ARTICLES
https://doi.org/10.1038/s41588-022-01171-3

nature genetics

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

Manuel Mattheisen[1,2,3,38] ✉, Jakob Grove[1,4,5,6,38], Thomas D. Als[1,4,5,38], Joanna Martin[7,38], Georgios Voloudakis[8,9,10], Sandra Meier[1,2], Ditte Demontis[1,4,5], Jaroslav Bendl[8,9,10], Raymond Walters[11,12], Caitlin E. Carey[11,12], Anders Rosengren[4,13], Nora I. Strom[1,3,14], Mads Engel Hauberg[8,9,10], Biao Zeng[8,9,10], Gabriel Hoffman[8,9,10], Wen Zhang[8,9,10], Jonas Bybjerg-Grauholm[4,15], Marie Bækvad-Hansen[4,15], Esben Agerbo[4,16,17], Bru Cormand[18,19,20,21], Merete Nordentoft[4,22,23,24], Thomas Werge[4,13,22,25], Ole Mors[4,26], David M. Hougaard[4,15], Joseph D. Buxbaum[10,27,28,29], Stephen V. Faraone[30,31], Barbara Franke[32,33], Søren Dalsgaard[16], Preben B. Mortensen[4,5,16,17], 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] and Anders D. Børglum[1,4,5] ✉

Attention-deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) are highly heritable neurodevelopmental conditions, with considerable overlap in their genetic etiology. We dissected their shared and distinct genetic etiology by cross-disorder analyses of large datasets. We identified seven loci shared by the disorders and five loci differentiating them. All five differentiating loci showed opposite allelic directions in the two disorders and significant associations with other traits, including educational attainment, neuroticism and regional brain volume. Integration with brain transcriptome data enabled us to identify and prioritize several significantly associated genes. The shared genomic fraction contributing to both disorders was strongly correlated with other psychiatric phenotypes, whereas the differentiating portion was correlated most strongly with cognitive traits. Additional analyses revealed that individuals diagnosed with both ASD and ADHD were double-loaded with genetic predispositions for both disorders and showed distinctive patterns of genetic association with other traits compared with the ASD-only and ADHD-only subgroups. These results provide insights into the biological foundation of the development of one or both conditions and of the factors driving psychopathology discriminatively toward either ADHD or ASD.

ADHD and ASD are among the most common neurodevelopmental disorders in children and often persist throughout adulthood[1]. ADHD and ASD are both highly heritable (60–93%)[2–4], and the mode of their inheritance is complex and polygenic. Despite high family-based heritability estimates, genome-wide association studies (GWAS) have only recently identified common variants robustly associated with each disorder[5–7]. Although the two disorders differ from one another with regard to core clinical symptoms, genetic studies have demonstrated substantial overlap between them, with a genetic correlation (r_G) from common variation of 0.36 (refs. [5,8]) and substantial sharing of rare genetic risk variants including large copy number variants[9] and protein-truncating variants[10]. These findings are consistent with clinical and epidemiological evidence showing overlap in phenotypic features[11], high comorbidity rates between ASD and ADHD[12,13] in both females and males[14], and familial coaggregation of the disorders, with increased risk of ADHD among relatives of ASD probands (odds ratios (ORs) of 17.8 for monozygotic twins, 4.3 for dizygotic twins, 4.6 for full siblings and 1.6 for full cousins)[15]. Identification of the genetic components that are shared or distinct between the two disorders may provide insights into the underlying biology and potentially provide information on subclassification, course and treatment.

Here we use large collections of genotyped samples from individuals with 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 individuals diagnosed with both disorders (comorbid cases) or individuals with just one of them (ASD-only and 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), including a total of 34,462 cases and 41,201 controls on 8.9 million single nucleotide polymorphism (SNP) allele dosages imputed from 1000 Genomes phase 3 (ref. [16]). Using linkage

A full list of affiliations appears at the end of the paper.

disequilibrium (LD) score regression (LDSC)[17], we found evidence for a strong polygenic signal 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, Fig. 1 and Supplementary Fig. 1). All but one of these loci showed associations with both of the disorders separately at *P* values below 1×10^{-4}; this exception was genome-wide significant in ADHD and had a *P* value of 0.009 in ASD. Overall, the findings corroborated previous results[8,18] but included two loci that had not been identified before as shared between ADHD and ASD. The new shared associations were located in a highly pleiotropic multigene locus on chromosome 1 (rs7538463) and on chromosome 4 (rs227293) in the gene encoding β-mannosidase (*MANBA*). Mutations in *MANBA* are associated with β-mannosidosis, a lysosomal storage disease that has a wide spectrum of neurological phenotypes, including intellectual disability (ID), hearing loss and speech impairment[19]. More details on the seven loci can be found in Table 1, and results of lookups in the OpenGWAS project database (https://gwas.mrcieu.ac.uk/about/) and comparisons with previous cross-disorder studies are available in the Supplementary Note and Supplementary Data 1 and 2, and as PheWAS plots in Supplementary Fig. 2.

To identify and prioritize putative causal shared genes, we performed a transcriptome-wide association study (TWAS), imputing genetically regulated gene expression using EpiXcan[20] and expression data from the PsychENCODE Consortium[21] for genes and isoforms detected in 924 samples from the dorsolateral prefrontal cortex (DLPFC). Applying a conservative significance threshold ($P<1.44\times10^{-6}$; corresponding to Bonferroni correction of all 34,646 genes and isoforms tested), we identified five genes or isoforms showing significant differential expression between the combined case group and controls and 177 genes or isoforms that were significant at a false discovery rate (FDR) < 0.05 (Fig. 1 and Supplementary Data 4). One of the five Bonferroni-significant transcripts, KRT8P46-201, was located in the chromosome 4 GWAS locus in an intron of *MANBA*, which was among the genes with FDR < 0.05 (Supplementary Fig. 3a). The other four top findings were the two genes *MOCS2* and *CCDC71* or their isoforms, which were not located in any of the identified GWAS loci and thus represent additional candidate genes for shared ADHD and ASD risk.

Gene-based analysis using MAGMA v.1.08 (refs. [22,23]) largely corroborated the results from the GWAS and TWAS, highlighting, for example, *MANBA* (Supplementary Fig. 4a and Supplementary Data 5). Furthermore, two of the significant genes—sortilin-related VPS10 domain containing receptor 3 (*SORCS3*) and dual specificity phosphatase 6 (*DUSP6*)—were 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 (ADHDvsASD GWAS). Using LDSC[17], we found an intercept of 0.9863 and a SNP heritability of 0.4468 on the observed scale, the latter indicating that a substantial part of the variance in the phenotypic representation differentiating the two case groups can be explained by common variants (see Supplementary Note for more details). Five genome-wide significant loci were identified, three of which had not previously been identified in GWAS of either of the two disorders separately (although one has been reported as an ADHD–ASD differentiating locus[24]). All loci have been reported in related disorders and, remarkably, all but one are associated with cognitive abilities and/or neuroticism or neuroticism items (Table 2, Fig. 1 and Supplementary Data 2 and 7). The lead variants all show opposite directions of effects in the two disorders.

Two of the five lead SNPs have previously been found to be associated with educational attainment[25]. For the first SNP (rs3791033 on chromosome 1; $P=4.65\times10^{-23}$), the C allele confers an increased risk for ASD and increased cognitive performance, whereas the ADHD risk allele (T) is associated with decreased performance. Similarly, for the second SNP (rs9379833 on chromosome 6; $P=2.26\times10^{-8}$), the A allele confers an increased risk for ASD and increased cognitive performance, whereas the ADHD risk allele (C) is associated with decreased performance. Notably, this SNP (rs9379833) is located in the large histone gene cluster *HIST1* (ref. [26]) and has also been reported to be associated with regional brain volume, specifically that of the left globus pallidus[27] ($P=2.95\times10^{-8}$; the C allele confers an increased risk for ADHD and decreased volume, whereas the ASD risk allele (A) is associated with increased volume). It is also of note that the lead SNP on chromosome 8 (rs7821914) is associated with neuroticism[28] ($P=9.46\times10^{-21}$). For this SNP, the effect allele (C) in the neuroticism GWAS leads to an increased risk of ASD and a decreased risk of ADHD. Two additional lead SNPs were in LD ($r^2>0.6$) with SNPs that have previously been identified in neuroticism or one of its sub-dimensions (rs147420422 and rs9379833; Table 2). Results from additional lookups in the OpenGWAS project database (https://gwas.mrcieu.ac.uk/about/) are available in Supplementary Data 7 and as PheWAS plots in Supplementary Fig. 6.

TWAS using EpiXcan identified 11 Bonferroni-significant genes and/or isoforms and 96 significant transcripts at FDR < 0.05 with different imputed expression in DLPFC between ADHD and ASD cases (Fig. 1 and Supplementary Data 4). The *HIST1H2BD-201* isoform located in the chromosome 6 (*HIST1*) GWAS locus showed the strongest association ($P=2.08\times10^{-9}$), with higher expression in ADHD compared with ASD cases (Supplementary Fig. 3b). The other genes and/or isoforms showed associations that were orders of magnitude less significant, with *HIST1H2BD-201* as the top-ranking causal candidate in the locus. The remaining ten Bonferroni-significant genes and/or isoforms were located in the chromosome 8 GWAS locus or in two loci on chromosome 3 (Supplementary Fig. 3c,d, respectively), where all except the gene encoding the TRAF-interacting protein (*TRAIP*) were also genome-wide significant in gene-based analysis using MAGMA (Supplementary Fig. 4b and Supplementary Data 5).

Genetic correlations with other traits. To examine the polygenic architecture of the identified shared and differentiating genetic risk for the two 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 Hub[29] and LDSC[30]. Among the 258 previously reported GWAS, 30 (combined GWAS) and 32 (ADHDvsASD) traits showed significant correlations after Bonferroni correction for multiple testing (Supplementary Data 6 and Supplementary Fig. 7). The strongest correlations for the liability-differentiating ADHDvsASD GWAS were observed for cognitive traits including years of schooling ($r_G=-0.669$, correlation p-value (P_{corr}) $=3.68\times10^{-85}$) and childhood IQ ($r_G=-0.609$, $P_{corr}=2.78\times10^{-10}$), whereas the strongest correlations for the combined GWAS were with traits including depressive symptoms ($r_G=0.506$, $P_{corr}=2.08\times10^{-19}$) and the PGC cross-disorder GWAS ($r_G=0.433$, $P_{corr}=5.30\times10^{-25}$).

Tissue and cell-type enrichment analyses. We next tested whether genetic associations of shared and differentiating liabilities were enriched with respect to the transcriptomic profiles of human tissues. We found significant enrichment for shared liability in several brain tissues, most notably for the basal ganglia (Supplementary Fig. 8). Cell-type enrichment analyses revealed experiment-wide significant association (across all datasets tested) of the red nucleus (Supplementary Fig. 9c). Associations that were significant within

Table 1 | Results of combined GWAS (ADHD or ASD)

							Meta		ASD		ADHD			
SNP (CS)	CHR	BP	A1	A2	FRQ_{ca}	FRQ_{co}	OR	*P*	OR	*P*	OR	*P*	Genes	Other
rs7538463 (2/2/2)	1	44196416	A	T	0.707	0.721	0.928	7.26×10^{-10}	0.961	0.0091	0.914	1.00×10^{-9}	*PTPRF*, *KDM4A*, *ST3GAL3*, *MIR6079*	ADHD[a], *Many*[b]
rs4916723 (5/5/5)	5	87854395	A	C	0.558	0.573	0.935	1.52×10^{-9}	0.935	1.92×10^{-6}	0.925	1.81×10^{-8}	*MIR9-2* (58.3)	ALC[c], Neuroticism[d], ADHD[a], ADHD-CDG[e], CDG[f], *sexual partners*[g], *CDG*[h]
rs2391769 (2/2/2)	1	96978961	A	G	0.351	0.364	0.934	1.77×10^{-9}	0.926	1.14×10^{-7}	0.928	1.04×10^{-7}	–	ADHD-CDG[e], CDG[f], *CDG*[h]
rs9530773 (0/0/0)	13	78852243	T	G	0.674	0.689	0.935	1.14×10^{-8}	0.938	1.76×10^{-5}	0.933	1.78×10^{-6}	–	*ADHD*[a], *CDG*[h]
rs138696645 (4/4/4)	20	21154234	A	AAAG	0.644	0.659	0.937	1.27×10^{-8}	0.926	1.22×10^{-7}	0.940	1.11×10^{-5}	*PLK1S1*, *KIZ*, *XRN2*	*CDG*[f], *CDG*[h], *Many*[i]
rs227293 (0/0/0)	4	103623491	T	C	0.689	0.672	1.061	2.57×10^{-8}	1.061	7.02×10^{-5}	1.080	1.08×10^{-7}	*MANBA*	ADHD-CDG[j], Blood[k]
rs325506 (23/27/24)	5	104012303	C	G	0.441	0.428	1.064	2.66×10^{-8}	1.074	3.50×10^{-7}	1.070	8.40×10^{-7}	–	ASD-CDG[l], *ADHD*[a], ADHD-CDG[e], CDG[f], CDG[h], *Many*[m]

Results shown in the table are for three different GWAS. Meta refers to our combined ADHD or ASD GWAS described in the main text body of this manuscript, ADHD refers to results from the previously published GWAS on ADHD (PubMed Indexing Number (PMID) 30478444) and ASD refers to results from the previously published GWAS on ASD (PMID 30804558). Results from lookups in the OpenGWAS project database (https://gwas.mrcieu.ac.uk/about/, accessed 14 October 2020) are available in Supplementary Data 1 and as PheWAS plots in Supplementary Fig. 2. SNP (CS) denotes the marker name and number of reported GWAS where this marker is in the 95% credible set in FINEMAP/PAINTOR/CAVIARBF according to http://mulinlab.org/causaldb/; note that SNPs do not need to be genome-wide significant in those reported GWAS to be in the list of credible SNPs. SNPs representing new 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. ORs are based on the effect allele; *P* values are for association results (two-sided from logistic regression). 'Genes' indicates protein-coding genes and/or microRNAs in a LD region around the lead SNP ($r^2=0.6$) or, in cases where no protein-coding gene or microRNA is present in the region, the nearest protein-coding gene or microRNA within a 100-kb window around the LD region is provided together with the distance in kb (if there is no gene present, '–' is shown). 'Other' indicates previously reported associations with the lead SNP (underlined) or other SNPs (italics) in LD with the lead SNP ($r^2=0.6$); reported *P* values needed to be genome-wide significant to be listed. For the ASD and ADHD *P* values, these are the *P* values in the original GWAS. The ORs and *P* values reported for the ADHD and ASD GWAS include the comorbid cases (that is, in each of the two GWAS) as well as related individuals across studies. [a]ADHD (PMID 30478444). [b]Cross-disorder GWAS in the PGC (PMID 31835028), educational attainment (years of education; PMID 30038396), intelligence (multi-trait analysis of GWAS (MTAG); PMID 29326435), adventurousness (PMID 30643258), feeling worry (neuroticism item; 29500382), household income (PMID 31844048), balding type 1 (PMID 30595370), number of sexual partners (PMID 30643258). [c]Alcohol consumption (PMIDs 30643258, 31358974 and 30643251). [d]Neuroticism (PMID 29942085), worry (neuroticism item; PMID 29942085), [e]ADHD or cannabis use (PMID 30610198). [f]Cross-disorder GWAS in the PGC (PMID 31835028). [g]Number of sexual partners (PMID 30643258). [h]Cross-disorder GWAS for TS-ADHD-ASD (PMID 33714545), [i]Fat-free mass (PMID 30593698), appendicular lean mass (PMID 31761296), height (PMIDs 30595370 and 25282103). [j]Asthma and ADHD (PMID 31619474). [k]Blood protein levels (PMID 29875488). [l]Autism and major depressive disorder (MTAG; PMID 30804558). [m]Educational attainment (PMID 30038396), life satisfaction (PMID 30643256), well-being spectrum (multivariate analysis; PMIDs 30643256, 29292387), depressive symptoms (PMIDs 30643256 and 29292387), neuroticism (PMID 29292387), positive affect (PMID 30643256), loneliness (PMID 31518406), asthma and ADHD (PMID 31619474), asthma and major depressive disorder (PMID 31619474), insomnia (PMIDs 30804566 and 30804565), risk-taking tendency (four-domain PC model; PMID 30643258), body mass index (BMI) (PMIDs 31669095, 30595370 and 30239722), highest math class taken (PMID 30038396), hand grip strength (PMID 29691431), predicted visceral adipose tissue (PMID 31501611).

one of the three tested datasets individually, but not overall, were observed for several cell types, including, for example, dopaminergic and GABAergic neurons. For the disorder-differentiating analysis (ADHDvsASD), we observed no significant association with tissues or specific cell types after correction for multiple testing (Supplementary Figs. 9 and 10). We also intersected our genetic associations with a recent multiomics single-cell epigenetic catalog of the human brain[31]. Here, both the combined and differentiating GWAS results showed significant enrichment for several neuronal cell populations (Supplementary Fig. 11 and Supplementary Data 8), including excitatory and inhibitory neurons. The only difference in terms of significant associations between the combined and differentiating GWAS was seen for oligodendrocytes (which were not significant in the combined GWAS but were significant in the ADHDvsASD GWAS). Whereas aberrant myelination by oligodendrocytes resulting in disruption of white matter development has previously been reported in both ASD and ADHD[32,33], the degree of severity of this alteration might be a distinct pathophysiological factor[34].

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 showed, as expected, a significant association of the ASD-only subgroup with PRS for ASD ($P=6.89\times10^{-26}$) and of the ADHD-only subgroup with PRS for ADHD ($P=3.29\times10^{-23}$; Fig. 2). Both scores were trained with PGC-only GWAS results[5,35]. Strikingly, the ASD-PRS load of comorbid ASD+ADHD cases was similar to that of ASD-only cases ($P=0.77$); likewise, the ADHD-PRS load of the comorbid subgroup was similar to that of ADHD-only cases ($P=0.44$; Fig. 2), demonstrating that comorbid cases carry a load of both ADHD and ASD polygenic scores that are similar to the loads carried by the single-disorder cases of their respective disorder. In other words, comorbid cases are double-burdened with both ASD-PRS and ADHD-PRS. By contrast, the ASD-PRS load of ADHD-only cases was not different from that of controls ($P=0.79$), and the ADHD-PRS was only slightly increased in ASD-only cases compared with controls ($P=3.26\times10^{-3}$; Fig. 2).

Our leave-one-out framework analysis (including only the iPSYCH data in the training GWAS) showed similar results (Table 3). In this analysis, the ASD-PRS loads in ADHD-only cases and ASD-only cases were increased compared with controls. Furthermore, secondary analysis in the leave-one-out framework suggested that ADHD cases with ($n=625$) and without ($n=11{,}339$) mild ID 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 with those without mild ID ($n=8{,}681$) but did not differ in terms of PRS_{ADHD} (Table 3).

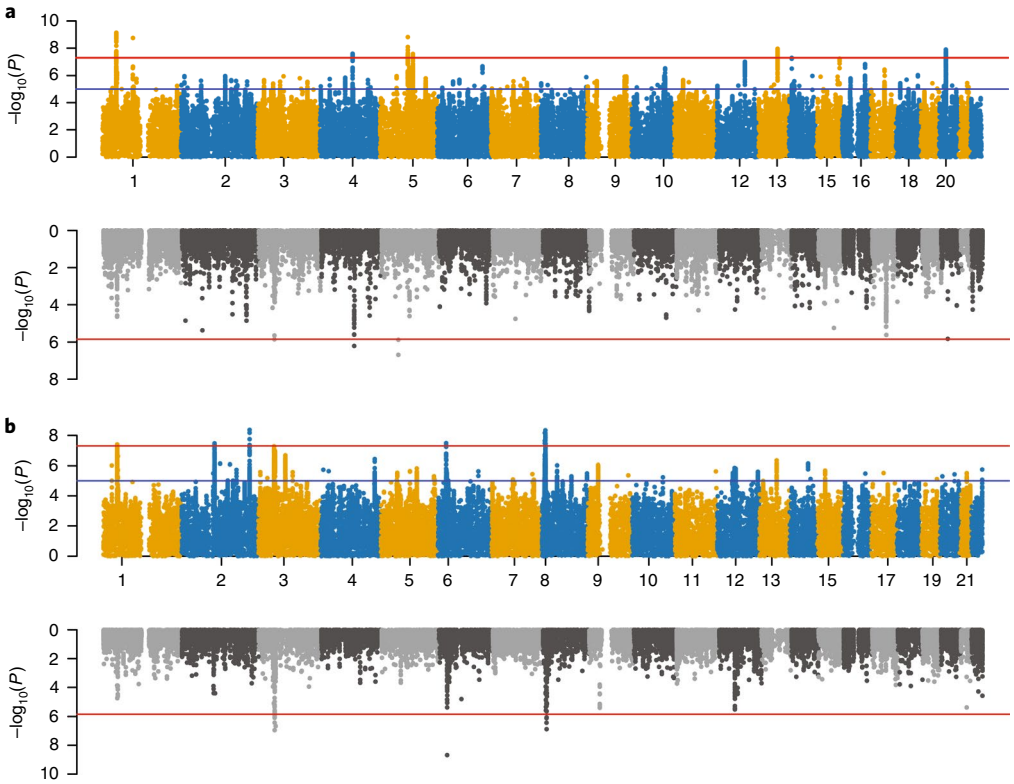

Fig. 1 | Manhattan plots for GWAS and TWAS results. a,**b**, Results for GWAS (top panels) and TWAS for DLPFC transcripts (bottom panels) for combined (**a**) and ADHD versus ASD (**b**) analyses. In the top panel, the blue line in the Manhattan plot indicates a *P* value of 1×10^{-5} and the red line indicates 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 imputed gene expression and isoform expression (features, represented by dots); two-tailed *P* values were derived from *z* scores (Wald statistic) of the gene–trait association. The red line indicates Bonferroni-corrected genome-wide significance within analyses (combined or ADHD versus ASD; $P<1.44\times10^{-6}$; corresponding to Bonferroni correction of all 34,646 features). We implemented an imputation R^2 filter (pred_perf_r2) of 0.01 in this study, which means that at least 10% of the variance in expression of each gene could be explained by *cis*-heritability. See also the results in Supplementary Data 4.

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[5,6,36]. Although PRS for schizophrenia and depression (and genetically related phenotypes) did not show substantially different loads across the subgroups, other traits showed compelling differences (Fig. 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. An item-level analysis of neuroticism also revealed specific patterns of associations across the subgroups (Supplementary Fig. 12). On average, ADHD-only cases showed much stronger association than ASD-only cases with items belonging to the depressed affect cluster (for example, the MOOD item) compared with the worry cluster. For comorbid cases, a distinct pattern was observed, with PRS loads either ranking between those of the ADHD-only and ASD-only cases (for example, for the MOOD item) or even exceeding those of the two single-disorder groups (for example, for the GUILT item).

In summary, we observed a genetic architecture of comorbid cases that presents as clearly distinct 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 distinguish them from the single-disorder cases, typically by carrying an intermediate load level but in some cases a load similar to 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 cases[5]. Here, we investigated the correlations across diagnostic subgroups of the disorders in the iPSYCH sample using genome-wide complex trait analysis genomic relatedness matrix restricted maximum likelihood (GCTA-GREML)[37]. For ASD and ADHD overall, we found $r_G=0.497$ (s.e.$=0.054$, $P=7.8\times10^{-19}$). Excluding the comorbid cases reduced the correlation to $r_G=0.397$ (s.e.$=0.056$, $P=6.3\times10^{-12}$). After excluding cases with ID, the correlations between ASD and ADHD were even stronger: $r_G=0.523$ (s.e.$=0.054$, $P=6.5\times10^{-21}$) and $r_G=0.425$ (s.e.$=0.056$, $P=1.7\times10^{-13}$) with and without comorbid cases, respectively (Supplementary Data 9 and Supplementary Fig. 13).

Correlations between ADHD and ICD-10 diagnostic subcategories of childhood autism (F84.0), atypical autism (F84.1), Asperger's syndrome (F84.5) and other/unspecified pervasive developmental disorders (other PDDs, F84.8–9) were similar to those for the ASD group overall, albeit with generally higher estimates for the groups with other PDDs and Asperger's syndrome (Supplementary Data 9 and Supplementary Fig. 14).

Genetic liability in comorbid cases. Guided by our results from the previously described analyses, we also performed a GWAS of the comorbid cases. Despite the small sample size (2,304 cases), we identified a genome-wide significant locus on chromosome 6 (rs1321614, $P=3.54\times10^{-9}$, OR$=0.8190$, minor allele frequency

Table 2 | Results of differentiating GWAS (ADHDvsASD)

							ADHDvsASD			ASD		ADHD		
SNP (CS)	CHR	BP	A1	A2	FRQ_{ADHD}	FRQ_{ASD}	OR	*P*	OR	*P*	OR	*P*	Genes	Other
rs13023832 (NA/NA/NA)	2	215219808	A	G	0.121	0.102	1.207	4.28×10^{-9}	0.956	0.0484	1.122	9.33×10^{-8}	*SPAG16*	ADHD[b], CDG[c]
rs7821914 (3/5/5)	8	10805015	T	C	0.584	0.556	1.127	4.58×10^{-9}	0.935	1.86×10^{-6}	1.022	0.1113	*XKR6*	Neuroticism[d], many[e]
rs147420422 (16/17/17)	2	104139422	CAT	C	0.529	0.502	1.118	3.37×10^{-8}	0.947	6.89×10^{-5}	1.036	0.0092	-	Neuroticism[f], many[g]
rs3791033[n] (6/7/6)	1	44134077	T	C	0.681	0.656	1.124	3.98×10^{-8}	0.979	0.1407	1.095	2.76×10^{-10}	*PTPRF, KDM4A, ST3GAL3, MIR6079*	EA[h], ADHD[i], many[j]
rs9379833 (58/59/58)	6	26207175	A	C	0.251	0.275	0.884	4.51×10^{-8}	1.041	0.0102	0.949	0.0007	HIST1[a]	
		EA[h], neuroticism[k], height[l], many[m]												

Results are for three different GWAS. ADHDvsASD refers to our ADHD versus 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 OpenGWAS project database (https://gwas.mrcieu.ac.uk/about/, accessed 14 October 2020) are available in Supplementary Data 7 and as PheWAS plots in Supplementary Fig. 11. SNP (CS) denotes the marker name and number of reported GWAS where this marker is in the 95% credible set in FINEMAP/PAINTOR/CAVIARBF according to http://mulinlab.org/causaldb/; note that SNPs do not need to be genome-wide significant in the reported GWAS to be in the list of credible SNPs. 'NA' indicates that the SNP has not been reported in a credible set before. SNPs highlighted in bold have not been identified in GWAS of ADHD and ASD before. FRQ_{ADHD}, frequency in iPSYCH ADHD-only cases; FRQ_{ASD}, frequency in iPSYCH-ASD-only cases. ORs are based on the effect allele; *P* values are for association results (two-sided from logistic regression). 'Genes' refers to protein-coding genes and/or microRNAs in an LD region around the lead SNP ($r^2=0.6$); in cases where no protein-coding gene or microRNA is present in the region, the nearest protein-coding gene or microRNA within a 100-kb window around the LD region is provided, together with the distance in kb (if there is no gene present, '–' is shown). 'Other' refers to 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. For ASD and ADHD, *P* values are those from the original GWAS. The ORs and *P* values reported for the ADHD and ASD GWAS include the ADHD–ASD comorbid cases (that is, in each of the two GWAS), as well as related individuals across studies. [a]Genes in the HIST1 region (PMID 12408966): *HIST1H1E, HIST1H2BD, HIST1H2BE, HIST1H4D, HIST1H3D, HIST1H2AD, HIST1H2BF, HIST1H4E, HIST1H2BG, HIST1H2AE, HIST1H3E, HIST1H1D, HIST1H4F, HIST1H4G, HIST1H3F, HIST1H2BH*. [b]ADHD GWAS (PMID 30478444). [c]Cross-disorder GWAS (PMID 31835028). [d]General factor of neuroticism (PMID 30867560), neuroticism (PMIDs 29255261 and 30643256). [e]Remission 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). [f]Worry (neuroticism item; PMID 29942085), feeling nervous (neuroticism item; PMID 29500382), anxiety/tension (special factor of neuroticism; PMID 30867560). [g]Smoking-related phenotypes (PMIDs 30617275, 30643251, 30643258, 30595370 and 30679032), number of sexual partners (PMID 30643258), age at first sexual intercourse (PMID 27089180), reaction time (PMID 29844566), risk-taking tendency (four-domain PC model; PMID 30643258), general risk tolerance (MTAG; PMID 30643258), BMI (PMID 30239722), pneumonia (PMID 28928442), photic sneeze reflex (PMID 27182965). [h]*Educational attainment* (PMID 30038396). [i]ADHD GWAS (PMID 30478444), ADHD or cannabis use (PMID 30610198). [j]*Highest 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), nenarche (age at onset; PMID 30595370), red blood cell count (PMID 30595370), height (PMID 30595370). [k]Worry too long after an embarrassing experience (neuroticism item; PMID 29500382). [l]Height (PMID 31562340). [m]Brain 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). [n]rs7538463(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).

(MAF) = 0.47 for the T allele). The lead SNP showed no association in the overall combined (ADHD+ASD) GWAS ($P=0.0261$), the differentiating GWAS ($P=0.2883$) or GWAS of the ADHD-only and ASD-only cases ($P=0.7721$ and $P=0.0086$, respectively). The liability scale SNP heritability for the GWAS using GCTA was 0.0557 (s.e. = 0.0088). Please see Supplementary Note for more information.

Discussion

This study dissects the genetic architecture of ADHD and ASD with respect to their shared and differentiating genetic underpinnings as well as across case subgroups. At the single-variant level, we identified new shared loci for the two disorders and five genome-wide significant loci differentiating the disorders, four of which were new. Integration with DLPFC transcriptomic data enabled us to identify and prioritize several possible causal genes (Supplementary Note). At the polygenic level, we found compelling differences across comorbid and single-disorder case groups.

The identified shared loci were generally highly pleiotropic and had previously been identified in GWAS of related disorders or cross-disorder studies including ADHD and/or ASD. However, considering only the eight major psychiatric disorders included in the most recent PGC cross-disorder study[8], three of the loci (rs4916723, rs2391769 and rs227293) appeared to be shared only between ADHD and ASD (Table 1 and Supplementary Data 2). For the other SNPs, only one (rs325506) showed support for involvement in more than one additional disorder. This is consistent with evidence from genomic 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].

In the ADHDvsASD GWAS, we identified five genome-wide significant loci, all showing opposite allelic directions in the separate GWAS of the two disorders, providing specific genetic insights into the biology that drives the pathophysiology toward developing one disorder or the other. Although one of the identified loci (rs3791033) supported the single ADHD–ASD differentiating locus reported previously[24] (using case-case GWAS (CC-GWAS) analysis on available summary statistics), the four new loci all showed supportive (but not statistically significant) results in the CC-GWAS study, except the histone 1 locus at the MHC region, which was not included in the CC-GWAS (Supplementary Data 2). The yield of more significant loci in our study compared with the CC-GWAS could (in addition to methodological differences) have been because we were able to remove comorbid ADHD+ASD cases, which were included in the GWAS results used in the CC-GWAS study, resulting in stronger analytical power in our study.

The top-ranking differentiating TWAS gene and/or isoform was *HIST1H2BD-201*, which was two orders of magnitude more significant than the second-ranking one (*CAMKV-210*) and was the only Bonferroni-significant transcript in the identified *HIST1* GWAS

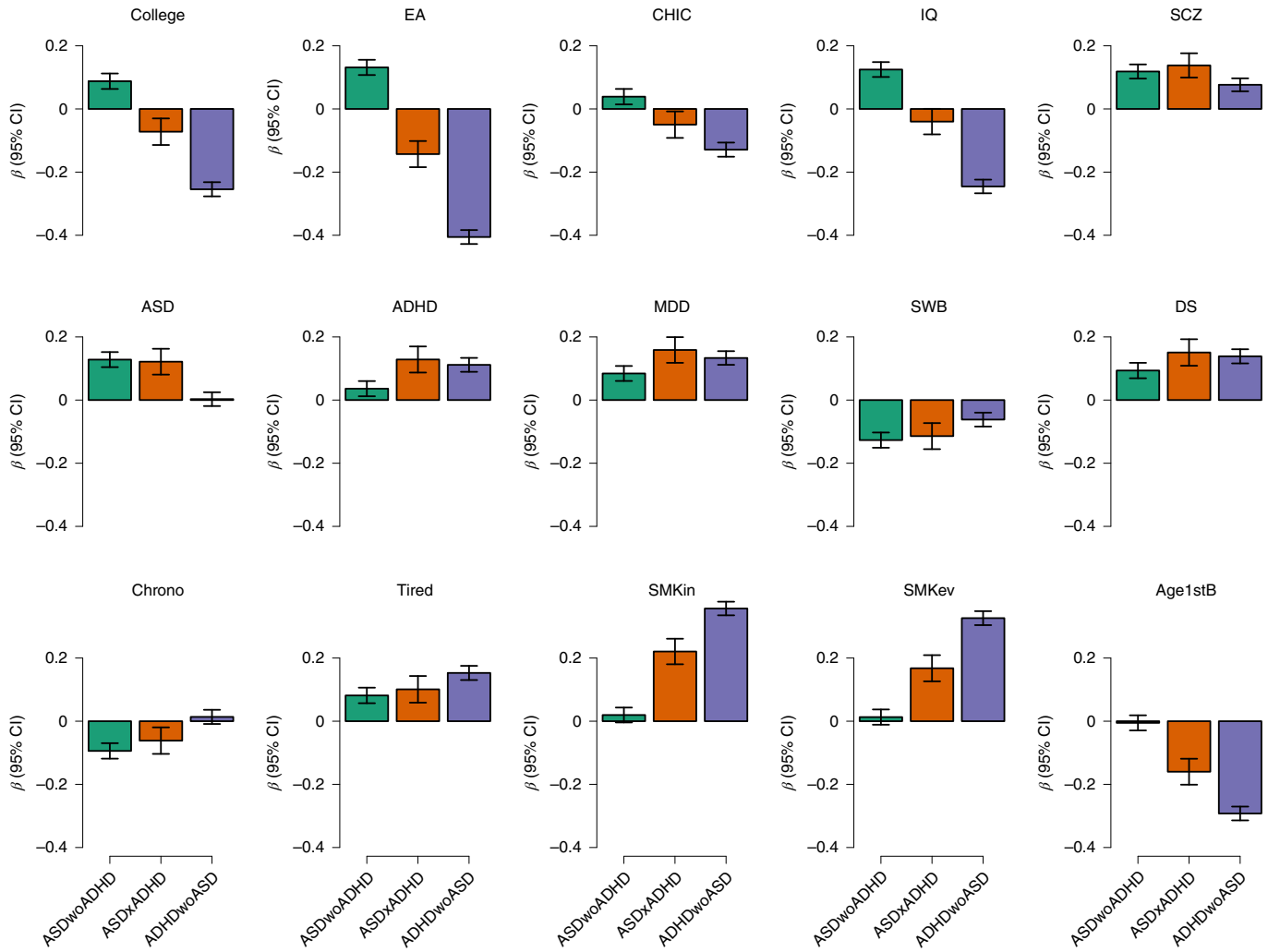

Fig. 2 | Comparison of PRS profiles across ADHD and ASD subtypes for 15 traits and/or phenotypes that have shown significant genetic correlations with ADHD and ASD in the past. Bars display regression coefficients from a multivariate regression of the 15 normalized polygenic scores on ASD–ADHD comorbidity classes ($n = 23{,}583$) and controls as reference ($n = 22{,}122$, not shown). Green represents ASD-only cases ($n = 9{,}315$; ASD without ADHD (ASDwoADHD)), orange depicts comorbid samples ($n = 2{,}304$; comorbid ASD and ADHD (ASDxADHD)) and purple represents ADHD-only cases ($n = 11{,}964$; ADHD without ASD (ADHDwoASD)). Error bars are 95% confidence intervals centered on the point estimate. ADHD, attention-deficit/hyperactivity disorder (PMID 20732625); ASD, autism spectrum disorder (PMID 30804558 without the iPSYCH sample); MDD, major depressive disorder (PMID 29700475 without DK or 23andMe); SWB, subjective well-being (PMID 27089181); DS, depressive symptoms (PMID 27089181); College, college completion (PMID 27046643); EA, educational attainment (PMID 30038396); CHIC, childhood IQ (PMID 23358156); IQ (PMID 29942086); SCZ, schizophrenia (PGC3 without DK); Chrono, chronotype (PMID 30696823); Tired, self-reported tiredness (PMID 28194004); SMKin, smoking initiation (PMID 30643251); SMKev, ever smoker (PMID 30643258); Age1stB, age at first birth (PMID 20418890).

locus. Deleterious de novo mutations in several histone-modifying or histone-interacting genes[38–40], as well as in core histone genes[39,41], 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). The ASD risk allele of the lead SNP in the locus was also associated with both increased educational performance[25] and increased volume of the left globus pallidus[27], whereas the opposite was the case for the ADHD risk allele. As part of the basal ganglia, the globus pallidus is involved in several functions relating to phenotypic domains affected in ASD and/or ADHD, including cognition, social interactions, speech, repetitive behaviors and tics[42]. 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, which may contribute—as one weak-acting factor among many—to driving the pathophysiology toward either ASD or ADHD and affecting 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 such as educational attainment when assessing common variants genome-wide[5,6,43]. Corroborating these reports, we found that the ADHDvsASD GWAS showed the strongest correlations for cognitive traits (Supplementary Data 6 and Supplementary Fig. 7). Moreover, two of the identified differentiating loci (on chromosomes 1 and 6) had lead SNPs that were genome-wide significant in educational attainment and showed opposite allelic effects, with increasing educational performance associated with the ASD risk alleles and decreasing educational performance associated with ADHD.

Table 3 | Results of ADHD and ASD PRS 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	1.3×10^{-207}	1.08	1.06	1.11	7.5×0^{-12}
ASD-only	Controls	1.10	1.07	1.13	3.1×10^{-13}	1.21	1.18	1.24	1.2×10^{-48}
Comorbid	Controls	1.32	1.25	1.39	2.8×10^{-25}	1.22	1.16	1.29	3.5×10^{-14}
Comorbid	ADHD-only	0.92	0.88	0.97	0.0015	1.13	1.08	1.19	4.7×10^{-7}
Comorbid	ASD-only	1.22	1.16	1.28	6.4×10^{-16}	1.01	0.96	1.06	0.68
ASD-only	ADHD-only	0.76	0.74	0.78	4.5×10^{-79}	1.12	1.09	1.15	1.2×10^{-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 PRS analyses. PRS_{ADHD}, analyses using a PRS trained on an ADHD phenotype; PRS_{ASD}, analyses using a PRS 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; LCI, lower boundary for 95% confidence interval; UCI, upper boundary for 95% confidence interval. 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 are two-sided from regression model and are without correction for multiple testing. Experiment-wide significance at 0.0042 (Bonferroni corrected for 2×6 tests). Additional secondary analyses also compared groups of individuals with ADHD or ASD with co-occurring mild ID (ADHD+ID and ASD+ID) with those without (ADHD-no-ID and ASD-no-ID).

We note that the chromosome 1 locus (at position 44 Mb) was identified, counterintuitively, in both the shared and differentiating GWAS, albeit with different lead SNPs (Tables 1 and 2). The locus covers a gene-rich 250-kb region of generally strong LD, but it also harbors variants with limited LD to the main haploblock (Supplementary Figs. 1a and 5d). The two lead SNPs are located 62 kb apart and showed low pairwise LD ($r^2=0.1687$; Table 2), indicating that the two SNPs are largely independent markers for association. This LD difference was also reflected in the different lists of other traits with previously reported associations for the lead SNPs or their LD proxies (Tables 1 and 2). Furthermore, this was the only locus that showed significant heterogeneity across cohorts in the recent ADHD GWAS[6], in which 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 enrichment of brain-expressed genes in the combined GWAS, implicating particularly the basal ganglia and cerebellum. Both structures have been found to be altered in both ASD[42,44] and ADHD[45–47], with evidence for reductions in basal ganglia volume being the most robustly observed finding in the neuroimaging literature for both ASD and ADHD. The cell-type enrichment results implicating the red nucleus in the 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 and jaw: both motor coordination and speech problems are frequent in both ASD and ADHD[48,49]. The red nucleus is strongly connected with many brain structures involved in ASD and ADHD, including the basal ganglia and the cerebellum[50].

Dissecting the polygenic architecture using PRS approaches, we observed remarkable differences across the comorbid and single-disorder (ADHD-only and ASD-only) case groups. The comorbid cases carried a double burden of ASD-PRS and ADHD-PRS, whereas the single-disorder cases largely had just a single burden 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. This was further highlighted by the identification of a genome-wide significant locus for comorbid cases (chromosome 6). It also supports pharmacological treatment of comorbid ADHD in individuals with ASD. In a recent meta-analysis, 25–32% of individuals with ASD were found to also fulfill criteria for ADHD[13], yet only 15–16% of such individuals are treated with ADHD medications[51,52], despite strong evidence of beneficial effects on the core symptoms of ADHD, potentially reduced risks of injuries[53], depression[54] and suicidal behavior[55], and improved academic performance[56]. Moreover, it indicates that pharmacological treatment of symptoms such as hyperactivity, inattention, impulsivity, aggression and tics in individuals diagnosed with either ADHD or ASD may be guided by the individual symptomatology regardless of the given diagnosis.

We recently reported a significant genetic correlation of $r_G=0.36$ between ASD and ADHD, using LDSC and results from GWAS that included multiple cohorts and comorbid cases[5]. This was a considerable increase from the previous estimate of $r_G=0.08$ (s.e.=0.10, $P=0.40$), which was based on much smaller GWAS samples without information on comorbid diagnoses[57]. 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 not driven by comorbid cases alone. Although we cannot exclude the possibility that underdiagnosis of comorbidity might exist, leading to an upwards bias of the genetic correlation estimate between the single-disorder cases, our result is corroborated by data from Swedish twin studies that support the distinction of ASD and ADHD but also suggest considerable co-occurrence of symptoms of both disorders in individuals only fulfilling diagnostic criteria for one of the two disorders[58,59].

In addition, the correlations increased when excluding cases with ID, indicating that individuals with ID are more genetically heterogeneous in terms of common variant risk between the two disorders than individuals 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 individuals with ID carry a higher load of (ultra)rare damaging risk variants compared with those without ID[10]. Consistent with this, our PRS analyses found lower ASD-PRS in the group of ASD cases with comorbid mild ID (IQ=50–70) compared with 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 (because the rare variant risk load is similar for the two disorders in the data available so far) and more extensively in cases with ID than without ID (because the common variant genetic correlation is lower when including cases with comorbid 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 shared and disorder-specific risk variants providing information on pathophysiology. In addition, we have revealed specific patterns of polygenic architecture that are characteristic of comorbid cases compared with single-disorder cases. The results advance our understanding of the complex etiologic basis of ASD and ADHD and the relationship between the two disorders, toward the long-term goals of better diagnosis and treatment of these disorders.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-022-01171-3.

Received: 1 June 2021; Accepted: 20 June 2022;
Published online: 26 September 2022

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. 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**, 761–770.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**, 1363–1396 (2011).
12. Zablotsky, 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. 1000 Genomes Project Consortiumet 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**, 291–295 (2015).
18. Yang, Z. et al. Investigating shared genetic basis across Tourette syndrome and comorbid neurodevelopmental disorders along the impulsivity-compulsivity spectrum. *Biol. Psychiatry* **90**, 317–327 (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**, eaat8464 (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. Peyrot, W. J. & Price, A. L. Identifying loci with different allele frequencies among cases of eight psychiatric disorders using CC-GWAS. *Nat. Genet.* **53**, 445–454 (2021).
25. 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).
26. Marzluff, W. F., Gongidi, P., Woods, K. R., Jin, J. & Maltais, L. J. The human and mouse replication-dependent histone genes. *Genomics* **80**, 487–498 (2002).
27. 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).
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**, 1236–1241 (2015).
31. 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).
32. 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).
33. Wu, Z. M. et al. White matter microstructural alterations in children with ADHD: categorical and dimensional perspectives. *Neuropsychopharmacology* **42**, 572–580 (2017).
34. Aoki, Y. et al. Association of white matter structure with autism spectrum disorder and attention-deficit/hyperactivity disorder. *JAMA Psychiatry* **74**, 1120–1128 (2017).
35. Neale, B. M. et al. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **49**, 884–897 (2010).
36. 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).
37. 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).
38. 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).
39. 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).
40. De Rubeis, S. et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
41. 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**, eabc9207 (2020).
42. Subramanian, K. et al. Basal ganglia and autism - a translational perspective. *Autism Res.* **10**, 1751–1775 (2017).
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**, 419–425 (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. 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.* **115**, 211–220 (2008).
49. 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).
50. Basile, G. A. et al. Red nucleus structure and function: from anatomy to clinical neurosciences. *Brain Struct. Funct.* **226**, 69–91 (2021).
51. 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**, 432–439 (2013).
52. 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**, 342–351 (2010).
53. 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).
54. 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).
55. Chang, Z. et al. Medication for attention-deficit/hyperactivity disorder and risk for suicide attempts. *Biol. Psychiatry* **88**, 452–458 (2020).
56. Keilow, M., Holm, A. & Fallesen, P. Medical treatment of attention deficit/hyperactivity disorder (ADHD) and children's academic performance. *PLoS ONE* **13**, e0207905 (2018).
57. Brainstorm Consortiumet al. Analysis of shared heritability in common disorders of the brain. *Science* **360**, eaap8757 (2018).
58. 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).
59. Ronald, A., Larsson, H., Anckarsater, H. & Lichtenstein, P. Symptoms of autism and ADHD: a Swedish twin study examining their overlap. *J. Abnorm Psychol.* **123**, 440–451 (2014).

[1]Department of Biomedicine - Human Genetics and the iSEQ Center, Aarhus University, Aarhus, Denmark. [2]Department of Community Health and Epidemiology & Department of Psychiatry, Dalhousie University, Halifax, NS, Canada. [3]Institute of Psychiatric Phenomics and Genomics (IPPG), University Hospital, LMU Munich, Munich, Germany. [4]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus, Denmark. [5]Center for Genomics and Personalized Medicine, Aarhus, Denmark. [6]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. [7]MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK. [8]Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [9]Icahn Institute for Data Science and Genomic Technology, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [10]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [11]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [12]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [13]Institute of Biological Psychiatry, Mental Health Services Copenhagen, Copenhagen University Hospital, Copenhagen, Denmark. [14]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden. [15]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark. [16]National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark. [17]Centre for Integrated Register-based Research, Aarhus University, Aarhus, Denmark. [18]Department of Genetics, Microbiology and Statistics, Faculty of Biology, University of Barcelona, Barcelona, Catalonia, Spain. [19]Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, Madrid, Spain. [20]Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Catalonia, Spain. [21]Institut de Recerca Sant Joan de Déu (IR-SJD), Esplugues de Llobregat, Catalonia, Spain. [22]Department of Clinical Medicine, Faculty of Health Science, University of Copenhagen, Copenhagen, Denmark. [23]Copenhagen Research Centre for Mental Health (CORE), Mental Health Centre Copenhagen, Copenhagen, Denmark. [24]University Hospital, Hellerup, Denmark. [25]GLOBE Institute, Center for GeoGenetics, University of Copenhagen, Copenhagen, Denmark. [26]Psychosis Research Unit, Aarhus University Hospital-Psychiatry, Aarhus, Denmark. [27]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [28]Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [29]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [30]Department of Psychiatry, State University of New York Upstate Medical University, Syracuse, NY, USA. [31]Department of Neuroscience and Physiology, State University of New York Upstate Medical University, Syracuse, NY, USA. [32]Department of Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands. [33]Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands. [34]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [35]Department of Psychiatry, JJ Peters VA Medical Center, Bronx, NY, USA. [36]Department of Medicine, Harvard Medical School, Boston, MA, USA. [37]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. [38]These authors contributed equally: Manuel Mattheisen, Jakob Grove, Thomas D. Als, Joanna Martin. ✉e-mail: manuel.mattheisen@gmail.com; anders@biomed.au.dk

Methods

Ethics and overview. This study was approved by the Regional Scientific Ethics Committee in Denmark and the Danish Data Protection Agency. We report results from different analyses all carried out in large-scale samples from the PGC and iPSYCH. We used samples included in the most recently published GWAS of ASD[5] and ADHD[6]. In this work, we refer to cases of individuals in the study cohort (most importantly in iPSYCH) that at the time of inclusion had only one of the two diagnoses registered (that is, ADHD or ASD) as ADHD-only and ASD-only cases, respectively. We refer to cases of 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 (that is, 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 iPSYCH case–cohort sample[60] can be found in the respective publications[5,6]. Here, we focus on differences in selection criteria in the iPSYCH cohort and additional quality control (QC) procedures.

The majority of inclusion and exclusion criteria for the original studies were also used in this study. The only difference compared with the original studies was an additional exclusion criterion that removed individuals with moderate to severe mental retardation (ICD-10: F71–F79) from both the case and control cohorts. Although this criterion was also used in the original ADHD GWAS[6], it was not used in the original ASD GWAS[5]. 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 (Table 3, Supplementary Fig. 14b and Supplementary Data 9).

Wave-wise preimputation 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]. As our analyses used a combined study cohort with samples from both the original ADHD and ASD GWAS, we performed some additional QC on the combined sample. Additional QC steps included the removal of related individuals across the original ADHD and ASD GWAS and a new principal components 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-hat > 0.2 (using PLINK's[61] identity-by-state analysis), and one subject of each pair was excluded at random (with a preference for keeping cases). PCA was carried out using smartPCA in the EIGENSOFT software package[62,63] using the Ricopili pipeline[64]. The original PGC datasets for ADHD and ASD did not include overlapping individuals; therefore, the original datasets and summary statistics were used. The final combined dataset across all samples comprised 34,462 cases (that is, 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, respectively. Thus, the proportion of ADHD among ASD cases in the iPSYCH cohort was 19.8%, and the proportion of ASD among ADHD cases was 16.1%.

Genome-wide association analyses. As in the original GWAS in ADHD and ASD, all processing and analyses for the individual GWAS and meta-analyses (see below) used the Ricopili pipeline[64]. More details on individual modules and steps can be found elsewhere[5,6,64]. We ran two main GWAS for our analyses. The first aimed to identify shared genetic risk for ADHD and ASD (combined GWAS), and the second aimed to identify differentiating genetic risk with an opposite direction of effects for ADHD and ASD (ADHDvsASD GWAS). All analyses of the iPSYCH sample and meta-analyses with the PGC samples were conducted at the secure national GenomeDK high-performance computing cluster in Denmark.

Combined GWAS. We first ran an analysis in the combined dataset, that is, on all 34,462 cases and 41,201 controls. The GWAS was conducted in each cohort (that is, in the wave-wise iPSYCH samples and the individual PGC cohorts) using logistic regression with the imputed additive genotype dosages. The first five principal components (PCs) were included as covariates to correct for population stratification (Supplementary Note), and variants with imputation INFO score <0.8 or MAF < 0.01 were excluded. The resulting summary statistics files were then meta-analyzed using an inverse-variance weighted fixed effects model[65]. Postprocessing of the summary statistics files through the Ricopili pipeline[64] was used to create Manhattan plots, individual regional association plots and forest plots. For a QQ-plot of the analysis, see Supplementary Fig. 14a.

ADHDvsASD 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, $n = 11{,}964$) against ASD-only (coded as 2, $n = 9{,}315$) cases in the iPSYCH cohort. This approach was in line with that of our recent study that compared the genetic risks of developing bipolar disorder and schizophrenia[66]. We excluded comorbid cases from this GWAS, and the GWAS was conducted wave-wise using logistic regression with imputed additive genotype dosages. The first five 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 statistics files were then meta-analyzed using an inverse-variance weighted fixed effects model[65], and results were visualized through the Ricopili pipeline[64] (see above). For a QQ-plot of the analysis, see Supplementary Fig. 14b.

Identification of previously reported associations for top findings. Different resources were used to identify previously reported associations of our top findings with other phenotypes and traits within and outside psychiatry. We assessed associations reported in the OpenGWAS project database (https://gwas.mrcieu.ac.uk/about/, accessed 14 October 2020; see Supplementary Data 1 and 7 for results) and used the GWAS ATLAS website[67] to visualize PheWAS analyses (Supplementary Figs. 2 and 6). We also used results from the GWAS Catalog[68] (Table 2). Finally, we also compared our results with those of previous cross-disorder studies in the field. These included the recent analyses of the cross-disorder group in the PGC[8], a study that used a new approach to study case–case associations in psychiatric disorders[24], and a study that used conditional analyses to highlight associations that might be specific to individual psychiatric disorders[69]. Results are available in the Supplementary Note and Supplementary Data 2.

Transcriptomic imputation model construction and TWAS. Transcriptomic imputation models were constructed as previously described[205] for DLPFC transcript levels[70]. The genetic dataset of the PsychENCODE cohort was uniformly processed for QC steps before genotype imputation. We restricted our analysis to samples from individuals of European ancestry as previously described[20]. Genotypes were imputed using the University of Michigan server[71] with the Haplotype Reference Consortium reference panel[72]. Gene expression information (at both gene and transcript levels) was derived from RNA sequencing counts, which were adjusted for known and hidden confounders, followed by quantile normalization[70]. 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[73]. EpiXcan was recently shown to increase power to identify genes under a causality model compared with TWAS approaches that do not integrate epigenetic information[74]. We used this model (924 samples from DLPFC) owing to power considerations[20]; by comparison, brain gene expression imputation models based on GTEx v.8 (ref. [75]) are trained on 205 or fewer samples. Based on only samples from DLPFC, we acknowledge that ADHD and ASD are both associated with other brain regions and highlight this as a potential limitation of our study. We performed the transcript–trait association analysis for the traits in this study as previously described[20]. Briefly, we used the S-PrediXcan method[20] to integrate the GWAS summary statistics and the transcriptomic imputation models constructed as described above to obtain association results at both the gene and transcript levels.

Cell-type enrichment analysis. A major portion of cell-type-specific enrichment can be attributed to distal regulatory elements, as local regulatory events remain highly consistent across various tissues and cell types[76]. Therefore, we examined the overlap of common genetic variants of investigated traits (Supplementary Fig. 14 and Supplementary Data 8) and open chromatin from a single-cell assay for transposase accessible chromatin study[31] using the LD score partitioned heritability approach[77]. All regions of open chromatin were extended by 500 bp in either direction. The broad major histocompatibility complex (MHC) region (hg19 chr6:25–35 Mb) was excluded owing 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 different approaches combining in-house scripts and data with those available via the FUMA v1.3.6a[23] 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 the FUMA website. Finally, we used FUMA to perform cell-type specificity analyses[78] based on our summary statistics. For all the 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 Supplementary Note. Supplementary Data 10 contains results from standard FUMA-based analyses, such as expression quantitative trait loci and chromatin interaction mapping.

Gene-based analysis. We also used FUMA v1.3.6a[23] 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 Supplementary Note.

Our results in the 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 PGC[8] and a recent analysis based on genomic structural equation modeling of 11 major psychiatric disorders[79]. We also compared our results with those of recent analyses that aimed to identify disorder-specific SNPs for psychiatric disorders[24,69].

PRS analyses. To examine potential polygenic heterogeneity across ADHD and ASD subtypes, we investigated how PRS trained on different phenotypes was 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[5,6,36] 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 standardized PRS variables and can superficially be viewed as running a linear regression for each score on the ADHD and ASD subgroups simultaneously. The regression coefficients can be interpreted as the mean value of the PRS relative to the value in controls. The framework allows us to compare the average PRS across subgroups for scores from several phenotypes while accounting for the inherent correlation between scores and adjusting for necessary covariates. This enables testing of a whole array of hypotheses, with comparisons both between subgroups and between PRSs. We can compare groups that are too small for GWAS and gauge genetic correlation with groups that are too small for LDSC, as was the case with the comorbid ASD–ADHD group. Polygenic scores were generated by clumping and thresholding employing standard Ricopili settings as explained[5] and using summary statistics from the GWAS[5,35,80–89].

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 ≤ 50) cases versus controls and ASD-only (excluding comorbid ADHD or severe ID) cases versus controls, using independent (split) controls. For each set of discovery results, LD-clumping was run in PLINK v.1.9 (ref. [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 *P* value threshold used for SNP selection was $P < 0.5$. Asymmetric and/or ambiguous SNPs (AT, TA, CG, GC), indels, multi-allelic SNPs and duplicate position SNPs were excluded. SNPs with MAF < 0.01 or INFO < 0.8 and those present in less than half of the sample were filtered out. PRS for ADHD and ASD were calculated by scoring the number of effect alleles weighted by the log(OR) across the set of independent, clumped, meta-analyzed SNPs in PLINK. PRS were derived from 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 five PCs were run to test for associations of PRS with each of the outcomes within each wave, as follows: (1) ADHD-only cases versus controls; (2) ASD-only cases versus controls; (3) comorbid cases versus controls; (4) ADHD-only cases versus ASD-only cases; (5) ADHD-only cases versus comorbid cases; and (6) ASD-only cases versus comorbid cases. Cases were coded as 1 and controls as 0, except that comorbid cases were coded as 1 in case–case comparisons and the ASD-only cases in analysis (d) 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 and 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 Supplementary Data 9). Several sensitivity tests were also run (including sex as a covariate, excluding cases and controls with mild ID).

Genetic correlations (LD Hub). The genetic correlations of our various 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, as new phenotypes had been added to LD Hub after publication of the original analyses. We also uploaded summary statistics for the two analyses described above, that is, the combined GWAS and the ADHDvsASD GWAS, to assess correlations with the identified shared and differentiating genetic liability, respectively. We used all available phenotypes in LD Hub[29] but performed analyses for the UK Biobank (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 for major depressive disorder[85] and bipolar disorder[91]. Levels of experiment-wide significance (Bonferroni correction for number of tests applied) were also established separately within the two groups, that is, in the UKBB traits ($P < 8.38 \times 10^{-5}$) 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; SNP-h^2) was estimated in the iPSYCH sample using the GREML approach of GCTA[37] for ADHD versus ASD and for ADHD versus each of the ASD subphenotypes (see below). The genetic relationship matrix (GRM) between all pairwise combinations of individuals was estimated using all case–control samples. The strict best-guess genotypes (that is, SNPs with INFO > 0.8, missing rate <0.01 and MAF > 0.05, indels removed) were used for GRM estimation. GCTA-GREML accounts for LD[92], and the GRM estimation was performed on a non-LD-pruned dataset. Estimation of the phenotypic variance explained by the SNPs was performed for each of the subphenotypes listed in Supplementary Data 9, with PCs 1–20 included as continuous covariates and waves 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 (Supplementary Data 9) 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 proportions of 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 Bland[93].

GCTA-GREML analyses were conducted for ADHD versus ASD as the main diagnosis (Supplementary Fig. 5a) by (1) excluding individuals with both phenotypes (comorbid) and (2) by randomly splitting comorbid cases into either ADHD or ASD. In addition, GCTA analyses were conducted for ADHD versus four ASD subphenotypes by (1) excluding individuals with both phenotypes (comorbid) and (2) randomly splitting comorbid cases into either the ADHD or ASD subphenotype. These analyses were conducted both including and excluding individuals with ID. See Supplementary Data 9 and Supplementary Fig. 5 for an overview of comparisons.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Summary statistics from this publication are available at http://ipsych.au.dk/downloads/. Summary statistics for the original ADHD and ASD GWAS analyses are available at the same site. For access to genotypes from the PGC samples and the iPSYCH sample, researchers should contact the lead PIs E.R. and/or A.B. (https://pgc.unc.edu/for-researchers/working-groups/autism-working-group/) for PGC-ASD; A.B. for iPSYCH-ASD; B.N. and/or B.F. (https://pgc.unc.edu/for-researchers/working-groups/adhd-working-group/) for PGC-ADHD; and A.B. for iPSYCH-ADHDy. Data used for generation of the brain transcriptome model are available from PsychENCODE (overview of available datasets at http://resource.psychencode.org/); genotypes are controlled data and access instructions are provided at https://www.synapse.org/#!Synapse:syn4921369/wiki/477467. Note that some datasets were indirectly accessed at the respective analytical websites (for example, GSE76381 through the FUMA website). Please refer to these websites (for example, for FUMA, https://fuma.ctglab.nl/links and https://fuma.ctglab.nl/tutorial#datasets) for availability of datasets used in the respective follow-up analyses and/or lookups (for example, GSE76381).

Code availability

Please refer to individual sections of the methods above for published code (for example, for EpiXcan or Ricopili). As the in-house scripts used for data processing and analysis of the iPSYCH data on the GenomeDK HPC infrastructure are highly dependent on that context, they can only be obtained from the authors upon request. This way we can ensure the proper context is explained in dialog with the interested parties.

References

60. 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).
61. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

62. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
63. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
64. Lam, M. et al. RICOPILI: Rapid imputation for COnsortias PIpeLIne. *Bioinformatics* **36**, 930–933 (2020).
65. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
66. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell* **173**, 1705–1715.e16 (2018).
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. Byrne, E. M. et al. Conditional GWAS analysis to identify disorder-specific SNPs for psychiatric disorders. *Mol. Psychiatry* **26**, 2070–2081 (2021).
70. Gandal, M. J. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018).
71. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
72. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
73. Roadmap Epigenomics Consortiumet al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
74. Cao, C. et al. Power analysis of transcriptome-wide association study: implications for practical protocol choice. *PLoS Genet.* **17**, e1009405 (2021).
75. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
76. 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).
77. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
78. 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).
79. Grotzinger, A. D. et al. Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. *Nat. Genet.* **54**, 548–559 (2022).
80. Davies, G. et al. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (*N*=112 151). *Mol. Psychiatry* **21**, 758–767 (2016).
81. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
82. Benyamin, B. et al. Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Mol. Psychiatry* **19**, 253–258 (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 Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (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**, 624–633 (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 and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42**, 441–447 (2010).
90. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (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. USA* **113**, E4579–E4580 (2016).
93. Altman, D. G. & Bland, J. M. How to obtain the confidence interval from a P value. *BMJ* **343**, d2090 (2011).

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 A.D.B.) 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 A.D.B.). G.V. was funded by NIMH grant K08MH122911 and the 2020 NARSAD Young Investigator Grant no. 29350 from the Brain & Behavior Research Foundation. P.R. was funded by NIMH grants R01MH125246, U01MH116442 and R01MH109677. Support for this research was further received by B.C. 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 667302 (CoCA), 643051 (MiND) and 728018 (Eat2beNICE), and the ECNP network 'ADHD across the lifespan'. S.V.F. is supported by the European Union's Horizon 2020 research and innovation programme under grant agreements 667302 and 965381; NIMH grants U01MH109536-01, U01AR076092-01A1, R0MH116037 and 5R01AG06495502; Oregon Health and Science University, Otsuka Pharmaceuticals and Supernus Pharmaceutical Company. We thank all participants in the cohorts included in this analysis.

Author contributions

M.M., J.G. and A.D.B. designed the study. M.M., J.G., T.D.A., J.M., G.V., S.M., D.D., J.B., R.W., C.E.C., A.R., N.I.S., W.Z., M.E.H., B.Z. and G.H. conducted data analysis. P.B.M., E.B.R., P.R., B.M.N., M.J.D. and A.D.B. supervised data analysis. J.B.-G., M.B.-H., E.A., J.D.B., M.N., T.W., O.M., D.M.H., P.B.M., B.M.N., M.J.D. and A.D.B. provided data. M.M., J.G., T.D.A., J.M., S.M. and A.D.B. wrote the paper. M.M., J.G., T.D.A., J.M., B.C., E.B.R., S.V.F., B.F., S.D. and A.D.B. formed the core revision group. A.D.B. directed the study. All authors discussed the results and approved the final version of the manuscript.

Competing interests

B.F. has received educational speaking fees from Medice. In the past year, S.V.F. has 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, S.V.F. has US patent US20130217707 A1 for the use of sodium–hydrogen exchange inhibitors in the treatment of ADHD. S.V.F. 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*). S.V.F. is Program Director of www.adhdinadults.com. B.M.N. is a member of the scientific advisory board at Deep Genomics and Neumora (formerly Neumora) and consultant for Camp4 Therapeutics, Takeda Pharmaceutical and Biogen. M.J.D. is a founder of Maze Therapeutics and is on the Scientific Advisory Board of RBNC Therapeutics. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-022-01171-3.

Correspondence and requests for materials should be addressed to Manuel Mattheisen or Anders D. Børglum.

Peer review information *Nature Genetics* thanks E. Byrne, Andreas G. Chiocchetti and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

nature portfolio

Corresponding author(s): Manuel Mattheisen and Anders Borglum

Last updated by author(s): Apr 23, 2020

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	This is fully described in the Online Methods, associated Supplementary Information and original GWAS publications. In brief, individual level imputed data from previous publications on ADHD (https://www.nature.com/articles/s41588-018-0269-7) and ASD (https://www.nature.com/articles/s41588-019-0344-8) in iPSYCH and PGC cohorts were re-evaluated for a combined and differentiating phenotype.
Data analysis	This is fully described in the Online Methods and associated Supplementary Information. In brief: Additional quality control, association analyses, and polygenic risk scoring was done using the Ricopili (v 1118b) pipeline: https://github.com/Nealelab/ricopili, which relies on the following software: Eigensoft 6.0.1 (incl. smartPCA), Plink 1.9, METAL 2011-03-25. For gene-based and gene-set analyses we used FUMA v1.3.6a and MAGMA 1.08. We used the S-PredXcan v0.6.5 method to perform transcriptomic imputation modelling and analyses. We used EpiXcan v1 for imputation of the genetically regulated gene expression. SNP heritability, partitioning of the heritability and genetic correlations were estimated using LD score regression (v 1.0.1; https://github.com/bulik/ldsc) and LD hub (accessed Oct 15th 2020; http://ldsc.broadinstitute.org/) for the large samples. Genetic correlation between subtypes of ADHD and ASD were estimated using GCTA v1.93. R v3.4 was used in general for statistical analyzes and plotting (https://www.Rproject.org) Analyses for PGC analyses were conducted on the SurfSara Lisa computing infrastructure and for iPSYCH and the joined analyses on iPSYCH secure servers in Denmark (GenomeDK).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Summary statistics from this publication are available at http://ipsych.au.dk/downloads/. Summary statistics for the original ADHD and ASD GWAS analyses are also available at this site. For access to genotypes from the PGC samples and the iPSYCH sample, researchers should contact the lead PIs Elise Robinson / Anders Børglum (https://pgc.unc.edu/for-researchers/working-groups/autism-working-group/) and Anders Børglum for PGC-ASD and iPSYCH-ASD, respectively and Benjamin Neale / Barbara Franke (https://pgc.unc.edu/for-researchers/working-groups/adhd-working-group/) and Anders Børglum for PGC-ADHD and iPSYCH-ADHD, respectively. Data used for brain transcriptome model generation are available from PsychENCODE (overview of available data sets at http://resource.psychencode.org/); of note, genotypes are controlled data and access instructions are provided at https://www.synapse.org/#!Synapse:syn4921369/wiki/477467. Note that some datasets have been indirectly accessed at the respective analytical websites (e.g., GSE76381 through the FUMA website). Please refer to these websites (e.g., for FUMA https://fuma.ctglab.nl/links and https://fuma.ctglab.nl/tutorial#datasets) for availability of datasets used in the respective follow-up analyses / lookups (e.g., GSE76381).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was made. Previous studies of psychiatric disorders that are very polygenic (e.g. schizophrenia) have demonstrated that high numbers of cases and controls (in line with the sample size analyzed in this study) yield enough power to detect common risk variants with low effect sizes. Also of note the original ASD and ADHD GWAS publications have strongly suggested current sample size to be sufficient to identify shared and differentiating risk loci for ADHD and ASD.
Data exclusions	Within each analyzed cohort we aimed at analyzing genetically homogeneous samples of unrelated individuals. Related individuals were excluded based on Identity by State analyses (pseudo controls were used for trios) and genetic outliers were excluded based on principal component analyses.
Replication	Consistency was checked internally between the 23 batches in iPSYCH and between iPSYCH and PGC where possible (see for example the leave-one-out PRS analyses). In addition, we checked for consistency between analyses that included or excluded cases of intellectual disability. Finally, we compared our results with similar analyses using the same (or related) datasets from published work (see Supplementary Tables 1a-c). We found good consistency across the board and have no reason to believe that our results were significantly influenced by technical artifacts or other errors.
Randomization	It is an observational study comparing everybody in the selected birth cohorts with and ASD diagnosis as cases, and a random sample from the complement in said cohort as controls. Where split controls where required those subgrouping in controls was done randomly.
Blinding	In iPSYCH, diagnoses are drawn from registries. These are administrative data bases populated by data from the clinicians long before the current study. The blood samples are pulled from a biobank. Hence, the study participants and diagnosing clinicians are blinded with respect to this study. Genotyping is done on a massive scale on 85.000 individuals on 500.000 variables for the overall iPSYCH experiment (which by imputation is expanded to ~10 million variables), and the data is generated without a specif goal or effect in mind except for an overall goal of investigating the genetic and environmental effects on psychiatric disorders. So although it is in principle possible for analysts in the lab to look up crude diagnostic data for a sample, it will not change the genotyping. - In the meta analysis we include data from the Psychiatric Genetics Consortium (PGC) which has been reported in an earlier publication. Their design was different, but analyses analogous.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	In the meta-analysis we included samples from the Psychiatric Genetics Consortium (PGC) and the iPSYCH sample. The iPSYCH sample was processed in 23 batches (genotyping, qc and imputation was done separately for theses batches) of approximately 3,500 individuals each. All analyzes were adjusted for batch, and principal components included to control for population stratification. The PGC samples are combinations of case control and trio samples. There was no need to adjust for populations stratification for the trio samples. In the case-control part the same approach as in the iPSYCH samples was selected.
Recruitment	In iPSYCH, diagnoses are drawn from national registries and the blood samples are pulled from a the Danish Neonatal Screening Biobank. Hence, it is a population sample and bias from self-selection is impossible. Recruitment details for PGC samples in ADHD and ASD have been described in the original GWAS publications and no changes due to additional inclusion or exclusion criteria have been imposed, as such sample characteristics remain the same.
Ethics oversight	The study was approved by the Regional Scientific Ethics Committee in Denmark and the Danish Data Protection Agency.

Note that full information on the approval of the study protocol must also be provided in the manuscript.