Localizing regions in the genome contributing to ADHD, aggressive and antisocial behavior

Running title: Genetic overlap between ADHD, aggression and antisocial behavior

Mariana Lizbeth Rodríguez López[1], Barbara Franke[1,2]*, Marieke Klein[1,3]

[1] Radboud university medical center, Donders Institute for Brain, Cognition and Behaviour, Department of Human Genetics, Nijmegen, The Netherlands

[2] Radboud university medical center, Donders Institute for Brain, Cognition and Behaviour, Department of Psychiatry, Nijmegen, The Netherlands

[3] University Medical Center Utrecht, UMC Utrecht Brain Center, Department of Psychiatry, Utrecht, the Netherlands

*Corresponding author

Corresponding author

Barbara Franke | Telephone: +31243610181 | Email: barbara.franke@radboudumc.nl | Radboudumc, Department of Human Genetics, route 855 | Postbus 9101, 6500 HB Nijmegen, The Netherlands

Conflict of interest

Barbara Franke received educational speaking fees from Shire and Medice. The other authors declare no conflicts of interests.

Acknowledgements

This work was partly carried out on the Dutch national e-infrastructure with the support of SURF Foundation. BF and MK are supported by funding for the Dutch National Science Agenda NeurolabNL project (grant 400-17-602). This work was also supported by the European Community's Horizon 2020 Programme (H2020/2014 – 2020) under grant agreements n° 667302 (CoCA).

Abstract

Background: Attention-Deficit/Hyperactivity Disorder (ADHD) is a common neurodevelopmental disorder. ADHD co-morbidities cover a broad range of traits, including aggressive behavior (AGG) and antisocial behavior (ASB). Genetics has a high percentage on the heritability of the three traits, mainly attributed to large numbers of common variants (SNPs) with very small effect sizes. We believe that some of this common variants could give insights to other elements overlapping between ADHD, AGG and ASB. Therefore, we worked at regions where these variants group are, tested how some of these regions contribute and looked for candidate genes within these.

Methods: We tested for genome-wide genetic correlations using LD score regression analysis (LDSC), followed by analyses for local SNP-based heritability and cross-trait analyses for local genetic correlations. Propective genes were followed up by further investigations of gene expression patterns and genetic associations with structural subcortical brain volumes.

Results: ADHD, AGG and ASB showed global positive genetic correlations among them, but also with neuroticism. Negative global genetic correlations were observed between ADHD, AGG, ASB and Primary Sclerosing Cholangitis. At the locus level, large heterogeneity between regions was observed. Cross-trait comparisons revealed 53 shared loci of interest and prioritized 20 genes which are expressed in the human brain and showed association with subcortical brain volumes.

Conclusions: This study emphasizes the importance of investigating shared genetic effects between correlated traits at the locus level. Converging evidence from different cross-trait analyses approaches highlights novel candidate genes underlying the shared biological mechanisms of ADHD, AGG and ASB.

Keywords: genetic correlation, shared genetic loci, ADHD, aggressive behavior, antisocial behavior

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a common neurodevelopmental disorder with a world-wide prevalence of 5% in children and 2.5% in adults [Polanczyk, G., de Lima, M.S.,et al, 2007], [Simon, V., Czobor, P., et al., 2009]. ADHD covers a broad spectrum of heterogeneous symptoms and impairments which can be present throughout the lifespan [Biederman, J. et al, 2000], [Faraone, S.V., et al 2006]. According to the DSM-5, three main presentations of ADHD can occur: predominantly inattentive, predominantly hyperactive-impulsive or a combined subype [American Psychiatric Association. 2013.]. Twin and family studies revealed that ADHD is highly heritable, with heritability estimates ranging between 70-80% [Franke, B. et al., 2012], [Asherson, P. & Gurling, H., 2012][Larsson, H. et al 2014]. Genetic studies in ADHD mainly focused on common genetic variants (> 1% in the general population) and most common genetic variants showed small effect sizes [Faraone, S.V. et al., 2015]. Recently, the largest ADHD genome-wide association study (GWAS) obtained heritability estimates from single nucleotide polymorphisms (SNPs) and this so-called SNP-based heritability was estimated to be 22% [Demontis D. et al., 2019]. Patients with ADHD also often suffer from co-occurring psychiatric disorders, such as autism spectrum disorder (ASD), schizophrenia, depression, substance use disorders and anxiety disorders [Biederman, J., et al, 1991], [Cross-Disorder Group PGC, *Lancet* 2013], [Faraone, S.V. et al., 2015]. However, presence of co-morbid disorders largely varies between patients.

Aggressive behavior (AGG) is defined as the intended act (physical, verbal or psychological), of causing harm to others or to oneself [Liu, J., 2004], [Ukukuro, S. et al., 2013]. Different types of AGG (i.e. different intentions) have been described in different categories, of which the most common the bimodal category is, with its reactive (lead by impulsivity) and proactive (deliberated and planed) subtypes [Crick & Dodge, 1996]. AGG can also being considered as a natural behavior, with defensive or offensive subtypes [Weinshenker & Siegel, 2002]. Depending on the severity (measured by means of questionnaires and structured [diagnostic] interviews), it can be used as a behavioral criterion for other disorders, like intermittent explosive disorder [American Psychiatric Association. 2013] or conduct disorder [American Psychiatric Association. 2013]. ASB refers to actions and attitudes that affects others and violates societal norms. [Lahey & Waldman, 2003], [Burt & Neiderhiser,2009]. Similar to AGG, ASB are only a set of acts, only when the behavior becomes persistent, it is diagnosed as Antisocial Personality Disorder [American Psychiatric Association. 2013.]. ASB includes behaviors with an aggressive connotation, like bullying or mugging, but also includes non-aggressive ones like stealing, truancy, and lying [Lahey & Waldman, 2003] (i.e. AGG is a form of ASB, but not all ASB is AGG).

However, this does not mean that AGG and ASB are completely the same. Aggressiveness in ASB could involve similar, but also different behavioral patterns, etiologic mechanisms and genetics. [Burt & Neiderhiser,2009], [Davis, M. H., 2018]. Both AGG and ASB are moderately to highly heritable, with heritability estimates for AGG ranging between ~50-70% [Miles & Carey, 1997], [Hudziak et al., 2003] and between ~50-60% for ASB [Rhee & Waldman, 2002][Ferguson, 2010]. Only a few GWASs for AGG and ASB are currently available, all with limited sample sizes. A GWAS on AGG in children did not identify any genome-wide significant hits, but reported SNP-based heritability estimates between 10-54% [Pappa et al., 2016]. One GWAS on ASB in general population, which included the aggressive subtype, reported a SNP-based heritability of only 5% [Tielbeek, et al 2017].

As we mentioned above, AGG and ASB are usually included as behavioral criterion for other behavioral, psychiatric and neurodevelopmental traits, including ASD [Mandy W. et al., 2013], bipolar disorder [Látalová K., 2009], oppositional defiant disorder [American Psychiatric Association. 2013], substance use disorders [Séguin, J.R & Tremblay, R.E., 2013], and ADHD. [Biederman, J., Newcorn, J., Sprich S., 1991], [Barkley R.A. et al., 2004][Ralf, K.H. et al., 2015]. So far, mainly family and twin studies investigated the genetic mechanisms of AGG and ASB within the ADHD spectrum. Results of such studies showed that ADHD with AGG or ASB could be a different categorical type of ADHD due to the different development, persistence and riskiness seen in the patients. Differences can be even more specific, given the differences between AGG and non-AGG ASB subtypes, showing AGG to have greater relationship possible due to impulsiveness, with reactive aggression as the most present one. [Doyle & Faraone, 2002], [Christiansen et al., 2008][Retz & Roster, 2009]. Results of common genetic variant studies hint towards similar observations and point out that this relationship can be caused by a greater load of genetic factors [Hamshere et al., 2013]. Therefore, ADHD is not inevitably associated with AGG or ASB, but patients with co-occurring symptoms, overall have a higher persistence rate and risk, possibly driven by the complex combination of genetic factors.

Neurodevelopmental disorders and psychiatric traits are commonly known to have a complex etiology, with both genetic and environmental factors involved. Previous studies showed that clinical disorders and related behavioral symptoms can largely overlap in terms of their underlying genetics [Demontis D. et al., 2019]. However, these global approaches to assess the genetic correlations between two traits could over-generalize the relationship between traits and thereby lack specificity in terms of the actual causal genes.

In this study, with the use of GWAS summary statistics data, we do not only assess the genome-wide genetic correlations between ADHD, AGG and ASB, but also aimed to localize

the genetic effects. By defining the contribution of shared loci between to traits/disorders, these loci could give us specific insights into the genetic complexity underlying ADHD, AGG and ASB. Finally, we explored gene expression patterns in relevant tissues and prioritized candidate genes for future studies.

Methods

GWAS summary statistics

GWAS summary statistics (GWAS-ss) data were obtained from publicly available independent meta-analysis studies. Data files were formatted according to the requirements for the analysis method. All data had to meet general criteria to be included in the analyses. Participants must be of European ancestry. Variants had to be annotated in human genome assembly hg19, had an identifier column, supplied at least the effect allele with their respective signed summary statistic (beta, OR, Z-score, etc), provided with p-values. We only considered data of autosomal chromosomes. Because of its unusual LD structure, we removed the Major Histocompatibility Complex (MHC) region from analysis, as defined by the Genome Reference Consortium (GRC) for human genome assembly hg19 to be in chromosome 6 from 28,477,797bp to 33,448,354bp.

More details on the quality control (QC) steps, conditions, criteria, software or pipelines used by the individual studies, can be found in their original papers (Demontis et al., 2019; Pappa et al., 2016; Tielbeek et al., 2017; Okbay et al., 2016; Sun Gou et al., 2017). A short summary with the main characteristics of each of these studies can be found below.

ADHD

The ADHD GWAS-ss file was obtained from the study published by Demontis et al. in 2019. A genome-wide meta-analysis including 12 cohorts of the Psychiatric Genomics Consortium (PGC) and the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH) included a total of 19,099 cases and 34,194 controls with European ancestry, across children and adults (https://www.med.unc.edu/pgc/results-and-downloads). Cases were diagnosed through tests, interviews and reports by psychiatrists (different criteria between cohorts). DNA was extracted of blood samples and genotyped. The QC processing was already done by the meta-analysis study for each of the cohort's datasets. Details on imputation of the dataset can be found in the original article (Demontis et al. 2019). Individual cohorts were meta-analyzed

with an inverse-weighted fixed effects model. Finally, markers with an effective sample size > 70%, an imputation quality (INFO score) > 0.8 and a minor allele frequency (MAF) > 0.01 were kept for subsequent analyses.

Aggressive behavior

The aggressive behavior (AGG) GWAS-ss data was provided from the study by Pappa et al. in 2016, performed in the framework of the Early Genetics and Lifecourse Epidemiology (EAGLE) consortium (http://www.tweelingenregister.org/EAGLE/). There, they meta-analyzed nine cohorts, with a total of 18,988 participants and all of them were unrelated children between the age of 3 to 15 years. Aggressive behavior was measured on a continuous scale and assessed through well-validated questionnaires (correlated and interchangeable) and filled in by the children's parents. DNA was extracted from whole blood or buccal cells and genotyped. Details on imputation of the dataset can be found in the original article (Pappa et al. 2016). Data of individual cohorts were meta-analyzed using a sample size-weighted z-score method and only variants with MAF > 0.05 and INFO score > 0.4 were kept for further analyses.

Antisocial behavior

The GWAS-ss, for antisocial behavior (ASB) was obtained from the study by Tielbeek et al. in 2017 through the Broad Antisocial Behavior Consortium (http://broadabc.ctglab.nl/summary_statistics). The meta-analysis consisted of five discovery cohorts, giving a total of 16,400 individuals (mean age 6.7 - 43.8). A continuous scale for ASB was used and each one of the cohorts applied different antisocial measures (based also on age), including aggressive and nonaggressive domains of ASB (for detailed overview see Tielbeek et al 2017). Details on imputation of the data set can be found in the original article (Tielbeek et al 2017). Individual cohorts were meta-analyzed using a sample size-weighted z-score fixed effect model and only variants with a total sample size > 10,000, MAF > 0.01 and INFO score > 0.6 were kept for further analyses.

Additional GWAS summary statistics

The two additional GWAS-ss included in this study, had to meet the same general requirements as stated above, were downloaded and formatted for subsequent analysis. The GWAS-ss for Neuroticism (Neurot), was obtained from the study published by Okbay et al. in 2016 (https://www.thessgac.org/data). This is a study with 170,910 participants and meta-analyzed GWAS data, using a sample size-weighted fixed-effects model, from the UKBiobank

data and other previous GWASs of neuroticism. The second GWAS-ss, Primary Sclerosing Cholangitis (PSC), was provided by Sun Gou et al in 2017. The data comprises a total of 14,890 participants (sum of 2,871 cases and 12,019 controls) and it is a meta-analysis of four cohorts by using a linear mixed model (http://www.ipscsg.org/). Additional details of both studies can be found in their original articles (Okbay et al, 2016; Sun Gou et al, 2017)

SNP heritability (SNP h^2) estimates

For all traits of interest, we assessed the contribution that SNPs have on the phenotypic variance. We estimated the SNP-based heritability using the LD score regression method, performed by the command line tool LDSCore (LDSC) version 1.0.0 (Bulik-Sullivan et al 2015, a). This method was originally introduced with the aim of estimating the proportion of inflation in a GWAS dataset due to the actual polygenic signal and not from confounding factors, such as population stratification or cryptic relatedness. LDSC takes into account the properties of variants in linkage disequilibrium (LD) with causal variants, by performing χ^2 association statistic regression against these and evaluating the intercept. Of the most interest for this study in the regression is the slope, which can be interpreted as an estimator of heritability.

In order to use LDSC, GWAS-ss input had to meet the following minimum formatting requirements: presence of a column containing a SNP identifier, effect allele, other allele, p-value, sample size and a signed summary statistics. Not all GWAS-ss provided the sample size per genetic variant. In this case all variants were given the same sample size, namely the maximum sample size of the dataset. Columns for INFO score and MAF were kept and included when provided in the original data. Data processing in LDSC includes an automatic QC step. Here the imputation quality (if the INFO column is present) is controlled by excluding variants with INFO score < 0.9 and those that do not match with HapMap3 SNPs reference panel (considering to have well imputed variants and checking if SNP IDs are matching their respective alleles). Variants with a MAF < 0.01, small insertions and deletions (indels), strand ambiguous markers, duplicated SNPs, genetic variants with out of bounds p-values or small sample sizes are excluded. LD score estimates computed based on the 1000 Genome European data are already pre-calculated and can be downloaded here: https://data.broadinstitute.org/alkesgroup/LDSCORE/. For AGG, ASB and Neuroticism we calculated the SNP-based h2 on the observed scale and for ADHD and PSC we obtained the SNP-based h2 on the liability scale, applying a reported 5% population prevalence (Polanczyk, G. et al. 2007) and 36% sample prevalence for ADHD, while for PSC a reported 0.01% population prevalence (Boonstra, K et al, 2012) and 19.4% sample prevalence.

Genome wide genetic correlations (r_g)

LDSC expanded their method introducing the cross-trait LD regression technique, which estimates the genetic correlation between traits (Bulik-Sullivan et al 2015, b). Based on the same assumption as in the SNP h2 section above, variants with highest LD scores are more likely to be in LD with the causal variants, giving higher χ^2 association statistics. The explained χ^2 can also be represented as Z^2 which can be split into the product of two Z-scores, each one of them belonging to one of the two traits. In this case, the regression slope, is now the genetic covariance. Finally, normalizing the covariance (ρ_g) by the two SNP heritabilities ($SNPh1^2$,$SNPh2^2$), gives the genetic correlation (r_g), (Eq. 1).

$$r_g = \frac{\rho_g}{\sqrt{SNPh1^2 * SNPh2^2}} \qquad \text{Equation 1.}$$

We performed three principal cross-trait analyses between our traits of interest, ADHD vs AGG, ADHD vs ASB and AGG vs ASB, to estimate the genetic correlation between those traits.

Genome wide genetic correlations with other traits

We uploaded the GWAS-ss of AGG and ASB to LD Hub (version 1.9.0) (Zheng et al, 2017, http://ldsc.broadinstitute.org/ldhub/). LD Hub is a big database of publicly available GWAS-ss and at the same time, a web interface for automate running LDSC SNP h^2 estimates and r_g scores between your trait of interest against the ones available on the platform. The available GWAS-ss are publicly available and are shared by different projects and cover diverse categories of traits.

Input format requirements and QC were very much alike as described in the SNP h^2 section since LD Hub also uses LDSC. Now, all analyses were run automatically by the web interface and only a few additional adjustments for the input files were needed: GWAS-ss was required to have a Z-score column and the input file had to be zipped. All traits already present in the database, were equally formatted with a series of requirements listed in the LD Hub web page. One at a time, our GWAS-ss of interest (AGG and ASB) were uploaded to the web platform. For the genetic correlation analyses we selected all the 234 traits from 25 categories listed and excluded 597 traits from the UK Biobank. The Bonferroni corrected threshold for significance was set to $p < 0.05/234$.

From all 234 traits from LD Hub, we selected two traits for subsequent analyses: Neuroticism (Neurot) and Primary Sclerosing Cholangitis (PSC). We confirmed LD Hub genetic correlation results by running LDSC manually with combinations between ADHD, AGG and ASB against Neuroticism and PSC.

Local SNP heritability estimates

To assess which genomic regions are contributing most to the genome-wide SNP-based heritability estimates of the traits of interest, we used the software package Heritability Estimation from Summary Statistics (HESS) version 0.5.4-beta (Shi et al, 2016). Working with GWAS-ss, HESS estimates, for each one of the loci defined in a genome partition file, the variance explained by all the typed SNPs at a single locus while taking into account the LD pattern between the SNPs. While traditional models assessing the global SNP-based heritability assume that all SNPs contribute to the trait (and this model is invalid at most risk loci), HESS uses a more robust approach as it does not assume any distribution for the effect sizes at causal variants.

We run HESS to estimate the local SNP-based heritability for all five traits of interest (ADHD, AGG, ASB, Neurot and PSC). The GWAS-ss had to meet minimum requirements and these were compatible with the previously used LDSC format; only two additional columns were needed: chromosome number and base pair position for the individual SNPs. HESS uses a genome partition file to define the LD-independent loci and this file was generated from a study in 2015 by Berisa and Pickrell, in which they defined a total of 1,703 approximately LD-independent loci with an average length of ~1.6 Mb for the European population (http://bitbucket.org/nygcresearch/ldetect-data); additional information can be found in the original article (Berisa & Pickrell, 2015). However, for the current study some adjustments to the genome partition file were made: we merged three loci with low number of SNPs with neighboring regions and removed six loci inside or partially inside the MHC, resulting in a total of 1,694 loci. For LD-pattern data, since we did not have the in-sample LD information, we used the 1000 Genomes Project for European population reference panel (1000 Genomes Project Consortium, 2012). An automated QC was run by HESS prior to the analyses, by removing SNPs with either duplicate IDs or position. Ideally, analysis is run by not included a pre-calculated SNP-based heritability estimate, however in the case of a GWAS sample size < 50,000 it is recommended to calculate this prior to the analyses. Most of the GWASs used in this study have sample sizes < 50,000, the total SNP-based heritability estimates obtained with LDSC (**Fig. 1**) were used for the HESS analyses as well. Regions were considered to be significant with a Bonferroni corrected p-value threshold at $p < 0.05/1,694$.

Local genetic covariances and correlations

To estimate the local genetic covariance between a pair of traits and an extension to HESS method, known as ρ-HESS, was used (Shi et al 2017). Similar to HESS, we used the adjusted genome partition file and LD-pattern data, and uploaded a pair of GWAS-ss (with same previous requirements) to perform eight cross-trait analyses between the five traits of interest (ADHD, AGG, ASB, Neurot. and PSC). Local genetic covariance was considered to be significant if $p < 0.05/1{,}694$ (Bonferroni corrected). Since the covariance is a value that changes depending on the original scale of the traits, this makes it difficult for comparisons. Therefore, we calculate the local genetic correlation, the standardized version of the local genetic covariance, for each of the 1,694 loci across all the eight cross-trait analyses. The calculation looked similar to the one described by Bulik-Sullivan, 2015 (Eq. 1); but was adapted by Shi and collegesto obtain the genetic correlation for the ith region (Shi, et al., 2017). Thus, the genetic covariance ρ_g (ρ-HESS results) and both trait's SNP-based heritability estimates (SNPh1, SNPh2; HESS results) used in the new equation were also calculated for the same i^{th} region (Equation 2). This was done for all 1,694 loci.

$$r_{g,local,i} = \frac{\rho_{g,local,i}}{\sqrt{SNPh1_{local,i}{}^2 * SNPh2_{local,i}{}^2}} \qquad \text{Equation 2.}$$

Localizing regions of interest

To reduce the number of loci for subsequent analyses, we only considered loci with a local genetic correlation score between minus one and one [-1,1], which are the expected values for a correlation analysis. Since Eq. 2 can return values over that range, meaning near zero heritability estimates and high standard errors for correlation scores, we removed those loci in order to reduce noise and false positives.

Next, the eight cross-trait analyses were grouped into three groups: Group 1 included the cross-trait analyses of the three principal traits of this study which are directly genetically correlated among each other, namely ADHD-AGG & ADHD-ASB. Group 2 included all cross-trait analyses in which Neuroticism was present (indirect association via a shared positively genetically correlated trait), namely ADHD-Neurot, AGG-Neurot & ASB-Neurot. Group 3 included all cross-trait analyses in which PSC was one of the traits of interest (indirect association via a shared negatively genetically correlated trait), namely ADHD-PSC, AGG-PSC & ASB-PSC. Finally, for each one of the three groups, we selected the loci, emerging from the local genetic correlation analyses, that were overlapping between all the cross-trait analyses within each group (Group 1-3).

Gene-based analyses

From the identified overlapping loci, we set out to pinpoint the potentially most contributing genes included within these loci. First, we annotated all genes that were completely or partially located within the previously selected loci in each of the three cross-trait analyses groups by uploading the positions of the loci into the UCSC Genome Browser (https://genome.ucsc.edu) tables tool (Karolchik et al., 2004). Second, we selected all genes that showed a gene-based association with the traits included in each group. For this, we used the gene-based association analysis implemented in Multi-marker Analysis of GenoMic Annotation software (MAGMA) version 1.05 (de Leeuw, et al., 2015). Gene-based association analysis in MAGMA includes all genetic markers (SNPs), while considering the LD between them and map them to genes, and then test the association of each genes with a phenotype of interest. MAGMA can also perform gene-based cross-trait meta-analysis, which means combining the Z-scores for each gene from two gene-based analyses (i.e., a pair of traits) (de Leeuw, et al., 2015). The following files are needed as input for MAGMA are: GWAS-ss, a reference LD data file, which we obtained from the 1000 Genomes Project phase 3 for the European population. For the gene-based meta-analysis, the results of the gene-based association analyses are used as input. First gene-based association analyses for each of the five traits were performed. Then, we run gene meta-analyses for the eight cross-trait pairs described above. Next, we selected all genes with a nominally significant gene-based association p-value for each cross-trait meta-analysis results and kept the overlapping genes between the cross-trait pairs from each group (Group 1-3). Finally, we selected all genes overlapping between the two-step processes mentioned above. We assessed for Bonferroni corrected p-values, using different thresholds specific for each group ($p < 0.05$/number of genes from group). We considered a gene to be significant if at least one of the cross-trait analyses got a p-value under their respective threshold.

Biological annotation and functional significance of candidate genes

To retrieve more information on functional significance and biological annotation and to prioritize the selected set of genes, we used three complementary approaches.

First, to assess mRNA expression profiles of our candidate genes, we required RNA-sequencing data from 53 tissues, provided by Genotype-Tissue Expression (GTEx) Project V7 (https://gtexportal.org/home/). From all 53 tissues, we focused 12 brain related tissues: amygdala, anterior cingulate cortex, caudate, cerebellar hemisphere, cerebellum, cortex, frontal cortex, hippocampus, hypothalamus, nucleus accumbens, putamen and substantia

nigra (GTEx Consortium, 2017). For each gene, we obtained the median Transcripts Per Million (TPM) in each tissue.

Second, we assessed spatio-temporal transcription patterns of the selected candidate genes, including 16 brain regions: cerebellar cortex, mediodorsal nucleus of the thalamus, striatum, amygdala, hippocampus and 11 areas of the neocortex (add reference Kang et al here). Fifteen periods of development were defined spanning from embryonic development to late adulthood. We used the expression module definition from the original publication (add reference here) to investigate whether our candidate genes present in one of the 29 modules. These modules describe co-expressed genes, which are differentially expressed or enriched in a region during defined periods of development. We performed a look-up of our genes whether they are present within at least one spatial-temporal module or enriched in a specific region.

Third, we performed gene-based association analyses by using data from the ENIGMA consortium (add reference Hibar et al 2015, Nature). GWAS summary statistics were downloaded from http://enigma.ini.usc.edu/research/download-enigma-gwas-results/ and gene-based association analyses were performed as described above (refer to relevant section above here) for all subcortical brain regions (accumbens, amygdala, caudate, hippocampus, intracranial volume (ICV), pallidum, putamen, and thalamus). For each group (Group 1-3), genes were considered significant at a Bonferroni-corrected significance level of $p < 0.05$ / (no. genes within group x 8).

Further prioritization of our candidate genes was accomplished by a scoring system, which incorporates the information from the three approaches described above. Each of the genes was scored with one point for (1) each of the 12 tissues within the top five highest median TMP position and half a point if the genes was present within the top ten. (2) for differential expression of a candidate gene in at least one region or presence within at least one module, (3) for a significant gene-based association with one of the subcortical brain regions and half a point if the association was nominally significant. Additionally, each significant gene-based association of the relevant cross-trait analyses within a group (Group 1-3) was considered as one point. We summed up all individual points per gene and prioritized those with a score above four.

Biological annotations for all candidate genes were obtained by the GeneCards Human Gene Database (https://www.genecards.org/) a database which collects information from other bioinformatic databases (Ensembl, Gene Ontology, NCBI Entrez Gene, UniProtKB/Swiss-Prot, etc., Stelzer, et al., 2016).

Results

General view on the genetic architecture

Since this work is mainly focusing on genetic factors, as an important component in the relationship between the three traits of interest, we first estimated the overall contribution of common genetic to these traits, by assessing the SNP-based heritability. Next, we determined their genetic correlations on a global genome-wide scale.

SNP Heritability

First, we assessed the phenotypic variance in our traits explained by common variants through estimating the SNP-based heritability using LD score regression (Bulik-Sullivan et al, 2015,a). For ADHD, we were able to reported this on liability scale, assuming a 5% population prevalence (Polanczyk, G. et al. 2007) and obtaining a SNP h^2 = 0.2167 (SE = 0.014, p = 4.84E-54), similar to the one reported by the original study (Demontis et al, 2019). For AGG and ASB, we reported the heritability on an observed scale, resulting in SNP h^2 = 0.051 (SE = 0.025, p = 0.041) and SNP h^2 = 0.0537 (SE = 0.0265, p = 0.043) respectively. In the case of AGG, the original study estimated SNP h^2 (using the GCTA tool) only on three of their 9 cohorts, and therefore reported different results (for the largest sample cohort SNP h^2 = 0.10 (SE = 0.06, p = 0.04) (Pappa et al, 2016)). While in ASB, using also LDSC, Tielbeek and colleagues reported a SNP h^2 = 0.05 (SE = 0.027, p = 0.03) (Tielbeek et al, 2017).

Genome Wide Genetic Correlations

We assessed and confirmed that common genetic variants have a role in the global genetic correlation between the three traits. We ran genetic correlation analysis in pairs, i.e. ADHD with AGG, ADHD with ASB and AGG with ASB, using LDSC (Bulik-Sullivan et al, 2015b).

We observed that the genetic correlation between AGG and ASB was low and non-significant (r_g = 0.24, SE = 0.3585, p = 0.5031, **Fig. 1**). On the other side, despite having obtained very low SNP h^2 estimates in AGG and ASB and low r_g score, the genetic correlations of AGG and ASB with ADHD were strong and significant ($p < 0.05/3$). The genetic correlation between ADHD and AGG was highest (r_g = 0.7007, SE = 0.1812, p = 0.0001) and between ADHD and ASB still high (r_g = 0.5112, SE = 0.1862, p = 0.006, **Fig. 1**). This indicated that there is genetic overlap between these three traits (with ADHD having a key role) and these findings forms the basis for the following studies.

Additional genetically correlated traits

In the next step we set out to search for additional traits and/or disorders that are also genetically correlated with ADHD, AGG and ASB by using the web platform LD Hub (Zheng et al, 2017).

From the LD Hub database, we selected a total of 234 available traits from 26 categories (**Suppl. Table 1**) and assessed the genetic correlation statistics. Considering the low heritability estimates which we obtained for AGG and ASB, it was not surprising that none of the genetic correlations were significant (r_g score p-values did not survive the Bonferroni correction ($p < 0.05/234$)). Therefore, we decided to perform a more exploratory analysis and we selected traits with r_g scores with at least a nominal p-value ($p < 0.05$). Next, we looked for overlapping and genetically correlated traits between AGG and ASB (nine traits in total). Finally, the selected traits were grouped depending on whether they showed a positive or negative genetic correlation (r_g score) with our traits of interest. A schematic overview of this procedure can be found in **Fig. 2**.

For the group with positive genetic correlations, a total of three traits were selected and for group of negative correlations, we selected a total of six traits (**Table 1**). The most significantly correlated traits from the positive genetic correlations group was Neuroticism (Neurot). Because Neuroticism has a cognitive biological background (in contrast with the other positive correlating traits group), and a phenotypic relationship with ADHD, AGG and ASB, we decided to include neuroticism in subsequent analyses. From the negatively correlated traits, we selected primary sclerosing cholangitis (PSC) to be included in our subsequent analyses. Unlike the other negative genetically correlating traits, PSC has a medical diagnosis with biological causes, symptoms and complications making it a more biological perspective trait, ADHD has been known to also co-morbid with other autoimmune diseases (Nielsen, et al., 2017), and is a novel trait for studying co-morbidity with ADHD, AGG and ASB, making an expanded exploratory study that could lead to other biological factors.

For neuroticism the SNP-based heritability was estimated to be 8.8% and significant (h2 = 0.0877, SE 0 0.0067, p = 3.78E-39) and for PSC the SNP-based heritability was estimated to be 10.3% and significant (h2 = 0.1032, SE = 0.0214, p = 1.42E-06). Those heritability estimates are comparable to the ones that were reported previously, (for neuroticisms SNP h2 = 0.091 and for PSC SNP h2 = 0.148) (Okbay et al., 2016; Ji et al., 2017). Both Neuroticism and PSC were significantly genetically correlated with ADHD (ADHD-Neurot r_g = 0.264, SE = 0.0526, p

= 5.04E-07; ADHD-PSC r_g = -0.2639, SE = 0.066, p = 6.37E-05) and nominally significant with AGG and ASB (**Fig. 1**). Since AGG and ASB got a low SNP-based heritability and a low genetic correlation, with high standard errors, the cross-trait analysis between both (AGG-ASB) was excluded from the rest of analyses and we continued this study with eight cross-trait analyses.

These results showed us that globally, using information obtained from GWAS, it can be proven that there is significant genetic correlation between our primary traits of interest, ADHD, AGG and ASB. These traits are also (significantly in the case of ADHD) correlated with Neuroticism and PSC, which may allow generate a new perspective on investigating shared genetic effects between our three traits of interest.

Local view on the genetic architecture

Local SNP heritability

First, we estimated which regions contribute most to the phenotypic variance of a trait explained by the SNPs inside these loci. By using HESS (Shi, et al., 2016) we performed a local SNP-based heritability analysis for all five traits for 1,694 loci. None of the three initial traits of interest (ADHD, AGG, ASB) showed a significant locus ($p < 0.05/1{,}694$, **Suppl. Fig. 1**). Only PSC had two loci with high SNP h2 estimates that remained significant after correction for multiple testing (chr6:25684587-26791233, p = 1.0025E-11; chr6:26791233-28017819) (**Suppl. Fig. 1e**). ADHD showed two nominally significant loci on chromosomes 1 and 5 (**Suppl. Fig. 1a**), while neuroticism obtained the most results with a total of eight nominally significant loci on chromosomes 2, 6, 8, 11 and 18), and PSC showed two nominally significant loci on chromosome 3 (**Suppl. Fig. 1d,e**).

Local genetic covariance and correlation

Instead of considering only a genome-wide score as the absolute value for how two traits are genetically related, we examined how this global estimate is actually distributed across the genome. We used ρ-HESS to analyze the same previously defined eight trait pairs and estimated the local genetic covariance for 1,694 loci. After obtaining the genetic correlation estimates for each locus, we selected all loci with correlation estimates inside the expected range [-1,1] for subsequent analyses.

For the ADHD-AGG & ADHD-ASB cross-trait analyses we observed varying local genetic covariance and correlation estimates throughout the genome (**Suppl. Fig. 2a,b**), including both positively and negatively correlated loci with some having higher genetic correlation scores than others, contrary to the usually expected equal contribution from all loci. Similar patterns

of local correlations are observed for the other six rest cross-trait analyses (**Suppl. Fig. 2**). We did not observe any (nominally) significant local genetic covariances for the initial two cross-trait analyses (ADHD-AGG & ADHD-ASB). For the cross-trait analyses of ADHD + Neuroticism and ADHD + PSC, we found two and one loci, respectively, passing the threshold for nominal significance (**Suppl. Fig 2c,f**).

Candidate loci overlapping between traits

Next, we reduced the number of loci by comparing them between the different cross-trait analyses (**Fig. 3**). The eight cross-trait analyses were first clustered into three groups according to the key traits involved, (see methods section). We selected those loci, which were shared between the different cross-trait analyses within a specific group.

For Group1, we selected 31 loci, for Group2 we prioritized 28 and for Group3 with only identified one locus (**Fig. 3, Table 2**). The global genetic correlation analysis for Group1 and Group2 revealed a positive genetic correlation. Interestingly, for the individual loci this pattern can be different, e.g. in Group1, there are regions where both associations (ADHD-AGG & ADHD-ASB) have a local positive genetic correlation (e.g. chr4:6,773,043-7,539,692), while other loci showed either a positive or negative correlation (e.g. chr1:196,176,201-197,311,514) and again others showed overall negative genetic correlations (e.g. chr12:124,977,980-126,445,505) (**Table 2**). In contrast to the global (negative) genetic correlation between traits in Group 3, the only overlapping locus, had a positive local genetic correlation in all three cross-trait analyses involved. It should also be noted that even when loci between cross-trait analyses within a group, had the same genetic correlation direction, some loci had a higher estimate with one association, while for the other, the estimate was lower (e.g. Group 1, chr3:65,273,538-66,270,447) (**Table 2**).

Genes from candidate loci

After selecting the overlapping loci, we annotated those to genes that (1) were located inside these selected loci (using the UCSC genome browser), and (2) showed a gene-based association with the traits in question (using gene-based cross-trait [meta-] analysis). This process is illustrated in **Fig. 4** and from both methods we derived a list of genes for subsequent analyses for each of the three groups.

A total of 34 genes were selected for Group1, 39 for Group2 and only two for Group3 (**Fig. 4C, Table 3**). All of these genes showed nominally significant gene-based association and nine genes were significant after correction for multiple testing in Group1, 21 for Group2 and one for Group3 (**Table 3**). However, some of these significant genes were located within the same loci. Finally, we explored if there were genes shared between the three groups and identified

three genes overlapping between Group1 and Group2: *F13B*, *GATA4* and *HMGCS1*, although only *GATA4* showed strongest gene-based association for the meta-analyses of Group2.

Biological annotation and functional significance of candidate genes

Next, we developed a pipeline to investigate the biological characteristics and to prioritize brain-relevant genes out of all overlapping genes identified in the previous step. For this, we retrieved information from three different databases/datasets for each gene from the three cross-trait analyses groups: (1) assessment whether genes are expressed in relevant brain regions, (2) if they were differentially expressed or co-expressed in a specific region and period of development, and (3) if they showed significant gene-based association with relevant subcortical brain regions.

With the exception of genes *F13B*, *DEFB136* and *GIP* all genes were expressed in all 12 brain-related regions (**Suppl. Table 2**). Several genes were highly expressed in a number of brain tissues:*TMEM132B* (Group 1), *IGLON5* (Group 2) and *RCAN3* (Group 3) (**Suppl. Table 2**). When analyzing co-expression patterns, we found that only *SHOX2* and *NXPH4* (Group 2) were enriched in the mediodorsal nucleus of the thalamus and cerebellar cortex respectively (**Suppl. Table 3**). Looking at relevant modules of genes temporally and spatially co-expressed, the most commonly found modules in all three groups were M2 and M20 (**Suppl. Table 3**). Also, the module M1 for *NOX4* and M9 for *SHOX2* were found, and *SHOX2* was found to be a hub gene (genes with high degree of connectivity within the module). Two genes in Group 1 were significantly associated with the putamen, *KCNT2* (p=0.00012407) and *TMEM132B* (p=0.0001765, **Suppl. Table 4**). In Group 2 and 3, none of the genes showed significant gene-based association with one of the brain volumes. Across all eight brain structures *GATA4* showed most nominally significant results. Multiple genes showed nominally significant gene-based association in more than one brain region and overall most nominally significant gene-based associations were fund for the accumbens (**Suppl. Table 4**).

In order to integrate the previous findings and to converge the evidence for the most promising candidate genes, we applied a scoring system and prioritized genes with the most relevant and consistent results across studies. Using a cut-off of four points, we selected a total of eight genes for Group 1: *PYGO2*, *ZBTB41*, *ZNF131*, *NKAIN2*, *BLOC1S2*, *ATRNL1*, *KCNK18* and *TMEM132B*. For Group 2 we included 12 genes: *DAB1*, *RSRC1*, *MSRA*, *FAM167A*, *GATA4*, *NEIL2*, *CTSB*, *DEFB134*, *KIAA1456*, *NXPH4* and *IGLON5*. One gene was selected for Group 3: *RCAN3* (**Suppl. Table 5**).

Biological annotations were retrieved for all 21 candidate genes. **Suppl. Table 5** provides an overview on their biological processes or pathways, and importance for diseases. Many of the selected genes have regulatory functions, although in different processes and pathways. Some of the prioritized genes have functions related to neurobiology and have been associated with mental disorders (**Suppl. Table 5**).

Discussion

In this study we applied a locus-based approach to characterize and localized shared genetic effects between three neurodevelopmental and behavioral traits: ADHD, AGG and ASB. We identified 53 shared loci across the genome, addressing the global positive genetic relationship between our three traits and showed how some genetic elements are actually contributing to this relationship. Moreover, we also identified shared loci and candidate genes relevant for pleiotropic effects on ADHD, AGG, ASB, neuroticism and PSC.

In the genome-wide approach, we observed a low and non-significant genetic correlation between AGG and ASB, which was unexpected. However, this may be related to the different ages of the participants. It has been known that heritability is susceptible to age [Wray & Visscher], and while the ASB data included different age groups, AGG only included childhood samples [Tielbeek, et al. 2017][Pappa, et al., 2016]. Alternatively, the different types of measurements may affect the results. Whereas the AGG data was obtained from reports filled out by mothers, possibly biasing diagnosed participants, data for ASB, in an effort to increment its sample size, included different measurements to allow the inclusion of broad ASB subtypes [Pappa et al., 2016][Tielbeek, et al., 2017]. Nonetheless, we included both studies and assessed the genetic relationship between them further by reaching out to additional shared genetic traits. Given this pleiotropy, as seen in the three groups of cross-trait analyses, we used ADHD, Neuroticism and PSC as hubs in these groups. Neuroticism is a fundamental domain of personality, characterized by people who are easy to respond negatively (with anxiety, sadness, guilt, fear and anger) to environmental stress and associated with other mood, psychopathologic and neurodevelopmental disorders [Griffith, et al., 2010][Okbay et al., 2016]. It was the most phenotypically related trait we identified. PSC on the other hand, is a chronic autoimmune liver disease leading to destruction of the bile duct and ending principally with cirrhosis and liver failure [Bowlus, et al., 2017][Fricker & Lichtenstein 2019]. Although it is not a neurodevelopmental or a behavioral trait, due to its strong negative genetic correlation estimates with ADHD, AGG and ASB, we saw it as an opportunity to explore the genetic complexity further. Besides that, there exist some antecedents between gastrointestinal or

autoimmune diseases with psychological and neurodevelopmental disorders. [Shah, et al., 2014][Jeppesen & Benros, 2019][Nielsen, et al., 2017].

From the local-wide perspective, Shi and colleagues made the distinction that a genome-wide genetic covariance estimate does not give enough information into the genetic relationship between traits because it does not reflect the actual distribution and quantity effect sizes have across regions. For example, the genome-wide estimate could be the sum of many, but small same directional covariances or instead few loci with a bigger estimate [Shi et al., 2017]. This heterogeneity on loci was seen when comparing across loci results in a single cross-trait analysis. When even a global genetic correlation estimate dictated an unique directionality, not all loci represented the same direction. Adding these differences between loci, it refutes the idea that all genetic components across the genome are equally contributing to the trait and consequently to its correlation against other traits. These differences in directionality and magnitude for loci are also dependent for the traits studied here. When comparing a single locus within two cross-trait analyses, even if they shared a same genome-wide genetic correlation directionality and a common assessed trait, local genetic correlations could be highly different. Unfortunately, we could only report local genetic correlation scores on 53 loci instead of all 1,694, since these 53 were the ones with heritability and covariance estimates that allowed a correlation score inside of what is the expected range and were also present between cross-traits analyses from one of the three pleiotropic groups (Groups 1-3). A surprising result was PSC, since its global genetic correlation estimate was highly negative against ADHD, AGG and ASB, but the only one left locus marked positive local genetic correlation in all cross-trait analyses on that group. However, the few numbers of loci were expected on Group 3 since PSC was a non-conventional trait. Furthermore, another cause of lack of results could be attributed to the removed of MHC region in this study. MHC region it is known for its strongest genetic association with various autoimmune diseases, including PSC, and also showing genetic associations with ADHD too [Fernando et al., 2008][Bowlus, C., 2017][Nielsen et al., 2017]. Lack of resulting loci could also being explained by the relatively low sample sizes the AGG, ASB and PSC GWAS studies have. Contrarily, ADHD and Neuroticism, which were the biggest sample sizes GWAS, and the cross-trait analyses between them outputted the largest number of loci with genetic correlation estimates within the expected -1 to 1 range.

Instead of following the usual SNP-to-gene approach, we opted for a more robust approach knowing the caveats limited statistical power of some of our GWAS data, and looked for genes inside the selected candidate loci. This resulted in 72 different genes with a nominally associated p-value supporting their selection. Nevertheless, the results have to be interpreted

with caution, even if the loci are LD independent, it does not take away the possibility of retrieving false positive results at the gene level. Based on the association p-value with subcortical brain volume and gene expression in human brain tissue we selected a total of 21 genes (8 in Group 1, 11 in Group 2 and 2 in Group 3) with a biological basis in three brain-related aspects: expression, enrichment (spatially and/or temporally) and phenotypic effect (volume changes). We were able to obtain brain-related aspects that could also be useful for further studies in the relationship between these traits. The most implicated brain tissues were the cerebellar hemisphere, accumbens and putamen; three brain regions characterized by motor and movement functions and a possible behavioral relationship [Ivanov, I. et al., 2014][Floresco, Stan B., 2015][Korponay, C., et al., 2017]. Due to low number of genes, we were not able to apply a gene-enrichment analyses. However, when we retrieved additional information of the 21 candidate genes, we found that most of them are related with DNA or RNA binding, transcription regulation, signaling pathways and developmental processes. This could explain their enrichment within the prenatally and postnatally spatio-temporal modules most of the genes, periods of development sensitive to these processes. Some genes were more related with neurobiological processes, e.g., from Group 1, *ZNF131*, which is a zinc finger involved in transcription factor activities with a possible role in the developing central nervous system and organogenesis [Trappe R., et al., 2002] and *BLOC1S2*, a protein forming part of the BLOC-1 involved on the biogenesis of lysosomes. This complex has also been identified in the vesicles cargos in neurons [Chen X. et al., 2017][Di Pietro, S.M, et al., 2006]. From Group 2 we can highlight *DAB1* an adapter molecule involved in the correct laminar migration of neurons during development [Franco, S. et al., 2011] and *NXPH4*, which has a possible role in the neuropeptide signaling pathway and binding to alpha-neurexins [Missler, M & Sudhof, T., 1998]. The only gene in Group 3, *RCAN3* is a calcineurin regulator (calcium dependent protein involved in neuronal functions), but also it belongs to the group of Down Syndrome Critical Refion-1 proteins, suggesting a role in neurodevelopment and brain [Porta, S. et al., 2007].

These findings can be seen in light of strengths and limitations. Taking into account the differences between GWAS studies, some datasets were hampered by the relatively low sample sizes. Knowing that our traits are an accumulation of many small effect variants, is difficult to find these in small sample sizes [Visscher PM, 2012]. However, ADHD, one of our traits of interest with a reasonable GWAS sample size, was the trait showing the best and significant results throughout all this work, opening the possibility of obtaining better powered results as soon as larger sample sizes for AGG and ASB become available. Moreover, having significant genome-wide genetic correlations between most of our traits shows a highly probable existing genetic relationship between these and opens the doors to study their

relationship at locus level. Nevertheless, the analyses are still based on single SNP variants, which do not explain the complete heritability of our traits in questions. Future studies may want to also integrate information from other common genetic variants and rare variants.
Many of our analyses were motivated by a large sensitivity, meaning that we were not too strict in applying significance thresholds when selecting our candidate loci and genes. To limit the number of false positive findings, we decided to use complementary analyses strategies to filter out and prioritize the most relevant genes involved. For this, multiple lines of evidence were needed across the different cross-trait analyses, e.g., by using other association analyses (gene-based analysis) and looking for biological related presence or effect (brain expression data). Moreover, we only focused on protein-coding genes, because these are the most basic and widely studied elements in the genome. However, also other genetic elements may be included in the shared candidate loci and future studies may want to look into elements like non-coding genes, promoters or even incorporate epigenetic data, which could provide alternative explanations to the pleiotropic effects between our traits.

In sum, at a genome-wide perspective, we observed a positive and highly genetic correlation between ADHD with AGG and ASB and identified share local effects. This study emphasizes the importance of investigating shared genetic effects between correlated traits at the locus level. Converging evidence from different cross-trait analyses approaches highlights novel candidate genes underlying the shared biological mechanisms of ADHD, AGG and ASB.

References

Asherson, P. and H. Gurling (2012). "Quantitative and molecular genetics of ADHD." Curr Top Behav Neurosci **9**: 239-272.

Association, A. P. (2013). Diagnostic and Statistical Manual of Mental Disorders: DSM-v. Washington, DC, Author.

Barkley, R. A., et al. (2004). "Young adult follow-up of hyperactive children: antisocial activities and drug use." J Child Psychol Psychiatry **45**(2): 195-211.

Berisa, T. and J. K. Pickrell (2016). "Approximately independent linkage disequilibrium blocks in human populations." Bioinformatics (Oxford, England) **32**(2): 283-285.

Biederman, J., et al. (2000). "Age-dependent decline of symptoms of attention deficit hyperactivity disorder: impact of remission definition and symptom type." Am J Psychiatry **157**(5): 816-818.

Biederman, J., et al. (1991). "Comorbidity of attention deficit hyperactivity disorder with conduct, depressive, anxiety, and other disorders." Am J Psychiatry **148**(5): 564-577.

Boonstra, K., et al. (2012). "Epidemiology of primary sclerosing cholangitis and primary biliary cirrhosis: a systematic review." J Hepatol **56**(5): 1181-1188.

Bowlus, C. L. (2017). "Primary sclerosing cholangitis." Handbook of Liver Disease **24**(5): 233-242.

Bulik-Sullivan, B., et al. (2015). "An atlas of genetic correlations across human diseases and traits." Nat Genet **47**(11): 1236-1241.

Bulik-Sullivan, B. K., et al. (2015). "LD Score regression distinguishes confounding from polygenicity in genome-wide association studies." Nat Genet **47**(3): 291-295.

Burt, S. A. and J. M. Neiderhiser (2009). "Aggressive Versus Nonaggressive Antisocial Behavior: Distinctive Etiological Moderation by Age." Developmental Psychology **45**(4): 1164-1176.

Chen, X., et al. (2017). "The BLOC-1 Subunit Pallidin Facilitates Activity-Dependent Synaptic Vesicle Recycling." eNeuro **4**(1).

Christiansen, H., et al. (2008). "Co-transmission of conduct problems with attention-deficit/hyperactivity disorder: familial evidence for a distinct disorder." J Neural Transm (Vienna) **115**(2): 163-175.

Collins, A. L. and P. F. Sullivan (2013). "Genome-wide association studies in psychiatry: what have we learned?" Br J Psychiatry **202**(1): 1-4.

Consortium, C.-D. G. o. t. P. G. (2013). "Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis." Lancet **381**(9875): 1371-1379.

Consortium, G. T., et al. (2017). "Genetic effects on gene expression across human tissues." Nature **550**: 204.

Crick, N. R. and K. A. Dodge (1996). "Social information-processing mechanisms in reactive and proactive aggression." Child Dev **67**(3): 993-1002.

Davis, M. H. (2018). Aggression and Antisocial Behavior.

de Leeuw, C. A., et al. (2015). "MAGMA: generalized gene-set analysis of GWAS data." PLoS Comput Biol **11**(4): e1004219.

Demontis, D., et al. (2019). "Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder." Nat Genet **51**(1): 63-75.

Di Pietro, S. M., et al. (2006). "BLOC-1 interacts with BLOC-2 and the AP-3 complex to facilitate protein trafficking on endosomes." Mol Biol Cell **17**(9): 4027-4038.

Doyle, A. E. and S. V. Faraone (2002). "Familial links between attention deficit hyperactivity disorder, conduct disorder, and bipolar disorder." Curr Psychiatry Rep **4**(2): 146-152.

Faraone, S. V., et al. (2015). "Attention-deficit/hyperactivity disorder." Nat Rev Dis Primers **1**: 15020.

Faraone, S. V., et al. (2006). "The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies." Psychol Med **36**(2): 159-165.

Ferguson, C. J. (2010). "Genetic contributions to antisocial personality and behavior: a meta-analytic review from an evolutionary perspective." J Soc Psychol **150**(2): 160-180.

Fernando, M. M., et al. (2008). "Defining the role of the MHC in autoimmunity: a review and pooled analysis." PLoS Genet **4**(4): e1000024.

Floresco, S. B. (2015). "The nucleus accumbens: an interface between cognition, emotion, and action." Annu Rev Psychol **66**: 25-52.

Franco, S. J., et al. (2011). "Reelin regulates cadherin function via Dab1/Rap1 to control neuronal migration and lamination in the neocortex." Neuron **69**(3): 482-497.

Franke, B., et al. (2012). "The genetics of attention deficit/hyperactivity disorder in adults, a review." Mol Psychiatry **17**(10): 960-987.

Griffith, J. W., et al. (2010). "Neuroticism as a common dimension in the internalizing disorders." Psychol Med **40**(7): 1125-1136.

Hamshere, M. L., et al. (2013). "High loading of polygenic risk for ADHD in children with comorbid aggression." Am J Psychiatry **170**(8): 909-916.

Hibar, D. P., et al. (2015). "Common genetic variants influence human subcortical brain structures." Nature **520**: 224.

Hudziak, J. J., et al. (2003). "Individual differences in aggression: genetic analyses by age, gender, and informant in 3-, 7-, and 10-year-old Dutch twins." Behav Genet **33**(5): 575-589.

Ivanov, I., et al. (2014). "Cerebellar morphology and the effects of stimulant medications in youths with attention deficit-hyperactivity disorder." Neuropsychopharmacology **39**(3): 718-726.

Jeppesen, R. and M. E. Benros (2019). "Autoimmune Diseases and Psychotic Disorders." Front Psychiatry **10**: 131.

Ji, S. G., et al. (2017). "Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease." Nat Genet **49**(2): 269-273.

Kang, H. J., et al. (2011). "Spatio-temporal transcriptome of the human brain." Nature **478**: 483.

Karolchik, D., et al. (2004). "The UCSC Table Browser data retrieval tool." Nucleic Acids Res **32**(Database issue): D493-496.

Korponay, C., et al. (2017). "Impulsive-antisocial dimension of psychopathy linked to enlargement and abnormal functional connectivity of the striatum." Biol Psychiatry Cogn Neurosci Neuroimaging **2**(2): 149-157.

Kuja-Halkola, R., et al. (2015). "Codevelopment of ADHD and externalizing behavior from childhood to adulthood." J Child Psychol Psychiatry **56**(6): 640-647.

Lahey, B. B. and I. D. Waldman (2003). A Developmental Propensity Model of the Origins of Conduct Problems during Childhood and Adolescence. B. B. Lahey, T. E. Moffitt and A. Caspi. New York, Guilford Publications**:** 76-117.

Larsson, H., et al. (2014). "The heritability of clinically diagnosed attention deficit hyperactivity disorder across the lifespan." Psychol Med **44**(10): 2223-2229.

Latalova, K. (2009). "Bipolar disorder and aggression." Int J Clin Pract **63**(6): 889-899.

Lee, S. H., et al. (2013). "Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs." Nat Genet **45**(9): 984-994.

Liu, J. (2004). "Concept analysis: aggression." Issues Ment Health Nurs **25**(7): 693-714.

Mandy, W., et al. (2014). "Three dimensions of oppositionality in autism spectrum disorder." J Abnorm Child Psychol **42**(2): 291-300.

Miles, D. R. and G. Carey (1997). "Genetic and environmental architecture of human aggression." J Pers Soc Psychol **72**(1): 207-217.

Missler, M. and T. C. Sudhof (1998). "Neurexophilins form a conserved family of neuropeptide-like glycoproteins." J Neurosci **18**(10): 3630-3638.

Nielsen, P. R., et al. (2017). "Associations Between Autoimmune Diseases and Attention-Deficit/Hyperactivity Disorder: A Nationwide Study." J Am Acad Child Adolesc Psychiatry **56**(3): 234-240.e231.

Okbay, A., et al. (2016). "Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses." Nat Genet **48**(6): 624-633.

Pappa, I., et al. (2016). "A genome-wide approach to children's aggressive behavior: The EAGLE consortium." Am J Med Genet B Neuropsychiatr Genet **171**(5): 562-572.

Philip David, Z., et al. Aggression and Antisocial BehaviorA Developmental Perspective, Oxford University Press.

Polanczyk, G., et al. (2007). "The worldwide prevalence of ADHD: a systematic review and metaregression analysis." Am J Psychiatry **164**(6): 942-948.

Porta, S., et al. (2007). "Differential expression of members of the RCAN family of calcineurin regulators suggests selective functions for these proteins in the brain." Eur J Neurosci **26**(5): 1213-1226.

Retz, W. and M. Rosler (2009). "The relation of ADHD and violent aggression: What can we learn from epidemiological and genetic studies?" Int J Law Psychiatry **32**(4): 235-243.

Rhee, S. H. and I. D. Waldman (2002). "Genetic and environmental influences on antisocial behavior: a meta-analysis of twin and adoption studies." Psychol Bull **128**(3): 490-529.

Shah, E., et al. (2014). "Psychological disorders in gastrointestinal disease: epiphenomenon, cause or consequence?" Ann Gastroenterol **27**(3): 224-230.

Shi, H., et al. (2016). "Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data." Am J Hum Genet **99**(1): 139-153.

Shi, H., et al. (2017). "Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits." Am J Hum Genet **101**(5): 737-751.

Simon, V., et al. (2009). "Prevalence and correlates of adult attention-deficit hyperactivity disorder: meta-analysis." Br J Psychiatry **194**(3): 204-211.

Solomon, U., et al. (2013). "Aggressive behavior: A comprehensive review of its Neurochemical mechanisms and management." Agression and Violent Behavior **18**: 195-203.

Solovieff, N., et al. (2013). "Pleiotropy in complex traits: challenges and strategies." Nature Reviews Genetics **14**: 483.

Stelzer, G., et al. (2016). "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses." Curr Protoc Bioinformatics **54**: 1.30.31-31.30.33.

Tielbeek, J. J., et al. (2017). "Genome-Wide Association Studies of a Broad Spectrum of Antisocial Behavior." JAMA Psychiatry **74**(12): 1242-1250.

Weinshenker, N. J. and A. Siegel (2002). "Bimodal classification of aggression: Affective defense and predatory attack." Aggression and Violent Behavior **7**(3): 237-250.

Zheng, J., 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 (Oxford, England) **33**(2): 272-279.

Tables

Table 1 | Traits which are genetically correlated with both AGG and ASB.

Positive correlations		
LD-HUB trait	**AGG rg (p)**	**ASB rg (p)**
Neuroticism	**0.3595 (0.017)**	**0.2703 (0.0197)**
Waist to hip ratio	0.2413 (0.0496)	0.2569 (0.0474)
Number of children ever born	0.3566 (0.0396)	0.4177 (0.0261)
Negative correlations		
LD-HUB trait	**AGG rg (p)**	**ASB rg (p)**
Primary Sclerosing Cholangitis	**-0.4076 (0.0415)**	**-0.3997 (0.0432)**
Age of first birth	-0.4998 (0.0041)	-0.4305 (0.008)
College completion	-0.513 (0.0087)	-0.5148 (0.0171)
Intelligence	-0.4831 (0.0039)	-0.439 (0.0032)
Years of schooling (proxy cognitive performance)	-0.5252 (0.0189)	-0.4342 (0.0243)
Years of schooling 2013	-0.5817 (0.0077)	-0.4888 (0.0155)

Final list of traits from LDHub database, that are shared between AGG and ASB. Results were obtained by cross-trait genetic correlation analyses using LDSC automated by LDHub. Results were filtered by p value < 0.05. Highlighted traits, Neuroticism and Primary Sclerosing Cholangitis, were selected traits of interest to be further investigated in subsequent analyses.

Table 2 | Final list of all candidate loci with genetic correlation estimates between range [-1,1] and shared between groups of cross-trait analyses.

Overlapping loci between cross-traits analyses				
Group1				
chr	start	end	local rg ADHD-AGG	local rg ADHD-ASB
1	11777841	12779466	-0.152803121	0.420775744
1	59890409	61922365	0.067999574	0.042971589
1	154770403	156336133	0.483837511	0.354359797
1	196176201	197311514	0.988262402	-0.960761239
2	110572432	113921856	-0.052472426	0.225361764
3	65273538	66270447	0.035095413	0.960313445
3	77508835	79024541	-0.25031436	0.422913061
4	6773043	7539692	0.706325908	0.59543495
4	8152235	9326479	-0.045066369	0.68953158
4	27343722	27965868	0.003321683	-0.157937228
4	97540703	99424067	-0.348038358	0.030567005
4	127275477	127830293	-0.465601294	0.738451107
5	11940	982252	0.000550138	-0.38702438
5	41888710	43983499	-0.335177034	0.870562973
5	125785649	127344604	0.373532123	-0.003067617
6	94118142	94441175	0.688925442	-0.730273061
6	123856181	125424383	-0.253986859	0.618657966
8	11278998	13491775	0.499317422	0.383989982
8	42773823	46841315	0.884556756	0.938747441
8	119685457	121201700	0.309364949	0.105498468
8	133351144	134570001	0.528742598	-0.230621225
10	61891409	62659679	-0.647348266	0.974585721
10	100668400	102949239	0.13818094	0.551075746
10	116421406	119523934	0.940949363	-0.547744192

11	89208936	90966490	-0.689860672	-0.647230002
12	124977980	126445505	-0.125739598	-0.533918054
13	104066710	104844114	-0.138924264	-0.862517486
14	28466533	29972145	-0.721808643	0.299104259
17	36809344	38877404	0.317404181	-0.035964118
18	26875587	27866478	0.003438467	0.615754626
22	37570269	39307894	0.295994237	0.026228558

Group2

Chr	start	end	local rg ADHD-Neurot	local rg AGG-Neurot	local rg ASB-Neurot
1	57021728	58865399	0.895655652	0.40940298	0.003654553
1	182755356	184595513	0.147569851	-0.605614459	-0.13088678
1	189904130	191868930	-0.79273896	0.573530675	0.454281586
1	196176201	197311514	0.421699254	0.483415541	0.154436476
2	151731462	152486782	0.77577421	-0.455132944	-0.517349667
3	38356116	40221298	-0.871309416	0.012233931	0.676884355
3	77508835	79024541	-0.578917764	0.719323976	0.469687819
3	156008700	157312028	-0.447881897	0.773766535	-0.630568367
3	157312028	159477890	-0.268545081	0.771133872	-0.107704541
5	41888710	43983499	0.06848284	0.739404581	0.060205319
5	50163398	52134622	0.224628202	-0.206489898	-0.969030438
5	163771582	165642395	-0.488401703	-0.204683006	-0.24773838
8	9640787	10463197	0.169699	-0.295003919	-0.659509979
8	11278998	13491775	-0.775999929	0.125179505	-0.076076701
8	42773823	46841315	-0.914451824	-0.221785326	0.720648741
9	121321537	122260297	-0.123595254	0.327942747	0.569600028
11	89208936	90966490	0.435600808	-0.765910878	0.713596841
12	57548860	59308666	-0.333055241	-0.359004676	0.224973577
12	106436213	106958748	-0.6118229	0.561601355	0.924332907
13	54682864	55817131	-0.74901171	-0.469713568	0.187283437

13	84133144	85610916	0.427620302	-0.667910412	0.330227632
14	19002084	21589402	0.217441812	0.211845643	0.639189592
14	28466533	29972145	0.51477033	-0.007855277	0.126456928
14	50507760	51493572	0.142178135	0.51334262	0.790283473
17	45876022	47517400	-0.407063792	-0.509394249	0.796810598
19	9238393	11284028	0.166645818	-0.546635249	-0.325070885
19	44744108	46102697	0.296934793	0.850856823	0.070800353
19	51532567	52985290	-0.876976131	-0.353721632	-0.293779239
Group3					
Chr	**start**	**end**	**local rg ADHD-PSC**	**local rg AGG-PSC**	**local rg ASB-PSC**
1	23920590	25516845	0.351451182	0.086066883	0.864071011

Table is divided by groups (Group1-3). First three columns describe loci position (chr, start, end), subsequent columns contain the local genetic correlation estimate (rg) from each of the cross-trait analyses belonging to the subtitled group.

Table 3 | A list of all genes obtained by mapping and gene-based association analyses.

Results of gene-based meta-analysis for all three groups			
Group1			
Loci	**Genes**	**ADHD-AGG gene-based p-value**	**ADHD-ASB gene-based p-value**
chr1:154770403-156336133	*PBXIP1*	0.0033044	0.017379
	PYGO2	0.00062789*	0.00036099*
chr1:196176201-197311514	*KCNT2*	0.036276	0.021367
	CFHR1	0.037945	0.042817
	F13B	0.033877	0.02492
	ZBTB41	0.0074475	0.00098922*
chr2:110572432-113921856	*ZC3H8*	0.014159	0.026319
	ZC3H6	0.012073	0.0045015
	TTL	0.024659	0.0022036

chr4:6773043-7539692	*TBC1D14*	0.0022914		0.00033126*
chr5:41888710-43983499	*C5orf51*	0.00035659*		0.021966
	FBXO4	0.00032249*		0.0096655
	GHR	0.0077015		0.0028768
	CCDC152	0.0014035*		0.0034075
	SEPP1	0.0008789*		0.011502
	ZNF131	0.017818		0.04255
	HMGCS1	0.031783		0.0090487
chr6:123856181-125424383	*NKAIN2*	0.012034		0.0014679*
chr8:11278998-13491775	*GATA4*	0.0031767		0.0091211
chr8:133351144-134570001	*TMEM71*	0.020976		0.044428
	TG	0.0033284		0.011278
	SLA	0.034599		0.047501
chr10:100668400-102949239	*CHUK*	0.039743		0.043994
	CWF19L1	0.012558		0.023098
	BLOC1S2	0.010662		0.045214
chr10:116421406-119523934	*ATRNL1*	0.0000223*		0.00018444*
	KCNK18	0.037176		0.005236
chr12:124977980-126445505	*TMEM132B*	0.0018939		0.0025947
chr17:36809344-38877404	*MLLT6*	0.030755		0.041041
	PCGF2	0.017368		0.0030954
chr22:37570269-39307894	*ANKRD54*	0.025402		0.047839
	TMEM184B	0.0044383		0.029471
	GTPBP1	0.042065		0.020954
	SUN2	0.026453		0.011303
Group2				
Loci	**Genes**	**ADHD-Neurot**	**AGG-Neurot**	**ASB-Neurot**

		gene-based p-value	gene-based p-value	gene-based p-value
chr1:57021728-58865399	*DAB1*	0.029433	0.038587	0.018406
chr1:196176201-197311514	*F13B*	0.040111	0.016431	0.013425
chr2:151731462-152486782	*RBM43*	0.010506	0.0021002	0.00056273*
	NMI	0.010662	0.0061572	0.00044518*
chr3:157312028-159477890	*SHOX2*	0.0092963	0.021085	0.021579
	RSRC1	0.00088814*	0.0038031	0.020339
chr5:41888710-43983499	*HMGCS1*	0.015796	0.024921	0.011553
	CCL28	0.028373	0.031887	0.021219
chr8:9640787-10463197	*MSRA*	1.13E-17*	7.54E-19*	3.22E-18*
	PRSS55	0.038606	0.0017147	0.00027131*
chr8:11278998-13491775	*FAM167A*	3.4E-11*	1.04E-11*	3.58E-12*
	BLK	1.52E-10*	2.17E-11*	4.30E-11*
	GATA4	2.33E-11*	4.76E-10*	1.85E-09*
	NEIL2	1.78E-06*	7.48E-06*	4.68E-06*
	FDFT1	2.12E-08*	6.84E-08*	2.38E-08*
	CTSB	6.39E-05*	4.43E-07*	5.57E-08*
	DEFB136	2.58E-07*	3.99E-07*	3.79E-06*
	DEFB134	6.56E-07*	4.29E-07*	1.25E-06*
	USP17L7	1.74E-05*	1.74E-05*	1.74E-05*
	KIAA1456	0.022893	0.010996	0.01714
chr11:89208936-90966490	*NOX4*	0.042464	0.035669	0.047489
chr12:57548860-59308666	*LRP1*	6.38E-05*	0.0029426	0.0013207
	NXPH4	1.45E-06*	2.48E-05*	9.56E-05*
	STAC3	0.0013196	0.0015251	0.0021667
	R3HDM2	0.009344	0.040141	0.018112
	INHBC	0.0062844	0.032011	0.018054

	INHBE	0.00015136*	0.0020457	0.0014326
chr14:19002084-21589402	*ARHGEF40*	0.0099376	0.00092659*	0.0049658
chr17:45876022-47517400	*ATP5G1*	0.00079013*	0.032271	0.027971
	UBE2Z	0.0041105	0.021864	0.012026
	SNF8	0.0018604	0.022341	0.019265
	GIP	0.00098551*	0.0046015	0.0053558
chr19:9238393-11284028	*C19orf66*	0.022639	0.0086664	0.010897
chr19:44744108-46102697	*BCL3*	0.021743	0.010464	0.039127
	KLC3	0.0023481	0.017855	0.0023252
	ERCC2	0.0061647	0.0046247	0.0036333
chr19:51532567-52985290	*IGLON5*	0.021855	0.0019036	0.0013062
	VSIG10L	0.007612	0.0027017	0.0065383
	ETFB	0.0059393	0.0010191*	0.00070212*
Group3				
Loci	**Genes**	**ADHD-PSC gene-based p-value**	**AGG-PSC gene-based p-value**	**ASB-PSC gene-based p-value**
chr1:23920590-25516845	*SRSF10*	0.03326	0.03326	0.025284
	RCAN3	0.0066229*	0.0032891*	0.048198

Table is divided by groups (Group1-3) and genes are grouped if they belong to the same loci. First column describes locus position where the gene was found (chr:start-end), second column contains the gene symbol and subsequent columns contain the gene-based association p-values from the cross-trait gene-based meta-analyses run with MAGMA. P-values with an asterisk indicate a Bonferroni corrected significant result ($p < 0.05$ / no. of genes from subtitled group).

Figures

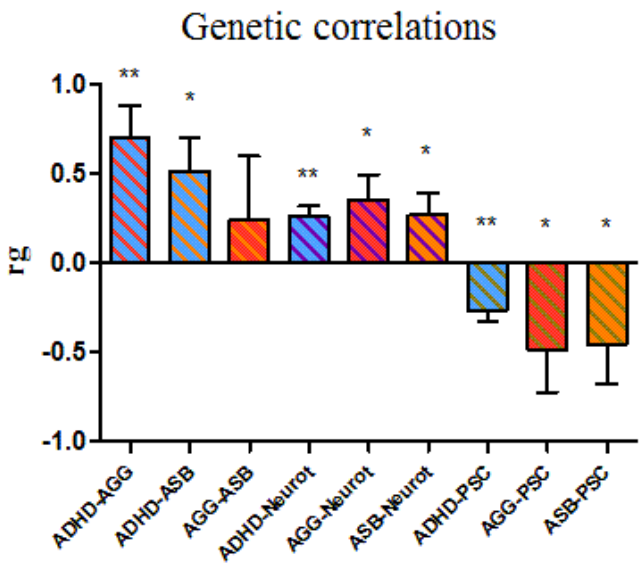

Figure 1 | Genome-Wide Genetic Correlations. All nine cross-trait genetic correlation analyses, using LDSC (https://github.com/bulik/ldsc) for each combination of traits.

***** $p < 0.05$

****** $p < 0.05/9$ (Bonferroni corrected p-value significance threshold)

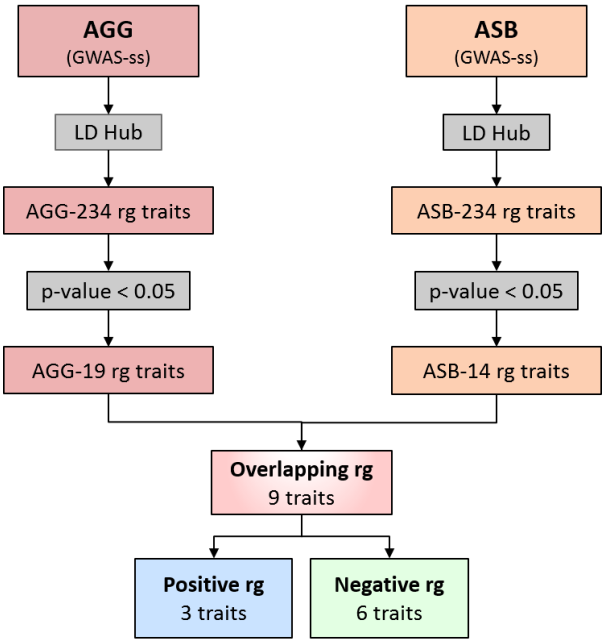

Figure 2 | Schematic overview of approach taken in this study to identify traits/disorders that are genetically correlated with AGG and/or ASB. The individual steps were: Use LD Hub (http://ldsc.broadinstitute.org/ldhub/) LDSC web interface, to ran two groups of cross-trait

genetic correlation analysis, one with AGG and the other with ASB, against 234 traits from LDHub database. Select correlated traits with p values < 0.05 in both groups of analyses. Look for overlapping traits. Finally, separate traits into those that have a positive genetic correlation from those with a negative genetic correlation.

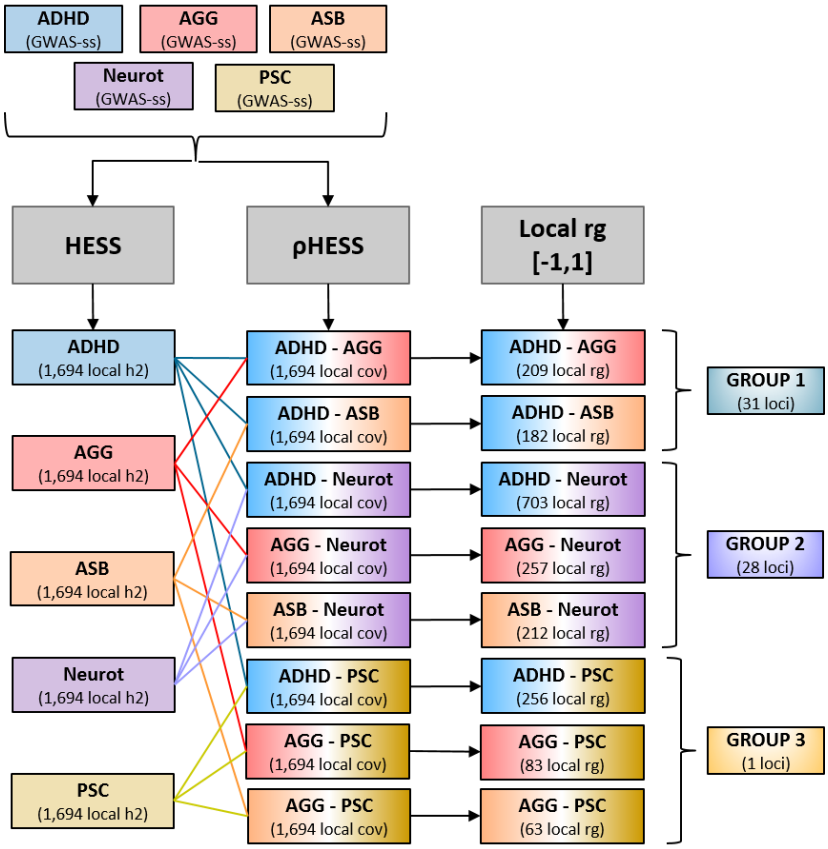

Figure 3 | Schematic overview of approach taken in this study to identify candidate loci overlapping between traits of interest. First, local SNP-based heritability estimates were obtained by analyzing GWAS-ss with HESS (https://github.com/huwenboshi/hess). Second, local genetic covariance estimates were obtained by using pHESS (https://github.com/huwenboshi/hess). Third, we estimated local genetic correlation for each of the 1,694 loci across the eight cross-trait analyses using the local SNP-h2 and local covariance results, but only kept loci with genetic correlation values inside expected range [-1,1]. Finally, we selected overlapping loci between cross-traits analyses and clustered results in three groups.

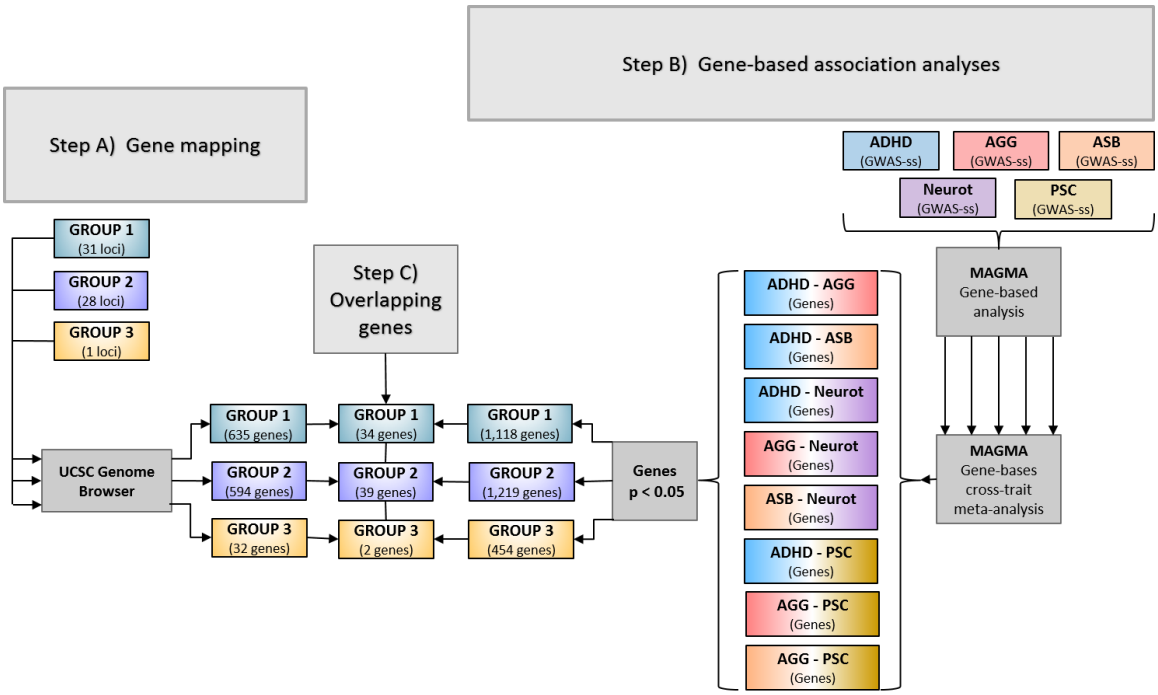

Figure 4 | Schematic overview of approach taken in this study to identify genes overlapping between traits of interest. Step A) Genes were mapped to selected loci from each one of the three groups of the cross-trait analyses (Group 1-3) by uploading loci positions to UCSC genome browser (https://genome.ucsc.edu). Step B) We performed gene-based association analyses by using MAGMA (https://ctg.cncr.nl/software/magma) followed by eight gene-based meta-analyses. All genes with gene-based p-value < 0.05 were selected and clustered according to the previously defined Groups (1-3) of the cross-trait analyses. Step C) We assessed the overlapping genes between the two methods for all three groups.