

Identification of shared and differentiating genetic risk for autism spectrum disorder, attention deficit hyperactivity disorder and case subgroups

*Manuel Mattheisen^{1,2,3}, *Jakob Grove^{1,4,5,6}, *Thomas D Als^{1,4,5}, *Joanna Martin⁷, Georgios Voloudakis^{8,9,10}, Sandra Meier^{1,2}, Ditte Demontis^{1,4,5}, Jaroslav Bendl^{8,9}, Raymond Walters^{11,12}, Caitlin E Carey^{11,12}, Anders Rosengren^{4,13}, Nora Strom^{1,3,14}, Mads Engel Hauberg^{8,9,10}, Biao Zeng¹⁵, Gabriel Hoffman^{8,9}, Jonas Bybjerg-Grauholm^{4,16}, Marie Bækvad-Hansen^{4,16}, Esben Agerbo^{4,17,18}, Bru Cormand^{19,20,21,22}, Merete Nordentoft^{4,23,24}, Thomas Werge^{4,13,23,25}, Ole Mors^{4,26}, David M Hougaard^{4,16}, Joseph D Buxbaum^{10,27,28,29}, Stephen V Faraone^{30,31}, Barbara Franke^{32,33}, Søren Dalsgaard¹⁷, Preben B Mortensen^{4,5,17,18}, Elise B Robinson^{11,12,34}, Panos Roussos^{8,9,10,29,35}, Benjamin M Neale^{11,12}, Mark J Daly^{11,12,36,37}, Anders D Børghlum^{1,4,5}

*contributed equally

Affiliations:

¹Department of Biomedicine - Human Genetics and the iSEQ Center, Aarhus University, Aarhus, Denmark

²Department of Psychiatry and Department of Community Health and Epidemiology, Dalhousie University, Halifax, NS, Canada

³Institute of Psychiatric Phenomics and Genomics (IPPG), University Hospital, LMU Munich, Munich, Germany.

⁴The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark

⁵Center for Genomics and Personalized Medicine, Aarhus, Denmark

⁶Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark

⁷MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK

⁸Pamela Sklar Division of Psychiatric Genomics, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

⁹Icahn Institute for Data Science and Genomic Technology, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

¹⁰Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

¹¹Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts

¹²Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts.

¹³Institute of Biological Psychiatry, Mental Health Services Copenhagen, Copenhagen University Hospital

¹⁴Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm Sweden

¹⁵Department of Genetics and Genomic Science and Institute for Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

¹⁶Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark

¹⁷National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark

¹⁸Centre for Integrated Register-based Research, Aarhus University, Aarhus, Denmark

¹⁹Department of Genetics, Microbiology and Statistics, Faculty of Biology, University of Barcelona, Catalonia, Spain.

²⁰Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, Madrid, Spain.

²¹Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Catalonia, Spain.

²²Institut de Recerca Sant Joan de Déu (IR-SJD), Esplugues de Llobregat, Catalonia, Spain.

²³Department of Clinical Medicine, Faculty of Health Science, University of Copenhagen, Denmark

²⁴Copenhagen Research Centre for Mental Health (CORE), Mental Health Centre Copenhagen, Copenhagen, University Hospital, Hellerup, Denmark.

²⁵GLOBE Institute, Center for GeoGenetics, University of Copenhagen, Copenhagen, Denmark

²⁶Psychosis Research Unit, Aarhus University Hospital-Psychiatry, Denmark

²⁷Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

²⁸Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

²⁹Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

³⁰Department of Psychiatry, State University of New York Upstate Medical University, Syracuse, NY, USA.

³¹Department of Neuroscience and Physiology, State University of New York Upstate Medical University, Syracuse, NY, USA.

³²Department of Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands.

³³Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands.

³⁴Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA.

³⁵Psychiatry, JJ Peters VA Medical Center, Bronx, New York, USA.

³⁶Department of Medicine, Harvard Medical School, Boston, MA, USA.

³⁷Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.

Please direct correspondence to:

Manuel Mattheisen: manuel.mattheisen@dal.ca

Anders D. Børglum: anders@biomed.au.dk

Abstract

Attention deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) are highly heritable neurodevelopmental disorders with a considerable overlap in their genetic etiology. We dissected their shared and distinct genetic architecture by cross-disorder analyses of large data sets, including samples with information on comorbid diagnoses. We identified seven loci shared by the disorders and the first five genome-wide significant loci differentiating the disorders. All five differentiating loci showed opposite allelic directions in the two disorders separately as well as significant associations with variation in other traits e.g. educational attainment, items of neuroticism and regional brain volume. Integration with brain transcriptome data identified and prioritized several significantly associated genes. Genetic correlation of the shared liability across ASD-ADHD was strong for other psychiatric phenotypes while the ASD-ADHD differentiating liability correlated most strongly with cognitive traits. Polygenic score analyses revealed that individuals diagnosed with both ASD and ADHD are double-burdened with genetic risk for both disorders and show distinctive patterns of genetic association with other traits when compared to the ASD-only and ADHD-only subgroups. The results provide novel insights into the biological foundation for developing just one or both of the disorders and for driving the psychopathology discriminatively towards either ADHD or ASD.

Introduction

Attention deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) are among the most common neurodevelopmental disorders in children and often persist throughout adulthood¹. ADHD and ASD are both highly heritable (60-93%)²⁻⁴ and the mode of their inheritance is complex and polygenic. Despite the high family-based heritability estimates, genome-wide association studies (GWAS) have only recently identified common variants robustly associated with each disorder⁵⁻⁷. Although differing from one another with regard to core clinical symptoms, genetic studies have demonstrated significant overlap between the two disorders, with a genetic correlation (r_G) from common variation of 0.36^{5,8} and substantial sharing of rare genetic risk variants such as large copy number variants⁹ and protein-truncating variants¹⁰. These findings are consistent with clinical and epidemiological evidence showing overlap in phenotypic features¹¹, high comorbidity rates between ASD and ADHD^{12,13} in both females and males¹⁴, and familial co-aggregation of the disorders with increased risk of ADHD among relatives of ASD probands (odds ratios monozygotic twins: 17.8, dizygotic twins: 4.3; full-siblings: 4.6, full cousins: 1.6)¹⁵. Identification of the genetic components that are shared or distinct for the disorders may provide insights into the underlying biology and potentially inform on sub-classification, course and treatment.

Here, we utilize large collections of genotyped samples of ADHD and ASD from the Psychiatric Genomics Consortium (PGC) and the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH) to address two questions: 1) What specific variants and genes are shared by, or differentiate, ASD and ADHD? 2) Are there distinct genetic signatures in terms of polygenic burden for subgroups within these disorders such as cases diagnosed with both disorders (comorbid cases) or with just one of them (ASD-only, ADHD-only cases)?

Results

Shared genetic liability to ADHD and ASD

We performed a GWAS of diagnosed ADHD and/or ASD combined into a single phenotype ("combined GWAS"), totaling 34,462 cases and 41,201 controls on 8.9 million SNP allele dosages imputed from 1000 genomes phase 3¹⁶. Using LD score regression (LDSC)¹⁷ we found evidence for a strong polygenic signal for this GWAS with an intercept of 1.0134 (ratio = 0.0558) and calculated the liability scale SNP-heritability to be 0.128 (for an assumed population prevalence of 0.055). We identified 263 genome-wide significant SNPs in seven distinct loci (**Table 1, Figure 1, Supplemental Figure S1**). All these loci showed associations with both of the disorders separately at p-values below 1×10^{-4} except one, which is genome-wide significant in ADHD and has a p-value of 0.009 in ASD. Overall, the findings corroborate previous results^{8,18}, but two loci have not been identified before as shared between ADHD and ASD. The novel shared associations are located in a highly pleiotropic multigene locus on chromosome 1 (rs7538463) and on chromosome 4 (rs227293) in *MANBA* (which encodes beta-D-mannoside mannohydrolase). Mutations in *MANBA* are associated with beta-mannosidosis, a lysosomal storage disease that has a wide spectrum of neurological phenotypes, including intellectual disability, hearing loss and speech impairment¹⁹. More details on the seven loci can be found in **Table 1** and results from lookups in the open GWAS project database (<https://gwas.mrcieu.ac.uk/about/>, accessed Oct 14th 2020) are available in **Supplemental Table S1** and as PheWAS plots in **Supplemental Figure S2**.

To identify and prioritize putative causal shared genes we performed a transcriptome-wide association study (TWAS), imputing the genetically regulated gene expression using EpiXcan²⁰ and expression data from the PsychENCODE Consortium²¹ for genes as well as isoforms detected in 924 samples from the dorsolateral prefrontal cortex (DLPFC). Applying a conservative significance threshold ($p < 1.44 \times 10^{-6}$; corresponding to Bonferroni correction of all the 34,646 genes and isoforms tested), we identified five genes/isoforms showing significant differential expression between the combined case group and controls,

and 177 genes/isoforms significant at a false discovery rate (FDR) < 0.05 (**Supplemental Table S2** and **Figure 1**). One of the five Bonferroni significant transcripts, the *KRT8P46-201* isoform, is located in the identified chromosome 4 GWAS locus in an intron of *MANBA* (which itself is among the genes with an FDR < 0.05) as illustrated in **Supplemental Figure S3a**. The four other top findings are the two genes *MOCS2* and *CCDC71* or their isoforms, which are not located in any of the identified GWAS loci and thus represent additional novel candidate genes for shared ADHD and ASD risk.

Gene-based analysis using MAGMA v 1.08²² with default settings as implemented in FUMA²³ largely corroborated the results from the GWAS and TWAS, highlighting e.g. *MANBA* (**Supplemental Figure S4a** and **Supplemental Table S3**). Furthermore, two of the significant genes (*SORCS3* and *DUSP6*) are located in regions that were not identified in the GWAS, suggesting these as additional shared loci.

Differentiating genetic liability to ADHD and ASD

To identify loci with divergent effects on ADHD and ASD, we performed an association analysis comparing 11,964 ADHD-only cases with 9,315 ASD-only cases from the iPSYCH cohort, excluding all 2,304 comorbid cases (“ADHD vs ASD GWAS”). Using LDSC¹⁷ we found an intercept of 0.9863 and a SNP-heritability of 0.4468 on the observed scale. Five genome-wide significant loci were identified, three of which have not previously been identified in GWAS of either of the two disorders separately, while all have been reported in related disorders and, remarkably, all but one are associated with cognitive abilities and/or neuroticism or neuroticism sub-items (**Table 2, Figure 1, Supplemental Table S5**). The lead variants all show opposite directions of effects in the two disorders.

Two of the five lead SNPs have previously been found associated with educational attainment²⁴. For the first SNP (rs3791033 on chromosome 1; $p = 4.65 \times 10^{-23}$) the C allele confers an increased risk for ASD and increased cognitive performance while the ADHD risk allele (T) is associated with decreased performance. Similarly, for the second SNP (rs9379833 on chromosome 6; $p = 2.26 \times 10^{-8}$) the A allele confers an increased risk for ASD and increased cognitive performance while the ADHD risk allele (C) is associated with

decreased performance. Notably, this SNP (rs9379833) is located in the large histone gene cluster HIST1²⁵ and has previously been associated with regional brain volume, specifically of the left globus pallidus²⁶ ($p = 2.95 \times 10^{-8}$; the C allele confers an increased risk for ADHD and a decreased volume while the ASD risk allele (A) is associated with an increased volume). Globus pallidus is part of the basal ganglia, which are involved in both motor and non-motor functions, including higher order cognition, social interactions, speech, repetitive behaviors and tics²⁷. It is also of note, that the lead SNP on chromosome 8 (rs7821914) is associated with neuroticism²⁸ ($p = 9.46 \times 10^{-21}$). For this SNP, the effect allele (C) in the neuroticism GWAS leads to an increased risk for ASD and a decreased risk for ADHD. An additional two of our lead SNPs are in LD ($r^2 > 0.6$) with SNPs that have previously been identified in neuroticism or one of its subdimensions (rs147420422 and rs9379833; see **Table 2**). Results from additional lookups in the open GWAS project database (<https://gwas.mrcieu.ac.uk/about/>, accessed Oct 14th 2020) are available in **Supplemental Table S5** and as PheWAS plots in **Supplemental Figure S6**.

TWAS using EpiXcan identified 11 Bonferroni significant genes/isoforms and 96 significant transcripts at FDR < 0.05 with different imputed expression in DLPFC between ADHD and ASD cases (**Supplemental Table S2** and **Figure 1**). The *HIST1H2BD-201* isoform located in the chromosome 6 (HIST1) GWAS locus showed the strongest association ($p = 2.08 \times 10^{-9}$) with higher expression in ADHD compared to ASD cases (**Supplemental Figure S3b**). The other genes/isoforms showed orders of magnitude less significant association, appointing *HIST1H2BD-201* as the top-ranking causal candidate in the locus. The remaining 10 Bonferroni significant genes/isoforms were located in the chromosome 8 GWAS locus (*SLC35G5-201*, *AF131215.5*, *AF131215.5-201*, *FAM167A* and *TDH-204*) or in two loci on chromosome 3 (3p21.1: *RFT1-204* and 3p21.31: *CAMKV-210*, *MON1A-201*, *RBM6-210* and *TRAIIP*; **Supplemental Figures S3c + 3d**, respectively) where all except *TRAIIP* were also genome-wide significant in gene-based analysis using MAGMA (**Supplemental Figure S4b** and **Supplemental Table S3**).

Genetic correlations with other traits

To examine the polygenic architecture of the identified shared and differentiating genetic risk for the disorders we investigated the genetic correlations with 258 traits from a manually curated list of previously published GWAS and 597 traits from the UK Biobank making use of LD Hub²⁹ and LDSC³⁰. Among the 258 previously reported GWAS, 30 (combined GWAS) and 32 (ADHD vs ASD) traits showed significant correlations after Bonferroni correction for multiple testing (**Supplemental Table S4**). The strongest correlations for the liability differentiating ADHD vs ASD GWAS were observed for cognitive traits such as *years of schooling* ($r_G = -0.669, p_{\text{corr}} = 3.68 \times 10^{-85}$) and *childhood IQ* ($r_G = -0.609, p_{\text{corr}} = 2.78 \times 10^{-10}$), while the strongest correlations for the combined GWAS were with traits such as *depressive symptoms* ($r_G = 0.506, p_{\text{corr}} = 2.08 \times 10^{-19}$) and the cross-disorder analyses of the PGC ($r_G = 0.433, p_{\text{corr}} = 5.30 \times 10^{-25}$). Unsurprisingly, the largest difference in r_G ($\text{abs}(\Delta r_G)$) for the original ADHD and ASD GWAS was identified for a series of cognitive traits (largest $\text{abs}(\Delta r_G) = 0.733$ for *years of schooling*) (**Supplemental Table S4**). For a scatter plot of all genetic correlations that showed $Z > 2$ in the original ADHD and ASD GWAS please refer to **Supplemental Figure S7**.

Tissue and cell-type enrichment analyses

We next tested whether genetic associations of shared and differentiating genetic underpinnings were enriched with respect to the transcriptomic profiles of human tissues. We found that the transcriptomic profiles related to brain tissues were significantly associated with the shared ADHD-ASD genetics (**Supplemental Figure S8**). At the general tissue level, these associations were with brain, pituitary and testis tissue. At the individual tissue level, the most significant association was observed for the basal ganglia (putamen), followed by the cerebellum. Cell-type enrichment analyses revealed experiment-wide significant association (across all data sets tested) of the red nucleus in the midbrain in human adult brain samples reported in La Manno et al. 2016³¹ (**Supplemental Figure S9c**). Associations that were significant within one of the three tested data sets individually, but not overall, were observed for several cell types, including e.g. dopaminergic and GABAergic neurons. For the disorder-differentiating analysis (ADHD vs

ASD) we observed no significant association with tissues or specific cell-types after correction for multiple testing (**Supplemental Figure S10 and S11**). We also intersected our genetic associations with a recent multi-omic single-cell epigenetic catalogue of the human brain (obtained from GEO GSE147672³²). Here both the combined and differentiating GWAS results showed significant enrichment for several neuronal cell populations (see **Supplemental Figure S12 and Supplemental Table S6**), including excitatory and inhibitory neurons. Interestingly, the only difference in terms of significant associations between the combined and ADHD vs ASD GWAS was seen for oligodendrocytes (which were not significant in the combined but in the ADHD vs ASD GWAS). While aberrant myelination by oligodendrocytes resulting in disruption of white matter development has previously been reported in both ASD and ADHD^{33,34}, the degree of severity of this alteration might be a distinct pathophysiological factor³⁵.

Polygenic characterization of case subgroups

We used two complementary polygenic risk score (PRS) approaches to investigate differences in polygenic load for ADHD, ASD and related phenotypes in the iPSYCH data across the three phenotypic subgroups: ASD-only, ADHD-only and comorbid cases. The multivariate PRS framework revealed significant association of the ASD-only subgroup with PRS for ASD ($p = 6.89 \times 10^{-26}$) and the ADHD-only subgroup with PRS for ADHD ($p = 3.29 \times 10^{-23}$). Both scores were trained with PGC-only GWAS results^{5,36}. Strikingly, the ASD-PRS load on comorbid ASD+ADHD cases was similar to that on ASD-only cases ($p = 0.77$) and, likewise, the ADHD-PRS load on the comorbid subgroup was similar to that on ADHD-only cases ($p = 0.44$, **Figure 2**), demonstrating that the comorbid cases carry a load of both ADHD and ASD polygenic scores that are similar to the load carried by the single-disorder cases of their respective disorder PRS. In other words, comorbid cases are double-burdened with both ASD and ADHD PRS. In contrast, the ASD-PRS load on ADHD-only cases was not different from controls ($p = 0.79$) and the ADHD-PRS was only slightly increased in ASD-only cases compared to controls ($p = 3.26 \times 10^{-3}$, **Figure 2**).

Results from our leave-one-out framework analysis (including both PGC and iPSYCH data in the training GWAS) showed similar results (**Table 3**), adding further support to the observation of comorbid cases being double-burdened with both ASD and ADHD PRS. We note that in this analysis, the ASD-PRS load on ADHD-only cases as well as the ADHD-PRS load on ASD-only cases were increased compared to controls. Furthermore, secondary analysis in the leave-one-out framework suggested that ADHD cases with ($n = 625$) and without mild ID ($n = 11,339$) did not differ in terms of PRS for either ADHD or ASD. On the other hand, ASD cases with ID ($n = 634$) had lower PRS_{ASD} (OR = 0.89 [0.81-0.97], $p = 0.0072$) compared to those without mild ID ($n = 8,681$) but did not differ in terms of PRS_{ADHD} (**Table 3**).

To further dissect the genetic architecture across the ASD and ADHD subgroups we examined the relative burden of PRS for phenotypes and traits that have shown significant genetic correlation with ADHD and ASD^{5,6,37} using the multivariate framework analysis. While PRS for SZ and depression (and genetically related phenotypes) did not show substantially different loads across the subgroups, other traits showed compelling differences (**Figure 2**). For instance, years of education, IQ, age at first birth, tiredness, and smoking showed differences between ADHD-only and ASD-only cases with the comorbid cases at an intermediate level. These differences were most compelling for the cognitive phenotypes displaying PRS loads in opposite directions for the single-disorder cases and intermediate loads for comorbid cases. In addition, analyses for the chronotype trait showed similar PRS loads in ASD-only and comorbid cases without evidence for loading in ADHD-only cases, reflecting that chronotype is genetically correlated to ASD but not ADHD^{5,6}.

An item-level analysis of neuroticism revealed specific patterns of associations for the ASD-only and ADHD-only groups that were mostly consistent with previously described patterns³⁷ (**Supplemental Figure S13**). On average, ADHD-only cases showed much stronger association than ASD-only cases with items belonging to the depressed affect cluster (e.g. the *MOOD* item) compared to the worry cluster (for a definition of the clusters see **Supplemental Figure S13**). For comorbid cases a distinct pattern was

observed with PRS loads either ranking between the ADHD- and ASD-only cases (e.g. for the *MOOD* item) or even exceeding the two single-disorder groups (e.g. for the *GUILT* item).

Summarizing, we observe a genetic architecture of comorbid cases that presents itself in clear distinction from the ADHD and ASD single-disorder cases. Showing burden of both ASD and ADHD genetic risk, the comorbid cases also carry polygenic load profiles across other phenotypes that distinguishes them from their single-disorder cases, typically by carrying an intermediate load level but in some cases a load similar to just one of the single-disorder groups.

Genetic correlation and heritability across case subgroups

We recently reported an LDSC genetic correlation of 0.36 between ASD and ADHD using the largest GWAS meta-analyses of the two disorders, including multiple cohorts and comorbid cases⁵. Here we investigated the correlations across diagnostic subgroups of the disorders in the iPSYCH sample using the GREML approach of GCTA³⁸. For ASD and ADHD overall, we found $r_G = 0.497$ (SE = 0.054, $p = 7.8 \times 10^{-19}$). Excluding the comorbid cases reduced the correlation to $r_G = 0.397$ (SE = 0.056, $p = 6.3 \times 10^{-12}$). After excluding cases with intellectual disability (ID), the correlations between ASD and ADHD were even stronger: $r_G = 0.523$ (SE = 0.054, $p = 6.5 \times 10^{-21}$) and $r_G = 0.425$ (SE = 0.056, $p = 1.7 \times 10^{-13}$) with and without comorbid cases, respectively. All the GCTA results on genetic correlations and SNP heritability estimates can be found in **Supplemental Table S7** and **Supplemental Figure S14**.

Correlations between ADHD and ICD-10 diagnostic subcategories of childhood autism (F84.0, $n = 3,273$), atypical autism (F84.1, $n = 1,472$), Asperger's syndrome (F84.5, $n = 4,363$), and other/unspecified pervasive developmental disorders (other PDDs, F84.8-9, $n = 3,794$), reducing to nonoverlapping groups when performing pairwise comparisons (**Supplemental Table S7**), were mostly similar to those for the ASD group overall, albeit with generally higher estimates for the groups with other PDDs and Asperger's syndrome (**Supplemental Table S7** and **Supplemental Figure S14**).

Discussion

This study dissects the genetic architecture for shared and differentiating genetic underpinnings of ADHD and ASD as well as across case subgroups. At the single variant level, we identified novel shared loci for the two disorders and, to the best of our knowledge, the first genome-wide significant loci differentiating the disorders. Integration with DLPFC transcriptomic data identified and prioritized several possibly causal genes (**Box 1**). At the polygenic level, we revealed compelling differences across comorbid and single-disorder case groups.

The identified shared loci are generally highly pleiotropic and have previously been identified in GWAS of related disorders or cross-disorder studies including ADHD and/or ASD. However, considering only psychiatric disorders, the two novel shared loci (on chromosome 1 and 4) appear to be shared only between ADHD and ASD (**Table 1**). This is consistent with evidence from structural equation modeling of eight major psychiatric disorders, showing that ASD and ADHD cluster together in a group of early-onset neurodevelopmental disorders along with Tourette syndrome ⁸.

Analyzing genetic correlation with other traits, the combined GWAS results showed strong correlations of shared ADHD-ASD genetics with other psychiatric phenotypes, suggesting that additional shared loci that may be discovered with increasing sample sizes in future studies will likely show a high degree of overlap with other psychiatric disorders, similarly to the shared loci reported here.

In the ADHD vs ASD GWAS we identified five genome-wide significant loci, all showing opposite allelic directions in the separate GWAS of the two disorders, providing the first specific genetic clues to understanding the biology that drives the pathophysiology towards developing one or the other disorder.

The top-ranking TWAS gene/isoform was *HIST1H2BD-201*, which was two orders of magnitude more significant than the second-ranking (*CAMKV-210*) and the only Bonferroni significant transcript in the identified histone 1 GWAS locus. Deleterious *de novo* mutations in several histone modifying or interacting genes ³⁹⁻⁴¹ as well as in core histone genes ^{40,42} have been associated with autism and developmental delay

with autistic features. The haploinsufficiency resulting from these de novo mutations is consistent with our TWAS result showing reduced expression of *HIST1H2BD-201* in ASD (relative to ADHD). Intriguingly, the ASD risk allele of the lead SNP in the locus is also associated with both increased educational performance²⁴ and increased volume of the left globus pallidus²⁶ while the opposite is the case for the ADHD risk allele. As part of the basal ganglia, globus pallidus is involved in several functions relating to phenotypic domains affected in ASD and/or ADHD such as cognition, social interactions, speech, repetitive behaviors and tics²⁷. Taken together our results suggest that the identified ADHD-ASD differentiating locus on chromosome 6 has downstream effects involving differential expression of the histone isoform *HIST1H2BD-201* and volumetric changes of the left globus pallidus in driving the pathophysiology towards either ASD or ADHD and impacting key phenotypic domains such as educational performance, social interaction and motor impairments.

Previous studies found ASD and ADHD to display opposite genetic correlations with cognitive traits like educational attainment when assessing common variants genome-wide^{5,6,43}. Corroborating these reports, we found that the ADHD vs ASD GWAS showed the strongest correlations for cognitive traits among the multiple phenotypes tested (**Supplemental Table S4 and Supplemental Figure S7**). Moreover, two of the identified differentiating loci (on chromosome 1 and 6) have lead SNPs that are genome-wide significant in educational attainment and show opposite allelic effects with increasing and decreasing educational performance for the ASD and ADHD risk alleles, respectively.

We note that the chromosome 1 locus (at position 44Mb) was identified, counterintuitively, in both the shared and differentiating GWAS albeit with different lead SNPs (**Table 1 and 2**). The locus covers a gene-rich 250kb region of generally strong linkage disequilibrium (LD) but it also harbors variants with limited LD to the main haploblock (**Supplemental Figure S1a and S5d**). The two lead SNPs are located 62kb apart and show low pairwise LD ($r^2 = 0.1687$, **Table 2**), indicating that the two SNPs are largely independent markers for association. This LD difference is also reflected in the different lists of other traits with

previously reported associations for the lead SNPs or their LD proxies (**Table 1 and 2**). Furthermore, the locus was the single locus showing significant heterogeneity across cohorts in the recent ADHD GWAS⁶ where the 23andMe sample provided no support for the otherwise consistently supported locus and, also in contrast to the other cohorts, exhibited limited genetic correlation with educational attainment.

Our analyses revealed the expected enrichment of brain-expressed genes for the combined GWAS. In particular, the basal ganglia and the cerebellum seem to be implicated. Both structures have been found to be altered in both ASD^{27,44} and ADHD⁴⁵⁻⁴⁷, with evidence for reductions in basal ganglia volume the most robustly observed finding in the neuroimaging literature for both ASD and ADHD. In addition to the brain, enrichment of genes expressed in the pituitary gland and the testes was also observed for the combined GWAS results. This finding may suggest the involvement of the (hypothalamic-)pituitary-gonadal axis and potentially estrogen signaling, which is known to play a role in psychiatric disorders, cognition, and neuroprotection via several neurotransmitter systems, such as the dopaminergic, serotonergic, and glutamatergic system (e.g. Hwang et al.⁴⁸). The cell-type enrichment result implicating the red nucleus in midbrain is also consistent with our knowledge of phenotypic sharing between ASD and ADHD, as it relates to skilled movements and motor control in the limbs as well as jaw: both motor coordination problems and speech problems are frequent in both ASD and ADHD^{49,50}. The red nucleus is strongly connected with many brain structures involved in ASD and ADHD, including the basal ganglia and the cerebellum⁵¹.

Dissecting the polygenic architecture using PRS approaches we observed remarkable differences across comorbid and single-disorder case groups. The comorbid cases carry a double burden of ASD- and ADHD-PRS, whereas the single-disorder cases were largely just (single-)burdened for the respective disorder. Thus, cases diagnosed with both disorders have on average a similar level of genetic liability to each disorder as the single-disorder cases, providing strong biological support for the change in diagnostic guidelines from DSM-IV to DSM-5 allowing for diagnoses of both disorders in the same person. It also supports pharmacological treatment of comorbid ADHD in individuals with ASD. In a recent meta-analysis,

25-32% of individuals with ASD also fulfill criteria for ADHD¹³, yet only 15-16% are treated with ADHD medications^{52,53}, despite strong evidence of beneficial effects on the core symptoms of ADHD and potentially also reduced risk of injuries⁵⁴, depression⁵⁵, suicidal behavior⁵⁶ and improved academic performance⁵⁷. Moreover, it indicates that pharmacological treatment of symptoms like hyperactivity, inattention, impulsivity, aggression and tics in cases diagnosed with either ADHD or ASD may be guided by the individual symptomatology regardless of the given diagnosis.

The multivariate PRS analysis also revealed clear differences across the case subgroups for PRS from several other traits, particularly cognition-related traits, again highlighting the opposite relationship for the two disorders with e.g. educational attainment and IQ, which, unsurprisingly, was expanded for the single-disorder cases (with exclusion of the comorbid cases), while the PRS load in the comorbid case group was placed in-between but dominated by the strong negative correlation observed for ADHD cases.

We recently reported a significant genetic correlation of $r_G = 0.36$ between ASD and ADHD, using LDSC and results from GWASs that included multiple cohorts and comorbid cases⁵. This was a considerable increase from the previous estimate of $r_G = 0.08$ ($SE = 0.10$, $p = 0.40$), which was based on much smaller GWAS sample sizes without information on comorbid diagnoses⁵⁸. Here we analyzed exclusively the iPSYCH cohort, which is relatively homogeneous and has information on all diagnoses given to each individual. We found a higher correlation ($r_G = 0.497$), which remained substantial when excluding the comorbid cases ($r_G = 0.397$), demonstrating that the genetic overlap between the disorders is far from driven by comorbid cases alone. This is corroborated by data from Swedish twin studies that supports the distinction of ASD and ADHD, but also suggests considerable co-occurrence of symptoms of both disorders in individuals only fulfilling diagnostic criteria for one of the two disorders^{59,60}.

In addition, the correlations increased when excluding cases with ID, indicating that cases with ID are more genetic heterogeneous in common variant risk between the two disorders than cases without ID. A recent exome sequencing study of ASD and ADHD (also in the iPSYCH cohort) showed that the disorders have

substantial overlap in rare variant risk and that cases with ID carry a higher load of (ultra)rare damaging risk variants compared to cases without ID¹⁰. Consistent with this, our PRS analyses found that there was lower ASD PRS in the group of ASD cases with comorbid mild ID (IQ=50-70) compared to those without mild ID. Taken together, these observations are consistent with the notion that the genetics differentiating the two disorders may be driven primarily by common variants and more extensively for cases with ID than without ID. However, larger sample sizes for both GWAS and sequencing studies are needed to clarify this.

In conclusion, we have disentangled the shared and differentiating genetic liability underlying ASD and ADHD, identifying novel shared as well as disorder-specific risk variants informing on the pathophysiology. In addition, we have revealed specific patterns of polygenic architecture that are characteristic for comorbid cases compared to single-disorder cases. The results advance the understanding of the complex etiologic basis and relationship between ASD and ADHD towards the long term goals of better diagnosis and treatment of these disorders.

Methods

We report results from a framework of different analyses all carried out in large-scale samples from the Psychiatric Genomics Consortium (PGC) and the Lundbeck initiative of integrative psychiatric research (iPSYCH) samples. We used samples included in the most recent GWAS of ASD⁵ and ADHD⁶. For the purpose of this manuscript we will refer to individuals in the study cohort (most importantly in the iPSYCH cohort) that at the time of inclusion only had one of the two diagnoses registered (i.e. ADHD or ASD) as *ADHD-only* and *ASD-only* cases, respectively. We refer to individuals that during their lifetime and up to the time of inclusion had both an ADHD and ASD diagnosis registered as *comorbid* cases. Furthermore, we refer to these three groups of cases (i.e. *ADHD-only*, *ASD-only*, and *comorbid*) as *ASD and ADHD subgroups*.

Sample description and additional quality control

Details about study specific case and control selection criteria and how individuals were drawn from the overall population-based iPSYCH case-cohort sample⁶¹ can be found in the respective publications^{5,6}. Here we focus on important differences in the case and control selection criteria in the iPSYCH cohort as well as additional quality control (QC) procedures necessary for the current study.

Almost all of the inclusion and exclusion criteria for the original studies were also used in this study. The only difference compared to the original studies was an additional exclusion criterion that removed individuals with a moderate to severe mental retardation (ICD10: F71-F79) from both the case and control cohorts. While this criterion was also used in the original ADHD GWAS⁶, it was, however, not used in the original ASD GWAS⁵. The rationale for this decision lies in the interpretability of our results where we treated ADHD and ASD consistently. We address the potential impact of this decision through different analyses (see **Table 3**, **Supplementary Figure S14b**, and **Supplementary Table S7**).

Wave-wise pre-imputation QC and imputation of the iPSYCH case-cohort sample were taken from the original ADHD and ASD GWAS, respectively. Details about the respective steps and filters can be found elsewhere^{5,6}. Since our analysis framework used a combined study cohort with samples from both the original ADHD and ASD GWAS we performed some additional QC on the combined sample. The additional QC steps included the removal of related individuals across the original ADHD and ASD GWAS as well as a new principal component analysis (PCA) on the combined sample after exclusion of these related individuals. Following the same procedures as in the original studies, pairs of subjects were identified with $\pi\text{-hat} > 0.2$ (using PLINK's⁶² identity by state analysis) and one subject of each such pair was excluded at random (with a preference for keeping cases). PCA was carried out using smartPCA in the EIGENSOFT software package^{63,64} using the framework of the Ricopili pipeline⁶⁵. The original PGC datasets for ADHD and ASD did not include overlapping individuals and therefore the original datasets and summary statistics were used. The final combined dataset across all samples comprised 34,462 cases (i.e. individuals with an ADHD and/or ASD diagnosis) and 41,201 controls. We only included samples of European ancestry from the original ADHD and ASD GWAS. Among the cases in the iPSYCH cohort 11,964 had an ADHD-only diagnosis, 9,315 had an ASD-only diagnosis, and 2,304 individuals had a comorbid diagnosis.

Genome-wide association analyses

Like with the original GWAS in ADHD and ASD, all processing and analyses for the individual GWAS and meta-analyses (see below) used the framework of the Ricopili pipeline⁶⁵. More details on individual modules and steps can be found elsewhere^{5,6,65}. We ran two main GWAS within our framework of analyses. The first one aimed to identify shared genetic risk for ADHD and ASD (*combined GWAS*) and the second one aimed to identify differentiating genetic risk with an opposite direction of effects for ADHD and ASD (*ADHD vs ASD GWAS*). All analyses of the iPSYCH sample and meta-analyses with the PGC samples were conducted at the secured national GenomeDK high-performance computing cluster in Denmark. The study

was approved by the Regional Scientific Ethics Committee in Denmark and the Danish Data Protection Agency.

Combined GWAS

We first ran an analysis in the combined dataset, i.e. on all 34,462 cases and 41,201 controls. The GWAS was conducted in each cohort (i.e. in the wave-wise iPSYCH samples and the individual PGC cohorts) using logistic regression with the imputed additive genotype dosages. The first 5 principal components (PCs) were included as covariates to correct for population stratification (**Supplementary Information**), and variants with imputation INFO score < 0.8 or minor allele frequency (MAF) < 0.01 were excluded. The resulting summary statistic files were then meta-analyzed using an inverse-variance weighted fixed effects model⁶⁶. Post-processing of the summary statistics files through the Ricopili pipeline⁶⁵ created Manhattan plots, individual regional associations plots, and forest plots. For a QQ-plot of the analysis please refer to **Supplemental Figure S15a**.

ADHD vs ASD GWAS

To identify unique genetic risk loci or loci with opposite direction of effects for ADHD and ASD we ran a case-only analysis for the ADHD-only (coded as 1, i.e. “pseudo-cases”; $n = 11,964$) against ASD-only cases (coded as 2, i.e. “pseudo-controls”; $n = 9,315$) in the iPSYCH cohort. We excluded the comorbid cases from this GWAS. Similar to the analysis in the combined sample (see above) GWAS was conducted wave-wise using logistic regression with the imputed additive genotype dosages. The first 5 PCs were included as covariates to correct for population stratification, and variants with imputation INFO score < 0.8 or MAF < 0.01 were excluded. The resulting summary statistic files were then meta-analyzed using an inverse-variance weighted fixed effects model⁶⁶ and visualization of results was achieved through the Ricopili pipeline⁶⁵. Post-processing of the summary statistics files through the Ricopili pipeline⁶⁵ created

Manhattan plots, individual regional associations plots, and forest plots. For a QQ-plot of the analysis please refer to **Supplemental Figure S15b**.

Identification of previously reported associations for top findings

Different resources were used to look up previously reported associations of our top findings with other phenotypes and traits within and outside of psychiatry. We assessed associations reported in the open GWAS project database (<https://gwas.mrcieu.ac.uk/about/>, accessed Oct 14th 2020; see **Supplementary Table S1** and **Supplementary Table S5** for results) and also used the GWAS ATLAS website ⁶⁷ to visualize PheWAS analyses (see **Supplementary Figures S2** and **S11**). Finally, we also used results from the GWAS Catalog ⁶⁸ (see **Table 2**).

Transcriptomic imputation model construction and transcriptome-wide association study (TWAS).

Transcriptomic imputation models were constructed as previously described ²⁰ for dorso-lateral prefrontal cortex (DLPFC) transcript levels ⁶⁹. The genetic dataset of the PsychENCODE cohort was uniformly processed for quality control (QC) steps before genotype imputation. We restricted our analysis to samples of European ancestry as previously described ²⁰. Genotypes were imputed using the University of Michigan server ⁷⁰ with the Haplotype Reference Consortium (HRC) reference panel ⁷¹. Gene expression information (both at the level of gene and transcript) was derived from RNA-seq counts which are adjusted for known and hidden confounds, followed by quantile normalization ⁶⁹. For the construction of the transcriptomic imputation models we used EpiXcan ²⁰, an elastic net based method, which weighs SNPs based on available epigenetic annotation information ⁷². EpiXcan was recently shown to increase power to identify genes under a causality model when compared to TWAS approaches that don't integrate epigenetic information ⁷³. We use this model (924 samples from DLPFC) due to power considerations ²⁰; in comparison, brain gene expression imputation models based on GTEx V8 ⁷⁴ are trained in 205 or fewer samples. We performed the transcript-trait association analysis for the traits in this study as previously described ²⁰. Briefly, we

applied the S-PrediXcan method²⁰ to integrate the GWAS summary statistics and the transcriptomic imputation models constructed above to obtain association results at both the level of genes and transcripts.

Cell-type enrichment analysis

A major portion of cell type specific enrichment is attributed to distal regulatory elements, as local regulatory events remain highly consistent across various tissues and cell types⁷⁵. Therefore, we examined an overlap of common genetic variants of investigated traits (see **Supplemental Figure S14** and **Supplemental Table S6**) and open chromatin from scATAC-seq study (single-cell assay for transposase accessible chromatin)³² using the LD-score partitioned heritability approach⁷⁶. All regions of open chromatin were extended by 500 base pairs in either direction. The broad MHC-region (hg19 chr6:25-35MB) was excluded due to its extensive and complex LD structure, but otherwise default parameters were used for the algorithm.

Additional functional characterization and annotation of main findings

We used a number of different approaches combining in house scripts and data with those available via the FUMA v1.3.6a²³ website (<http://fuma.ctglab.nl>) for downstream functional characterization and annotation of our findings. For FUMA we uploaded our summary statistics from the individual analyses. We also used FUMA to perform tissue expression analyses on data available through their website. Finally, we used FUMA to perform cell-type specificity analyses⁷⁷ based on our summary statistics. For all above mentioned analyses default settings were applied. More detailed information about the individual third-party datasets (available through FUMA) included in the analyses as well as individual aspects of the FUMA analyses can be found in the **Supplemental information. Supplemental Table S8** contains results from standard FUMA-based analyses, such as eQTL and chromatin interaction mapping.

Gene-based analysis

We also used FUMA v1.3.6a²³ to perform gene-based analysis. Genome-wide significance was assessed through Bonferroni correction for the number of genes tested. More detailed information about the individual third-party datasets (available through FUMA) included in the analyses as well as individual aspects of the gene-based analyses can be found in the **Supplemental information**.

Our results in context of other findings

Since the publication of the original ADHD and ASD results a few studies have investigated the shared and unique risk architecture of these disorders. We compared our results with the findings of the cross disorder working group of the PGC⁸ and a recent analysis based on structural equation modelling of 11 major psychiatric disorders [<https://doi.org/10.1101/2020.09.22.20196089>]. We also compared our results with recent analyses that aimed at identifying disorder-specific SNPs for psychiatric disorders^{78,79}.

Polygenic risk score (PRS) analyses

To examine potential polygenic heterogeneity across ADHD and ASD subtypes, we investigated how PRS trained on different phenotypes were distributed across ADHD-only, ASD-only and comorbid subgroups in the iPSYCH data through two complementary analysis frameworks: multivariate PRS and leave-one-out PRS. These two approaches have different strengths and limitations, allowing for robust interrogation of differences in ADHD and ASD subgroups in terms of polygenic burden for ADHD and ASD, as well as genetically related phenotypes.

Multivariate PRS analyses

To examine the relative burden of PRS for phenotypes and traits that have shown significant genetic correlation with ADHD and ASD in the past^{5,6,37} across ADHD and ASD subgroups in the iPSYCH data, we ran a multivariate regression of the scores on these subgroups, adjusting for PCs and batch. For details, see Grove et al.⁵. In brief, this is a regression of multiple outcome variables and can superficially be viewed

as running a linear regression for each score on the ADHD and ASD subgroups. It allows us to compare the average PRS across subtypes for scores from a number of phenotypes while accounting for the inherent correlation between scores and adjusting for necessary covariates. This enables testing a whole array of hypotheses comparing both subtypes and PRS. Polygenic scores were generated by clumping and thresholding employing standard Ricopili settings as explained in ⁵ and using summary statistics from the GWASs ^{5,36,80-89}.

Leave-one-out PRS analyses

As a complementary approach, a leave-one-wave-out approach within the iPSYCH data was used to maximize power and maintain independent target and discovery samples for PRS analyses. Meta-analyses were run in METAL (using inverse-variance weighted fixed effects models with the STDERR scheme), including the per-wave GWAS summary results from all but one wave of data, for each combination of waves. Separate meta-analyses were run for GWAS of ADHD-only (excluding comorbid ASD or severe ID, defined as $IQ \leq 50$) cases vs. controls and ASD-only (excluding comorbid ADHD or severe ID) cases vs. controls, using independent (split) controls. For each set of discovery results, LD-clumping was run in PLINK v.1.9 ⁹⁰ (with the parameters `--clump-kb 500 --clump-r2 0.3`) to obtain a relatively independent set of SNPs, while retaining the most significant SNP in each LD block. The SNP selection p-value threshold used was $p < 0.5$. Asymmetric/ambiguous SNPs (AT, TA, CG, GC), indels, multi-allelic and duplicate position SNPs were excluded. SNPs with $MAF < 0.01$, $INFO < 0.8$ or present in less than half of the sample were filtered out. PRS for ADHD and ASD were calculated for each individual in each target wave by scoring the number of effect alleles weighted by the log(odds ratio [OR]) across the set of independent clumped, meta-analyzed SNPs in PLINK. PRS were derived in best guess imputed data after filtering out SNPs with $MAF < 0.05$ and $INFO < 0.8$. The PRS were standardized using z-score transformations; ORs can be interpreted as the increase in risk of the outcome, per standard deviation in PRS. Logistic regression analyses including 5 PCs were run to test for association of PRS with each of the outcomes within each wave, as follows: a) ADHD-only cases vs. controls, b) ASD-only cases vs. controls, c) comorbid cases vs.

controls, d) ADHD-only cases vs. ASD-only cases, e) ADHD-only cases vs. comorbid cases, and f) ASD-only cases vs. comorbid cases. Cases were coded as 1 and controls as 0, except that comorbid cases were coded as 1 in case-case comparisons and in analysis (d), the ASD-only cases were coded as 1. Overall meta-analyses of these per-wave analyses were performed in R using the ‘metafor’ package. As secondary tests, we stratified the ADHD-only and ASD-only cases by presence of mild ID (defined as IQ between 50-70). We also examined differences across several ASD hierarchical subtypes (childhood autism, atypical autism, Asperger’s, and pervasive developmental disorders mixed; see Grove et al ⁵ and **Supplemental Table S7**). Several sensitivity tests were also run (including sex as a covariate, excluding cases and controls with mild ID).

Genetic correlations (LDhub)

The genetic correlations of our different datasets with other phenotypes were evaluated using LD Score regression (LDSC) ³⁰ and the LD Hub ²⁹ website (<http://ldsc.broadinstitute.org/ldhub/>). In brief, we re-ran analyses of the original GWAS of ADHD and ASD ^{5,6} in the European-only datasets since new phenotypes have been added to LD Hub after publication of the original analyses. We also uploaded summary statistics for the two analyses described above, i.e. the GWAS in the combined sample (combined GWAS) and the pseudo case-control analysis (ADHD vs ASD GWAS). We used all available phenotypes in LD Hub ²⁹ but performed analyses for the UKBB traits ($n = 597$) and the remaining individual phenotypes ($n = 257$) separately. For ADHD ⁶ and ASD ⁵ the most recent summary statistics replaced corresponding summary statistics in LD Hub as these had not been included at the date of analysis. The same was true for the summary statistics of major depressive disorder (MDD ⁸⁵) and bipolar disorder (BD ⁹¹). Levels of experiment-wide significance (Bonferroni correction for number of tests applied) were also established separately within the two groups, i.e. in the UKBB traits ($p < 8.38 \times 10^{-3}$) and the remaining individual phenotypes ($p < 0.00019$), respectively.

GCTA-GREML analyses across subgroups

The additive variance explained by our GWAS dataset (SNP-based heritability; $\text{SNP-}h^2$) was estimated in the iPSYCH sample using the GREML approach of GCTA³⁸ for ADHD versus ASD and for ADHD versus each of the ASD sub-phenotypes listed in **Table 1**. The genetic relationship matrix (GRM) between all pairwise combinations of individuals was estimated using all case-control samples. The strict best-guess-genotypes (i.e. SNPs with $\text{INFO} > 0.8$, missing rate < 0.01 and $\text{MAF} > 0.05$, INDELs removed) were used for GRM estimation. GCTA-GREML accounts for linkage disequilibrium (LD)⁹², and the GRM estimation was therefore performed on a non-LD-pruned dataset. Estimation of the phenotypic variance explained by the SNPs was performed for each of the phenotypes listed in **Supplemental Table S7**, with PCs 1-20 included as continuous covariates and wave (1-23) as categorical dummy variables. ADHD prevalence of 0.05 and ASD prevalence of 0.01 was assumed to estimate the variance explained on the liability scale. Prevalence was estimated for hierarchical ASD phenotypes based on the estimate for the overall ASD phenotype and the proportion of each hierarchical phenotype over all ASD cases observed in our sample. Genetic covariance between pairs of traits (**Supplemental Table S7**) was estimated using the bivariate approach implemented in GCTA, by randomly splitting controls into two groups, one for each trait, in proportions corresponding to the proportion of the cases for each of the two traits in the total sample. PCs 1-20 and dummy variables for wave 1-23 were included as covariates in the bivariate analyses. Two-tailed p-values were obtained for r_G point estimates based on the standard error estimated by GCTA using the approach by Altman and Bland⁹³.

GCTA-GREML analyses were conducted for ADHD versus ASD main diagnosis (**Supplemental Figure S5a**), by (1) excluding individuals with both phenotypes (comorbid) and (2) by randomly splitting comorbid cases into either ADHD or ASD. GCTA analyses were, in addition, conducted for ADHD versus four ASD sub-phenotypes, by (1) excluding individuals with both phenotypes (comorbid) and (2) by randomly splitting comorbid cases into either the ADHD or ASD sub-phenotype. These analyses were conducted both including and excluding individuals with intellectual disability. Please see **Supplemental Table S7** and **Supplemental Figure S5** for an overview of comparisons.

Acknowledgements

The iPSYCH team was supported by grants from the Lundbeck Foundation (R102-A9118, R155-2014-1724, and R248-2017-2003), the EU H2020 Program (Grant No. 667302, “CoCA”), NIMH (1U01MH109514-01 to ADB) and the Universities and University Hospitals of Aarhus and Copenhagen. The Danish National Biobank resource was supported by the Novo Nordisk Foundation. High-performance computer capacity for handling and statistical analysis of iPSYCH data on the GenomeDK HPC facility was provided by the Center for Genomics and Personalized Medicine and the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark (grant to ADB). Dr. Valodakis was funded by NIMH grant K08MH122911 and the 2020 NARSAD Young Investigator Grant #29350 from the Brain & Behavior Research Foundation. Support for this research was further received by Dr. Cormand from the Spanish ‘Ministerio de Ciencia, Innovación y Universidades’ (RTI2018-100968-B-100), ‘Ministerio de Economía y Competitividad’, ‘AGAUR/Generalitat de Catalunya’ (2017-SGR-738), the European Union H2020 Program [H2020/2014-2020] under grant agreements n°667302 (CoCA), 643051 (MiND) and 728018 (Eat2beNICE), and the ECNP network “ADHD across the lifespan”. Dr. Faraone is supported by the European Union’s Horizon 2020 research and innovation programme under grant agreements No 667302 and 965381; NIMH grants U01MH109536-01, U01AR076092-01A1, R0MH116037 and 5R01AG06495502; Oregon Health and Science University, Otsuka Pharmaceuticals and Supernus Pharmaceutical Company.

The authors would like to deeply thank all participants in the cohorts included in this analysis.

Author contributions

MM, JG and ADB designed the study. MM, JG, TDA, JM, GV, SM, DD, JB, RW, CEC, AR, NS, MEH, BZ and GH conducted data analysis. PBM, EBR, PR, BMN, MJD and ADB supervised data analysis. JB-G, MBH, EA, MN, TW, OM, DMH, PBM, BMN, MJD and ADB provided data. MM, JG, TDA, JM, SM and ADB wrote the paper. MM, JG, TDA, JM, BC, EBR, SVF, BF, SD and ADB comprised the core

revision group. ADB directed the study. All authors discussed the results and approved the final version of the manuscript.

Competing interests

Barbara Franke has received educational speaking fees from Medice. In the past year, Dr. Faraone received income, potential income, travel expenses continuing education support and/or research support from Takeda, OnDosis, Tris, Otsuka, Arbor, Ironshore, Rhodes, Akili Interactive Labs, Sunovion, Supernus and Genomind. With his institution, he has US patent US20130217707 A1 for the use of sodium-hydrogen exchange inhibitors in the treatment of ADHD. He also receives royalties from books published by Guilford Press: Straight Talk about Your Child's Mental Health, Oxford University Press: Schizophrenia: The Facts and Elsevier: ADHD: Non-Pharmacologic Interventions. He is Program Director of www.adhdinadults.com. The other authors declare no competing interests.

References

1. Dalsgaard, S. *et al.* Incidence Rates and Cumulative Incidences of the Full Spectrum of Diagnosed Mental Disorders in Childhood and Adolescence. *JAMA Psychiatry* **77**, 155-164 (2020).
2. Faraone, S.V. & Larsson, H. Genetics of attention deficit hyperactivity disorder. *Mol Psychiatry* **24**, 562-575 (2019).
3. Pettersson, E. *et al.* Genetic influences on eight psychiatric disorders based on family data of 4 408 646 full and half-siblings, and genetic data of 333 748 cases and controls. *Psychol Med* **49**, 1166-1173 (2019).
4. Sandin, S. *et al.* The Heritability of Autism Spectrum Disorder. *JAMA* **318**, 1182-1184 (2017).
5. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* **51**, 431-444 (2019).
6. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* **51**, 63-75 (2019).
7. Matoba, N. *et al.* Common genetic risk variants identified in the SPARK cohort support DDHD2 as a candidate risk gene for autism. *Transl Psychiatry* **10**, 265 (2020).
8. Cross-Disorder Group of the Psychiatric Genomics Consortium. Electronic address, p.m.h.e. & Cross-Disorder Group of the Psychiatric Genomics, C. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469-1482 e11 (2019).
9. Martin, J. *et al.* Biological overlap of attention-deficit/hyperactivity disorder and autism spectrum disorder: evidence from copy number variants. *J Am Acad Child Adolesc Psychiatry* **53**, 761-70 e26 (2014).
10. Satterstrom, F.K. *et al.* Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. *Nat Neurosci* **22**, 1961-1965 (2019).
11. Rommelse, N.N., Geurts, H.M., Franke, B., Buitelaar, J.K. & Hartman, C.A. A review on cognitive and brain endophenotypes that may be common in autism spectrum disorder and attention-deficit/hyperactivity disorder and facilitate the search for pleiotropic genes. *Neurosci Biobehav Rev* **35**, 1363-96 (2011).
12. Zablotzky, B., Bramlett, M.D. & Blumberg, S.J. The Co-Occurrence of Autism Spectrum Disorder in Children With ADHD. *J Atten Disord* **24**, 94-103 (2020).
13. Lai, M.C. *et al.* Prevalence of co-occurring mental health diagnoses in the autism population: a systematic review and meta-analysis. *Lancet Psychiatry* **6**, 819-829 (2019).
14. Ottosen, C. *et al.* Sex Differences in Comorbidity Patterns of Attention-Deficit/Hyperactivity Disorder. *J Am Acad Child Adolesc Psychiatry* **58**, 412-422 e3 (2019).
15. Ghirardi, L. *et al.* The familial co-aggregation of ASD and ADHD: a register-based cohort study. *Mol Psychiatry* **23**, 257-262 (2018).
16. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
17. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).

18. Yang, Z. *et al.* Investigating Shared Genetic Basis Across Tourette Syndrome and Comorbid Neurodevelopmental Disorders Along the Impulsivity-Compulsivity Spectrum. *Biol Psychiatry* (2021).
19. Sabourdy, F. *et al.* A MANBA mutation resulting in residual beta-mannosidase activity associated with severe leukoencephalopathy: a possible pseudodeficiency variant. *BMC Med Genet* **10**, 84 (2009).
20. Zhang, W. *et al.* Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. *Nat Commun* **10**, 3834 (2019).
21. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**(2018).
22. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).
23. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
24. Lee, J.J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* **50**, 1112-1121 (2018).
25. Marzluff, W.F., Gongidi, P., Woods, K.R., Jin, J. & Maltais, L.J. The human and mouse replication-dependent histone genes. *Genomics* **80**, 487-98 (2002).
26. Zhao, B. *et al.* Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat Genet* **51**, 1637-1644 (2019).
27. Subramanian, K. *et al.* Basal ganglia and autism - a translational perspective. *Autism Res* **10**, 1751-1775 (2017).
28. Baselmans, B.M.L. *et al.* Multivariate genome-wide analyses of the well-being spectrum. *Nat Genet* **51**, 445-451 (2019).
29. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279 (2017).
30. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
31. La Manno, G. *et al.* Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell* **167**, 566-580 e19 (2016).
32. Corces, M.R. *et al.* Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat Genet* **52**, 1158-1168 (2020).
33. Graciarena, M., Seiffe, A., Nait-Oumesmar, B. & Depino, A.M. Hypomyelination and Oligodendroglial Alterations in a Mouse Model of Autism Spectrum Disorder. *Front Cell Neurosci* **12**, 517 (2018).
34. Wu, Z.M. *et al.* White Matter Microstructural Alterations in Children with ADHD: Categorical and Dimensional Perspectives. *Neuropsychopharmacology* **42**, 572-580 (2017).
35. Aoki, Y. *et al.* Association of White Matter Structure With Autism Spectrum Disorder and Attention-Deficit/Hyperactivity Disorder. *JAMA Psychiatry* **74**, 1120-1128 (2017).

36. Neale, B.M. *et al.* Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* **49**, 884-97 (2010).
37. Nagel, M., Watanabe, K., Stringer, S., Posthuma, D. & van der Sluis, S. Item-level analyses reveal genetic heterogeneity in neuroticism. *Nat Commun* **9**, 905 (2018).
38. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
39. Satterstrom, F.K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584 e23 (2020).
40. Duffney, L.J. *et al.* Epigenetics and autism spectrum disorder: A report of an autism case with mutation in H1 linker histone HIST1H1E and literature review. *Am J Med Genet B Neuropsychiatr Genet* **177**, 426-433 (2018).
41. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209-15 (2014).
42. Bryant, L. *et al.* Histone H3.3 beyond cancer: Germline mutations in Histone 3 Family 3A and 3B cause a previously unidentified neurodegenerative disorder in 46 patients. *Sci Adv* **6**(2020).
43. Clarke, T.K. *et al.* Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. *Mol Psychiatry* **21**, 419-25 (2016).
44. Traut, N. *et al.* Cerebellar Volume in Autism: Literature Meta-analysis and Analysis of the Autism Brain Imaging Data Exchange Cohort. *Biol Psychiatry* **83**, 579-588 (2018).
45. Hoogman, M. *et al.* Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: a cross-sectional mega-analysis. *Lancet Psychiatry* **4**, 310-319 (2017).
46. Shaw, P. *et al.* A multicohort, longitudinal study of cerebellar development in attention deficit hyperactivity disorder. *J Child Psychol Psychiatry* **59**, 1114-1123 (2018).
47. Wolfers, T. *et al.* Individual differences v. the average patient: mapping the heterogeneity in ADHD using normative models. *Psychol Med* **50**, 314-323 (2020).
48. Hwang, W.J., Lee, T.Y., Kim, N.S. & Kwon, J.S. The Role of Estrogen Receptors and Their Signaling across Psychiatric Disorders. *Int J Mol Sci* **22**(2020).
49. Fliers, E. *et al.* Motor coordination problems in children and adolescents with ADHD rated by parents and teachers: effects of age and gender. *J Neural Transm (Vienna)* **115**, 211-20 (2008).
50. Franke, B. *et al.* Live fast, die young? A review on the developmental trajectories of ADHD across the lifespan. *Eur Neuropsychopharmacol* **28**, 1059-1088 (2018).
51. Basile, G.A. *et al.* Red nucleus structure and function: from anatomy to clinical neurosciences. *Brain Struct Funct* **226**, 69-91 (2021).
52. Dalsgaard, S., Nielsen, H.S. & Simonsen, M. Five-fold increase in national prevalence rates of attention-deficit/hyperactivity disorder medications for children and adolescents with autism spectrum disorder, attention-deficit/hyperactivity disorder, and other psychiatric disorders: a Danish register-based study. *J Child Adolesc Psychopharmacol* **23**, 432-9 (2013).
53. Rosenberg, R.E. *et al.* Psychotropic medication use among children with autism spectrum disorders enrolled in a national registry, 2007-2008. *J Autism Dev Disord* **40**, 342-51 (2010).

54. Dalsgaard, S., Leckman, J.F., Mortensen, P.B., Nielsen, H.S. & Simonsen, M. Effect of drugs on the risk of injuries in children with attention deficit hyperactivity disorder: a prospective cohort study. *Lancet Psychiatry* **2**, 702-709 (2015).
55. Chang, Z., D'Onofrio, B.M., Quinn, P.D., Lichtenstein, P. & Larsson, H. Medication for Attention-Deficit/Hyperactivity Disorder and Risk for Depression: A Nationwide Longitudinal Cohort Study. *Biol Psychiatry* **80**, 916-922 (2016).
56. Chang, Z. *et al.* Medication for Attention-Deficit/Hyperactivity Disorder and Risk for Suicide Attempts. *Biol Psychiatry* **88**, 452-458 (2020).
57. Keilow, M., Holm, A. & Fallesen, P. Medical treatment of Attention Deficit/Hyperactivity Disorder (ADHD) and children's academic performance. *PLoS One* **13**, e0207905 (2018).
58. Brainstorm, C. *et al.* Analysis of shared heritability in common disorders of the brain. *Science* **360**(2018).
59. Polderman, T.J., Hoekstra, R.A., Posthuma, D. & Larsson, H. The co-occurrence of autistic and ADHD dimensions in adults: an etiological study in 17,770 twins. *Transl Psychiatry* **4**, e435 (2014).
60. Ronald, A., Larsson, H., Anckarsater, H. & Lichtenstein, P. Symptoms of autism and ADHD: a Swedish twin study examining their overlap. *J Abnorm Psychol* **123**, 440-51 (2014).
61. Pedersen, C.B. *et al.* The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol Psychiatry* **23**, 6-14 (2018).
62. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
63. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet* **2**, e190 (2006).
64. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
65. Lam, M. *et al.* RICOPIILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics* **36**, 930-933 (2020).
66. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
67. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* **51**, 1339-1348 (2019).
68. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-D1012 (2019).
69. Gandal, M.J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**(2018).
70. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).
71. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).
72. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).

73. Cao, C. *et al.* Power analysis of transcriptome-wide association study: Implications for practical protocol choice. *PLoS Genet* **17**, e1009405 (2021).
74. Consortium, G.T. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330 (2020).
75. Liu, X. *et al.* Functional Architectures of Local and Distal Regulation of Gene Expression in Multiple Human Tissues. *Am J Hum Genet* **100**, 605-616 (2017).
76. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
77. Watanabe, K., Umicevic Mirkov, M., de Leeuw, C.A., van den Heuvel, M.P. & Posthuma, D. Genetic mapping of cell type specificity for complex traits. *Nat Commun* **10**, 3222 (2019).
78. Byrne, E.M. *et al.* Conditional GWAS analysis to identify disorder-specific SNPs for psychiatric disorders. *Mol Psychiatry* (2020).
79. Peyrot, W.J. & Price, A.L. Identifying loci with different allele frequencies among cases of eight psychiatric disorders using CC-GWAS. *Nat Genet* (2021).
80. Davies, G. *et al.* Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). *Mol Psychiatry* **21**, 758-67 (2016).
81. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539-42 (2016).
82. Benyamin, B. *et al.* Childhood intelligence is heritable, highly polygenic and associated with FBNP1L. *Mol Psychiatry* **19**, 253-8 (2014).
83. Sniekers, S. *et al.* Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat Genet* **49**, 1107-1112 (2017).
84. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-7 (2014).
85. Wray, N.R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* **50**, 668-681 (2018).
86. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet* **48**, 624-33 (2016).
87. Jones, S.E. *et al.* Genome-Wide Association Analyses in 128,266 Individuals Identifies New Morningness and Sleep Duration Loci. *PLoS Genet* **12**, e1006125 (2016).
88. Deary, V. *et al.* Genetic contributions to self-reported tiredness. *Mol Psychiatry* **23**, 609-620 (2018).
89. Tobacco & Genetics, C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* **42**, 441-7 (2010).
90. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
91. Stahl, E.A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* **51**, 793-803 (2019).
92. Yang, J., Lee, S.H., Wray, N.R., Goddard, M.E. & Visscher, P.M. GCTA-GREML accounts for linkage disequilibrium when estimating genetic variance from genome-wide SNPs. *Proc Natl Acad Sci U S A* **113**, E4579-80 (2016).
93. Altman, D.G. & Bland, J.M. How to obtain the confidence interval from a P value. *BMJ* **343**, d2090 (2011).

Table 1: Results of combined (*ADHD* or *ASD*) GWAS

SNP (#CS)	CHR	BP	A1	A2	META				ASD		ADHD		GENES	OTHER
					FRQ _{ca}	FRQ _{co}	OR	P	OR	P	OR	P		
rs7538463 (2/2/2)	1	44196416	A	T	0.707	0.721	0.928	$7.26 \cdot 10^{-10}$	0.961	0.0091	0.914	$1.00 \cdot 10^{-9}$	<i>PTPRF, KDM4A, ST3GAL3, MIR6079</i>	<u>ADHD</u> ¹ , <i>Many</i> ²
rs4916723 (5/5/5)	5	87854395	A	C	0.558	0.573	0.935	$1.52 \cdot 10^{-9}$	0.935	$1.92 \cdot 10^{-6}$	0.925	$1.81 \cdot 10^{-8}$	<i>MIR9-2 (58.3)</i>	<u>ALC</u> ³ , <u>Neuroticism</u> ⁴ , <u>ADHD</u> ¹ , <u>ADHD- CDG</u> ⁵ , <u>CDG</u> ⁶ , <i>sexual partners</i> ⁷ , <u>CDG</u> ⁸
rs2391769 (2/2/2)	1	96978961	A	G	0.351	0.364	0.934	$1.77 \cdot 10^{-9}$	0.926	$1.14 \cdot 10^{-7}$	0.928	$1.04 \cdot 10^{-7}$	-	<u>ADHD-CDG</u> ⁵ , <u>CDG</u> ⁶ , <u>CDG</u> ⁸
rs9530773 (0/0/0)	13	78852243	T	G	0.674	0.689	0.935	$1.14 \cdot 10^{-8}$	0.938	$1.76 \cdot 10^{-5}$	0.933	$1.78 \cdot 10^{-6}$	-	<i>ADHD</i> ¹ , <u>CDG</u> ⁸
rs138696645 (4/4/4)	20	21154234	A	AAAG	0.644	0.659	0.937	$1.27 \cdot 10^{-8}$	0.926	$1.22 \cdot 10^{-7}$	0.940	$1.11 \cdot 10^{-5}$	<i>PLK1S1, KIZ, XRN2</i>	<u>CDG</u> ⁶ , <u>CDG</u> ⁸ , <i>Many</i> ⁹
rs227293 (0/0/0)	4	103623491	T	C	0.689	0.672	1.061	$2.57 \cdot 10^{-8}$	1.061	$7.02 \cdot 10^{-5}$	1.080	$1.08 \cdot 10^{-7}$	<i>MANBA</i>	<u>ADHD-CDG</u> ¹⁰ , <i>Blood</i> ¹¹
rs325506 (23/27/24)	5	104012303	C	G	0.441	0.428	1.064	$2.66 \cdot 10^{-8}$	1.074	$3.50 \cdot 10^{-7}$	1.070	$8.40 \cdot 10^{-7}$	-	<u>ASD-CDG</u> ¹² , <u>ADHD</u> ¹ , <u>ADHD- CDG</u> ⁵ , <u>CDG</u> ⁶ , <u>CDG</u> ⁸ , <i>Many</i> ¹³

SNP (#CS) – marker name and number of reported GWASs were this marker is in the 95% credible set in (FINEMAP/PAINTOR/CAVIARBF) according to <http://mulinlab.org/causaldb/>, please note that SNPs do not need to be genome-wide significant in those reported GWASs to be in the list of credible SNPs. SNPs representing novel shared loci for ASD and ADHD are highlighted in bold; CHR – chromosome; BP – Base pair position on the chromosome; A1 – effect allele; A2 – other allele; FRQ_{ca} – Frequency in the cases; FRQ_{co} – Frequency in the controls; OR – Odds ratio based on effect allele; P – *P* value for association results; GENES – protein coding genes and/or microRNAs in a LD region around lead SNP ($r^2 = 0.6$), in case no protein coding gene or microRNA is present in the region the nearest protein coding gene or microRNA within a 100kb window around the LD region is provided together with the distance in kb (if there is no gene present a “-“ will be shown); OTHER – previously reported associations with the lead SNP (underlined letters) or other SNPs (italic letters) in LD with the lead SNP ($r^2 = 0.6$), reported *P* values needed to be genome-wide significant to be listed. In case of the ASD and ADHD *P* values these are the *P* values in the original GWAS. Please note that the OR and the *P* values reported for the ADHD and ASD GWASs both times include the comorbid cases (i.e. in each of the two GWASs) as well as related individuals across studies. Markers highlighted in bold letters indicate previously unidentified associations with either of the two disorders (ADHD / ASD). ¹ADHD (PMID 30478444); ²Educational attainment (years of education, PMID 30038396), Intelligence (MTAG, PMID 29326435), Adventurousness (PMID 30643258), Feeling worry (neuroticism item; 29500382), Household income (PMID 31844048), Balding type 1 (PMID 30595370), Number of sexual partners (30643258); ³Alcohol consumption (PMIDs 30643258, 31358974, 30643251); ⁴Neuroticism (PMID 29942085), Worry (neuroticism item; PMID 29942085); ⁵Attention deficit hyperactivity disorder or cannabis use (PMID 30610198), ⁶Cross Disorder GWAS in the PGC (PMID 31835028); ⁷Number of sexual partners (PMID 30643258); ⁸Cross disorder GWAS for TS-ADHD-ASD (PMID 33714545), ⁹Fat-free mass (PMID 30593698), Appendicular lean mass (PMID 31761296), Height (PMIDs 30595370 and 25282103); ¹⁰Asthma and attention deficit hyperactivity disorder (PMID 31619474); ¹¹Blood protein levels (PMID 29875488); ¹²Autism and major depressive disorder (MTAG, PMID 30804558); ¹³Educational attainment (PMID 30038396), Life satisfaction (PMID 30643256), Well-being spectrum (multivariate analysis, PMIDs 30643256 and 29292387), Depressive symptoms (PMIDs 30643256 and 29292387), Neuroticism (PMID 29292387), Positive affect (PMID 30643256), Loneliness (PMID 31518406), Asthma and attention deficit hyperactivity disorder (PMID 31619474), Asthma and major depressive disorder (PMID 31619474), Insomnia (PMIDs 30804566 and 30804565), Risk-taking tendency (4-domain principal component model, PMID 30643258), BMI (PMIDs 31669095, 30595370, 30239722), Highest math class taken (PMID 30038396), Hand grip strength (PMID 29691431), Predicted visceral adipose tissue (PMID 31501611).

Table 2: Results of differentiating GWAS (ADHD vs ASD)

SNP (#CS)	CHR	BP	A1	A2	ADHDvsASD				ASD		ADHD		GENES	OTHER
					FRQ _{ADHD}	FRQ _{ASD}	OR	P	OR	P	OR	P		
rs13023832 (NA/NA/NA)	2	215219808	A	G	0.121	0.102	1.207	$4.28 \cdot 10^{-9}$	0.956	0.0484	1.122	$9.33 \cdot 10^{-8}$	SPAG16	ADHD ² , CDG ³
rs7821914 (3/5/5)	8	10805015	T	C	0.584	0.556	1.127	$4.58 \cdot 10^{-9}$	0.935	$1.86 \cdot 10^{-6}$	1.022	0.1113	XKR6	Neuroticism ⁴ , Many ⁵
rs147420422 (16/17/17)	2	104139422	CAT	C	0.529	0.502	1.118	$3.37 \cdot 10^{-8}$	0.947	$6.89 \cdot 10^{-5}$	1.036	0.0092	-	Neuroticism ⁶ , Many ⁷
rs3791033 ¹⁰ (6/7/6)	1	44134077	T	C	0.681	0.656	1.124	$3.98 \cdot 10^{-8}$	0.979	0.1407	1.095	$2.76 \cdot 10^{-10}$	PTPRF, KDM4A, ST3GAL3, MIR6079	EA ⁸ , ADHD ⁹ , Many ¹⁰
rs9379833 (58/59/58)	6	26207175	A	C	0.251	0.275	0.884	$4.51 \cdot 10^{-8}$	1.041	0.0102	0.949	0.0007	HIST1 ¹	EA ⁸ , Neuroticism ¹¹ , Height ¹² , Many ¹³

Results shown in the table are for three different GWAS: ADHDvsASD refers to our ADHD vs ASD GWAS described in the main text body of this manuscript, ADHD refers to results from the previously published GWAS on ADHD (PMID 30478444) and ASD refers to results from the previously published GWAS on ASD (PMID 30804558). Results from lookups in the open GWAS project database (<https://gwas.mrcieu.ac.uk/about/>, accessed Oct 14th 2020) are available in **Supplemental Table S2b** and as PheWAS plots in **Supplemental Figure S11**. SNP (#CS) – marker name and number of reported GWASs were this marker is in the 95% credible set in (FINEMAP/PAINTOR/CAVIARBF) according to <http://mulinlab.org/causaldb/>, please note that SNPs do not need to be genome-wide significant in those reported GWASs to be in the list of credible SNPs. If instead of a number "NA" appears this means the SNP has not been reported in a credible set before. SNPs highlighted in bold have not been identified in GWASs of ADHD and ASD before; CHR – chromosome; BP – Base pair position on the chromosome; A1 – effect allele; A2 – other allele; FRQ_{ADHD} – Frequency in the iPSYCH ADHD only cases; FRQ_{ASD} – Frequency in the iPSYCH ASD only cases; OR – Odds ratio based on effect allele; P – P value for association result; GENES – protein coding genes and/or microRNAs in a LD region around lead SNP ($r^2 = 0.6$), in case no protein coding gene or microRNA is present in the region the nearest protein coding gene or microRNA within a 100kb window around the LD region is provided together with the distance in kb (if there is no gene present a "--" will be shown);

OTHER – previously reported associations with the lead SNP or other SNPs in LD with the lead SNP ($r^2 = 0.6$), reported P values needed to be genome-wide significant to be listed. In case of the ASD and ADHD P values these are the P values in the original GWAS. Please note that the OR and the P values reported for the ADHD and ASD GWAS both times include the ADHD/ASD comorbid cases (i.e. in each of the two GWASs) as well as related individuals across studies. ¹Genes in the HIST1 region (PMID 12408966): *HIST1H1E*, *HIST1H2BD*, *HIST1H2BE*, *HIST1H4D*, *HIST1H3D*, *HIST1H2AD*, *HIST1H2BF*, *HIST1H4E*, *HIST1H2BG*, *HIST1H2AE*, *HIST1H3E*, *HIST1H1D*, *HIST1H4F*, *HIST1H4G*, *HIST1H3F*, *HIST1H2BH*. ²ADHD GWAS (PMID 30478444); ³Cross disorder GWAS (PMID 31835028); ⁴General factor of Neuroticism (PMID 30867560), Neuroticism (PMIDs 29255261 and 30643256), ⁵Remission after SSRI treatment in MDD or neuroticism (PMID 29559929), Gene alcohol interaction for blood pressure (PMID 29912962), White matter microstructure (PMID 31666681), Estimated glomerular filtration rate (PMID 31152163); ⁶Worry (neuroticism item; PMID 29942085), Feeling nervous (neuroticism item; PMID 29500382), Anxiety/tension (special factor of neuroticism; PMID 30867560); ⁷Smoking related phenotypes (PMIDs 30617275, 30643251, 30643258, 30595370, 30679032), Number of sexual partners (PMID 30643258), Age at first sexual intercourse (PMID 27089180), Reaction time (PMID 29844566), Risk-taking tendency (4-domain principal component model, PMID 30643258), General risk tolerance (MTAG, PMID 30643258), BMI (PMID 30239722), Pneumonia (PMID 28928442), Photoc sneeze reflex (PMID 27182965); ⁸Educational Attainment (PMID 30038396); ⁹ADHD GWAS (PMID 30478444), Attention deficit hyperactivity disorder or cannabis use (PMID 30610198), ¹⁰Highest math class taken (PMID 30038396), Self-reported math ability (PMID 30038396), Cognitive ability, years of educational attainment or schizophrenia (pleiotropy, PMID 31374203), Intelligence (PMIDs 29326435 and 29942086), Educational attainment (years of education, PMID 27225129), General cognitive ability (PMIDs 29844566 and 29186694), Smoking related phenotypes (PMID 30643251), Household income (MTAG, PMID 31844048), C-reactive protein levels (PMID 31900758), Menarche (age at onset, PMID 30595370), Red blood cell count (PMID 30595370), Height (PMID 30595370); ¹¹Worry too long after an embarrassing experience (neuroticism item; PMID 29500382); ¹²Height (PMID 31562340) ; ¹³Brain region volumes (PMID 31676860), Smoking related phenotypes (PMID 30643251), Strenuous sports or other exercises (PMID 29899525), Height (PMIDs 28552196, 28270201, 23563607, 20881960, 25282103, 25429064, 18391950, 18391951, 19343178, 31217584), Body fat percentage (PMID 30593698), Predicted visceral adipose tissue (PMID 31501611), Hip circumference adjusted for BMI (PMID 25673412), Hip circumference (PMID 25673412), Waist circumference (PMID 25673412), Waist circumference adjusted for BMI (joint analysis main effects and physical activity interaction; PMID 28448500), Waist circumference adjusted for body mass (PMID 28448500), Body fat distribution (leg fat ratio, PMID 30664634), Birth weight (PMIDs 27680694 and 31043758); ¹⁰rs7538463(A) allele from Table 1 is correlated with rs3791033(C) allele in this table, $r^2=0.1687$, $D'=0.8989$ (LDpair Tool at LDlink website, EUR reference).

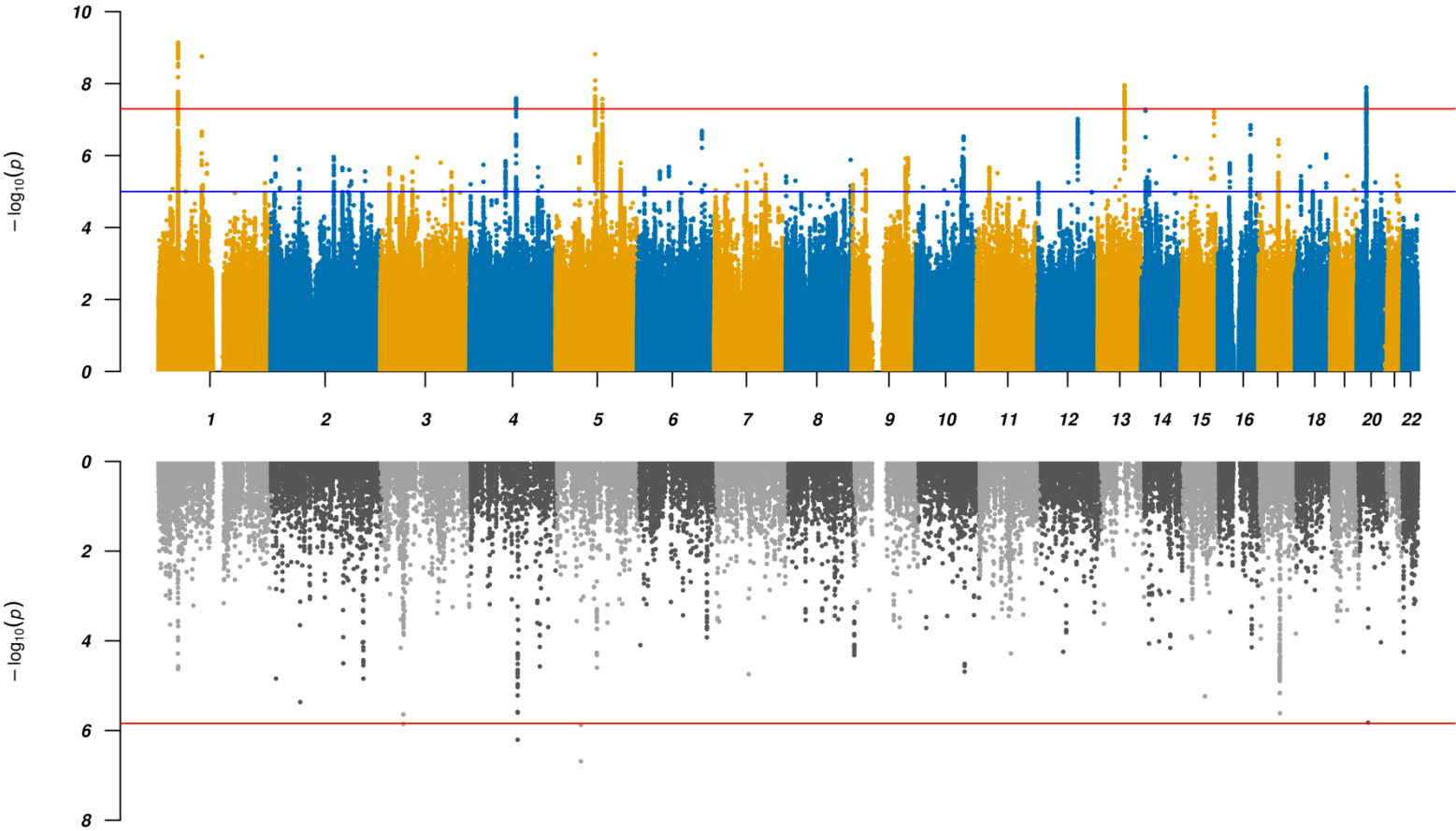
Table 3: Results of ADHD and ASD polygenic risk score analyses in the iPSYCH cohort using a leave-one-out analysis framework

Cases (coded as 1)	Comparison (coded as 0)	PRS _{ADHD}				PRS _{ASD}			
		OR	LCI	UCI	P	OR	LCI	UCI	P
ADHD-only	Controls	1.45	1.41	1.48	$1.3 \cdot 10^{-207}$	1.08	1.06	1.11	$7.5 \cdot 10^{-12}$
ASD-only	Controls	1.10	1.07	1.13	$3.1 \cdot 10^{-13}$	1.21	1.18	1.24	$1.2 \cdot 10^{-48}$
Comorbid	Controls	1.32	1.25	1.39	$2.8 \cdot 10^{-25}$	1.22	1.16	1.29	$3.5 \cdot 10^{-14}$
Comorbid	ADHD-only	0.92	0.88	0.97	0.0015	1.13	1.08	1.19	$4.7 \cdot 10^{-7}$
Comorbid	ASD-only	1.22	1.16	1.28	$6.4 \cdot 10^{-16}$	1.01	0.96	1.06	0.68
ASD-only	ADHD-only	0.76	0.74	0.78	$4.5 \cdot 10^{-79}$	1.12	1.09	1.15	$1.2 \cdot 10^{-15}$
ADHD+ID	ADHD-no-ID	0.97	0.88	1.06	0.46	0.94	0.86	1.03	0.19
ASD+ID	ASD-no-ID	1.03	0.93	1.12	0.58	0.89	0.81	0.97	0.0072

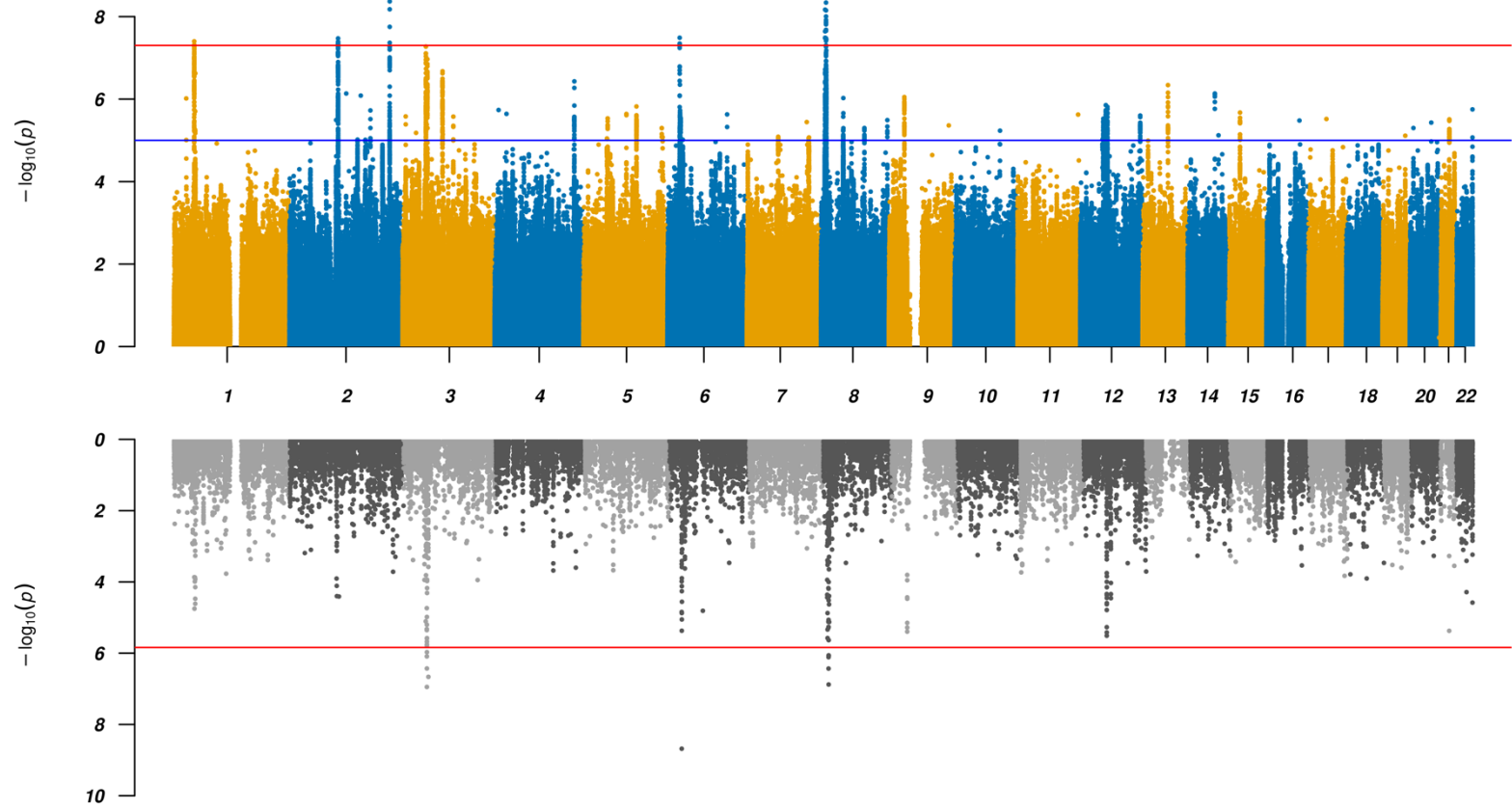
Results for per wave polygenic risk score analyses. PRS_{ADHD} – Analyses using a polygenic risk score trained on an ADHD phenotype, PRS_{ASD} – Analyses using a polygenic risk score trained on an ASD phenotype. Cases – group coded as 1 (cases) for the purpose of the analyses, Comparison – other group coded as 0 for the purpose of the analyses, OR – Odds ratio, LCI – lower boundary for 95% confidence interval, UCI – upper boundary for 95% confidence interval, P – P-value. Groups are as follows: ADHD-only – cases with ADHD diagnosis and without comorbid ASD diagnosis, ASD-only – cases with ASD diagnosis and without comorbid ADHD diagnosis, Comorbid – cases with comorbid ADHD and ASD diagnoses, Controls – individuals without ADHD and ASD diagnoses. P-values shown are without correction for multiple testing. Experiment-wide significant at 0.0042 (Bonferroni corrected for 2 x 6 tests). Additional secondary analyses also compare groups of individuals with ADHD or ASD with co-occurring mild intellectual disability (ADHD+ID and ASD+ID) to those without (ADHD-no-ID and ASD-no-ID).

Figure 1: Miami plots of GWAS and TWAS results.

A)

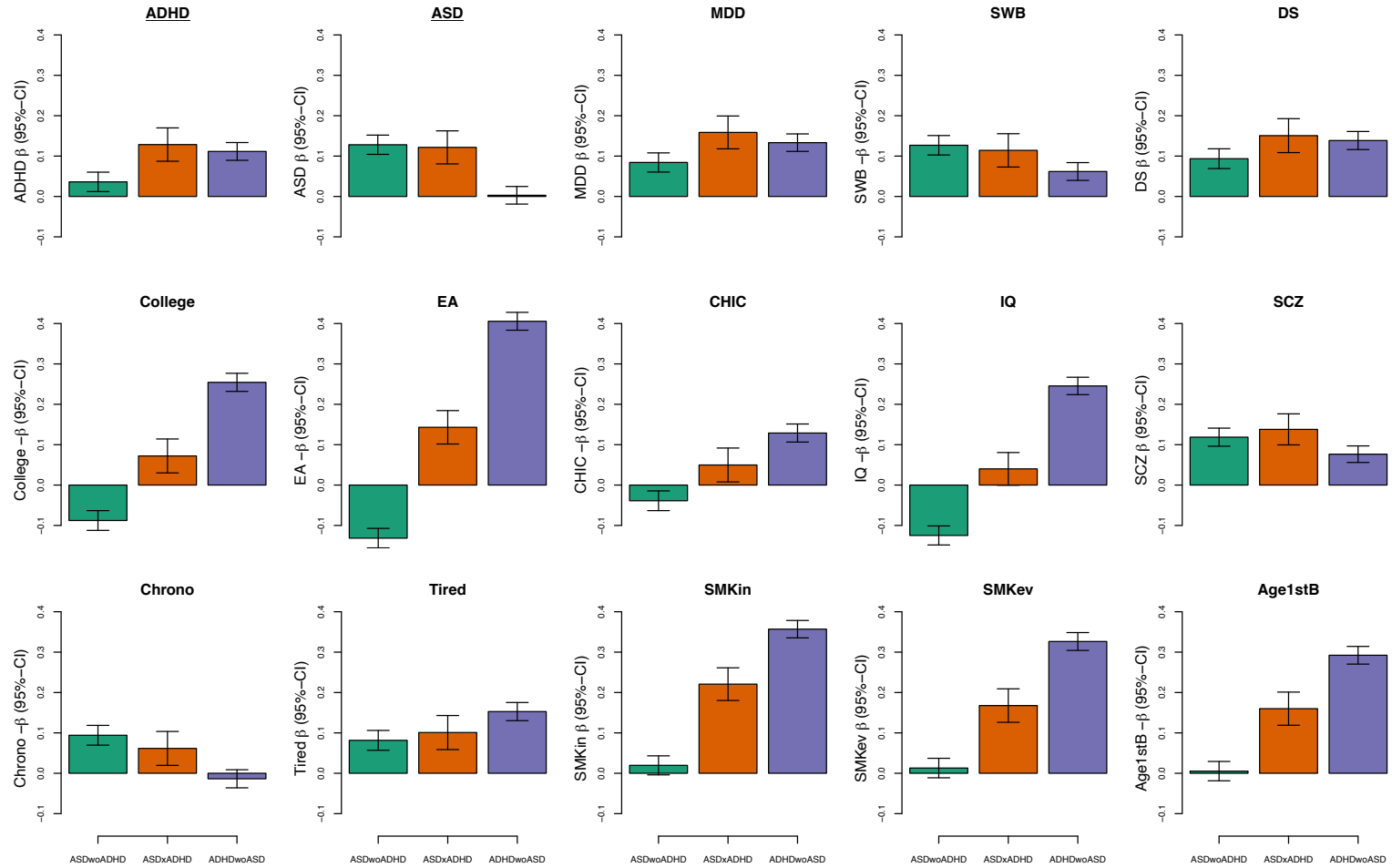


B)



Results for GWAS (top panels) and TWAS results for DLPFC transcripts (bottom panels) for (A) combined and (B) ADHD vs ASD GWAS. In the top panel a blue line in the Manhattan plot indicates a p-value of 1×10^{-5} , a red line a p-value of 5×10^{-8} (genome-wide significance). Each dot represents a tested SNP. In the bottom panel genes are represented by both gene expression and isoform expression (= features, represented by the dots). A red line indicates Bonferroni corrected genome-wide significance within analyses (combined or ADHD vs ASD; $p < 1.44 \times 10^{-6}$; corresponding to Bonferroni correction of all the 34,646 features).

Figure 2: Multivariate PRS analyses for 15 traits associated with ADHD and/or ASD.



Comparison of PRSs profiles across ADHD/ASD subtypes for 15 traits/phenotypes that have shown significant genetic correlation with ADHD and ASD in the past. Green bars represent ASD only cases, orange bars depict comorbid samples, and purple bars show average PRS for ADHD only cases. **ADHD** – attention-deficit/ hyperactivity disorder [PMID 20732625]; **ASD** – autism spectrum disorder [30804558 without the iPSYCH sample]; **MDD** – major depressive disorder [29700475 wo DK, wo 23am]; **SWB** – subjective well-being [27089181]; **DS** – depressive symptoms [27089181]; **College** – college completion [27046643]; **Edu** – educational attainment [30038396]; **CHIC** – childhood IQ [23358156]; **IQ** – IQ [29942086]; **SCZ** – schizophrenia [PGC3 woDK]; **Chrono** – chronotype [30696823]; **Tired** – self-reported tiredness [28194004]; **SMKos** – smoking initiation [30643251]; **SMKev** – ever smoker [30643258]; **Age1stB** – age of first birth [20418890].

Box 1 – Prioritized genes or transcripts that are (i) located in GWAS loci and/or are genome-wide significant in gene-wise analysis and (ii) Bonferroni significant in TWAS. Genes/transcripts showing increased imputed DLPFC expression in ADHD-ASD combined compared to controls are highlighted in red while those with decreased expression are blue. Genes/transcripts showing decreased imputed DLPFC expression in ASD compared to ADHD are highlighted in green while those with decreased expression in ADHD compared to ASD are purple.

Shared liability genes identified in the combined ADHD-ASD GWAS and TWAS:

Keratin 8 Pseudogene 46 (KRT8P46) (transcript, isoform 201): Located on chromosome 4 (**Supplemental Figure 3a**), *KRT8P46* is a pseudogene located in an intron of *MANBA*, a gene that has been previously associated with ADHD and asthma¹ as well as blood protein levels². Pseudogenes have recently been highlighted as regulators in health and disease^{3,4} amongst others through potential regulatory relationship with their parent genes. It is of note that another keratin 8 pseudogene (*KRT8P44*) is located in a region that has been identified to harbor a rare CNV associated with ADHD⁵.

Differentiating liability genes identified in the ADHD vs ASD GWAS and TWAS:

HIST1H2BD (transcript, isoform 201): Located in the cytogenetic band 6p22.2 as part of the histone gene cluster (more precisely the H1 histone family) (**Supplemental Figure 3b**). The gene (also known as *H2BC5*) encodes the Histone H2B type 1-D protein. Histone proteins in general are involved in the structure of chromatin in eukaryotic cells and play a central role in transcription regulation, DNA repair, DNA replication, and chromosomal stability. Little is known about the specific function of *HIST1H2BD*, however, deleterious de novo mutations in several histone modifying or interacting genes (albeit not including *HIST1H2BD*)⁶⁻⁸ as well as in core histone genes^{7,9} have been associated with autism and developmental delay with autistic features.

CAMKV (transcript, isoform 210): Located together with two other TWAS significant genes in a region on chromosome 3p21.31 (**Supplemental Figure 3c**), which has been reported to harbor CNVs in individuals with autism, intellectual disability and developmental delays¹⁰⁻¹³. The calmodulin kinase-like vesicle-associated (CaMKv) is a pseudokinase required for the activity-dependent maintenance of dendritic spines¹⁴.

RFT1 (transcript, isoform 204): Located on chromosome 3p21.1 (**Supplemental Figure 3d**) *RFT1* encodes an enzyme which catalyzes the translocation of the Man(5)GlcNAc (2)-PP-Dol intermediate from the cytoplasmic to the luminal side of the endoplasmic reticulum membrane in the pathway for the N-glycosylation of proteins. Mutations in *RFT1* cause recessive congenital disorder of glycosylation type 1N (CDG1N) [[MIM:612015](https://omim.org/entry/612015)], which presents with a wide variety of clinical features, including severe developmental delay, hypotonia, dysmorphic features and epilepsy.

FAMI67A (transcript, isoform 201): Located together with three other TWAS significant genes in the identified GWAS locus on chromosome 8 (**Supplemental Figure 3e**). Using data from a subset of our iPSYCH ASD samples, a previous study identified two differentially methylated positions (DMPs) in the same region that showed association with polygenic risk for ASD¹⁵. One of the DMPs was annotated to the Family With Sequence Similarity 167 Member A gene (*FAMI67A*), while the other was annotated to *RP1L1*.

1. Zhu, Z. *et al.* Shared genetics of asthma and mental health disorders: a large-scale genome-wide cross-trait analysis. *Eur Respir J* **54**(2019).
2. Sun, B.B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73-79 (2018).
3. Pink, R.C. *et al.* Pseudogenes: pseudo-functional or key regulators in health and disease? *RNA* **17**, 792-8 (2011).
4. Cheetham, S.W., Faulkner, G.J. & Dinger, M.E. Overcoming challenges and dogmas to understand the functions of pseudogenes. *Nat Rev Genet* **21**, 191-201 (2020).
5. Jarick, I. *et al.* Genome-wide analysis of rare copy number variations reveals PARK2 as a candidate gene for attention-deficit/hyperactivity disorder. *Mol Psychiatry* **19**, 115-21 (2014).
6. Satterstrom, F.K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584 e23 (2020).
7. Duffney, L.J. *et al.* Epigenetics and autism spectrum disorder: A report of an autism case with mutation in H1 linker histone HIST1H1E and literature review. *Am J Med Genet B Neuropsychiatr Genet* **177**, 426-433 (2018).
8. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209-15 (2014).
9. Bryant, L. *et al.* Histone H3.3 beyond cancer: Germline mutations in Histone 3 Family 3A and 3B cause a previously unidentified neurodegenerative disorder in 46 patients. *Sci Adv* **6**(2020).
10. Stuart, S.W., King, C.H. & Pai, G.S. Autism spectrum disorder, Klinefelter syndrome, and chromosome 3p21.31 duplication: a case report. *MedGenMed* **9**, 60 (2007).
11. Haldeman-Englert, C.R. *et al.* A 3.1-Mb microdeletion of 3p21.31 associated with cortical blindness, cleft lip, CNS abnormalities, and developmental delay. *Eur J Med Genet* **52**, 265-8 (2009).
12. Eto, K. *et al.* Microdeletions of 3p21.31 characterized by developmental delay, distinctive features, elevated serum creatine kinase levels, and white matter involvement. *Am J Med Genet A* **161A**, 3049-56 (2013).
13. Battaglia, A. *et al.* Confirmation of chromosomal microarray as a first-tier clinical diagnostic test for individuals with developmental delay, intellectual disability, autism spectrum disorders and dysmorphic features. *Eur J Paediatr Neurol* **17**, 589-99 (2013).
14. Liang, Z. *et al.* The pseudokinase CaMKv is required for the activity-dependent maintenance of dendritic spines. *Nat Commun* **7**, 13282 (2016).
15. Hannon, E. *et al.* Elevated polygenic burden for autism is associated with differential DNA methylation at birth. *Genome Med* **10**, 19 (2018).