a SNP is causal in each trait to be identical ( $1e-4$  is the recommended threshold for colocalization analysis in the context of GWAS (Giambartolomei et al., 2018; Wallace et al., 2019)) and the prior probability that a SNP is causal for both traits at  $5e-6$  (meaning that 5 out of 100 SNPs that are associated with one trait are also associated with the other). A given SNP was considered to colocalize if  $H4 > 0.5$  using COLOC and Hyprcoloc packages (Dobbyn et al., 2018; Wallace, 2019). A supplementary colocalization analysis was performed with  $H4 > 0.8$ . SNPs colocalizing between ASD and ADHD were considered as SNPs shared by ASD and ADHD. Among the SNPs not colocalizing between ASD and ADHD, those associated with a  $p$ -value below  $5e-8$  with ASD were classified as SNPs specific to ASD and those associated with a  $p$ -value below  $5e-8$  with ADHD were classified as SNPs specific to ADHD.

To estimate the percentage of SNPs that colocalized between both traits we performed a similar analysis but at the  $p$ -value threshold of  $1e-6$ , which enabled us to estimate this percentage using a larger number of colocalizing SNPs.

### **Functional genomic Analysis**

Functional annotation and analyses were performed separately on SNPs shared by ASD and ADHD (i.e. SNPs colocalizing between ASD and ADHD), and SNPs specific to each disorder using FUMA (Functional Mapping and Annotation of Genome-Wide Association Studies) (<http://fuma.ctglab.nl/>) (Watanabe et al., 2017). Using FUMA SNP2GENE, SNPs were mapped to genes based on (i) positional mapping (Ensembl genes (build 85) using ANNOVAR), deleteriousness score (CADD score $\geq$ 12.37) and 15-core chromatin state (13 brain tissue types from the GTEx.v8 project)), (ii) eQTL associations with 13 brain tissue types (FDR $\leq$ 0.05), and (iii) 3D chromatin interactions mapping (Hi-C data) of 2 tissue types: adult and foetal cortex (Giusti-Rodriguez & Sullivan, 2019), with annotate enhancer/promoter regions (Roadmap 111 epigenomes) and 15-core chromatin state from 13 brain tissues (FDR $<$ 10e-6).

GENE 2FUNC then annotates these genes in a biological context. An enrichment of differentially expressed gene (DEG, based on GTEx v6 RNA-seq data) in a certain tissue compared to all other tissue types is provided. Log-transformed  $p$ -values for the enrichment of DEG in each tissue are shown in an interactive heatmap for each gene, and globally for genes mapped to SNPs that are shared by ASD and ADHD and for those that are specific to ASD and ADHD. A supplementary analysis of the temporal expression in the brain (based on 11 general developmental stages of brain samples of BrainSpan data) was also performed in the 3 gene sets.

### **Bidirectional Mendelian Randomization Analyses**

We ran bidirectional summary data Mendelian randomization (MR) analysis to determine the effect of the liability for ADHD on ASD and the liability for ASD on ADHD (Pingault et al., 2016). Because only few SNPs associated with ASD and ADHD at genome-wide significance level ( $p < 5e-8$ ) are available in the current GWAS ((Grove et al., 2019) and

(Demontis et al., 2019)), MR analyses were conducted at a  $p$ -value threshold of  $1e-6$ . LD-independent SNPs ( $r^2 > 0.001$ ; window of 500kb) associated with ASD ( $n_{SNPs}=15$ ) were selected as instruments for ASD to estimate the effect of ASD on ADHD. Associations were also ascertained in the opposite direction, using  $n=40$  SNPs as instruments for ADHD.

We estimated the main effects using the inverse weighted variance (IVW) estimator, which consists of a linear regression of the instrument-outcome association estimates on the instrument-exposure association estimates, weighted by the inverse of the variance of the instrument-outcome association estimates.

We estimated the Cochran's Q statistic to detect the presence of heterogeneity. If individual SNPs are valid instruments, corresponding causal estimates should only vary by chance, so larger than chance heterogeneity is indicative of invalid SNPs. To deal with heterogeneity, we conducted a number of analyses. We first performed leave-one-out sensitivity analyses to check for a disproportionate influence of individual SNPs on overall effect estimates using the IVW method. In addition, we implemented MR-PRESSO (Pleiotropy RESidual Sum and Outlier), which provides: (i) a global test to detect horizontal pleiotropy, (ii) a correction for horizontal pleiotropy via outlier removal, and (iii) a distortion test to test whether the causal estimate is significantly different (distorted) after the MR-PRESSO outlier adjustment.

We implemented additional methods that can provide unbiased estimates when only some of the instruments are valid, including weighted median analysis, weighted mode analysis and MR contamination mixture method (Slob & Burgess, 2020). The MR contamination mixture method is implemented by constructing a likelihood function based on the SNP-specific ratio estimate (genetic association with exposure divided by genetic association with outcome). If a SNP is a valid instrument, then its ratio estimate is assumed to be normally distributed about the true causal effect. If a genetic variant is not a valid instrument, then its causal estimate will be normally distributed about zero with a standard deviation corresponding to

1.5 times the standard deviation of the ratio estimates. We implemented MR-Egger regression to test for unbalanced horizontal pleiotropy. We determined the strength of instrumental variables using the mean F-statistic (Bowden et al., 2019). MR RAPS (Robust Adjusted Profile Score) was implemented to deal with weak instrument bias (Zhao et al., 2018).

All LD-independent SNPs were included in bidirectional MR, irrespective of whether the SNPs have been identified as specific or shared between the 2 traits. A supplementary bidirectional MR analysis was performed excluding SNPs colocalizing between ASD and ADHD.

## RESULTS

### Genomic SEM

Genomic SEM was conducted on 6,971,687 SNPs that were present in both ASD (number of available SNPs=7,757,027) and ADHD (number of available SNPs=8,094,094) GWAS. SNP heritabilities of ASD and ADHD were 11.7% (SE=0.9%) and 21.4% (SE=1.3%) respectively. The genetic correlation between ASD and ADHD was 0.37 (0.05).

We conducted a GWAS on the general factor and examined the overlap in results with the two original GWAS. We assessed the overlap both ways by (i) testing whether SNPs identified in the general factor GWAS reached genome-wide significance ( $p < 5e-8$ ) in the original GWAS; and (ii) testing whether the SNPs identified in the original GWAS reached genome-wide significance in the general factor GWAS.

The Manhattan plot of the general factor GWAS is depicted in **Fig. 1**. We found 267 genome-wide significant SNPs in the general factor GWAS, which reduced to 8 LD-independent SNPs after clumping (27 when the  $p$ -value threshold was set at  $1e-6$ ). We then verified whether the 8 SNPs identified in the general factor GWAS were also identified in the original trait GWAS at genome-wide significance threshold. For ASD, 1 out of 8 (rs6047310) was

identified, and 3 out of 8 SNPs (rs1222063, rs4916723 and rs6584649) for ASD (See **Table 1**, General factor SNPs section, p-value columns for ASD and ADHD).

There were 74 genome-wide significant ( $p < 5e-8$ ) SNPs in the original ASD GWAS, which reduced to 2 LD-independent SNPs after clumping (16 when the  $p$ -value threshold was set at  $1e-6$ ). One of these two SNPs (rs910805) reached genome-wide significance in the general factor GWAS. Note that this SNP is in the same region 20p11 as the aforementioned SNP rs6047310, which was identified in the general factor GWAS and reached significance in the original ASD GWAS.

There were 275 genome-wide significant ( $p < 5e-8$ ) SNPs in the original ADHD GWAS, which reduced to 16 LD-independent SNPs after clumping (46 when the  $p$ -value threshold was set at  $1e-6$ ). Most of these SNPs ( $n=13$ ) were unique to ADHD and did not reach genome-wide significance in the general factor GWAS.

INSERT TABLE 1 and FIGURE 1 HERE

### **Colocalization Analyses**

Among the 8 SNPs associated with the general factor, all of them colocalized between ASD and ADHD (**Fig. 2** and *Supp. Fig 2-8*). When the  $p$ -value threshold was set at  $1e-6$ , the proportion of SNPs shared by ASD and ADHD as determined by genomic SEM that colocalized between ASD and ADHD GWAS was 66.7% (18/27).

Neither of the two SNPs (rs10099100 and rs910805) associated with ASD in the original GWAS colocalized with ADHD. When the  $p$ -value threshold was set at  $1e-6$ , the proportion of SNPs of ASD that colocalized with ADHD GWAS was 37.5% (6/16).

Among the 16 SNPs associated with ADHD in the original GWAS, 3 of them colocalized with the ASD GWAS (the same 3 SNPs that were identified in the general factor GWAS). When the  $p$ -value threshold was set at  $1e-6$ , the proportion of SNPs for ADHD that colocalized with ASD GWAS was 19.6% (9/46).

INSERT FIG. 2 HERE

### **Supplementary analysis with a higher H4 threshold**

When we performed our colocalisation analysis with  $H_4 > 0.8$  ( $H_4$ : probability of the hypothesis of colocalization between ASD and ADHD GWAS), we found that 48% (13/27) of SNPs associated with the general factor colocalized between ASD and ADHD GWAS. In addition, 18.8% (3/16) of the SNPs associated with ASD colocalized with the SNPs associated with ADHD and 15.2% (7/46) of the SNPs associated with ADHD colocalized with the SNPs associated with ASD ( $p$ -value threshold was set at  $1e-6$ ).

### **Functional genomic analyses of SNPs shared by ASD and ADHD**

We applied FUMA to the SNPs that both (i) colocalized between ASD and ADHD and (ii) were associated with the general factor at  $p < 5e-8$ . FUMA SNP2GENE prioritized 8 genes (*Supp. Table 1*): *MANBA* in region 4q24 around rs227378 (**Fig. 2**), *XRN2*, *INSM1*, *NKX2-4* and *PAX1* in region 20p11 around rs6047310 (*Supp. Fig. 2*), *SORCS3* in region 10q25 around rs6584649 (*Supp. Fig. 3*), *PTBP2* and *DPYD* (in region 1p21 around rs2391769 (*Supp. Fig. 4*)).

The prioritized genes included 6 genes which were reported in previous GWA and genomic SEM studies: *XRN2*, *SORCS3*, *PTBP2*, *NKX2-4*, *MANBA* and *DPYD* (Alonso-Gonzalez et al., 2019; Cross-Disorder Group of the Psychiatric Genomics Consortium et al., 2013; Demontis

et al., 2019; Grove et al., 2019), while *INSM1* and *PAX1* were novel genes. *INSM1* is known to play a key role in neurogenesis and neuroendocrine cell differentiation during embryonic and/or fetal development (**Fig. 3 (A)**). *PAX1* is a member of the paired box (PAX) family of transcription factors without specific expression in brain tissues. We did not find significant DEG values in any tissue for this set of genes (*Supp. Fig. 9 (A)*).

INSERT FIG. 3 HERE

### **Functional genomic analyses of SNPs Specific to ASD**

We applied FUMA to the SNPs specific to ASD (rs10099100 and rs910805). Positional mapping and chromatin interaction mapping of rs910805 prioritized two genes (*XRN2* and *NKX2-4*) which were previously reported as candidates by Grove et al. (Grove et al., 2019) (*Supp. Table 2*). The gene *XRN2* had already been mapped to one SNP (rs6047310) shared by ASD and ADHD. The SNPs rs910805 and rs6047310 are on the same region 20p11 and they are in low LD ( $R^2=0.17$ ). We did not find significant DEG values in any tissue for this set of genes (*Supp. Fig. 9 (B)*).

### **Functional genomic analyses of SNPs Specific to ADHD**

We applied FUMA to the 13 genome-wide significant SNPs specific to ADHD. Positional mapping (17 genes), eQTL mapping (5 genes), and chromatin interaction mapping of SNPs (26 genes) prioritized 36 unique genes (some genes identified by multiple mapping methods; *Supp. Table 3*). We used different mapping parameters on FUMA than the study by Demontis et al. (Demontis et al., 2019). Therefore only 19 out of the 36 genes were previously reported as candidate genes by Demontis et al. (Demontis et al., 2019).

This set of genes was mostly expressed in the nervous, muscle, uterus and heart tissues (*Supp. Fig. 9 (C)*).

### **Supplementary analysis: temporal expression in the brain**

Genes mapped to SNPs shared by ASD and ADHD (*Supp. Fig 10 (A)*) and those specific to ASD or ADHD (*Supp. Fig 10 (B and C)*) did not show a temporal specificity of gene expression in brain tissues.

### **Bidirectional Mendelian Randomization Analysis**

Using bidirectional Mendelian randomization, we found that the risk of ASD was associated with an increased risk of ADHD ( $\beta - \text{MR-IVW}=0.51$  (0.09),  $p\text{-value}<0.001$ , number of SNPs=15; **Fig. 4** and *Supp. Table 4*). Testing the reverse direction, we found that the risk of ADHD was associated with an increased risk of ASD ( $\beta - \text{MR-IVW}=0.33$  (0.05),  $p\text{-value}<0.001$ , number of SNPs=40).

There was substantial heterogeneity in our MR-IVW analyses ( $Q=51.4$  on 15 SNPs,  $p\text{-value}=3.5\text{e-}06$ , and  $Q=79.2$  on 40 SNPs,  $p\text{-value}=2\text{e-}04$ , respectively). Leave-one-out sensitivity analyses did not indicate a disproportionate influence of an individual SNP in any of our MR-IVW analyses (*Supp. Fig. 11*). Our two sets of instrumental variables had high mean strengths (mean F-statistic=29.4 (SD=6.7, range: 15.0 - 40.7) and 31.1 (SD=7.6, range: 21.0 - 62.6), respectively). MR-Egger did not show directional pleiotropy in any direction. Overall, MR estimates were consistent across sensibility analyses (*Supp. Table 4*). The MR-PRESSO global test detected horizontal pleiotropy in both MR-analyses ( $p<0.001$  for both) and only identified one outlier for the ASD exposure and one outlier for the ADHD exposure. The MR-PRESSO distortion tests were not significant ( $p=0.5$  for the exposure ASD and  $p=0.8$  for the exposure ADHD). MR contamination mixture method identified 6 outliers for the



exposure ASD and 12 outliers for the exposure ADHD (including the outliers identified by MR-PRESSO).

INSERT FIG. 4 HERE

### **Supplementary analysis: excluding SNPs colocalizing between ASD and ADHD**

Among the 15 SNPs identified when the exposure was ASD, 5 colocalized between ASD and ADHD. And among the 40 SNPs identified when the exposure was ADHD, 8 colocalized between ASD and ADHD. After excluding the SNPs that colocalized between ASD and ADHD from the bi-directional MR, our findings were still significant in both directions (*Supp. Table 5* and *Supp. Fig. 14*).

## **DISCUSSION**

Using powerful multivariate genomic approaches, we examined the complex relationships between ASD and ADHD. By combining genomic SEM with colocalization analysis, an approach that takes into account linkage disequilibrium between SNPs, we found that around one third of the common genetic variants associated with ASD were linked to ADHD and that about one fifth of the common genetic variants associated with ADHD were linked to ASD.

ASD and ADHD are highly comorbid neurodevelopmental disorders (Grzadzinski et al., 2016; Lai et al., 2019). Abundant evidence from family-based (e.g. Ghirardi et al., 2018; Septier et al., 2019) and twin-based (e.g. Constantino et al., 2003; Ronald et al., 2008) studies suggests that the comorbidity between the two disorders is largely explained by shared genetic factors. Previous studies have shown the role of rare genetic variants in contributing

to this comorbidity (Gudmundsson et al., 2019; Satterstrom et al., 2019). Prior research has also investigated the common genetic variants underlying several psychiatric disorders (Lee et al., 2019; Schork et al., 2019) by modelling shared variation between multiple disorders. Here, we focused on ASD and ADHD and we showed that a sizeable minority of common variants contributing to one disorder also contribute to the other disorder. Genes, including novel genes, mapped to these common variants point towards the existence of common pathophysiological pathways underlying these disorders. However, functional analyses neither showed that the genes underlying the comorbidity of these two neurodevelopmental disorders were preferentially expressed in the brain, nor that expression in the brain took place early in development. Rather, they were expressed throughout development, including in adulthood.

Importantly, a majority of identified SNPs were still specific to each disorder. Those SNPs were mapped to genes that were also expressed in the brain without clear developmental patterns. Fine-grained gene-phenotype mapping is now required to better understand the nature of these specific genetic influences. For example, it is unlikely that common genetic variants homogenously affect all symptoms within a disorder. Instead, the specific genetic variants and pathways may play a more prominent role in the symptoms of one disorder that rarely occur in patients affected by the other disorder. Such fine-grained mapping to multiple symptom dimensions or individual symptoms should provide a renewed understanding of the pathways from genetic variants to disease presentation.

From a methodological perspective, combining colocalization analysis and genomic SEM analysis appears useful to identify shared common genetic variants associated with two traits. Genomic SEM is a powerful approach that boost power to identify novel shared risk SNPs

that remain undetected in univariate GWAS of overlapping traits (Amare et al., 2019; Grotzinger et al., 2019), while colocalization allows identification of shared and specific SNPs after taking linkage disequilibrium (LD) into account. At the conventional genome-wide  $p$ -value threshold of  $5e-8$ , both approaches converge (all SNPs identified by genomic SEM colocalized between ASD and ADHD). However, there were some divergences between both approaches at a more lenient  $p$ -value threshold. Specifically, of the 27 SNPs that were found to be shared by ASD and ADHD in the genomic SEM ( $p$ -value threshold at  $1e-6$ ), 9 SNPs did not colocalize between ASD and ADHD GWAS. Converging findings in the present study provide more certainty that the SNPs that were identified by both methods are truly shared. However, a systematic comparison of genomic SEM analysis and colocalization methods, with simulation studies, would be valuable to further understand when the two methods converge or diverge, as well as their respective strengths and limitations.

Finally, results from the bidirectional Mendelian randomisation analyses indicated that the risk of ASD was associated with an increased risk of ADHD and vice versa. This bidirectional relationship was also found using multiple MR methods, including some that have been developed to control for bias due to horizontal pleiotropy, such as MR-PRESSO and MR contamination mixture methods (Slob & Burgess, 2020). When SNPs colocalizing between ASD and ADHD were excluded, the bidirectional relationship between both disorders prevailed. Importantly, we note that a strong shared heritable confounder can lead to MR findings consistent with bi-directional causal relationships, even in the absence of such relationships. Such a strong shared heritable confounder is entirely plausible for disorders as comorbid as ASD and ADHD, and consistent with our genomic SEM and colocalization

findings. Therefore, our MR findings, although largely consistent across methods, cannot definitively establish such reciprocal causal relationships at the phenotypic level.

### **Limitations**

First, although summary statistics for ASD and ADHD were obtained from the largest samples available, they enabled the identification of only a few SNPs at the conventional genome-wide  $p$ -value threshold of  $5e-8$  (and thus few genes, especially for ASD). This contrasts with the findings that many more genes have been linked to ASD or ADHD with regards to rare genetic variants (Gudmundsson et al., 2019; Satterstrom et al., 2019). Thus, GWAS of ASD and ADHD based on larger samples are needed to fully explore the genetic complexity of these disorders. Second, there were some sample overlap in controls that might have biased our findings. However, genomic SEM (Grotzinger et al., 2019) and colocalization (C. N. Foley et al., 2019) are unbiased by sample overlap. The extent to which MR findings are biased by sample overlap remains uncertain but the bias is likely to be small (Burgess et al., 2016).

### **Conclusion**

Our study combined multivariate genomic approaches to elucidate the comorbidity between two of the most common neurodevelopmental disorders. The present study uncovers common genetic variants and genes shared by ASD and ADHD and those specific to each disorder. About one fifth to one third of the common genetic variants associated with one disorder are also linked to the other disorder. From a methodological perspective, our study highlights the advantage of using multiple methods (genomic SEM, colocalization, and Mendelian randomization) to identify and interpret the contribution of common genetic variation to the comorbidity of the two disorders. Further investigation of the shared and specific genetic

factors of neurodevelopmental disorders may help identify pathophysiological pathways and new targets for treatment.

## **ACKNOWLEDGEMENTS**

**Author contributions:** HP and JBP designed the study. HP performed the analyses and drafted the manuscript. JBP supervised the findings of this work. All authors discussed the results and contributed to the writing of the final manuscript.

**Conflict of interest:** There is no conflict of interest.

**Funding:** JBP is supported by the Medical Research Foundation Emerging Leaders 1st Prize 2018 Adolescent Mental Health.

**Data and materials availability:** Summary statistics for ASD and ADHD were obtained from the European ancestry subgroup of the Psychiatric Genomics Consortium and iPSYCH (PGC + iPSYCH). Data are freely available on the PGC website (<https://www.med.unc.edu/pgc/download-results>). Any inquiries about analytical results or other information should be directed to Lead Contact, Hugo Peyre ([peyrehug@yahoo.fr](mailto:peyrehug@yahoo.fr)).

**Correspondence** to Hugo Peyre, MD, PhD: LSCP, Département d'Etudes Cognitives. Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France. Email: [peyrehugo@yahoo.fr](mailto:peyrehugo@yahoo.fr).

## REFERENCES

- Alonso-Gonzalez, A., Calaza, M., Rodriguez-Fontenla, C., & Carracedo, A. (2019). Novel Gene-Based Analysis of ASD GWAS : Insight Into the Biological Role of Associated Genes. *Frontiers in Genetics, 10*. <https://doi.org/10.3389/fgene.2019.00733>
- American Psychiatric Association. (2013). *Draft DSM-V Criteria for ADHD*. [www.adhdbasics.org/ref/DSM-V.pdf](http://www.adhdbasics.org/ref/DSM-V.pdf). [www.adhdbasics.org/ref/DSM-V.pdf](http://www.adhdbasics.org/ref/DSM-V.pdf)
- Baio, J. (2018). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR. Surveillance Summaries, 67*. <https://doi.org/10.15585/mmwr.ss6706a1>
- Bhutta, A. T., Cleves, M. A., Casey, P. H., Cradock, M. M., & Anand, K. J. S. (2002). Cognitive and behavioral outcomes of school-aged children who were born preterm : A meta-analysis. *JAMA, 288*(6), 728-737.
- Bowden, J., Del Greco M, F., Minelli, C., Zhao, Q., Lawlor, D. A., Sheehan, N. A., Thompson, J., & Davey Smith, G. (2019). Improving the accuracy of two-sample summary-data Mendelian randomization : Moving beyond the NOME assumption. *International Journal of Epidemiology, 48*(3), 728-742. <https://doi.org/10.1093/ije/dyy258>
- Burgess, S., Davies, N. M., & Thompson, S. G. (2016). Bias due to participant overlap in two-sample Mendelian randomization. *Genetic Epidemiology, 40*(7), 597-608. <https://doi.org/10.1002/gepi.21998>
- Christensen, J., Pedersen, L., Sun, Y., Dreier, J. W., Brikell, I., & Dalsgaard, S. (2019). Association of Prenatal Exposure to Valproate and Other Antiepileptic Drugs With Risk for Attention-Deficit/Hyperactivity Disorder in Offspring. *JAMA Network Open, 2*(1), e186606. <https://doi.org/10.1001/jamanetworkopen.2018.6606>

- Constantino, J. N., Hudziak, J. J., & Todd, R. D. (2003). Deficits in reciprocal social behavior in male twins : Evidence for a genetically independent domain of psychopathology. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*(4), 458-467. <https://doi.org/10.1097/01.CHI.0000046811.95464.21>
- Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V., Purcell, S. M., Perlis, R. H., Mowry, B. J., Thapar, A., Goddard, M. E., Witte, J. S., Absher, D., Agartz, I., Akil, H., Amin, F., Andreassen, O. A., Anjorin, A., Anney, R., Anttila, V., ... International Inflammatory Bowel Disease Genetics Consortium (IIBDGC). (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics, 45*(9), 984-994. <https://doi.org/10.1038/ng.2711>
- Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Churchhouse, C., Dumont, A., Eriksson, N., Gandal, M., Goldstein, J. I., Grasby, K. L., Grove, J., ... Neale, B. M. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genetics, 51*(1), 63-75. <https://doi.org/10.1038/s41588-018-0269-7>
- Dobbyn, A., Huckins, L. M., Boocock, J., Sloofman, L. G., Glicksberg, B. S., Giambartolomei, C., Hoffman, G. E., Perumal, T. M., Girdhar, K., Jiang, Y., Raj, T., Ruderfer, D. M., Kramer, R. S., Pinto, D., CommonMind Consortium, Akbarian, S., Roussos, P., Domenici, E., Devlin, B., ... Sieberts, S. K. (2018). Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS. *American Journal of Human Genetics, 102*(6), 1169-1184. <https://doi.org/10.1016/j.ajhg.2018.04.011>

- Foley, C. N., Staley, J. R., Breen, P. G., Sun, B. B., Kirk, P. D. W., Burgess, S., & Howson, J. M. M. (2019). A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *BioRxiv*, 592238.  
<https://doi.org/10.1101/592238>
- Foley, C., & Staley, J. (2019). *hyprcoloc : HyPrColoc in jrs95/hyprcoloc: Hypothesis Prioritisation in multi-trait Colocalisation (HyPrColoc) analyses*.  
<https://rdrr.io/github/jrs95/hyprcoloc/man/hyprcoloc.html>
- Fortune, M. D., Guo, H., Burren, O., Schofield, E., Walker, N. M., Ban, M., Sawcer, S. J., Bowes, J., Worthington, J., Barton, A., Eyre, S., Todd, J. A., & Wallace, C. (2015). Statistical colocalization of genetic risk variants for related autoimmune diseases in the context of common controls. *Nature Genetics*, 47(7), 839-846.  
<https://doi.org/10.1038/ng.3330>
- Ghirardi, L., Brikell, I., Kuja-Halkola, R., Freitag, C. M., Franke, B., Asherson, P., Lichtenstein, P., & Larsson, H. (2018). The familial co-aggregation of ASD and ADHD : A register-based cohort study. *Molecular Psychiatry*, 23(2), 257-262.  
<https://doi.org/10.1038/mp.2017.17>
- Giambartolomei, C., Zhenli Liu, J., Zhang, W., Hauberg, M., Shi, H., Boocock, J., Pickrell, J., Jaffe, A. E., CommonMind Consortium, Pasaniuc, B., & Roussos, P. (2018). A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics (Oxford, England)*, 34(15), 2538-2545.  
<https://doi.org/10.1093/bioinformatics/bty147>
- Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., Ip, H. F., Marioni, R. E., McIntosh, A. M., Deary, I. J., Koellinger, P. D., Harden, K. P., Nivard, M. G., & Tucker-Drob, E. M. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex



traits. *Nature Human Behaviour*, 3(5), 513-525. <https://doi.org/10.1038/s41562-019-0566-x>

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O. A., Anney, R., Awashti, S., Belliveau, R., Bettella, F., Buxbaum, J. D., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Christensen, J. H., ... Børglum, A. D. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics*, 51(3), 431-444. <https://doi.org/10.1038/s41588-019-0344-8>

Grzadzinski, R., Dick, C., Lord, C., & Bishop, S. (2016). Parent-reported and clinician-observed autism spectrum disorder (ASD) symptoms in children with attention deficit/hyperactivity disorder (ADHD) : Implications for practice under DSM-5. *Molecular Autism*, 7, 7. <https://doi.org/10.1186/s13229-016-0072-1>

Gudmundsson, O. O., Walters, G. B., Ingason, A., Johansson, S., Zayats, T., Athanasiu, L., Sonderby, I. E., Gustafsson, O., Nawaz, M. S., Jonsson, G. F., Jonsson, L., Knappskog, P.-M., Ingvarsdottir, E., Davidsdottir, K., Djurovic, S., Knudsen, G. P. S., Askeland, R. B., Haraldsdottir, G. S., Baldursson, G., ... Stefansson, K. (2019). Attention-deficit hyperactivity disorder shares copy number variant risk with schizophrenia and autism spectrum disorder. *Translational Psychiatry*, 9(1), 1-9. <https://doi.org/10.1038/s41398-019-0599-y>

Hollingdale, J., Woodhouse, E., Young, S., Fridman, A., & Mandy, W. (2019). Autistic spectrum disorder symptoms in children and adolescents with attention-deficit/hyperactivity disorder : A meta-analytical review. *Psychological Medicine*, 1-14. <https://doi.org/10.1017/S0033291719002368>

Lai, M.-C., Kasse, C., Besney, R., Bonato, S., Hull, L., Mandy, W., Szatmari, P., & Ameis, S. H. (2019). Prevalence of co-occurring mental health diagnoses in the autism

- population : A systematic review and meta-analysis. *The Lancet. Psychiatry*, 6(10), 819-829. [https://doi.org/10.1016/S2215-0366\(19\)30289-5](https://doi.org/10.1016/S2215-0366(19)30289-5)
- Lee, P. H., Anttila, V., Won, H., Feng, Y.-C. A., Rosenthal, J., Zhu, Z., Tucker-Drob, E. M., Nivard, M. G., Grotzinger, A. D., Posthuma, D., Wang, M. M.-J., Yu, D., Stahl, E. A., Walters, R. K., Anney, R. J. L., Duncan, L. E., Ge, T., Adolfsson, R., Banaschewski, T., ... Smoller, J. W. (2019). Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell*, 179(7), 1469-1482.e11. <https://doi.org/10.1016/j.cell.2019.11.020>
- Lord, C., Brugha, T. S., Charman, T., Cusack, J., Dumas, G., Frazier, T., Jones, E. J. H., Jones, R. M., Pickles, A., State, M. W., Taylor, J. L., & Veenstra-VanderWeele, J. (2020). Autism spectrum disorder. *Nature Reviews Disease Primers*, 6(1), 1-23. <https://doi.org/10.1038/s41572-019-0138-4>
- Nivard, M. (2019). *MichelNivard/GenomicSEM : Structural equation modeling based on GWAS summary statistics version 0.0.2 from GitHub*. <https://rdrr.io/github/MichelNivard/GenomicSEM/>
- Pingault, J.-B., Cecil, C., Murray, J., Munafò, M., & Viding, E. (2016). *Causal inference in psychopathology : A systematic review of Mendelian randomisation studies aiming to identify environmental risk factors for psychopathology*. <https://doi.org/10.17863/CAM.4366>
- Pingault, J.-B., O'Reilly, P. F., Schoeler, T., Plouhidis, G. B., Rijdsdijk, F., & Dudbridge, F. (2018). Using genetic data to strengthen causal inference in observational research. *Nature Reviews Genetics*, 19(9), 566-580. <https://doi.org/10.1038/s41576-018-0020-3>
- Ronald, A., Simonoff, E., Kuntsi, J., Asherson, P., & Plomin, R. (2008). Evidence for overlapping genetic influences on autistic and ADHD behaviours in a community

twin sample. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 49(5), 535-542. <https://doi.org/10.1111/j.1469-7610.2007.01857.x>

Satterstrom, F. K., Walters, R. K., Singh, T., Wigdor, E. M., Lescai, F., Demontis, D., Kosmicki, J. A., Grove, J., Stevens, C., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Palmer, D. S., Maller, J. B., Nordentoft, M., Mors, O., Robinson, E. B., Hougaard, D. M., Werge, T. M., Mortensen, P. B., ... Daly, M. J. (2019). Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. *Nature Neuroscience*, 22(12), 1961-1965. <https://doi.org/10.1038/s41593-019-0527-8>

Schorck, A. J., Won, H., Appadurai, V., Nudel, R., Gandal, M., Delaneau, O., Revsbech Christiansen, M., Hougaard, D. M., Bækved-Hansen, M., Bybjerg-Grauholm, J., Giørtz Pedersen, M., Agerbo, E., Bøcker Pedersen, C., Neale, B. M., Daly, M. J., Wray, N. R., Nordentoft, M., Mors, O., Børglum, A. D., ... Werge, T. (2019). A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment. *Nature Neuroscience*, 22(3), 353-361. <https://doi.org/10.1038/s41593-018-0320-0>

Septier, M., Peyre, H., Amsellem, F., Beggiato, A., Maruani, A., Poumeyreau, M., Amestoy, A., Scheid, I., Gaman, A., Bolognani, F., Honey, G., Bouquet, C., Ly-Le Moal, M., Bouvard, M., Leboyer, M., Bourgeron, T., & Delorme, R. (2019). Increased risk of ADHD in families with ASD. *European Child & Adolescent Psychiatry*, 28(2), 281-288. <https://doi.org/10.1007/s00787-018-1206-0>

Siewert, K. M., & Voight, B. F. (2018). Bivariate Genome-Wide Association Scan Identifies 6 Novel Loci Associated With Lipid Levels and Coronary Artery Disease. *Circulation. Genomic and Precision Medicine*, 11(12), e002239. <https://doi.org/10.1161/CIRCGEN.118.002239>

- Slob, E. A. W., & Burgess, S. (2020). A comparison of robust Mendelian randomization methods using summary data. *Genetic Epidemiology*, *44*(4), 313-329.  
<https://doi.org/10.1002/gepi.22295>
- Taylor, M. J., Charman, T., Robinson, E. B., Plomin, R., Happé, F., Asherson, P., & Ronald, A. (2013). Developmental associations between traits of autism spectrum disorder and attention deficit hyperactivity disorder : A genetically informative, longitudinal twin study. *Psychological Medicine*, *43*(8), 1735-1746.  
<https://doi.org/10.1017/S003329171200253X>
- Wallace, C. (2019). Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. *BioRxiv*, 838946. <https://doi.org/10.1101/838946>
- Wallace, C., Giambartolomei, C., & Plagnol, V. (2019). *coloc : Colocalisation Tests of Two Genetic Traits* (3.2-1) [Computer software]. <https://CRAN.R-project.org/package=coloc>
- Watanabe, K., Taskesen, E., Bochoven, A. van, & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, *8*(1), 1-11. <https://doi.org/10.1038/s41467-017-01261-5>
- Willcutt, E. G. (2012). The Prevalence of DSM-IV Attention-Deficit/Hyperactivity Disorder : A Meta-Analytic Review. *Neurotherapeutics*, *9*(3), 490-499.  
<https://doi.org/10.1007/s13311-012-0135-8>
- Zhao, Q., Wang, J., Bowden, J., & Small, D. (2018). Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Annals of Statistics*, *48*. <https://doi.org/10.1214/19-AOS1866>

**Key points**

- We identified genes mapped to SNPs that are shared by ASD and ADHD and those that are specific to each disorder.
- About one third of the common genetic variants associated with ASD were linked to ADHD.
- About one fifth of the common genetic variants associated with ADHD were linked to ASD.
- Bidirectional Mendelian randomization analyses indicated that greater ASD symptomatology increased the risk of ADHD and vice versa.