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

[*]Manuel Mattheisen[1,2,3], [*]Jakob Grove[1,4,5,6], [*]Thomas D Als[1,4,5], [*]Joanna Martin[7], Georgios Voloudakis[8,9,10], Sandra Meier[1,2], Ditte Demontis[1,4,5], Jaroslav Bendl[8,9], Raymond Walters[11,12], Caitlin E Carey[11,12], Anders Rosengren[4,13], Nora I Strom[1,3,14,15], Mads Engel Hauberg[8,9,10], Biao Zeng[16], Gabriel Hoffman[8,9], Wen Zhang[8,9,10], Jonas Bybjerg-Grauholm[4,17], Marie Bækvad-Hansen[4,17], Esben Agerbo[4,18,19], Bru Cormand[20,21,22,23], Merete Nordentoft[4,24,25], Thomas Werge[4,13,24,26], Ole Mors[4,27], David M Hougaard[4,17], Joseph D Buxbaum[10,28,29,30], Stephen V Faraone[31,32] , Barbara Franke[33,34], Søren Dalsgaard[18,24,35], Preben B Mortensen[4,5,18,19], Elise B Robinson[11,12,36], Panos Roussos[8,9,10,30,37], Benjamin M Neale[11,12], Mark J Daly[11,12,38,39], Anders D Børglum[1,4,5]

[*]contributed equally

Affiliations:

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

[2]Department of Psychiatry and Department of Community Health and Epidemiology, 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, 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]Pamela Sklar Division of Psychiatric Genomics, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, 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, New York, USA.

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

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

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

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

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

[15]Department of Psychology, Humboldt University of Berlin, Berlin, Germany

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

[17]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark

[18]National Centre for Register-Based Research, Business and Social Sciences, Dept. of Economics and Business, Aarhus University, Aarhus, Denmark

[19]Centre for Integrated Register-based Research, Aarhus University, Aarhus, Denmark

[20]Department of Genetics, Microbiology and Statistics, Faculty of Biology, University of Barcelona, Catalonia, Spain.

[21]Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, Spain.

[22]Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Catalonia, Spain.

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

[24]Department of Clinical Medicine, Faculty of Health Science, University of Copenhagen, Denmark

[25]Copenhagen Research Centre for Mental Health (CORE), Mental Health Centre Copenhagen, Copenhagen, University Hospital, Hellerup, Denmark.

[26]GLOBE Institute, Center for GeoGenetics, University of Copenhagen, Copenhagen, Denmark

[27]Psychosis Research Unit, Aarhus University Hospital-Psychiatry, Denmark

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

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

[30]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

[31]Department of Psychiatry, State University of New York Upstate Medical University, Syracuse, NY, USA.

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

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

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

[35]Mental Health Services of the Capital Region, Child and Adolescent Psychiatric Center, Glostrup, Denmark

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

[37]Mental Illness Research Education and Clinical Center (VISN 2 South), James J. Peters VA Medical Center, Bronx, NY, USA.

[38]Department of Medicine, Harvard Medical School, Boston, MA, USA.

[39]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.

Please direct correspondence to:
Manuel Mattheisen: manuel.mattheisen@dal.ca
Anders D. Børglum: anders@biomed.au.dk

Abstract

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

Introduction

Attention deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) are among the most common neurodevelopmental disorders in children and often persist throughout adulthood [1]. ADHD and ASD are both highly heritable (60-93%) [2-4] and the mode of their inheritance is complex and polygenic. Despite the high family-based heritability estimates, genome-wide association studies (GWAS) have only recently identified common variants robustly associated with each disorder [5-7]. Although differing from one another with regard to core clinical symptoms, genetic studies have demonstrated significant overlap between the two disorders, with a genetic correlation (r_G) from common variation of 0.36 [5,8] and substantial sharing of rare genetic risk variants such as large copy number 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 co-aggregation of the disorders with increased risk of ADHD among relatives of ASD probands (odds ratios monozygotic twins: 17.8, dizygotic twins: 4.3; full-siblings: 4.6, full cousins: 1.6) [15]. Identification of the genetic components that are shared or distinct for the disorders may provide insights into the underlying biology and potentially inform on sub-classification, course and treatment.

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

Results

Shared genetic liability to ADHD and ASD

We performed a GWAS of diagnosed ADHD and/or ASD combined into a single phenotype (”combined GWAS"), totaling 34,462 cases and 41,201 controls on 8.9 million SNP allele dosages imputed from 1000 genomes phase 3 [16]. Using LD score regression (LDSC) [17] we found evidence for a strong polygenic signal for this GWAS with an intercept of 1.0134 (ratio = 0.0558) and calculated the liability scale SNP-heritability to be 0.128 (for an assumed population prevalence of 0.055). We identified 263 genome-wide significant SNPs in seven distinct loci (**Table 1, Figure 1, Supplemental Figure S1**). All these loci showed associations with both of the disorders separately at p-values below $1^{\times}10^{-4}$ except one, which is genome-wide significant in ADHD and has a p-value of 0.009 in ASD. Overall, the findings corroborate previous results [8,18], but two loci have not been identified before as shared between ADHD and ASD. The novel shared associations are located in a highly pleiotropic multigene locus on chromosome 1 (rs7538463) and on chromosome 4 (rs227293) in *MANBA* (which encodes beta-D-mannoside mannohydrolase). Mutations in *MANBA* are associated with beta-mannosidosis, a lysosomal storage disease that has a wide spectrum of neurological phenotypes, including intellectual disability, hearing loss and speech impairment [19]. More details on the seven loci can be found in **Table 1** and results from lookups in the open GWAS project database (https://gwas.mrcieu.ac.uk/about/, accessed Oct 14th 2020) and comparisons with previous cross-disorder studies are available in the **Supplemental Material and Methods**, **Supplemental Table S1a,b** and as PheWAS plots in **Supplemental Figure S2**.

To identify and prioritize putative causal shared genes we performed a transcriptome-wide association study (TWAS), imputing the genetically regulated gene expression using EpiXcan [20] and expression data from the PsychENCODE Consortium [21] for genes as well as isoforms detected in 924 samples from the dorsolateral prefrontal cortex (DLPFC). Applying a conservative significance threshold ($p < 1.44^{\times}10^{-6}$; corresponding to Bonferroni correction of all the 34,646 genes and isoforms tested), we identified five

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

Gene-based analysis using MAGMA v 1.08 [22] with default settings as implemented in FUMA [23] largely corroborated the results from the GWAS and TWAS, highlighting, e.g., *MANBA* (**Supplemental Figure S4a** and **Supplemental Table S3**). Furthermore, two of the significant genes (*SORCS3* and *DUSP6*) are located in regions that were not identified in the GWAS, suggesting these as additional shared loci.

Differentiating genetic liability to ADHD and ASD

To identify loci with divergent effects on ADHD and ASD, we performed an association analysis comparing 11,964 ADHD-only cases with 9,315 ASD-only cases from the iPSYCH cohort, excluding all 2,304 comorbid cases ("ADHD vs ASD GWAS"). Using 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 (please also see the supplementary information for more details). Five genome-wide significant loci were identified, three of which have not previously been identified in GWAS of either of the two disorders separately (albeit 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 sub-items (**Table 2**, **Figure 1, Supplemental Tables 1b and S5**). The lead variants all show opposite directions of effects in the two disorders.

Two of the five lead SNPs have previously been found associated with educational attainment [25]. For the first SNP (rs3791033 on chromosome 1; $p = 4.65^{\times}10^{-23}$) the C allele confers an increased risk for ASD and increased cognitive performance while the ADHD risk allele (T) is associated with decreased performance. Similarly, for the second SNP (rs9379833 on chromosome 6; $p = 2.26^{\times}10^{-8}$) the A allele confers an increased risk for ASD and increased cognitive performance while the ADHD risk allele (C) is associated with decreased performance. Notably, this SNP (rs9379833) is located in the large histone gene cluster HIST1 [26] and has previously been associated with regional brain volume, specifically of the left globus pallidus [27] ($p = 2.95^{\times}10^{-8}$; the C allele confers an increased risk for ADHD and a decreased volume while the ASD risk allele (A) is associated with an increased volume). Globus pallidus is part of the basal ganglia, which are involved in both motor and non-motor functions, including higher order cognition, social interactions, speech, repetitive behaviors and tics [28]. It is also of note, that the lead SNP on chromosome 8 (rs7821914) is associated with neuroticism [29] ($p = 9.46^{\times}10^{-21}$). For this SNP, the effect allele (C) in the neuroticism GWAS leads to an increased risk for ASD and a decreased risk for ADHD. An additional two of our lead SNPs are in LD ($r^2 > 0.6$) with SNPs that have previously been identified in neuroticism or one of its subdimensions (rs147420422 and rs9379833; see **Table 2**). Results from additional lookups in the open GWAS project database (https://gwas.mrcieu.ac.uk/about/, accessed Oct 14th 2020) are available in **Supplemental Table S5** and as PheWAS plots in **Supplemental Figure S6**.

TWAS using EpiXcan identified 11 Bonferroni significant genes/isoforms and 96 significant transcripts at FDR < 0.05 with different imputed expression in DLPFC between ADHD and ASD cases (**Supplemental Table S2** and **Figure 1**). The *HIST1H2BD-201* isoform located in the chromosome 6 (HIST1) GWAS locus showed the strongest association ($p = 2.08^{\times}10^{-9}$) with higher expression in ADHD compared to ASD cases (**Supplemental Figure S3b**). The other genes/isoforms showed orders of magnitude less significant association, appointing *HIST1H2BD-201* as the top-ranking causal candidate in the locus. The remaining 10 Bonferroni significant genes/isoforms were located in the chromosome 8

GWAS locus (*SLC35G5-201, AF131215.5, AF131215.5-201, FAM167A* and *TDH-204*) or in two loci on chromosome 3 (3p21.1: *RFT1-204* and 3p21.31: *CAMKV-210, MON1A-201, RBM6-210* and *TRAIP*; **Supplemental Figures S3c + 3d**, respectively) where all except *TRAIP* were also genome-wide significant in gene-based analysis using MAGMA (**Supplemental Figure S4b** and **Supplemental Table S3**).

Genetic correlations with other traits

To examine the polygenic architecture of the identified shared and differentiating genetic risk for the disorders we investigated the genetic correlations with 258 traits from a manually curated list of previously published GWAS and 597 traits from the UK Biobank making use of LD Hub [30] and LDSC [31]. Among the 258 previously reported GWAS, 30 (combined GWAS) and 32 (ADHD vs ASD) traits showed significant correlations after Bonferroni correction for multiple testing (**Supplemental Table S4**). The strongest correlations for the liability differentiating ADHD vs ASD GWAS were observed for cognitive traits such as *years of schooling* (r_G = -0.669, $p_{corr} = 3.68^{\times}10^{-85}$) and *childhood IQ* ($r_G$ = -0.609, $p_{corr} = 2.78^{\times}10^{-10}$), while the strongest correlations for the combined GWAS were with traits such as *depressive symptoms* (r_G = 0.506, $p_{corr} = 2.08^{\times}10^{-19}$) and the cross-disorder analyses of the PGC (r_G = 0.433, $p_{corr} = 5.30^{\times}10^{-25}$). Unsurprisingly, the largest difference in r_G (abs(Δr_G)) for the original ADHD and ASD GWAS (both including the comorbid cases) was identified for a series of cognitive traits (largest abs(Δr_G) = 0.733 for *years of schooling*) (**Supplemental Table S4**). Of note, we identified three phenotypes in our UKBB analyses that show notable concordant correlations with ADHD and ASD. Positive correlations with ADHD and ASD were found for *Other disorders of nose and nasal sinuses* and *Tinnitus*, negative correlations with *Illnesses of siblings: None of the above (group 2).* For a scatter plot of all genetic correlations that showed Z > 2 in the original ADHD and ASD GWAS please refer to **Supplemental Figure S7**.

Tissue and cell-type enrichment analyses

We next tested whether genetic associations of shared and differentiating liabilities were enriched with respect to the transcriptomic profiles of human tissues. We found that the transcriptomic profiles related to brain tissues were significantly associated with the shared ADHD-ASD genetics (**Supplemental Figure S8**). At the general tissue level, these associations were with brain, pituitary and testis tissue. At the individual tissue level, the most significant association was observed for the basal ganglia (putamen), followed by the cerebellum. Cell-type enrichment analyses revealed experiment-wide significant association (across all data sets tested) of the red nucleus in the midbrain in human adult brain samples reported in La Manno et al. 2016 [32] (**Supplemental Figure S9c**). Associations that were significant within one of the three tested data sets individually, but not overall, were observed for several cell types, including, e.g., dopaminergic and GABAergic neurons. For the disorder-differentiating analysis (ADHD vs ASD) we observed no significant association with tissues or specific cell-types after correction for multiple testing (**Supplemental Figure S9 and S10**). We also intersected our genetic associations with a recent multi-omic single-cell epigenetic catalogue of the human brain (obtained from GEO GSE147672 [33]). Here both the combined and differentiating GWAS results showed significant enrichment for several neuronal cell populations (see **Supplemental Figure S11** and **Supplemental Table S6**), including excitatory and inhibitory neurons. Interestingly, the only difference in terms of significant associations between the combined and ADHD vs ASD GWAS was seen for oligodendrocytes (which were not significant in the combined but in the ADHD vs ASD GWAS). While aberrant myelination by oligodendrocytes resulting in disruption of white matter development has previously been reported in both ASD and ADHD [34,35], the degree of severity of this alteration might be a distinct pathophysiological factor [36].

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 ($p = 6.89 \times 10^{-26}$) and the ADHD-only subgroup with PRS for ADHD ($p = 3.29 \times 10^{-23}$; see **Figure 2**). Both scores were trained with PGC-only GWAS results[5,37]. The novel results of the multivariate PRS are those concerning the comorbid ASD+ADHD cases. Strikingly, the ASD-PRS load on comorbid ASD+ADHD cases was similar to that on ASD-only cases ($p = 0.77$) and, likewise, the ADHD-PRS load on the comorbid subgroup was similar to that on ADHD-only cases ($p = 0.44$, **Figure 2),** demonstrating that the comorbid cases carry a load of both ADHD and ASD polygenic scores that are similar to the load carried by the single-disorder cases of their respective disorder PRS. In other words, comorbid cases are double-burdened with both ASD and ADHD PRS. In contrast, the ASD-PRS load on ADHD-only cases was not different from controls ($p = 0.79$) and the ADHD-PRS was only slightly increased in ASD-only cases compared to controls ($p = 3.26 \times 10^{-3}$, **Figure 2**).

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

To further dissect the genetic architecture across the ASD and ADHD subgroups we examined the relative burden of PRS for phenotypes and traits that have shown significant genetic correlation with ADHD and ASD [5,6] [38] using the multivariate framework analysis. While PRS for SZ and depression (and genetically related phenotypes) did not show substantially different loads across the subgroups, other

traits showed compelling differences (**Figure 2**). For instance, years of education, IQ, age at first birth, tiredness, and smoking showed differences between ADHD-only and ASD-only cases with the comorbid cases at an intermediate level. Consistent with the LDSC results, these differences were most compelling for the cognitive phenotypes displaying PRS loads in opposite directions for the single-disorder cases and intermediate loads for comorbid cases. In addition, analyses for the chronotype trait showed similar PRS loads in ASD-only and comorbid cases without evidence for loading in ADHD-only cases, reflecting that chronotype is genetically correlated to ASD but not ADHD [5,6].

An item-level analysis of neuroticism revealed specific patterns of associations for the ASD-only and ADHD-only groups that were mostly consistent with previously described patterns [38] (**Supplemental Figure S12**). On average, ADHD-only cases showed much stronger association than ASD-only cases with items belonging to the depressed affect cluster (e.g., the *MOOD* item) compared to the worry cluster (for a definition of the clusters see **Supplemental Figure S12**). For comorbid cases a distinct pattern was observed with PRS loads either ranking between the ADHD- and ASD-only cases (e.g., for the *MOOD* item) or even exceeding the two single-disorder groups (e.g., for the *GUILT* item).

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

Genetic correlation and heritability across case subgroups

We recently reported an LDSC genetic correlation of 0.36 between ASD and ADHD using the largest GWAS meta-analyses of the two disorders, including multiple cohorts and comorbid cases [5]. Here we investigated the correlations across diagnostic subgroups of the disorders in the iPSYCH sample using the

GREML approach of GCTA [39]. For ASD and ADHD overall, we found $r_G = 0.497$ (SE = 0.054, $p = 7.8^{x}10^{-19}$). Excluding the comorbid cases reduced the correlation to $r_G = 0.397$ (SE = 0.056, $p = 6.3^{x}10^{-12}$). After excluding cases with intellectual disability (ID), the correlations between ASD and ADHD were even stronger: $r_G = 0.523$ (SE = 0.054, $p = 6.5^{x}10^{-21}$) and $r_G = 0.425$ (SE = 0.056, $p = 1.7^{x}10^{-13}$) with and without comorbid cases, respectively. All the GCTA results on genetic correlations and SNP heritability estimates can be found in **Supplemental Table S7** and **Supplemental Figure S13**.

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

Genetic liability in comorbid cases

Guided by our results from the previously described analyses, we also performed a GWAS of the comorbid cases. The sample size (n = 2,304 cases) falls significantly behind those for the other reported GWASs, however, we were able to identify a genome-wide significant association for a SNP on chromosome 6 (rs1321614, $p = 3.54^{x}10^{-9}$, OR = 0.8190, MAF = 0.47 for the T allele). This SNP showed only moderate evidence for association in a GWAS of the ASD-only cases ($p = 0.0086$, OR = 0.9622) and no evidence for association in a GWAS of ADHD-only cases ($p = 0.7721$, OR = 0.9960). In the overall combined GWAS (ADHD+ASD), this SNP showed a p-value of 0.0261 and a p-value of 0.2883 in the differentiating GWAS. Liability scale heritability for the GWAS using GCTA was 0.0557 (SE = 0.0088). Please refer to the supplementary material for information describing details of the analysis (**Supplemental Material and Methods**) and a follow-up for the SNPs identified in the combined and differentiating GWASs for GWASs of ASD-only, ADHD-only, and the comorbid cases (**Supplemental Table S1c**).

Discussion

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

The identified shared loci are generally highly pleiotropic and have previously been identified in GWAS of related disorders or cross-disorder studies including ADHD and/or ASD. However, considering only the eight major psychiatric disorders included in the most recent PGC cross-disorder study[8], three of the loci (rs4916723, rs2391769, and rs227293) appear to be shared only between ADHD and ASD (**Table 1**, **Supplemental Table S1b**). Interestingly, rs4916723, located close to MIR9-2, has been identified in a recent study as significant in ADHD-MDD and ASD-MDD case-case GWAS analyses. Similarly, rs227293, located in MANBA, has been found in the same study to show significantly divergent allele frequencies between ADHD and schizophrenia[24] (**Supplemental Table S1b**). For the other SNPs, only one of the SNPs (rs325506) shows support for involvement in more than one additional disorder. This is consistent with evidence from structural equation modeling of eight major psychiatric disorders, showing that ASD and ADHD cluster together in a group of early-onset neurodevelopmental disorders along with Tourette syndrome [8].

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

In the ADHD vs ASD GWAS we identified five genome-wide significant loci, all showing opposite allelic directions in the separate GWAS of the two disorders, providing specific genetic clues to understanding the biology that drives the pathophysiology towards developing one or the other disorder. While one of the identified loci (rs3791033) supported the single ADHD-ASD differentiating locus reported previously[24] (using CC-GWAS analysis on available summary statistics), the four novel 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 (**Supplemental Table S1b**). The yield of more significant loci in our study compared to the CC-GWAS could (in addition to methodological differences) be 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 relatively stronger analytical power in our study.

The top-ranking differentiating TWAS gene/isoform was *HIST1H2BD-201*, which was two orders of magnitude more significant than the second-ranking (*CAMKV-210*) and the only Bonferroni significant transcript in the identified histone 1 GWAS locus. Deleterious *de novo* mutations in several histone modifying or interacting genes [40-42] as well as in core histone genes [41,43] have been associated with autism and developmental delay with autistic features. The haploinsufficiency resulting from these de novo mutations is consistent with our TWAS result showing reduced expression of *HIST1H2BD-201* in ASD (relative to ADHD). Intriguingly, the ASD risk allele of the lead SNP in the locus is also associated with both increased educational performance [25] and increased volume of the left globus pallidus [27] while the opposite is the case for the ADHD risk allele. As part of the basal ganglia, globus pallidus is involved in several functions relating to phenotypic domains affected in ASD and/or ADHD such as cognition, social interactions, speech, repetitive behaviors and tics [28]. 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 towards

either ASD or ADHD and impacting key phenotypic domains such as educational performance, social interaction and motor impairments.

Previous studies found ASD and ADHD to display opposite genetic correlations with cognitive traits like educational attainment when assessing common variants genome-wide [5,6,44]. Corroborating these reports, we found that the ADHD vs ASD GWAS showed the strongest correlations for cognitive traits among the multiple phenotypes tested (**Supplemental Table S4 and Supplemental Figure S7**). Moreover, two of the identified differentiating loci (on chromosome 1 and 6) have lead SNPs that are genome-wide significant in educational attainment and show opposite allelic effects with increasing and decreasing educational performance for the ASD and ADHD risk alleles, respectively.

We note that the chromosome 1 locus (at position 44Mb) was identified, counterintuitively, in both the shared and differentiating GWAS albeit with different lead SNPs (**Table 1 and 2**). The locus covers a gene-rich 250kb region of generally strong linkage disequilibrium (LD) but it also harbors variants with limited LD to the main haploblock (**Supplemental Figure S1a and S5d**). The two lead SNPs are located 62kb apart and show low pairwise LD (r^2 = 0.1687, **Table 2**), indicating that the two SNPs are largely independent markers for association. This LD difference is also reflected in the different lists of other traits with previously reported associations for the lead SNPs or their LD proxies (**Table 1 and 2**). Furthermore, the locus was the single locus showing significant heterogeneity across cohorts in the recent ADHD GWAS [6] where the 23andMe sample provided no support for the otherwise consistently supported locus and, also in contrast to the other cohorts, exhibited limited genetic correlation with educational attainment.

Our analyses revealed the expected enrichment of brain-expressed genes for the combined GWAS. In particular, the basal ganglia and the cerebellum seem to be implicated. Both structures have been found to be altered in both ASD [28,45] and ADHD [46-48], with evidence for reductions in basal ganglia volume the

most robustly observed finding in the neuroimaging literature for both ASD and ADHD. In addition to the brain, enrichment of genes expressed in the pituitary gland and the testes was also observed for the combined GWAS results. This finding may suggest the involvement of the (hypothalamic-)pituitary-gonadal axis and potentially estrogen signaling, which is known to play a role in psychiatric disorders, cognition, and neuroprotection via several neurotransmitter systems, such as the dopaminergic, serotonergic, and glutamatergic system (e.g., Hwang et al. [49]). The cell-type enrichment result implicating the red nucleus in midbrain is also consistent with our knowledge of phenotypic sharing between ASD and ADHD, as it relates to skilled movements and motor control in the limbs as well as jaw: both motor coordination problems and speech problems are frequent in both ASD and ADHD [50,51]. The red nucleus is strongly connected with many brain structures involved in ASD and ADHD, including the basal ganglia and the cerebellum [52].

Dissecting the polygenic architecture using PRS approaches we observed remarkable differences across comorbid and single-disorder (ADHD-only and ASD-only) case groups. The comorbid cases carry a double burden of ASD- and ADHD-PRS, whereas the single-disorder cases were largely just (single-)burdened for the respective disorder. Thus, cases diagnosed with both disorders have on average a similar level of genetic liability to each disorder as the single-disorder cases, providing strong biological support for the change in diagnostic guidelines from DSM-IV to DSM-5 allowing for diagnoses of both disorders in the same person. This is further highlighted by the identification of a first genome-wide significant genetic 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 also fulfill criteria for ADHD [13], yet only 15-16% are treated with ADHD medications [53,54], despite strong evidence of beneficial effects on the core symptoms of ADHD and potentially also reduced risk of injuries [55], depression [56], suicidal behavior [57] and improved academic performance [58]. Moreover, it indicates that pharmacological treatment of symptoms like hyperactivity, inattention, impulsivity,

aggression and tics in cases diagnosed with either ADHD or ASD may be guided by the individual symptomatology regardless of the given diagnosis.

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

We recently reported a significant genetic correlation of $r_G = 0.36$ between ASD and ADHD, using LDSC and results from GWASs that included multiple cohorts and comorbid cases [5]. This was a considerable increase from the previous estimate of $r_G = 0.08$ (SE = 0.10, $p = 0.40$), which was based on much smaller GWAS sample sizes without information on comorbid diagnoses [59]. 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. While we cannot exclude that under-diagnosis of comorbidity might exist, leading to an upwards bias of the correlation estimate between the single-disorder cases, our result is corroborated by data from Swedish twin studies that supports the distinction of ASD and ADHD, but also suggests considerable co-occurrence of symptoms of both disorders in individuals only fulfilling diagnostic criteria for one of the two disorders [60,61].

In addition, the correlations increased when excluding cases with ID, indicating that cases with ID are more genetically heterogeneous in common variant risk between the two disorders than cases without ID. A recent exome sequencing study of ASD and ADHD (also in the iPSYCH cohort) showed that the disorders have substantial overlap in rare variant risk and that cases with ID carry a higher load of (ultra)rare damaging risk variants compared to cases without ID [10]. Consistent with this, our PRS analyses

found that there was lower ASD PRS in the group of ASD cases with comorbid mild ID (IQ=50-70) compared to those without mild ID. Taken together, these observations are consistent with the notion that the genetics differentiating the two disorders may be driven primarily by common variants (because the rare variant risk load is similar for the two disorders in the data available so far) and more extensively for cases with ID than without ID (because the common variant genetic correlation is lower for cases with ID). However, larger sample sizes for both GWAS and sequencing studies are needed to clarify this.

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

Methods

We report results from a framework of different analyses all carried out in large-scale samples from the Psychiatric Genomics Consortium (PGC) and the Lundbeck initiative of integrative psychiatric research (iPSYCH) samples. We used samples included in the most recent GWAS of ASD [5] and ADHD [6]. For the purpose of this manuscript we will refer to individuals in the study cohort (most importantly in the iPSYCH cohort) that at the time of inclusion only had one of the two diagnoses registered (i.e., ADHD or ASD) as *ADHD-only* and *ASD-only* cases, respectively. We refer to individuals that during their lifetime and up to the time of inclusion had both an ADHD and ASD diagnosis registered as *comorbid* cases. Furthermore, we refer to these three groups of cases (i.e., *ADHD-only*, *ASD-only*, and *comorbid*) as *ASD and ADHD subgroups*.

Sample description and additional quality control

Details about study specific case and control selection criteria and how individuals were drawn from the overall population-based iPSYCH case-cohort sample [62] can be found in the respective publications [5,6]. Here we focus on important differences in the case and control selection criteria in the iPSYCH cohort as well as additional quality control (QC) procedures necessary for the current study.

Almost all of the inclusion and exclusion criteria for the original studies were also used in this study. The only difference compared to the original studies was an additional exclusion criterion that removed individuals with a moderate to severe mental retardation (ICD10: F71-F79) from both the case and control cohorts. While this criterion was also used in the original ADHD GWAS [6], it was, however, 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 (see **Table 3**, **Supplementary Figure S14b**, and **Supplementary Table S7**).

Wave-wise pre-imputation QC and imputation of the iPSYCH case-cohort sample were taken from the original ADHD and ASD GWAS, respectively. Details about the respective steps and filters can be found elsewhere [5,6]. Since our analysis framework used a combined study cohort with samples from both the original ADHD and ASD GWAS we performed some additional QC on the combined sample. The additional QC steps included the removal of related individuals across the original ADHD and ASD GWAS as well as a new principal component analysis (PCA) on the combined sample after exclusion of these related individuals. Following the same procedures as in the original studies, pairs of subjects were identified with pi-hat> 0.2 (using PLINK's [63] identity by state analysis) and one subject of each such pair was excluded at random (with a preference for keeping cases). PCA was carried out using smartPCA in the EIGENSOFT software package [64,65] using the framework of the Ricopili pipeline [66]. The original PGC datasets for ADHD and ASD did not include overlapping individuals and therefore the original datasets and summary statistics were used. The final combined dataset across all samples comprised 34,462 cases (i.e., individuals with an ADHD and/or ASD diagnosis) and 41,201 controls. We only included samples of European ancestry from the original ADHD and ASD GWAS. Among the cases in the iPSYCH cohort 11,964 had an ADHD-only diagnosis, 9,315 had an ASD-only diagnosis, and 2,304 individuals had a comorbid diagnosis. 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

Like with the original GWAS in ADHD and ASD, all processing and analyses for the individual GWAS and meta-analyses (see below) used the framework of the Ricopili pipeline [66]. More details on individual modules and steps can be found elsewhere [5,6,66]. We ran two main GWAS within our framework of analyses. The first one aimed to identify shared genetic risk for ADHD and ASD (*combined GWAS*) and the second one aimed to identify differentiating genetic risk with an opposite direction of effects for ADHD and ASD (*ADHD vs ASD GWAS*). All analyses of the iPSYCH sample and meta-analyses with the

PGC samples were conducted at the secured national GenomeDK high-performance computing cluster in Denmark. The study was approved by the Regional Scientific Ethics Committee in Denmark and the Danish Data Protection Agency.

Combined GWAS

We first ran an analysis in the combined dataset, i.e., on all 34,462 cases and 41,201 controls. The GWAS was conducted in each cohort (i.e. in the wave-wise iPSYCH samples and the individual PGC cohorts) using logistic regression with the imputed additive genotype dosages. The first 5 principal components (PCs) were included as covariates to correct for population stratification (**Supplementary Information**), and variants with imputation INFO score < 0.8 or minor allele frequency (MAF) < 0.01 were excluded. The resulting summary statistic files were then meta-analyzed using an inverse-variance weighted fixed effects model [67]. Post-processing of the summary statistics files through the Ricopili pipeline [66] created Manhattan plots, individual regional associations plots, and forest plots. For a QQ-plot of the analysis please refer to **Supplemental Figure S14a**.

ADHD vs ASD GWAS

To identify unique genetic risk loci or loci with opposite direction of effects for ADHD and ASD we ran a case-only analysis for the ADHD-only (coded as 1, i.e., "pseudo-cases"; n = 11,964) against ASD-only cases (coded as 2, i.e., "pseudo-controls"; n = 9,315) in the iPSYCH cohort. This approach is in line with our recent study that compares the genetic risk to develop bipolar disorder and schizophrenia[68]. We excluded the comorbid cases from this GWAS. Similar to the analysis in the combined sample (see above) GWAS was conducted wave-wise using logistic regression with the imputed additive genotype dosages. The first 5 PCs were included as covariates to correct for population stratification, and variants with imputation INFO score < 0.8 or MAF < 0.01 were excluded. The resulting summary statistic files were then meta-analyzed using an inverse-variance weighted fixed effects model [67] and visualization of

results was achieved through the Ricopili pipeline [66]. Post-processing of the summary statistics files through the Ricopili pipeline [66] created Manhattan plots, individual regional associations plots, and forest plots. For a QQ-plot of the analysis please refer to **Supplemental Figure S14b**.

Identification of previously reported associations for top findings

Different resources were used to look up previously reported associations of our top findings with other phenotypes and traits within and outside of psychiatry. We assessed associations reported in the open GWAS project database (https://gwas.mrcieu.ac.uk/about/, accessed Oct 14th 2020; see **Supplementary Table S1a** and **Supplementary Table S5** for results) and also used the GWAS ATLAS website [69] to visualize PheWAS analyses (see **Supplementary Figures S2** and **S6**). We also used results from the GWAS Catalog [70] (see **Table 2**). Finally, we also compared our results with previous cross-disorder studies in the field. This 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[71], and a study that used conditional analyses to highlight associations that might be specific for individual psychiatric disorders[72]. Results are available in the **Supplemental Material and Methods** and **Supplemental Table S1b**.

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

Transcriptomic imputation models were constructed as previously described [20] for dorso-lateral prefrontal cortex (DLPFC) transcript levels [73]. The genetic dataset of the PsychENCODE cohort was uniformly processed for quality control (QC) steps before genotype imputation. We restricted our analysis to samples of European ancestry as previously described [20]. Genotypes were imputed using the University of Michigan server [74] with the Haplotype Reference Consortium (HRC) reference panel [75]. Gene expression information (both at the level of gene and transcript) was derived from RNA-seq counts which are adjusted for known and hidden confounds, followed by quantile normalization [73]. 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 [76]. EpiXcan was recently shown to increase power to identify genes under a causality model when compared to TWAS approaches that don't integrate epigenetic information [77]. We use this model (924 samples from DLPFC) due to power considerations [20]; in comparison, brain gene expression imputation models based on GTEx V8 [78] are trained in 205 or fewer samples. Using only samples from DLPFC, we acknowledge that ADHD and ASD are both also associated with other brain regions and would like to 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 applied the S-PrediXcan method [20] to integrate the GWAS summary statistics and the transcriptomic imputation models constructed above to obtain association results at both the level of genes and transcripts.

Cell-type enrichment analysis

A major portion of cell type specific enrichment is attributed to distal regulatory elements, as local regulatory events remain highly consistent across various tissues and cell types [79]. Therefore, we examined an overlap of common genetic variants of investigated traits (see **Supplemental Figure S14** and **Supplemental Table S6**) and open chromatin from scATAC-seq study (single-cell assay for transposase accessible chromatin) [33] using the LD-score partitioned heritability approach [80]. All regions of open chromatin were extended by 500 base pairs in either direction. The broad MHC-region (hg19 chr6:25-35MB) was excluded due to its extensive and complex LD structure, but otherwise default parameters were used for the algorithm.

Additional functional characterization and annotation of main findings

We used a number of different approaches combining in-house scripts and data with those available via the FUMA v1.3.6a [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 their website. Finally, we used FUMA to perform cell-type specificity analyses [81] based on our summary statistics. For all above mentioned analyses default settings were applied. More detailed information about the individual third-party datasets (available through FUMA) included in the analyses as well as individual aspects of the FUMA analyses can be found in the **Supplemental information**. **Supplemental Table S8** contains results from standard FUMA-based analyses, such as eQTL and chromatin interaction mapping.

Gene-based analysis

We also used FUMA v1.3.6a [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 **Supplemental information**.

Our results in context of other findings

Since the publication of the original ADHD and ASD results a few studies have investigated the shared and unique risk architecture of these disorders. We compared our results with the findings of the cross disorder working group of the PGC [8] and a recent analysis based on structural equation modelling of 11 major psychiatric disorders [https://doi.org/10.1101/2020.09.22.20196089]. We also compared our results with recent analyses that aimed at identifying disorder-specific SNPs for psychiatric disorders [24,72].

Polygenic risk score (PRS) analyses

To examine potential polygenic heterogeneity across ADHD and ASD subtypes, we investigated how PRS trained on different phenotypes were distributed across ADHD-only, ASD-only and comorbid subgroups in the iPSYCH data through two complementary analysis frameworks: multivariate PRS and

leave-one-out PRS. These two approaches have different strengths and limitations, allowing for robust interrogation of differences in ADHD and ASD subgroups in terms of polygenic burden for ADHD and ASD, as well as genetically related phenotypes.

Multivariate PRS analyses

To examine the relative burden of PRS for phenotypes and traits that have shown significant genetic correlation with ADHD and ASD in the past [5,6,38] 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 PRSs 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 a number of phenotypes while accounting for the inherent correlation between scores and adjusting for necessary covariates. This enables testing a whole array of hypotheses comparing both between subgroups and between PRSs. In particular we can compare groups that are too small for GWAS and gauge genetic correlation with groups that are too small for LDSC, as is the case with the comorbid ASD-ADHD group. Polygenic scores were generated by clumping and thresholding employing standard Ricopili settings as explained in [5] and using summary statistics from the GWASs [5,37,82-91].

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 vs. controls and ASD-only (excluding comorbid ADHD or severe ID) cases vs. controls, using independent (split) controls. For each set of discovery results, LD-clumping was run in PLINK v.1.9 [92] (with the parameters --clump-kb 500 --clump-r2 0.3) to obtain a relatively independent set of SNPs, while retaining the most significant SNP in each LD block. The SNP selection p-value threshold used was $p < 0.5$. Asymmetric/ambiguous SNPs (AT, TA, CG, GC), indels, multi-allelic and duplicate position SNPs were excluded. SNPs with MAF < 0.01, INFO < 0.8 or present in less than half of the sample were filtered out. PRS for ADHD and ASD were calculated for each individual in each target wave by scoring the number of effect alleles weighted by the log(odds ratio [OR]) across the set of independent clumped, meta-analyzed SNPs in PLINK. PRS were derived in best guess imputed data after filtering out SNPs with MAF < 0.05 and INFO < 0.8. The PRS were standardized using z-score transformations; ORs can be interpreted as the increase in risk of the outcome, per standard deviation in PRS. Logistic regression analyses including 5 PCs were run to test for association of PRS with each of the outcomes within each wave, as follows: a) ADHD-only cases vs. controls, b) ASD-only cases vs. controls, c) comorbid cases vs. controls, d) ADHD-only cases vs. ASD-only cases, e) ADHD-only cases vs. comorbid cases, and f) ASD-only cases vs. comorbid cases. Cases were coded as 1 and controls as 0, except that comorbid cases were coded as 1 in case-case comparisons and in analysis (d), the ASD-only cases were coded as 1. Overall meta-analyses of these per-wave analyses were performed in R using the 'metafor' package. As secondary tests, we stratified the ADHD-only and ASD-only cases by presence of mild ID (defined as IQ between 50-70). We also examined differences across several ASD hierarchical subtypes (childhood autism, atypical autism, Asperger's, and pervasive developmental disorders mixed; see Grove et al [5] and **Supplemental Table S7**). Several sensitivity tests were also run (including sex as a covariate, excluding cases and controls with mild ID).

Genetic correlations (LDhub)

The genetic correlations of our different datasets with other phenotypes were evaluated using LD Score regression (LDSC) [31] and the LD Hub [30] website (http://ldsc.broadinstitute.org/ldhub/). In brief, we re-reran analyses of the original GWAS of ADHD and ASD [5,6] in the European-only datasets since new phenotypes have been added to LD Hub after publication of the original analyses. We also uploaded summary statistics for the two analyses described above, i.e., the GWAS in the combined sample (combined GWAS) and the pseudo case-control analysis (ADHD vs ASD GWAS), to assess correlation with the identified shared and differentiating genetic liability, respectively. We used all available phenotypes in LD Hub [30] but performed analyses for the UKBB traits (n = 597) and the remaining individual phenotypes (n = 257) separately. For ADHD [6] and ASD [5] the most recent summary statistics replaced corresponding summary statistics in LD Hub as these had not been included at the date of analysis. The same was true for the summary statistics of major depressive disorder (MDD [87]) and bipolar disorder (BD [93]). Levels of experiment-wide significance (Bonferroni correction for number of tests applied) were also established separately within the two groups, i.e., in the UKBB traits ($p < 8.38 \times 10^{-5}$) 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 [39] for ADHD versus ASD and for ADHD versus each of the ASD sub-phenotypes listed in **Table 1**. The genetic relationship matrix (GRM) between all pairwise combinations of individuals was estimated using all case-control samples. The strict best-guess-genotypes (i.e. SNPs with INFO > 0.8, missing rate < 0.01 and MAF > 0.05, INDELs removed) were used for GRM estimation. GCTA-GREML accounts for linkage disequilibrium (LD) [94], and the GRM estimation was therefore performed on a non-LD-pruned dataset. Estimation of the phenotypic variance explained by the SNPs was performed for each of the phenotypes listed in

Supplemental Table S7, with PCs 1-20 included as continuous covariates and wave (1-23) as categorical dummy variables. ADHD prevalence of 0.05 and ASD prevalence of 0.01 was assumed to estimate the variance explained on the liability scale. Prevalence was estimated for hierarchical ASD phenotypes based on the estimate for the overall ASD phenotype and the proportion of each hierarchical phenotype over all ASD cases observed in our sample. Genetic covariance between pairs of traits (**Supplemental Table S7**) was estimated using the bivariate approach implemented in GCTA, by randomly splitting controls into two groups, one for each trait, in proportions corresponding to the proportion of the cases for each of the two traits in the total sample. PCs 1-20 and dummy variables for wave 1-23 were included as covariates in the bivariate analyses. Two-tailed p-values were obtained for r_G point estimates based on the standard error estimated by GCTA using the approach by Altman and Bland [95].

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

Acknowledgements

The iPSYCH team was supported by grants from the Lundbeck Foundation (R102-A9118, R155-2014-1724, and R248-2017-2003), the EU H2020 Program (Grant No. 667302, "CoCA"), NIMH (1U01MH109514-01 to ADB) and the Universities and University Hospitals of Aarhus and Copenhagen. The Danish National Biobank resource was supported by the Novo Nordisk Foundation. High-performance computer capacity for handling and statistical analysis of iPSYCH data on the GenomeDK HPC facility was provided by the Center for Genomics and Personalized Medicine and the Centre for

Integrative Sequencing, iSEQ, Aarhus University, Denmark (grant to ADB). Dr. Valodakis was funded by NIMH grant K08MH122911 and the 2020 NARSAD Young Investigator Grant #29350 from the Brain & Behavior Research Foundation. Support for this research was further received by Dr. Cormand from the Spanish 'Ministerio de Ciencia, Innovación y Universidades' (RTI2018-100968-B-100), 'Ministerio de Ciencia e Innovación~~Economía y Competitividad~~' (PID2021-1277760B-I00), 'AGAUR/Generalitat de Catalunya' (2017-SGR-738), the European Union H2020 Program [H2020/2014-2020] under grant agreements n°667302 (CoCA), 643051 (MiND) and 728018 (Eat2beNICE), and the ECNP network "ADHD across the lifespan". Dr. Faraone is supported by the European Union's Horizon 2020 research and innovation programme under grant agreements No 667302 and 965381; NIMH grants U01MH109536-01, U01AR076092-01A1, R0MH116037 and 5R01AG06495502; Oregon Health and Science University, Otsuka Pharmaceuticals and Supernus Pharmaceutical Company. Dr Roussos is supported by the National Institute of Mental Health, NIH grants, U01-MH116442 and R01-MH125246. The authors would like to deeply thank all participants in the cohorts included in this analysis.

Data and code availability

Summary statistics from this publication will be made available at http://ipsych.au.dk/downloads/ and https://www.med.unc.edu/pgc/ upon acceptance of the manuscript. Summary statistics for the original ADHD and ASD GWAS analyses are already available at these sites. For access to genotypes from the PGC samples and the iPSYCH sample, researchers should contact the lead PIs Elise Robinson / Anders Børglum and Anders Børglum for PGC-ASD and iPSYCH-ASD, respectively and Benjamin Neale / Barbara Franke and Anders Børglum for PGC-ADHD and iPSYCH-ADHD, respectively. In house scripts and code used for the processing and manipulation of the data can be obtained from the authors upon request.

Author contributions

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

Competing interests

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

References

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

18. Yang, Z. *et al.* Investigating Shared Genetic Basis Across Tourette Syndrome and Comorbid Neurodevelopmental Disorders Along the Impulsivity-Compulsivity Spectrum. *Biol Psychiatry* (2021).
19. Sabourdy, F. *et al.* A MANBA mutation resulting in residual beta-mannosidase activity associated with severe leukoencephalopathy: a possible pseudodeficiency variant. *BMC Med Genet* **10**, 84 (2009).
20. Zhang, W. *et al.* Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. *Nat Commun* **10**, 3834 (2019).
21. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**(2018).
22. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).
23. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
24. Peyrot, W.J. & Price, A.L. Identifying loci with different allele frequencies among cases of eight psychiatric disorders using CC-GWAS. *Nat Genet* (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**, 487-98 (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. Subramanian, K. *et al.* Basal ganglia and autism - a translational perspective. *Autism Res* **10**, 1751-1775 (2017).
29. Baselmans, B.M.L. *et al.* Multivariate genome-wide analyses of the well-being spectrum. *Nat Genet* **51**, 445-451 (2019).
30. 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).
31. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
32. La Manno, G. *et al.* Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell* **167**, 566-580 e19 (2016).
33. 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).
34. 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).
35. Wu, Z.M. *et al.* White Matter Microstructural Alterations in Children with ADHD: Categorical and Dimensional Perspectives. *Neuropsychopharmacology* **42**, 572-580 (2017).

36. Aoki, Y. *et al.* Association of White Matter Structure With Autism Spectrum Disorder and Attention-Deficit/Hyperactivity Disorder. *JAMA Psychiatry* **74**, 1120-1128 (2017).
37. Neale, B.M. *et al.* Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* **49**, 884-97 (2010).
38. 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).
39. 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).
40. 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).
41. 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).
42. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209-15 (2014).
43. Bryant, L. *et al.* Histone H3.3 beyond cancer: Germline mutations in Histone 3 Family 3A and 3B cause a previously unidentified neurodegenerative disorder in 46 patients. *Sci Adv* **6**(2020).
44. Clarke, T.K. *et al.* Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. *Mol Psychiatry* **21**, 419-25 (2016).
45. 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).
46. 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).
47. Shaw, P. *et al.* A multicohort, longitudinal study of cerebellar development in attention deficit hyperactivity disorder. *J Child Psychol Psychiatry* **59**, 1114-1123 (2018).
48. Wolfers, T. *et al.* Individual differences v. the average patient: mapping the heterogeneity in ADHD using normative models. *Psychol Med* **50**, 314-323 (2020).
49. Hwang, W.J., Lee, T.Y., Kim, N.S. & Kwon, J.S. The Role of Estrogen Receptors and Their Signaling across Psychiatric Disorders. *Int J Mol Sci* **22**(2020).
50. Fliers, E. *et al.* Motor coordination problems in children and adolescents with ADHD rated by parents and teachers: effects of age and gender. *J Neural Transm (Vienna)* **115**, 211-20 (2008).
51. 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).
52. Basile, G.A. *et al.* Red nucleus structure and function: from anatomy to clinical neurosciences. *Brain Struct Funct* **226**, 69-91 (2021).
53. Dalsgaard, S., Nielsen, H.S. & Simonsen, M. Five-fold increase in national prevalence rates of attention-deficit/hyperactivity disorder medications for children and adolescents with autism spectrum disorder, attention-deficit/hyperactivity disorder, and other psychiatric disorders: a Danish register-based study. *J Child Adolesc Psychopharmacol* **23**, 432-9 (2013).

54. Rosenberg, R.E. *et al.* Psychotropic medication use among children with autism spectrum disorders enrolled in a national registry, 2007-2008. *J Autism Dev Disord* **40**, 342-51 (2010).
55. 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).
56. 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).
57. Chang, Z. *et al.* Medication for Attention-Deficit/Hyperactivity Disorder and Risk for Suicide Attempts. *Biol Psychiatry* **88**, 452-458 (2020).
58. Keilow, M., Holm, A. & Fallesen, P. Medical treatment of Attention Deficit/Hyperactivity Disorder (ADHD) and children's academic performance. *PLoS One* **13**, e0207905 (2018).
59. Brainstorm, C. *et al.* Analysis of shared heritability in common disorders of the brain. *Science* **360**(2018).
60. 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).
61. Ronald, A., Larsson, H., Anckarsater, H. & Lichtenstein, P. Symptoms of autism and ADHD: a Swedish twin study examining their overlap. *J Abnorm Psychol* **123**, 440-51 (2014).
62. 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).
63. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
64. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet* **2**, e190 (2006).
65. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
66. Lam, M. *et al.* RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics* **36**, 930-933 (2020).
67. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
68. Bipolar, D., Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic address, d.r.v.e., Bipolar, D. & Schizophrenia Working Group of the Psychiatric Genomics, C. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell* **173**, 1705-1715 e16 (2018).
69. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* **51**, 1339-1348 (2019).
70. 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).
71. 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).

72. Byrne, E.M. *et al.* Conditional GWAS analysis to identify disorder-specific SNPs for psychiatric disorders. *Mol Psychiatry* (2020).
73. Gandal, M.J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**(2018).
74. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).
75. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).
76. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
77. Cao, C. *et al.* Power analysis of transcriptome-wide association study: Implications for practical protocol choice. *PLoS Genet* **17**, e1009405 (2021).
78. Consortium, G.T. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330 (2020).
79. 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).
80. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
81. 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).
82. Davies, G. *et al.* Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). *Mol Psychiatry* **21**, 758-67 (2016).
83. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539-42 (2016).
84. Benyamin, B. *et al.* Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Mol Psychiatry* **19**, 253-8 (2014).
85. 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).
86. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-7 (2014).
87. 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).
88. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet* **48**, 624-33 (2016).
89. 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).
90. Deary, V. *et al.* Genetic contributions to self-reported tiredness. *Mol Psychiatry* **23**, 609-620 (2018).
91. Tobacco & Genetics, C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* **42**, 441-7 (2010).
92. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).

93. Stahl, E.A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* **51**, 793-803 (2019).
94. Yang, J., Lee, S.H., Wray, N.R., Goddard, M.E. & Visscher, P.M. GCTA-GREML accounts for linkage disequilibrium when estimating genetic variance from genome-wide SNPs. *Proc Natl Acad Sci U S A* **113**, E4579-80 (2016).
95. Altman, D.G. & Bland, J.M. How to obtain the confidence interval from a P value. *BMJ* **343**, d2090 (2011).

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

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

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 (PMID 30478444) and ASD refers to results from the previously published GWAS on ASD (PMID 30804558). Results from lookups in the open GWAS project database (https://gwas.mrcieu.ac.uk/about/, accessed Oct 14th 2020) are available in **Supplemental Table S1a** and as PheWAS plots in **Supplemental Figure S2**. SNP (#CS) – marker name and number of reported GWASs where this marker is in the 95% credible set in (FINEMAP/PAINTOR/CAVIARBF) according to http://mulinlab.org/causaldb/, please note that SNPs do not need to be genome-wide significant in those reported GWASs to be in the list of credible SNPs. SNPs representing novel shared loci for ASD and ADHD are highlighted in bold; CHR – chromosome; BP – Base pair position on the chromosome; A1 – effect allele; A2 – other allele; FRQ_{ca} – Frequency in the cases; FRQ_{co} – Frequency in the controls; OR – Odds ratio based on effect allele; P – *P* value for association results; GENES – protein coding genes and/or microRNAs in a LD region around lead SNP ($r^2 = 0.6$), in case no protein coding gene or microRNA is present in the region the nearest protein coding gene or microRNA within a 100kb window around the LD region is provided together with the distance in kb (if there is no gene present a "-" will be shown); OTHER – previously reported associations with the lead SNP (underlined letters) or other SNPs (italic letters) in LD with the lead SNP ($r^2 = 0.6$), reported *P* values needed to be genome-wide significant to be listed. In case of the ASD and ADHD *P* values these are the *P* values in the original GWAS. Please note that the OR and the *P* values reported for the ADHD and ASD GWASs both times include the comorbid cases (i.e. in each of the two GWASs) as well as related individuals across studies. Markers highlighted in bold letters indicate previously unidentified associations with either of the two disorders (ADHD / ASD). [1]ADHD (PMID 30478444); [2]Cross Disorder GWAS in the PGC (PMID 31835028); Educational attainment (years of education, PMID 30038396), Intelligence (MTAG, PMID 29326435), Adventurousness (PMID 30643258), Feeling worry (neuroticism item; 29500382), Household income (PMID 31844048), Balding type 1 (PMID 30595370), Number of

sexual partners (30643258); [3]Alcohol consumption (PMIDs 30643258, 31358974, 30643251); [4]Neuroticism (PMID 29942085), Worry (neuroticism item; PMID 29942085); [5]Attention deficit hyperactivity disorder or cannabis use (PMID 30610198), [6]Cross Disorder GWAS in the PGC (PMID 31835028); [7]Number of sexual partners (PMID 30643258); [8]Cross disorder GWAS for TS-ADHD-ASD (PMID 33714545), [9]Fat-free mass (PMID 30593698), Appendicular lean mass (PMID 31761296), Height (PMIDs 30595370 and 25282103);[10]Asthma and attention deficit hyperactivity disorder (PMID 31619474); [11]Blood protein levels (PMID 29875488); [12]Autism and major depressive disorder (MTAG, PMID 30804558); [13]Educational attainment (PMID 30038396), Life satisfaction (PMID 30643256), Well-being spectrum (multivariate analysis, PMIDs 30643256 and 29292387), Depressive symptoms (PMIDs 30643256 and 29292387), Neuroticism (PMID 29292387), Positive affect (PMID 30643256), Loneliness (PMID 31518406), Asthma and attention deficit hyperactivity disorder (PMID 31619474), Asthma and major depressive disorder (PMID 31619474), Insomnia (PMIDs 30804566 and 30804565), Risk-taking tendency (4-domain principal component model, PMID 30643258), BMI (PMIDs 31669095, 30595370, 30239722), Highest math class taken (PMID 30038396), Hand grip strength (PMID 29691431), Predicted visceral adipose tissue (PMID 31501611).

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

					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	$4.28 \cdot 10^{-9}$	0.956	0.0484	1.122	$9.33 \cdot 10^{-8}$	*SPAG16*	ADHD[2], CDG[3]
rs7821914 (3/5/5)	8	10805015	T	C	0.584	0.556	1.127	$4.58 \cdot 10^{-9}$	0.935	$1.86 \cdot 10^{-6}$	1.022	0.1113	*XKR6*	Neuroticism[4], Many[5]
rs147420422 (16/17/17)	2	104139422	CAT	C	0.529	0.502	1.118	$3.37 \cdot 10^{-8}$	0.947	$6.89 \cdot 10^{-5}$	1.036	0.0092	-	Neuroticism[6], Many[7]
rs3791033[10] (6/7/6)	1	44134077	T	C	0.681	0.656	1.124	$3.98 \cdot 10^{-8}$	0.979	0.1407	1.095	$2.76 \cdot 10^{-10}$	*PTPRF*, *KDM4A*, *ST3GAL3*, *MIR6079*	EA[8], ADHD[9], Many[10]
rs9379833 (58/59/58)	6	26207175	A	C	0.251	0.275	0.884	$4.51 \cdot 10^{-8}$	1.041	0.0102	0.949	0.0007	HIST1[1]	EA[8], Neuroticism[11], Height[12], Many[13]

Results shown in the table are for three different GWAS: ADHDvsASD refers to our ADHD vs ASD GWAS described in the main text body of this manuscript, ADHD refers to results from the previously published GWAS on ADHD (PMID 30478444) and ASD refers to results from the previously published GWAS on ASD (PMID 30804558). Results from lookups in the open GWAS project database (https://gwas.mrcieu.ac.uk/about/, accessed Oct 14th 2020) are available in **Supplemental Table S5** and as PheWAS plots in **Supplemental Figure S11**. SNP (#CS) – marker name and number of reported GWASs where this marker is in the 95% credible set in (FINEMAP/PAINTOR/CAVIARBF) according to http://mulinlab.org/causaldb/, please note that SNPs do not need to be genome-wide significant in those reported GWASs to be in the list of credible SNPs. If instead of a number "NA" appears this means the SNP has not been reported in a credible set before. SNPs highlighted in bold have not been identified in GWASs of ADHD and ASD before; CHR – chromosome; BP – Base pair

position on the chromosome; A1 – effect allele; A2 – other allele; FRQ_{ADHD} – Frequency in the iPSYCH ADHD only cases; FRQ_{ASD} – Frequency in the iPSYCH ASD only cases; OR – Odds ratio based on effect allele; P – *P* value for association result; GENES – protein coding genes and/or microRNAs in a LD region around lead SNP ($r^2 = 0.6$)), in case no protein coding gene or microRNA is present in the region the nearest protein coding gene or microRNA within a 100kb window around the LD region is provided together with the distance in kb (if there is no gene present a "-" will be shown); OTHER – previously reported associations with the lead SNP or other SNPs in LD with the lead SNP ($r^2 = 0.6$), reported *P* values needed to be genome-wide significant to be listed. In case of the ASD and ADHD *P* values these are the *P* values in the original GWAS. Please note that the OR and the *P* values reported for the ADHD and ASD GWAS both times include the ADHD/ASD comorbid cases (i.e. in each of the two GWASs) as well as related individuals across studies. [1]Genes in the HIST1 region (PMID 12408966): *HIST1H1E*, *HIST1H2BD*, *HIST1H2BE*, *HIST1H4D*, *HIST1H3D*, *HIST1H2AD*, *HIST1H2BF*, *HIST1H4E*, *HIST1H2BG*, *HIST1H2AE*, *HIST1H3E*, *HIST1H1D*, *HIST1H4F*, *HIST1H4G*, *HIST1H3F*, *HIST1H2BH*. [2]ADHD GWAS (PMID 30478444); [3]Cross disorder GWAS (PMID 31835028); [4]General factor of Neuroticism (PMID 30867560), Neuroticism (PMIDs 29255261 and 30643256), [5]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); [6]Worry (neuroticism item; PMID 29942085), Feeling nervous (neuroticism item; PMID 29500382), Anxiety/tension (special factor of neuroticism; PMID 30867560); [7]Smoking related phenotypes (PMIDs 30617275, 30643251, 30643258, 30595370, 30679032), Number of sexual partners (PMID 30643258), Age at first sexual intercourse (PMID 27089180), Reaction time (PMID 29844566), Risk-taking tendency (4-domain principal component model, PMID 30643258), General risk tolerance (MTAG, PMID 30643258), BMI (PMID 30239722), Pneumonia (PMID 28928442), Photic sneeze reflex (PMID 27182965); [8]*Educational Attainment* (PMID 30038396); [9]ADHD GWAS (PMID 30478444), Attention deficit hyperactivity disorder or cannabis use (PMID 30610198), [10]*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), Menarche (age at onset, PMID 30595370), Red blood cell count (PMID 30595370), Height (PMID 30595370); [11]Worry too long after an embarrassing experience (neuroticism item; PMID 29500382); [12]Height (PMID 31562340) ; [13]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); [10]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).

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

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

Results for per wave polygenic risk score analyses. PRS_{ADHD} – Analyses using a polygenic risk score trained on an ADHD phenotype, PRS_{ASD} – Analyses using a polygenic risk score trained on an ASD phenotype. Cases – group coded as 1 (cases) for the purpose of the analyses, Comparison – other group coded as 0 for the purpose of the analyses, OR – Odds ratio, LCI – lower boundary for 95% confidence interval, UCI – upper boundary for 95% confidence interval, P – P-value. Groups are as follows: ADHD-only – cases with ADHD diagnosis and without comorbid ASD diagnosis, ASD-only – cases with ASD diagnosis and without comorbid ADHD diagnosis, Comorbid – cases with comorbid ADHD and ASD

diagnoses, Controls – individuals without ADHD and ASD diagnoses. P-values shown are without correction for multiple testing. Experiment-wide significant at 0.0042 (Bonferroni corrected for 2 x 6 tests). Additional secondary analyses also compare groups of individuals with ADHD or ASD with co-occurring mild intellectual disability (ADHD+ID and ASD+ID) to those without (ADHD-no-ID and ASD-no-ID).

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

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

Keratin 8 Pseudogene 46 (KRT8P46) (transcript, isoform 201): Located on chromosome 4 (**Supplemental Figure 3a**), *KRT8P46* is a pseudogene located in an intron of *MANBA*, a gene that has been previously associated with ADHD and asthma [1] as well as blood protein levels [2]. Pseudogenes have recently been highlighted as regulators in health and disease [3,4] amongst others through potential regulatory relationship with their parent genes. It is of note that another keratin 8 pseudogene (*KRT8P44*) is located in a region that has been identified to harbor a rare CNV associated with ADHD [5].

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

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

CAMKV (transcript, isoform 210): Located together with two other TWAS significant genes in a region on chromosome 3p21.31 (**Supplemental Figure 3c**), which has been reported to harbor CNVs in individuals with autism, intellectual disability and developmental delays [10-13]. The calmodulin kinase-like vesicle-associated (CaMKv) is a pseudokinase required for the activity-dependent maintenance of dendritic spines [14].

RFT1(transcript, isoform 204): Located on chromosome 3p21.1 (**Supplemental Figure 3d**) *RTF1* encodes an enzyme which catalyzes the translocation of the Man(5)GlcNAc (2)-PP-Dol intermediate from the cytoplasmic to the luminal side of the endoplasmic reticulum membrane in the pathway for the N-glycosylation of proteins. Mutations in *RFT1* cause recessive congenital disorder of glycosylation type 1N (CDG1N) [MIM:612015], which presents with a wide variety of clinical features, including severe developmental delay, hypotonia, dysmorphic features and epilepsy.

FAM167A (transcript, isoform 201): Located together with three other TWAS significant genes in the identified GWAS locus on chromosome 8 (**Supplemental Figure 3e**). Using data from a subset of our iPSYCH ASD samples, a previous study identified two differentially methylated positions (DMPs) in the same region that showed association with polygenic risk for ASD [15]. One of the DMPs was annotated to the Family With Sequence Similarity 167 Member A gene (*FAM167A*), while the other was annotated to *RP1L1*.

1. Zhu, Z. *et al.* Shared genetics of asthma and mental health disorders: a large-scale genome-wide cross-trait analysis. *Eur Respir J* **54**(2019).
2. Sun, B.B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73-79 (2018).
3. Pink, R.C. *et al.* Pseudogenes: pseudo-functional or key regulators in health and disease? *RNA* **17**, 792-8 (2011).
4. Cheetham, S.W., Faulkner, G.J. & Dinger, M.E. Overcoming challenges and dogmas to understand the functions of pseudogenes. *Nat Rev Genet* **21**, 191-201 (2020).
5. Jarick, I. *et al.* Genome-wide analysis of rare copy number variations reveals PARK2 as a candidate gene for attention-deficit/hyperactivity disorder. *Mol Psychiatry* **19**, 115-21 (2014).
6. Satterstrom, F.K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584 e23 (2020).
7. Duffney, L.J. *et al.* Epigenetics and autism spectrum disorder: A report of an autism case with mutation in H1 linker histone HIST1H1E and literature review. *Am J Med Genet B Neuropsychiatr Genet* **177**, 426-433 (2018).
8. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209-15 (2014).
9. Bryant, L. *et al.* Histone H3.3 beyond cancer: Germline mutations in Histone 3 Family 3A and 3B cause a previously unidentified neurodegenerative disorder in 46 patients. *Sci Adv* **6**(2020).

10. Stuart, S.W., King, C.H. & Pai, G.S. Autism spectrum disorder, Klinefelter syndrome, and chromosome 3p21.31 duplication: a case report. *MedGenMed* **9**, 60 (2007).

11. Haldeman-Englert, C.R. *et al.* A 3.1-Mb microdeletion of 3p21.31 associated with cortical blindness, cleft lip, CNS abnormalities, and developmental delay. *Eur J Med Genet* **52**, 265-8 (2009).

12. Eto, K. *et al.* Microdeletions of 3p21.31 characterized by developmental delay, distinctive features, elevated serum creatine kinase levels, and white matter involvement. *Am J Med Genet A* **161A**, 3049-56 (2013).

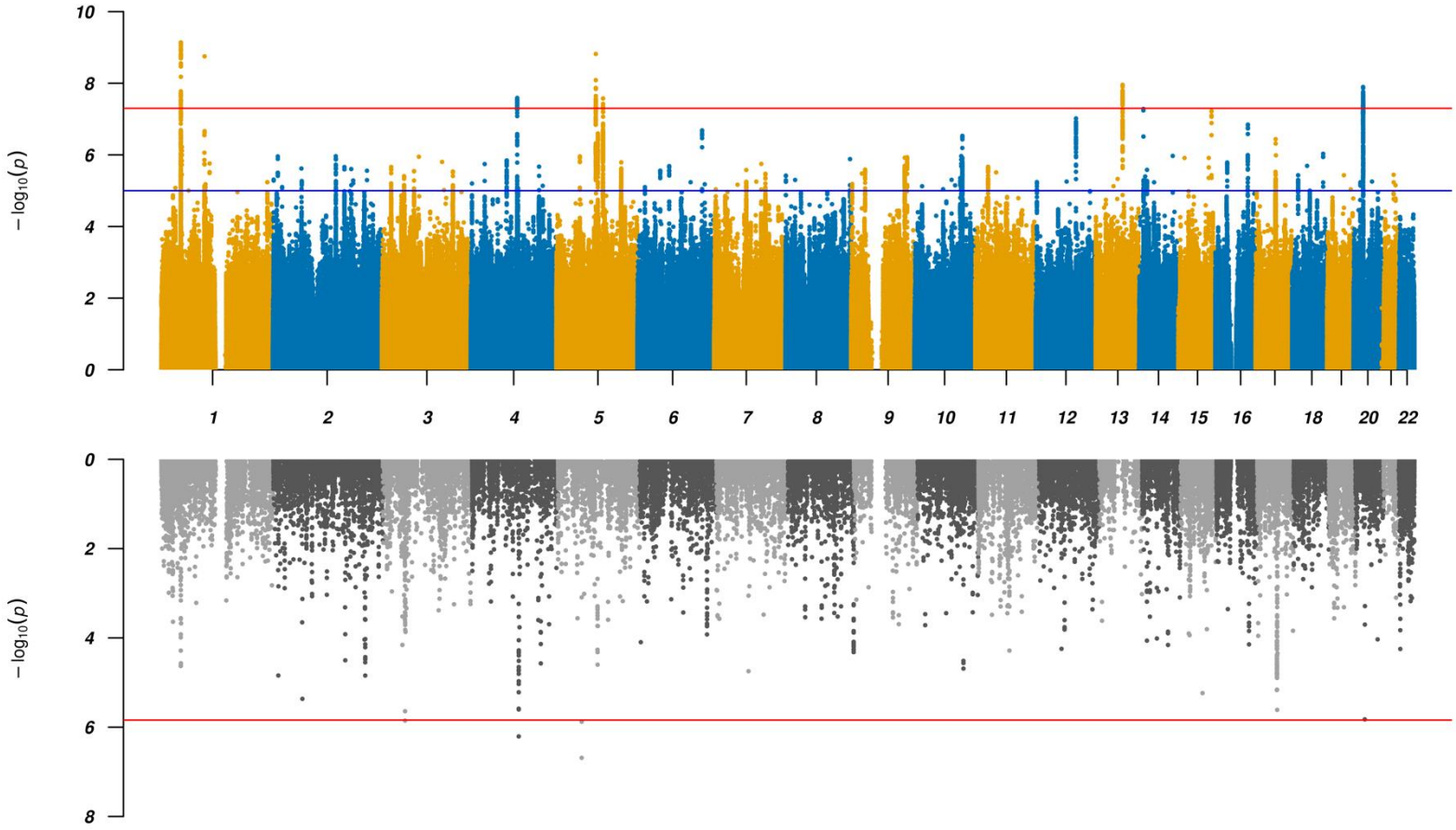
13. Battaglia, A. *et al.* Confirmation of chromosomal microarray as a first-tier clinical diagnostic test for individuals with developmental delay, intellectual disability, autism spectrum disorders and dysmorphic features. *Eur J Paediatr Neurol* **17**, 589-99 (2013).

14. Liang, Z. *et al.* The pseudokinase CaMKv is required for the activity-dependent maintenance of dendritic spines. *Nat Commun* **7**, 13282 (2016).

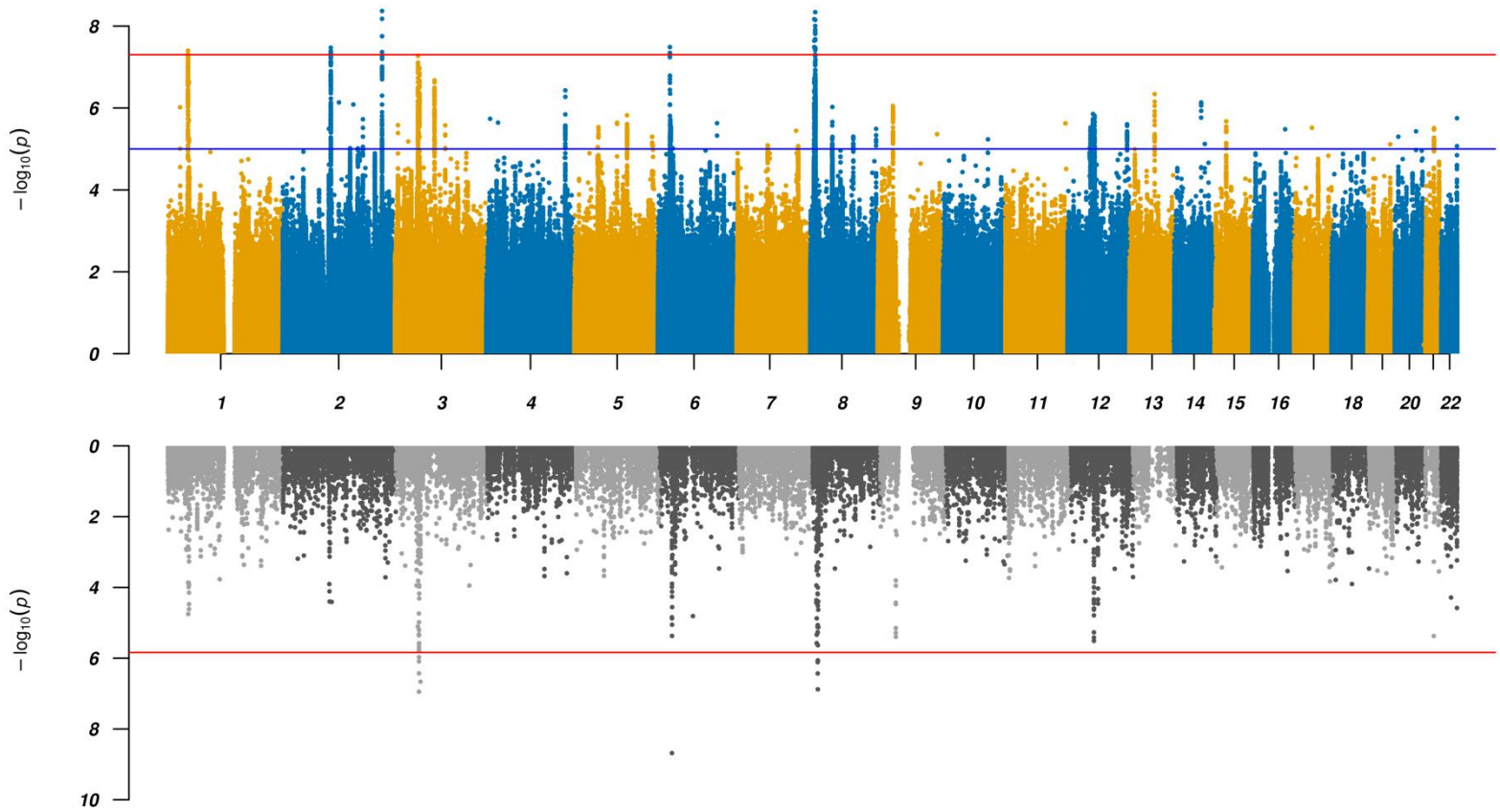
15. Hannon, E. *et al.* Elevated polygenic burden for autism is associated with differential DNA methylation at birth. *Genome Med* **10**, 19 (2018).

Figure 1: Miami plots of GWAS and TWAS results.

A)

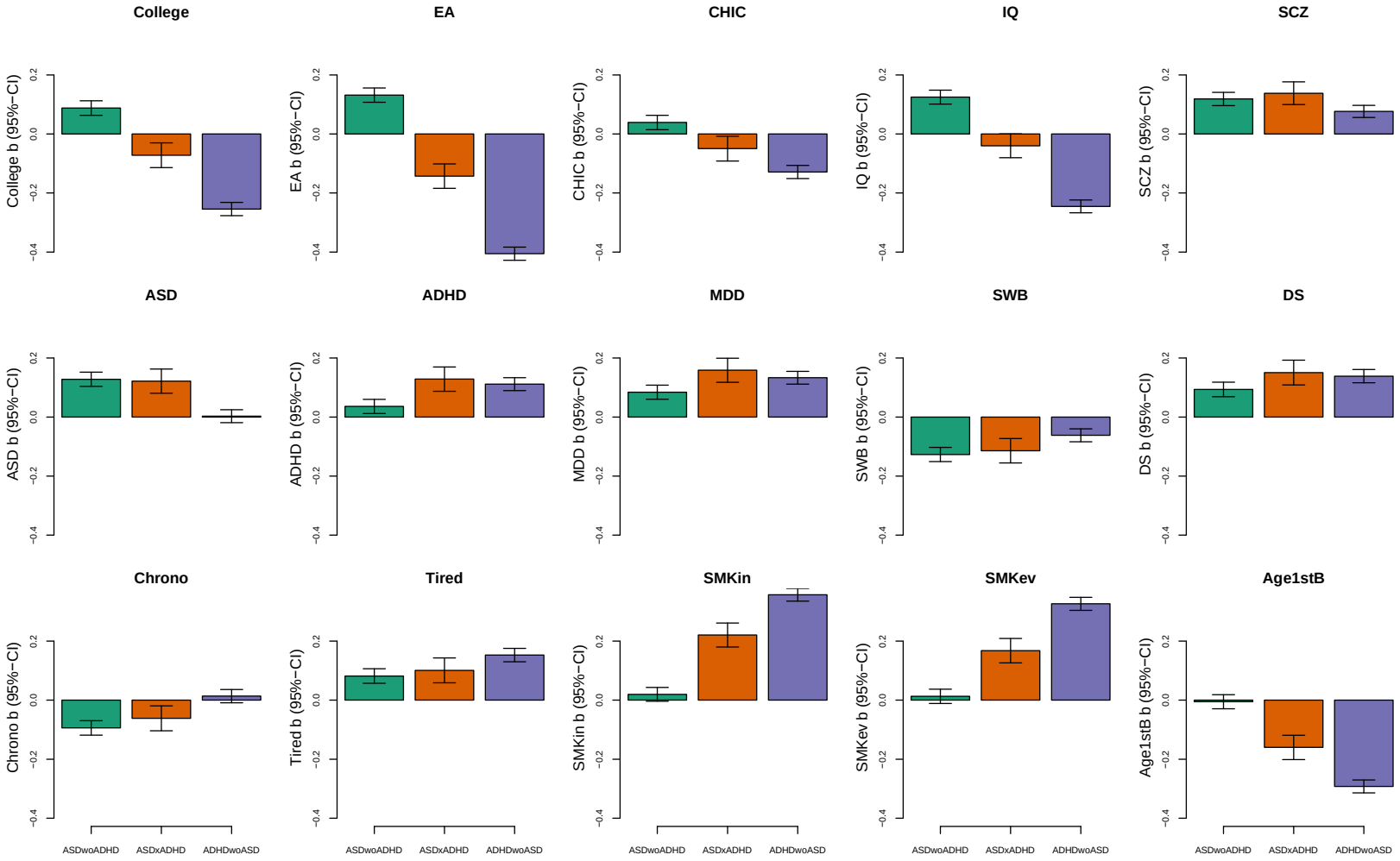


B)


8
6
4
2
0
− log10(p)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
18
20
22
0
2
4
6
8
10
− log10(p)

Results for GWAS (top panels) and TWAS results for DLPFC transcripts (bottom panels) for (A) combined and (B) ADHD vs ASD GWAS. In the top panel a blue line in the Manhattan plot indicates a p-value of $1x10^{-5}$, a red line a p-value of $5x10^{-8}$ (genome-wide significance). Each dot represents a tested SNP. In the bottom panel genes are represented by both gene expression and isoform expression (= features, represented by the dots). A red line indicates Bonferroni corrected genome-wide significance within analyses (combined or ADHD vs ASD; $p < 1.44x10^{-6}$; corresponding to Bonferroni correction of all the 34,646 features). We implement an imputation r2 filter (pred_perf_r2) of 0.01 in this study which means that at least 10% of the variance in expression of each gene can be explained by cis-heritability. Please also refer to the results in Supplemental Table S2.

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



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

Supplemental Information to:

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

Mattheisen, Grove, Als, Martin et al.

Table of Contents

Supplemental Material and Methods ***3***

A note on case-case comparisons and what they might mean **3**

Preliminary findings for a GWAS of comorbid cases. **4**

Functional characterization and annotation of main findings **5**

eQTL mapping 5

Chromatin Interaction mapping 6

Single Cell Analyses 6

A note on our approaches leveraging single-cell datasets to explore cell type specificity **7**

Supplemental Figure S1: Regional association plot for combined meta-analysis. ***9***

Supplemental Figure S2: PheWAS plots for associated SNPs from combined GWAS ***12***

Supplemental Figure S3: Regional Miami plots for combined and ADHD vs ASD GWASs. ***14***

Supplemental Figure S4: Manhattan Plot for gene-based analyses in main GWAS comparisons. ***17***

Supplemental Figure S5: Regional association plots for ADHD vs ASD GWAS. ***18***

Supplemental Figure S6: PheWAS plots for associated SNPs from ADHD vs ASD GWAS ***20***

Supplemental Figure S7: Genetic correlations between ADHD and ASD with other traits and disorders. ***22***

Supplemental Figure S8 MAGMA tissue expression analysis for combined GWAS. ***24***

Supplemental Figure S9 FUMA single-cell analyses for ASD, ADHD, combined, and differentiating GWAS. ***25***

Supplemental Figure S10 MAGMA tissue expression analysis for ADHD vs ASD GWAS. ***27***

Supplemental Figure S11: Single cell enrichment analysis for epigenomic peaks. ***28***

Supplemental Figure S12: Multivariate PRS analyses for Neuroticism subitems ***29***

Supplemental Figure S13: GCTA-based heritability estimates and genetic correlation for ASD (with subtypes) and ADHD ***30***

Supplemental Figure S14: QQ plots for combined and ADHD vs ASD GWASs ***32***

References ***33***

Supplemental Material and Methods

A note on case-case comparisons and what they might mean

A GWAS for a categorical trait (such as a psychiatric disorders), in its typical form (over-simplified) compares allele frequencies for markers across the genome between groups of individuals, usually cases and controls for categorical phenotypes. With regards to that, the categorical "phenotype" can be thought of as the variance in phenotypic representation between these two groups (i.e. between cases and controls). Heritability is usually (and again over-simplified) considered to be the proportion of that variance that can be explained by (additive) genetic effects. In case of a reported SNP heritability, it is the proportion of that variance that can be explained by (additive) genetic effects in common variants.

In at least one of our analyses (the "differentiating" GWAS), we compare allele frequencies between two case groups. Such case-case GWASs are not new, see e.g. [1,2]. In our analysis we compare a group of individuals that are ascertained for an ADHD diagnosis with a group of individuals that are ascertained for an ASD diagnosis. The "phenotype" as such is probably best interpreted as the variance in phenotypic representation between these two groups. Based on the ascertainment it is assumed that this variance (similar to a case-control analysis) is mostly explained by the factor(s) that determine(s) the ascertainment, i.e. the two diagnoses involved. Consequently, the heritability is therefore to be interpreted as the proportion of that variance that can be explained by (additive) genetic effects.

What might be less intuitive is to which degree the two variances in phenotypic representation (i.e. between cases and controls on one hand and between two case cohorts on the other hand) are similar. It can be generally assumed that some of the variance (and the genetic effects that cause them) are the same and some are unique. Translated into a GWAS context this means that the allele frequencies for some of the associated markers in a case-case GWAS will be different between all three groups when compared to their corresponding single disorder GWASs (here ADHD cases, ASD cases and controls) and for (the) other markers they will be the same for controls and one of the case groups but not the other (note that there will always be a difference between the two case groups if the marker is associated).

It is of note that heritability can be reported on different scales, usually an observed and a liability scale. We actually found the SNP heritability to be as high as 44% on the observed scale for the differentiating

(case-case) GWAS. We did not convert this into a liability scale heritability estimate (like for the combined GWAS SNP heritability) as such a conversion requires an estimate of population prevalence, which is not available for an abstract phenotype such as the difference between two case groups. Consequently, it is difficult to compare the observed scale heritability in this analysis with the liability scale heritability of our other analyses. Nevertheless, our analysis suggests that a substantial part of what differentiates the phenotypic representation in ADHD and ASD cases can be explained by common variants.

Preliminary findings for a GWAS of comorbid cases.

Based on the findings in our other analyses, especially the multivariate PRS analyses, we decided to run a GWAS of comorbid cases in the iPSYCH cohort against a subset of the available controls. There were 2,304 comorbid cases in the iPSYCH cohort, i.e. individuals with a lifetime diagnosis in the Danish register for ADHD and ASD. Controls were randomly selected from the full control cohort to roughly match a 1:4 ratio in cases and controls. Please note that as with the other analyses described in our manuscript, we excluded individuals with a moderate to severe mental retardation (ICD10: F71-F79) from both the case and control cohort. The main purpose of our analysis was to provide insight into individual disease associations for markers and regions that had been identified in our combined and differentiating GWASs. While reasonably sized compared to early day underpowered GWASs in psychiatry, we would like to refrain from reporting the full scan and instead focus on some preliminary findings. All analytical and processing steps followed the same protocols as with our other analyses.

We identified a genome-wide significant association on chromosome 6 (rs1321614, p = 3.54e-09, OR = 0.819, MAF = 0.47 for the T allele). This SNP showed only moderate evidence for association in the ASD only GWAS (p = 0.0086) and no evidence for association in the ADHD only GWAS (p = 0.7721). In both cases the direction of effects was the same as in the GWAS for the comorbid cases. In the overall combined GWAS (ADHD+ASD) this SNP showed a p-value of 0.0261 and a p-value of 0.2883 in the differentiating GWAS. A second locus (rs142703496) showed evidence for genome-wide significant association, however, based on MAF (0.02) and other factors we believe this signal to be a false positive association. Liability scale heritability for the GWAS per GCTA was 0.0557 (se = 0.0088). For completeness we would like to note that the locus identified in the GWAS for the comorbid cases was reported with a p-value of 7.46×10^{-4} in the PGC CDG GWAS (no23andMe version), a p-value of 1.67×10^{-4} in the ASD mtCOJO analysis (no other disorder showed a p-value less than 0.3)

and no case-case comparison in the CC-GWAS study reported a significant association for this SNP (with on average the strongest signals obtained for analyses that separated ASD from the other disorders).

While it is of note that the sample size of the comorbid GWAS (n = 2,304 cases) is substantially smaller compared to the ADHD and ASD only GWASs (n > 11,000 and 9,000 cases, respectively) point estimates of the effect sizes for the loci identified in the combined GWAS would indicate that for some of the associated SNPs the effects are indeed stronger for the comorbid GWAS compared to the other two GWASs (e.g. for the two lead SNPs on chromosome 5; see **Supplemental Table S1c**), but for other associated SNPs this is not the case (e.g. for lead SNP rs7538463 on chromosome 1; **Supplemental Table S1c**). For the SNPs identified in the differentiating GWAS some of the SNPs show effect sizes for the comorbid GWAS between effect sizes for the other two GWASs (e.g. lead SNP rs7821914 on chromosome 8; **Supplemental Table S1c**) while for others the effect sizes of the comorbid GWAS are more similar with one of the other two GWASs and substantially different to the other (e.g. lead SNP rs13023832 on chromosome 2; **Supplemental Table S1c**). These observations on the individual SNP level give further support to the results in our genetic correlation analyses, that not only the comorbid cases contribute to the genetic correlation between ADHD and ASD.

Functional characterization and annotation of main findings

We used the FUMA v1.3.6a [3] website (http://fuma.ctglab.nl) for downstream functional characterization and annotation of our findings. For all analyses mentioned in the manuscript default settings were applied. More detailed information on available datasets and analytical approaches are available on the website for FUMA (https://fuma.ctglab.nl/tutorial). Please also see Supplemental Table S8 for more details on respective default settings.

eQTL mapping

For eQTL mapping the following datasets available in FUMA were included (https://fuma.ctglab.nl/tutorial#eQTLs):

eQTLcatalogue/BrainSeq_ge_brain.txt.gz,

PsychENCODE/PsychENCODE_eQTLs.txt.gz,
scRNA_eQTLs/PBMC.txt.gz,
CMC/CMC_SVA_cis.txt.gz, CMC/CMC_SVA_trans.txt.gz, CMC/CMC_NoSVA_cis.txt.gz,
CMC/CMC_NoSVA_trans.txt.gz,
BRAINEAC/CRBL.txt.gz, BRAINEAC/FCTX.txt.gz, BRAINEAC/HIPP.txt.gz, BRAINEAC/MEDU.txt.gz,
BRAINEAC/OCTX.txt.gz, BRAINEAC/PUTM.txt.gz,
BRAINEAC/SNIG.txt.gz, BRAINEAC/TCTX.txt.gz, BRAINEAC/THAL.txt.gz,
BRAINEAC/WHMT.txt.gz, BRAINEAC/aveALL.txt.gz,
GTEx/v8/Cells_EBV-transformed_lymphocytes.txt.gz, GTEx/v8/Whole_Blood.txt.gz,
GTEx/v8/Brain_Amygdala.txt.gz, GTEx/v8/Brain_Anterior_cingulate_cortex_BA24.txt.gz,
GTEx/v8/Brain_Caudate_basal_ganglia.txt.gz, GTEx/v8/Brain_Cerebellar_Hemisphere.txt.gz,
GTEx/v8/Brain_Cerebellum.txt.gz, GTEx/v8/Brain_Cortex.txt.gz, GTEx/v8/Brain_Frontal_Cortex_BA9.txt.gz,
GTEx/v8/Brain_Hippocampus.txt.gz,
GTEx/v8/Brain_Hypothalamus.txt.gz, GTEx/v8/Brain_Nucleus_accumbens_basal_ganglia.txt.gz,
GTEx/v8/Brain_Putamen_basal_ganglia.txt.gz, GTEx/v8/Brain_Spinal_cord_cervical_c-1.txt.gz,
GTEx/v8/Brain_Substantia_nigra.txt.gz

No filtering (e.g. based on CADD scores or other available information) was applied.

Chromatin Interaction mapping

For chromatin interaction mapping the following datasets available in FUMA were used (https://fuma.ctglab.nl/tutorial#chromatin-interactions):

EP/PsychENCODE/EP_links_oneway.txt.gz, HiC/PsychENCODE/Promoter_anchored_loops.txt.gz,
HiC/Giusti-Rodriguez_et_al_2019/Adult_Cortex.txt.gz,
HiC/Giusti-Rodriguez_et_al_2019/Fetal_Cortex.txt.gz, HiC/GSE87112/Dorsolateral_Prefrontal_Cortex.txt.gz,
HiC/GSE87112/Hippocampus.txt.gz,
HiC/GSE87112/Neural_Progenitor_Cell.txt.g

Again, no posterior filtering was applied.

Single Cell Analyses

General details for single cell analyses within the FUMA framework can be found on the developer's website (https://fuma.ctglab.nl/tutorial#celltype). We used three of the available datasets within FUMA (https://fuma.ctglab.nl/tutorial#datasets):

PsychENCODE data for human developmental and adult brain samples [4].
GSE76381 data for human brain samples (ventral midbrain from 6-11 weeks embryos) [5].

For naming conventions on different cell types used in the three datasets please see the original publications [4,5]. In brief for the PsychENCODE data: *Ex1 to Ex9* and *In1 to In8* - excitatory and inhibitory neurons; *OPC* - oligodendrocyte progenitor cells, *IPC - intermediate progenitor cells; NEP* - neuroepithelial cells; *trans* - transient cell type. For GSE76381: *DA0-2* - dopaminergic neurons; *Endo* - endothelial cells; *Gaba* - GABAergic neurons; *Mgl* - microglia; *NProg* - neuronal progenitor; NbGaba - neuroblast gabaergic; *NbM* - medial neuroblast; *NbML1+5* - mediolateral neuroblasts; *OMTN* - oculomotor and trochlear nucleus; *OPC* - oligodendrocyte precursor cells. *Peric* - pericytes; *Prog* - progenitor medial floorplate (FPM), lateral floorplate (FPL), midline (M), basal plate (BP); *RN* - red nucleus; *Rgl1-3* - radial glia-like cells; *Sert* – serotonergic.

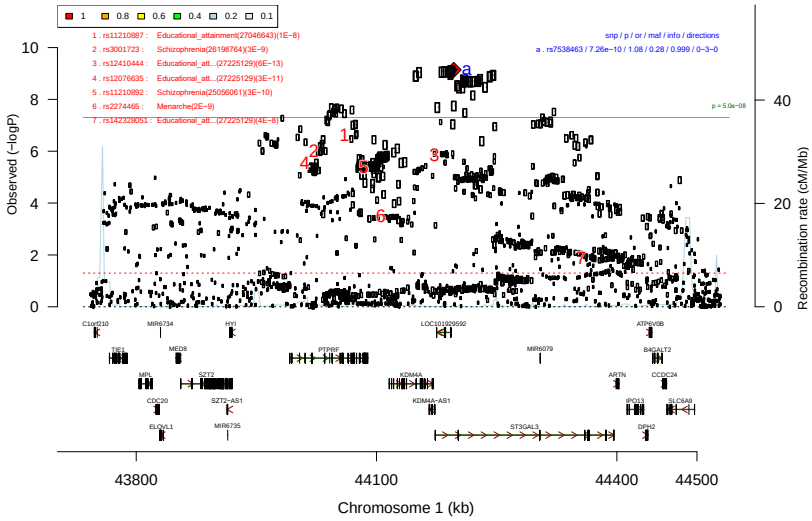
A note on our approaches leveraging single-cell datasets to explore cell type specificity

We performed two complementary approaches leveraging single-cell datasets to explore cell type specificity: (a) MAGMA-based gene expression specificity analysis with FUMA[3] and (b) LDSC-based chromatin accessibility specificity analysis[6]. One of the limitations of the first approach is that SNPs have to be assigned to genes by MAGMA[7] to derive gene-based p-values; even though this is helpful for downstream applications such as gene set enrichment analyses, we have to make assumptions that may not be true. Specifically, in the FUMA implementation, we are assigning to a particular gene all the SNPs between a gene's transcription start and stop sites as well as a 1kb window at both sides; however, important SNPs may be driving gene-mediated heritability from a larger distance and SNPs within a specific gene may predominantly regulate other genes more strongly, especially in cases where they are closer to the other gene's TSS. On the other hand the LDSC-based chromatin accessibility approach doesn't require assignment to genes and explores enrichment for cell-type specific epigenetic peaks; this in turn allows us to study a greater portion of the non-coding genome and e.g. identify enrichments of GWAS SNPs in distal enhancers This is an

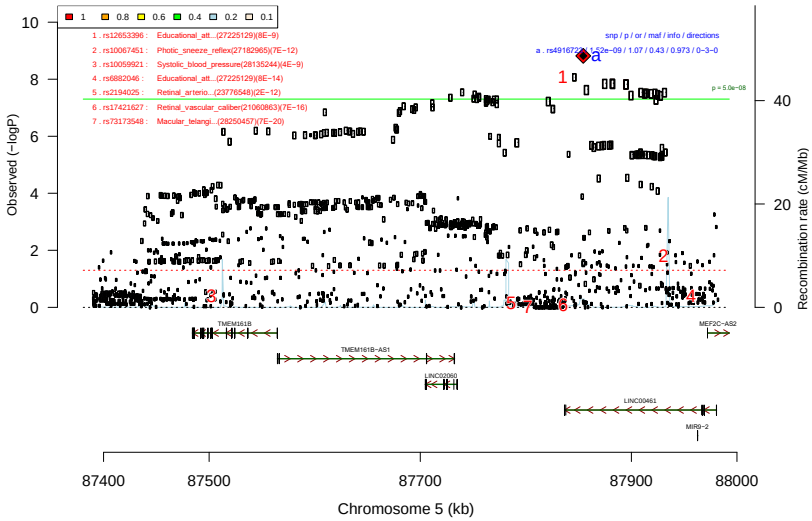
important point because a major portion of cell type-specific enrichment is attributed to distal regulatory elements, as local regulatory events remain highly consistent across various tissues and cell types [PMID: 28343628]. Regarding the statistical method, given a similar analysis (e.g. testing gene expression cell-type specificity), the MAGMA regression model results to more significant trait-cell type associations compared to LDSC[8] which may come at the cost of a higher false positive rate - there is no ground truth to objectively evaluate classification performance.

Supplemental Figure S1: Regional association plot for combined meta-analysis.

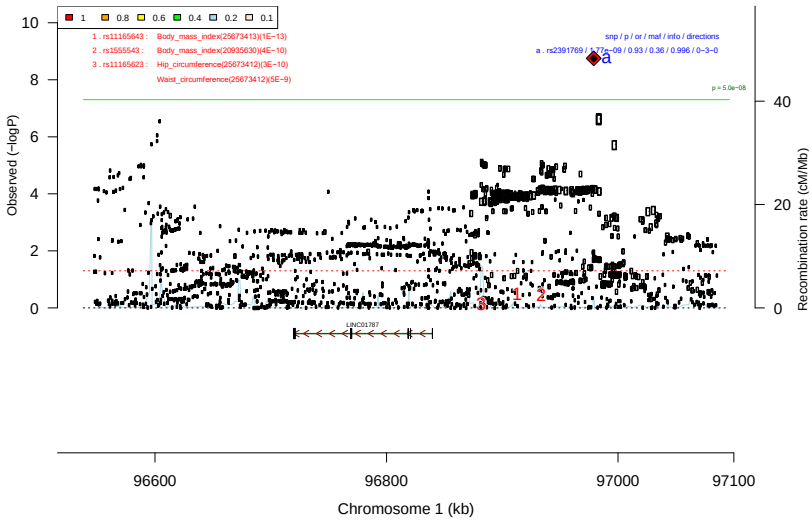
a) rs7538463



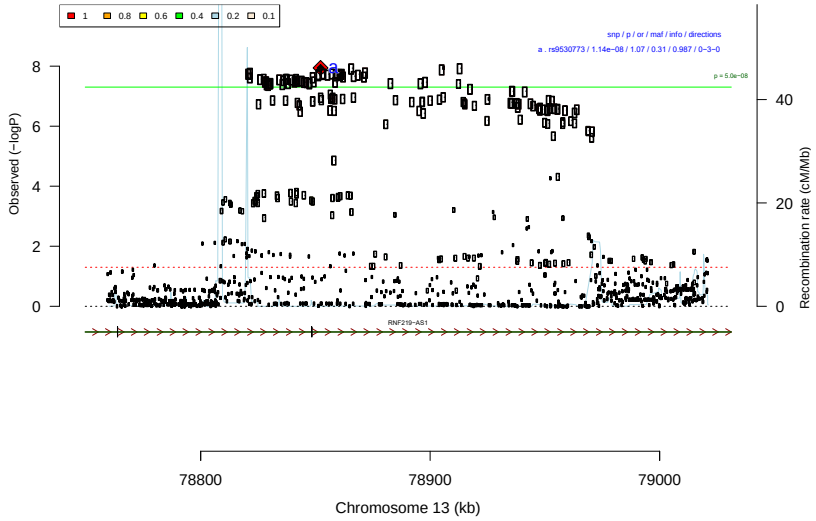
b) rs4916723



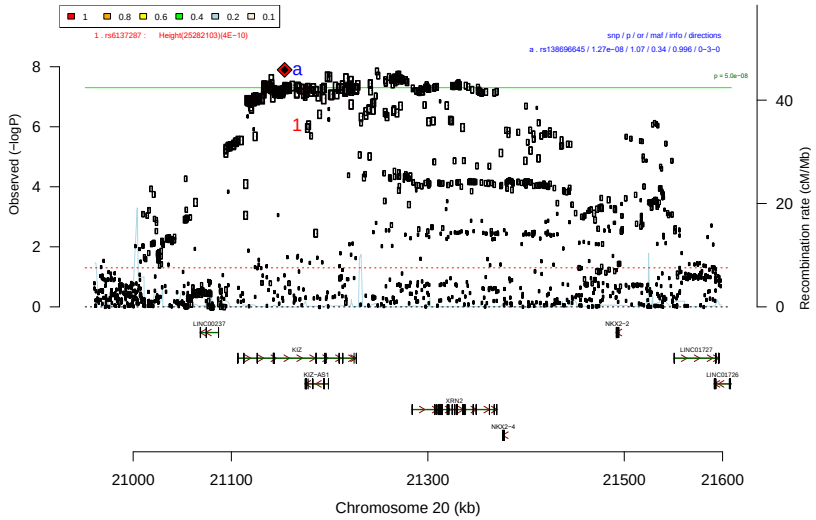
c) rs2391769



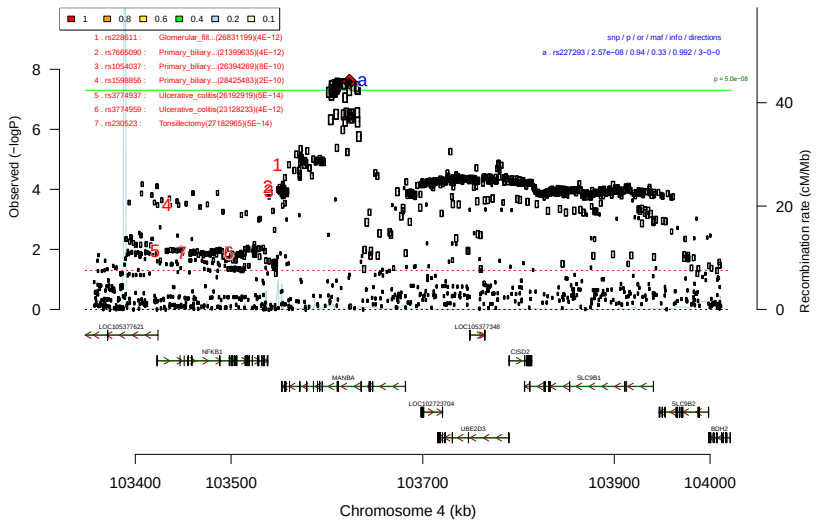
d) rs9530773



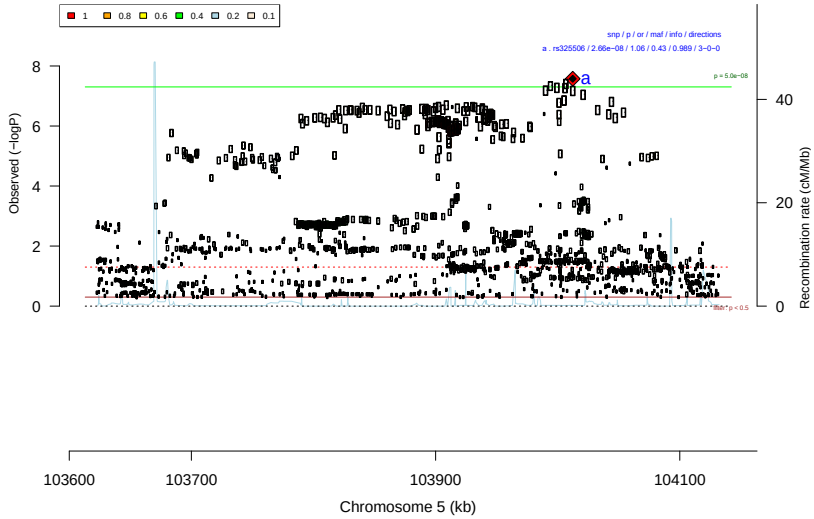
e) rs138696645



f) rs227293



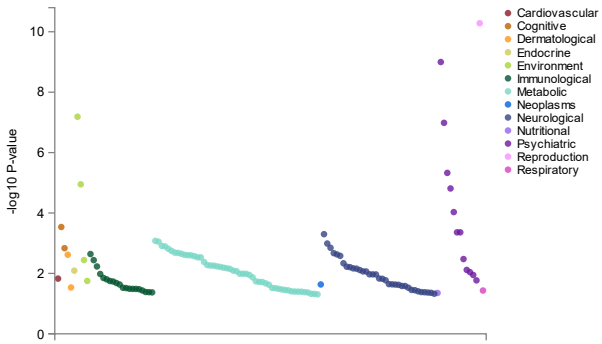
g) rs325506



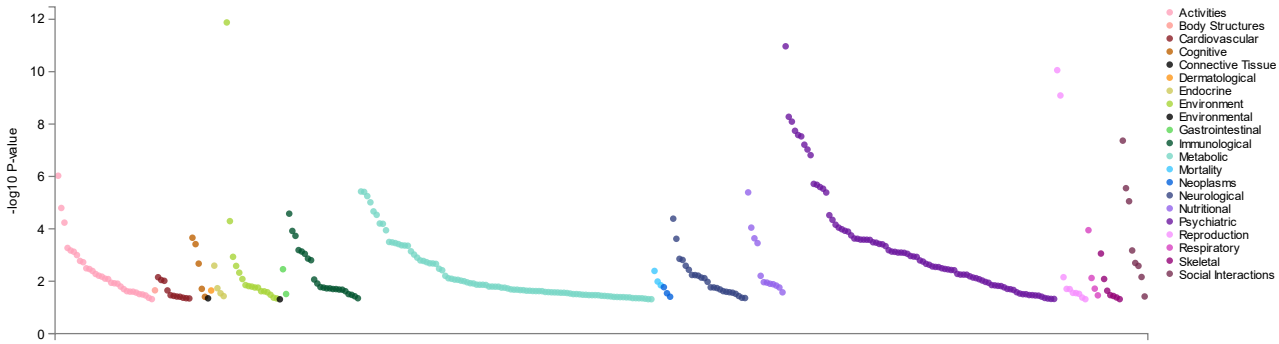
Regional association plots showing association significances for the top seven linkage disequilibrium (LD)-independent index SNPs and all markers within a region of strong LD. SNPs are color coded according to strength of LD with respect to lead SNP (black diamond with red corners) in each region (defined by r^2 statistic). Estimated recombination rates from HapMap phase3 CEU reference panel are depicted as blue lines along the physical position of each region. Genes are drawn in the bottom quarter of the plot (unless in a region devoid of genes) with vertical bars denoting positions of exons. LD-independent genome-wide significant hits are labeled with lower case letters and a list of main characteristics is provided (**snp** – marker name, **p** – P-value of association, **or** – Odds ratio for association, **maf** – Minor allele frequency, **info** – INFO score obtained through PLINK for associated marker, directions – brief table of direction of effects). We used data from the GWAS catalog (as of Oct 2017) to annotate region with known GWAS hits (if there are any), please refer to Supplemental Table S1 and **Supplemental Figure S2** for a more detailed overview. In the annotations, numbers are used to highlight previously associated markers within the plot and a corresponding table is provided. In one of the regional association plots **(g)** only SNPs below a P-value ($p < 0.5$) are shown, in all other instances all SNPs in the region are plotted.

Supplemental Figure S2: PheWAS plots for associated SNPs from combined GWAS

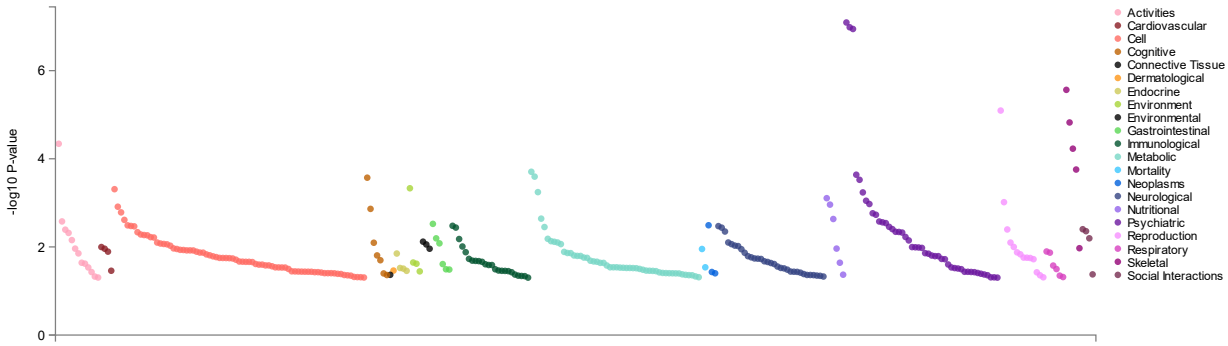
a) rs7538463



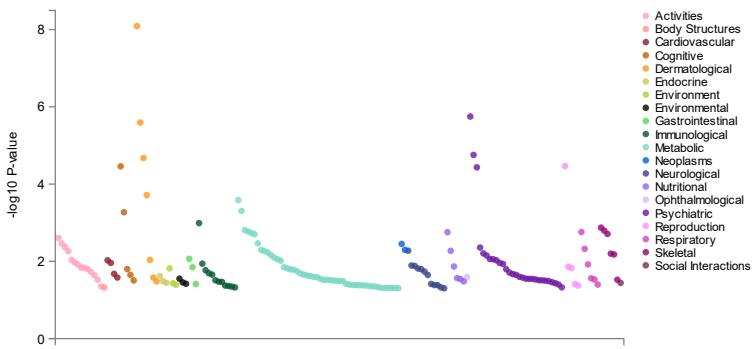
b) rs4916723



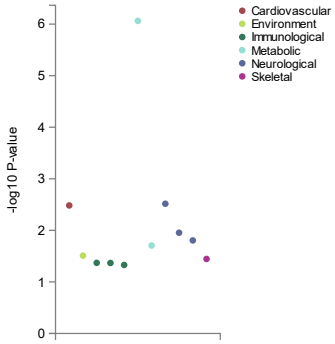
c) rs2391769



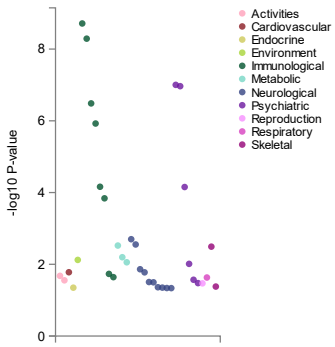
d) rs9530773



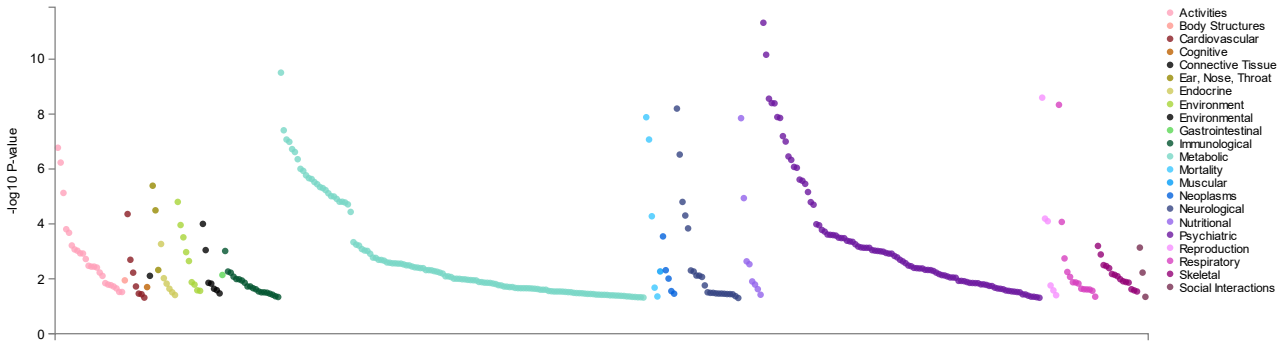
e) rs138696645



f) rs227293



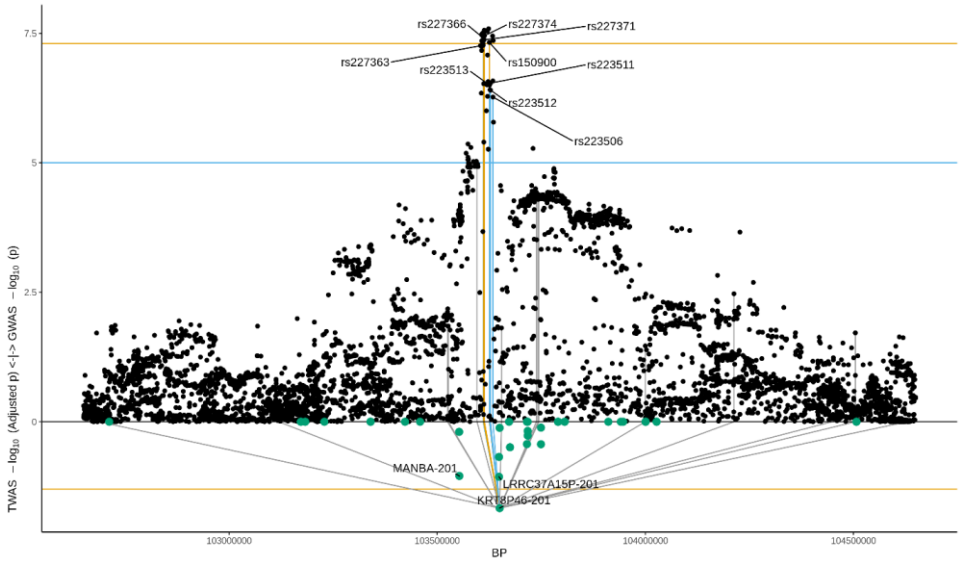
g) rs325506



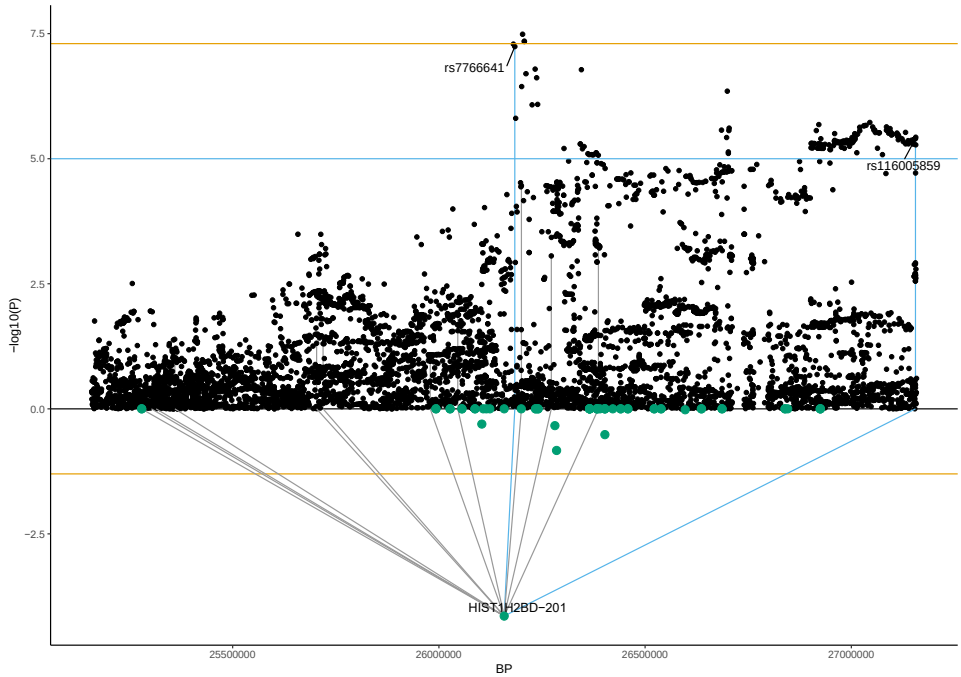
PheWAS analyses with gwasATLAS [9]. Default p-value cutoff at 0.05, traits ordered by domain and p-value. Overall number of GWASs considered for these analyses: 4,756. This also includes GWASs in which the searched SNP was not tested (Bonferroni corrected P-value: $p = 1.05^{x}10^{-5}$). Please note that the information in the corresponding Supplementary Table S1 is based on the OpenGWAS project and might in some instances deviate from results presented in this Figure (due to different data enrolled in both resources). For a comprehensive picture of previous associations, please refer to both tools in tandem.

Supplemental Figure S3: Regional Miami plots for combined and ADHD vs ASD GWASs.

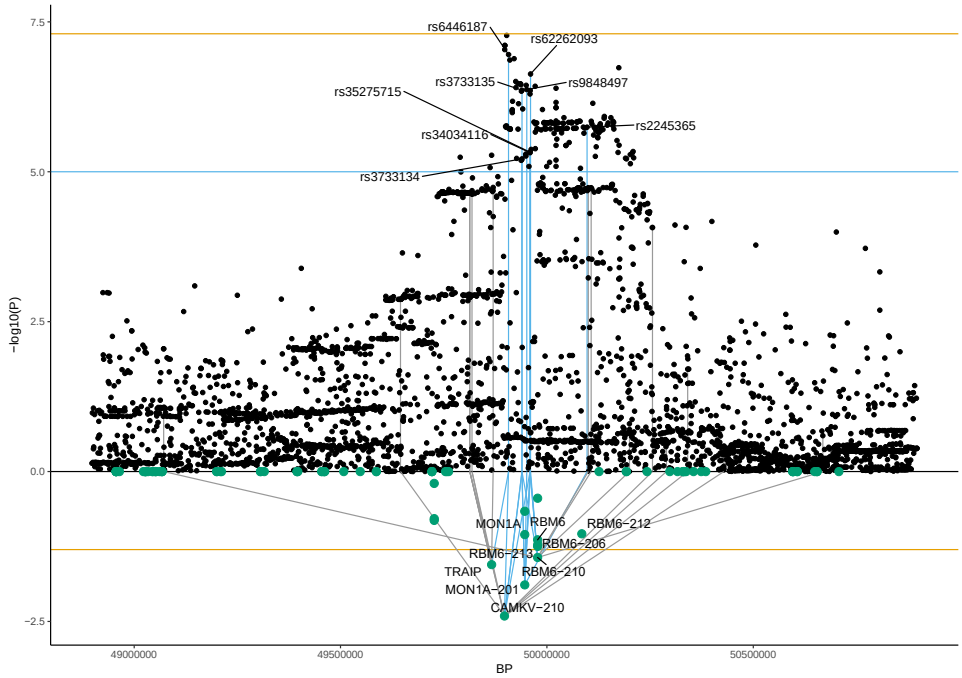
a) KRT8P46 (chromosome 4; transcript: KRT8P46-201)



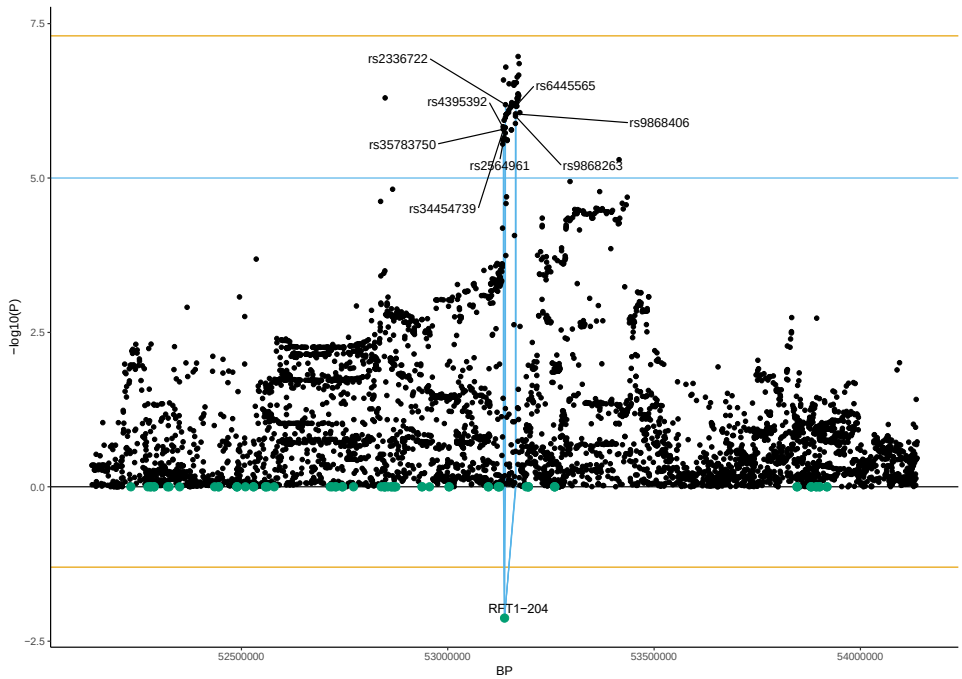
b) HIST1H2BD (chromosome 6; transcript: HIST1H2BD-201)



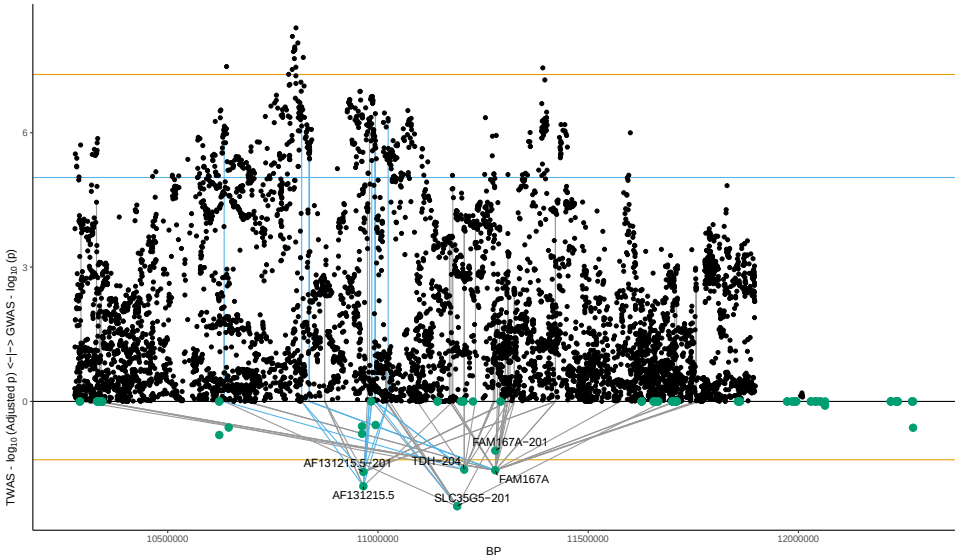
c) CAMKV (chromosome 3; transcript: CAMKV-210)



d) RFT1 (chromosome 3; transcript: RFT1-204)



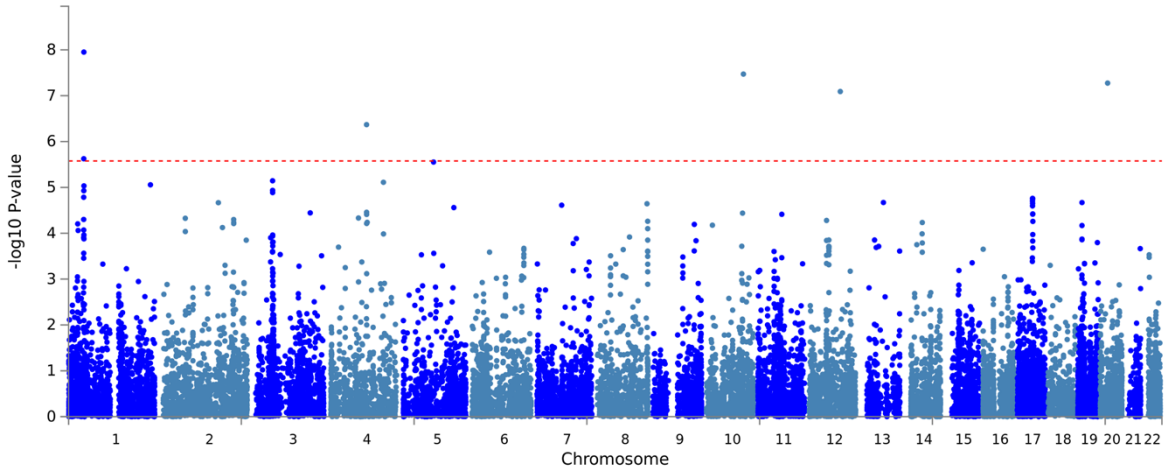
e) FAM167A (chromosome 8; gene)



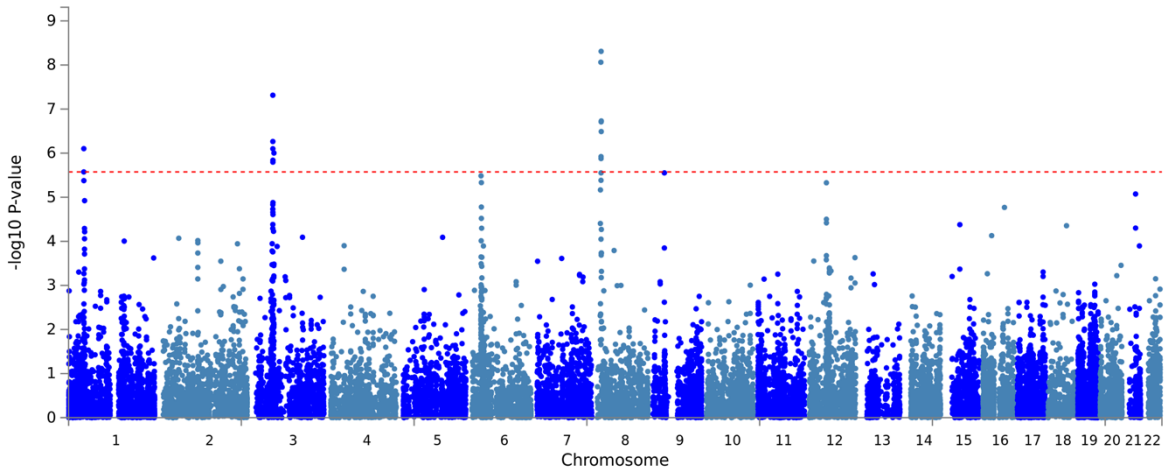
Regional Miami plots for **(a)** combined and **(b-e)** ADHD vs ASD GWASs corresponding to the genomic region of the respective transcript (1Mbp window from start site). Please also refer to Supplemental Table S2 for details. **Top panel** shows the GWAS results (***black dots***); ***blue line*** corresponds to $p = 1^{x}10^{-5}$, ***orange line*** to $p = 5^{x}10^{-8}$ (genome-wide significance). **Bottom panel** shows the TWAS results (***green dots***; only the transcripts with Bonferroni-adjusted $p < 0.1$ are labelled for clarity) for different transcripts (genes are represented by both gene expression and isoform expression); ***orange line*** corresponds to Bonferroni-adjusted $p = 0.05$. Each transcript that is Bonferroni-significant in the region is connected with lines to the SNPs that contribute to its transcriptomic imputation model; lines are ***grey*** when the SNPs have a $p > 1^{x}10^{-5}$, ***blue*** when $p < 1^{x}10^{-5}$ but $> 5^{x}10^{-8}$ and ***orange*** when $p < 5^{x}10^{-8}$. The SNPs that are above the blue line and contribute to the transcriptomic imputation models of significant transcripts are labelled.

Supplemental Figure S4: Manhattan Plot for gene-based analyses in main GWAS comparisons.

a)



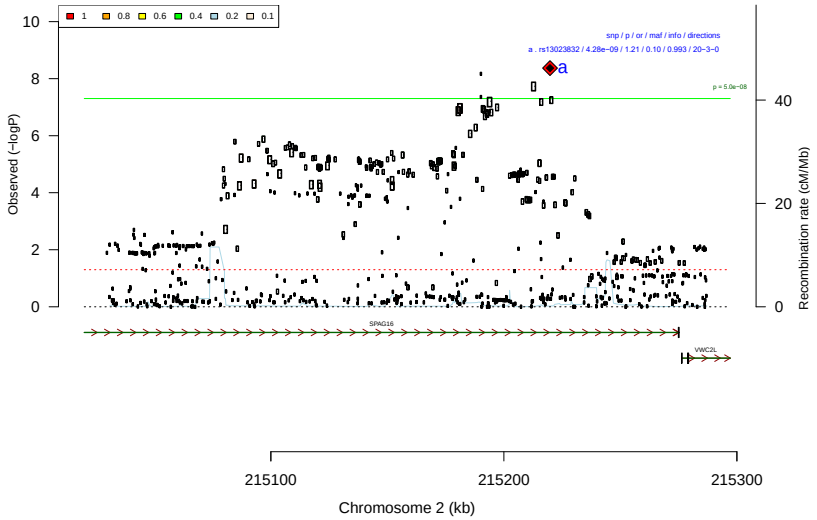
b)



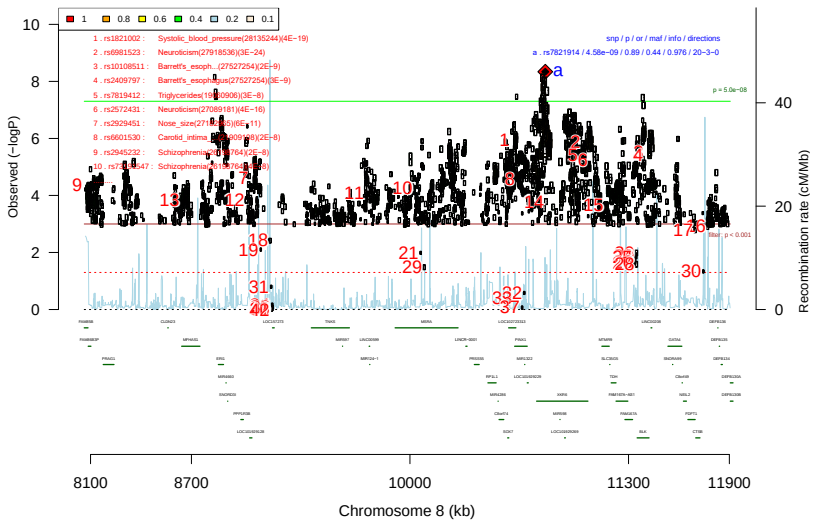
Results from analyses using MAGMA v 1.08 [7] with default settings (and without using a padding sequence) as implemented in FUMA [3]. The x-axis in both sub-plots shows the position in the genome (chromosomes 1–22) and the y-axis the statistical significance as –log10 (P). Red line indicates genome-wide significance. **(a)** Results of genome-wide analyses for combined GWAS (34,462 cases and 41,201 controls). Each dot represents one of the 18,837 genes tested in the analysis. Two of the genes (*SORCS3* and *DUSP6*) are located in regions that were not identified in the GWAS, suggesting these as additional shared loci. **(b)** Results of genome-wide analyses for ADHD vs ASD GWAS (11,964 ADHD only cases and 9,315 ASD only cases). Each dot represents one of the 18802 genes tested in the analysis. There were 14 genome-wide significant ($p < 2.66^{x}10^{-6}$) gene-based associations detected. Nine of these genes are novel associations not previously identified in gene-wise analyses of the disorders separately in the largest GWASs published to date [10,11] (Supplemental Table S3). However, three of the 9 genes are located in the chromosome 8 region that also harbors *SOX7*, *XRK6*, and *BLK*, genes that have been found associated with ASD before [10]. The remaining six genes are located on chromosome 3 in a locus that was not genome-wide significant in the GWAS.

Supplemental Figure S5: Regional association plots for ADHD vs ASD GWAS.

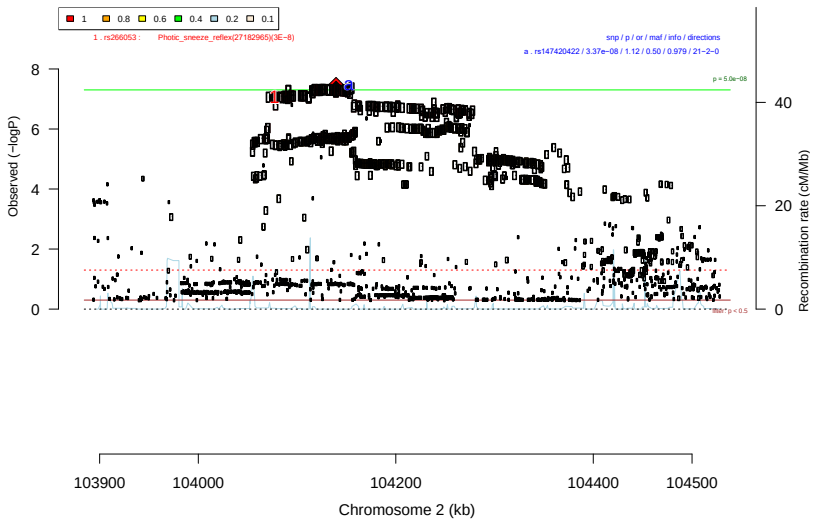
a) rs13023832



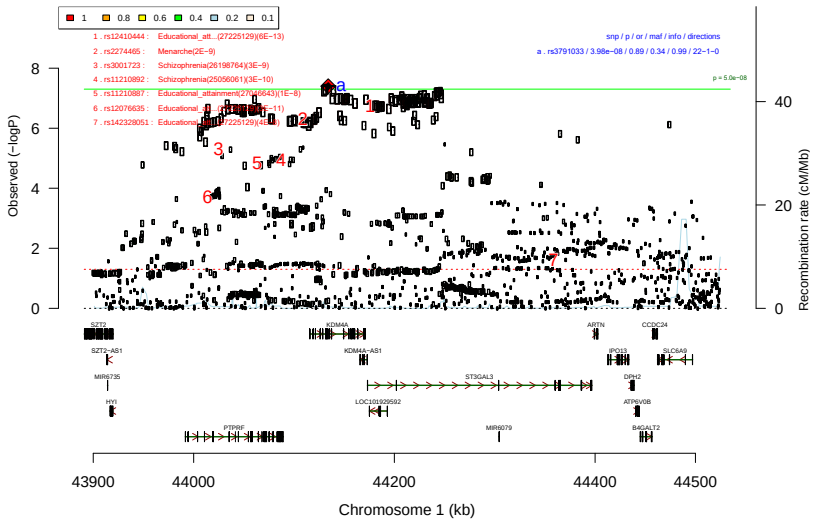
b) rs7821914



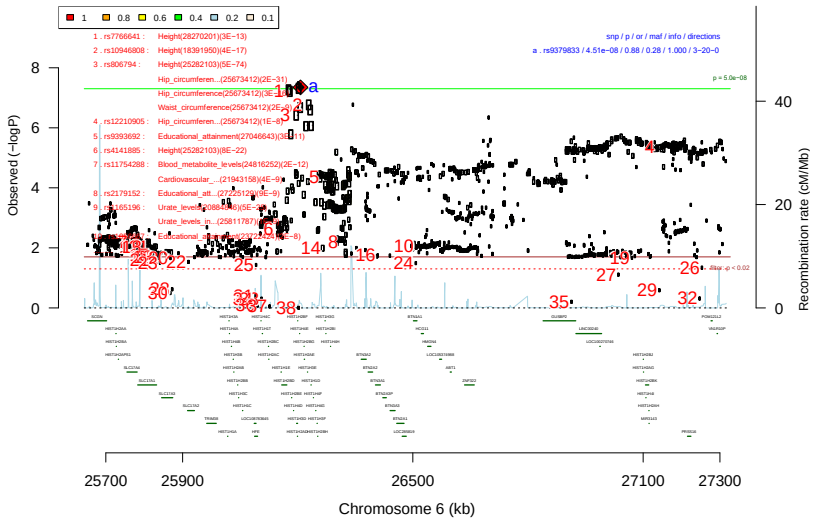
c) rs147420422



d) rs3791033



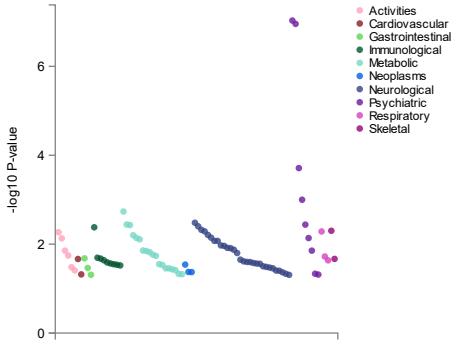
e) rs9379833



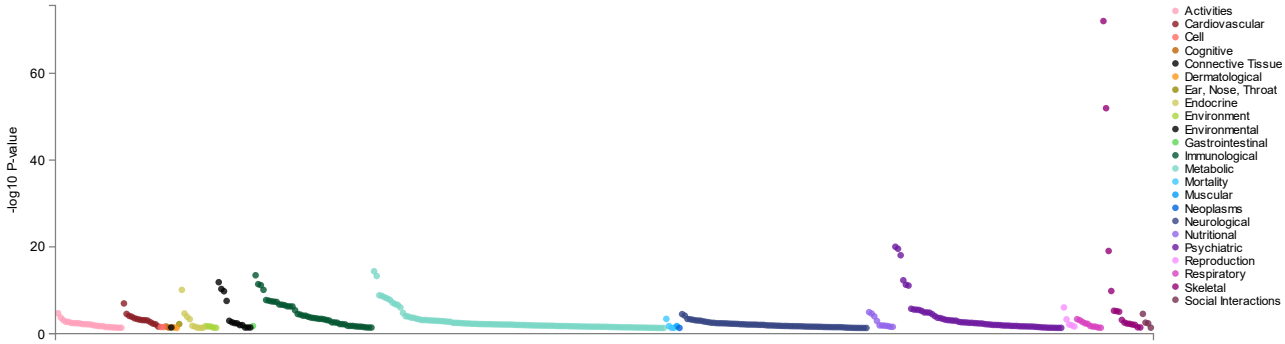
Regional association plots showing association significances for the top five linkage disequilibrium (LD)-independent index SNPs and all markers within a region of strong LD. SNPs are color coded according to strength of LD with respect to lead SNP (black diamond with red corners) in each region (defined by r^2 statistic). Estimated recombination rates from HapMap phase3 CEU reference panel are depicted as blue lines along the physical position of each region. Genes are drawn in the bottom quarter of the plot (unless in a region devoid of genes) with vertical bars denoting positions of exons. LD-independent genome-wide significant hits are labeled with lower case letters and a list of main characteristics is provided (snp – marker name, p – P-value of association, or – Odds ratio for association, maf – Minor allele frequency, info – INFO score obtained through PLINK for associated marker, directions – brief table of direction of effects). We used data from the GWAS catalog (as of Oct 2017) to annotate region with known GWAS hits (if there are any), please refer to Supplemental Table S5 and **Supplemental Figure S6** for a more detailed overview. In the annotations, numbers are used to highlight previously associated markers within the plot and a corresponding table is provided. In one of the regional association plots **(b)** only SNPs below a P-value ($p < 0.001$) are shown, or if they have beenb previously identified to be associated with a trait listed in the GWAS catalog. In regional association plot **(c)** only SNPs below a P-value ($p < 0.5$) are shown, in regional association plot **(e)** only SNPs below a P-value ($p < 0.02$) are shown and regional association plots **(a)** and **(d)** all SNPs in the region are plotted.

Supplemental Figure S6: PheWAS plots for associated SNPs from ADHD vs ASD GWAS

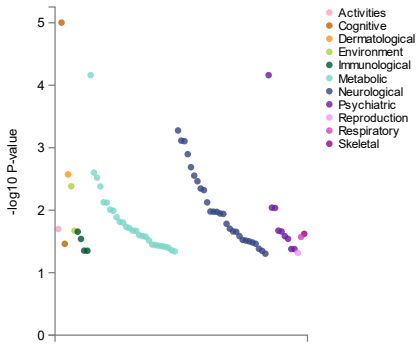
a) rs13023832



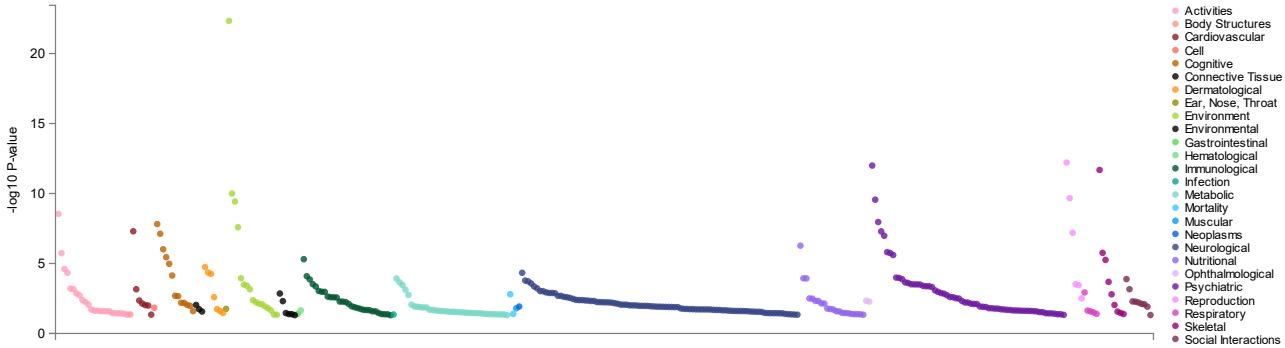
b) rs7821914



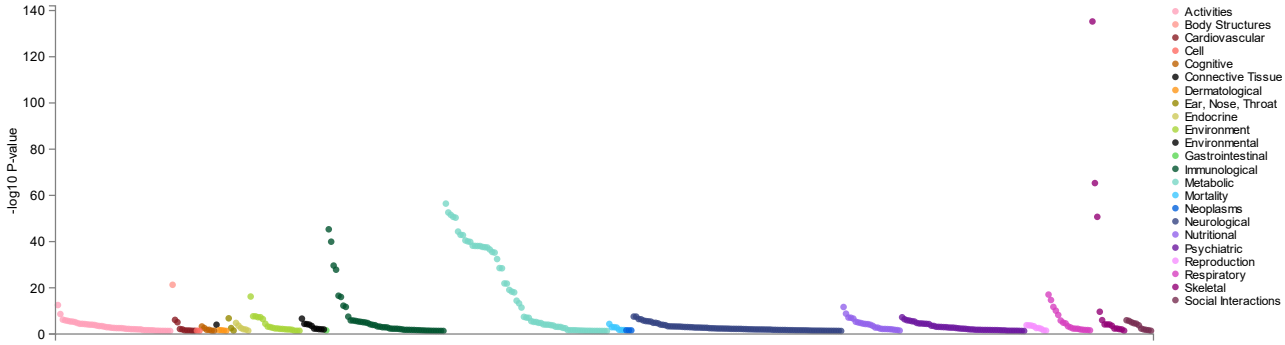
c) rs147420422



d) rs3791033



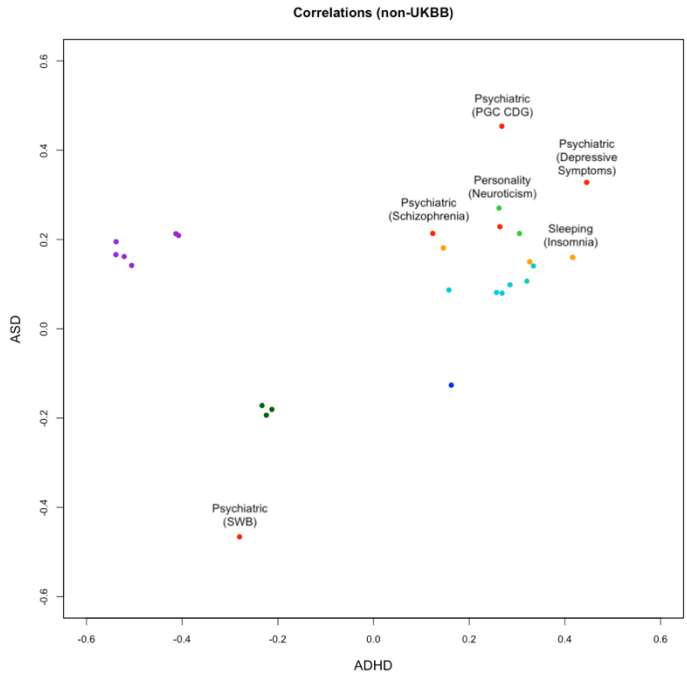
e) rs9379833



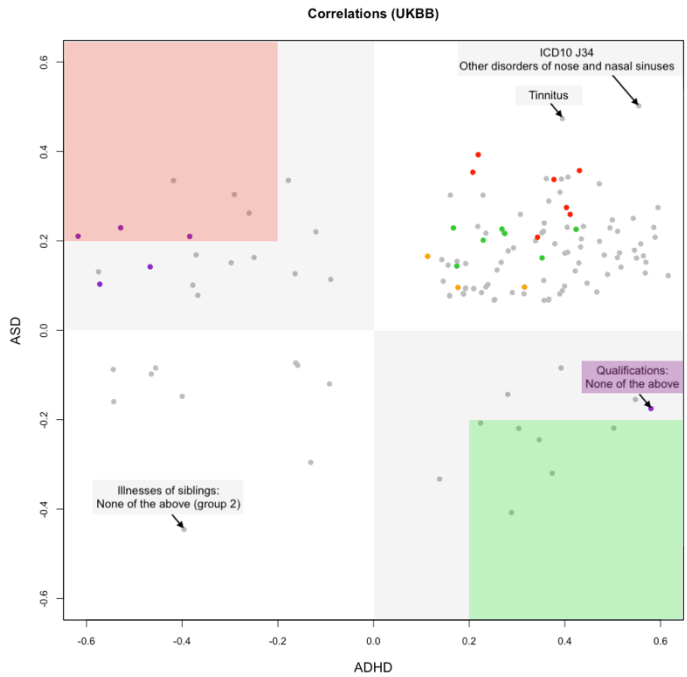
PheWAS analyses with gwasATLAS [9]. Default p-value cutoff at 0.05, traits ordered by domain and p-value. Overall number of GWASs considered for these analyses: 4,756. This also includes GWASs in which the searched SNP was not tested (Bonferroni corrected P-value: $p = 1.05^{x}10^{-5}$).

Supplemental Figure S7: Genetic correlations between ADHD and ASD with other traits and disorders.

a)



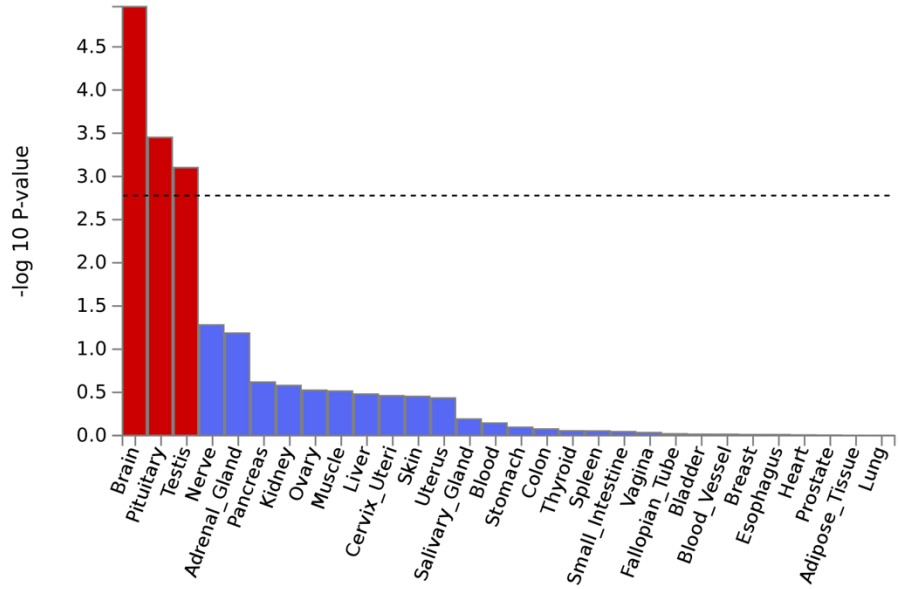
b)



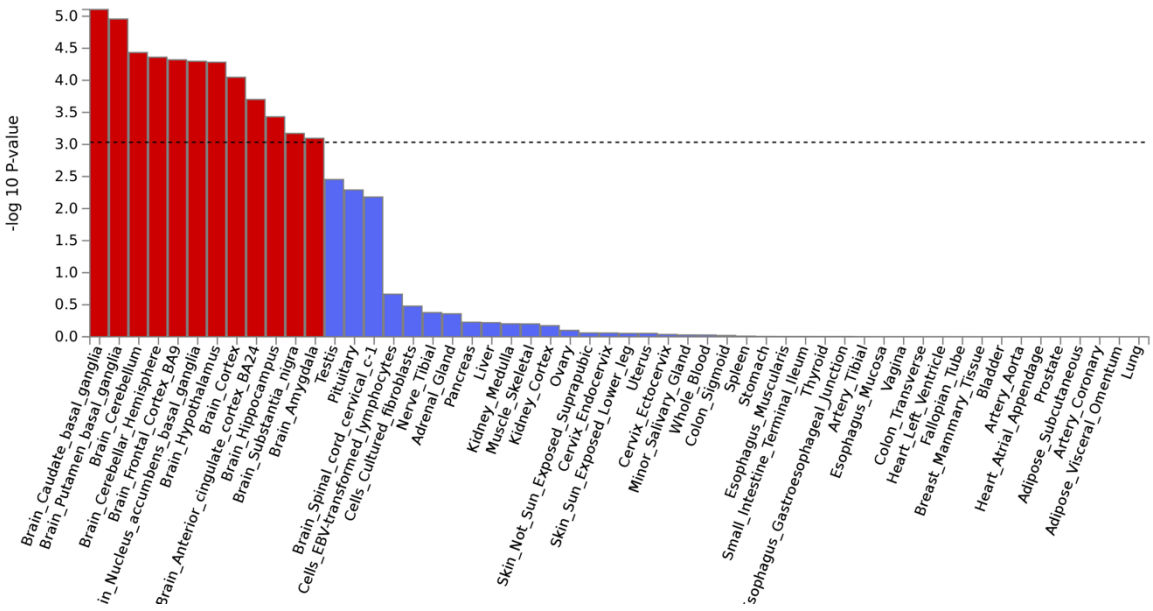
Genetic correlations for ADHD (PMID 30478444) and ASD (PMID 30804558) with other traits as calculated by LDhub (PMID 27663502). Please refer to Supplemental Table S4 for an overview of all genetic correlations. In **(a)** "Non-UKBB" (*n*= 28) and **(b)** "UKBB" (*n*= 136) only correlations with a *Z score* > 2 in both ADHD and ASD are shown ("*pass"* in Supplemental Table S4 – OVERVIEW for *z2_flag_asd* and *z2_flag_adhd*; *n*= 164 in total). **(a)** Non-UKBB: Color is coded as follows for different categories (please note that results for cross correlations for ADHD or ASD GWASs are not shown): personality (Neuroticism with two GWASs) – green, psychiatric (*Subjective well-being (SWB); PGC cross-disorder analysis (PGC CDG); Depressive Symptoms, Schizophrenia, Major Depressive Disorder*) – red, education/ cognitive (*Years of schooling (proxy cognitive performance), Years of schooling 2016; Years of schooling 2013; College completion; Childhood IQ; Intelligence*) – purple, autoimmune (*Rheumatoid Arthritis*) – blue, sleeping (*2 Insomnia GWASs; Excessive daytime sleepiness*) – orange, anthropometric (*3 Obesity GWASs (class 1, 2, 3); Waist circumference; Body Mass Index; Hip circumference*) – turquoise, metabolites (*Concentration of large HDL particles; Phospholipids in large HDL; Total cholesterol in HDL*) – dark green. **(b)** UKBB: Areas are shaded in grey if they contain rGs for ADHD and ASD that are positive for one and negative for the other. The areas are colored in light reen if they have a rG > 0.2 for ADHD and one for ASD < -0.2 and in light red if they have a rG < -0.2 for ADHD and one for ASD > 0.2. There are 6 UKBB traits in either of the two colored areas: *light green* - *Job involves mainly walking or standing* (ADHD: 0.50, ASD: -0.22), *Transport type for commuting to job workplace: Car/motor vehicle* (0.37, -0.32), *Weight change compared with 1 year ago* (0.35, -0.24), *Duration of vigorous activity* (0.30, -0.22), *Number of children fathered* (0.29, -0.41), *Prospective memory result* (0.22, -0.21); *light red* - *Qualifications: A levels/AS levels or equivalent* (ADHD: -0.62, ASD 0.21), *Qualifications: College or University degree* (-0.53, 0.23), *Transport type for commuting to job workplace: Public transport* (-0.42, 0.34), *Fluid intelligence score* (-0.38, 0.21), *Types of transport used (excluding work): Public transport* (-0.29, 0.30), *Transport type for commuting to job workplace: Walk* (-0.26, 0.26). personality (*Frequency of tenseness / restlessness in last 2 weeks, Fed-up feelings, Guilty feelings, Tense / highly strung, Irritability, Sensitivity / hurt feelings, Neuroticism score*) – green, psychiatric (*Number of depression episodes, Loneliness_ isolation, Illnesses of siblings: Severe depression, Miserableness, Frequency of depressed mood in last 2 weeks, Ever unenthusiastic/disinterested for a whole week, Ever depressed for a whole week*) – red, education/ cognitive (*Qualifications: A levels/AS levels or equivalent, Qualifications: College or University degree, Qualifications: None of the above, Qualifications: O levels/GCSEs or equivalent, Qualifications: Other professional qualifications e.g.: nursing_ teaching, Fluid intelligence score*) – purple, sleeping (*Sleeplessness / insomnia, Nap during day, Daytime dozing / sleeping (narcolepsy)*) – orange.

Supplemental Figure S8 MAGMA tissue expression analysis for combined GWAS.

a)



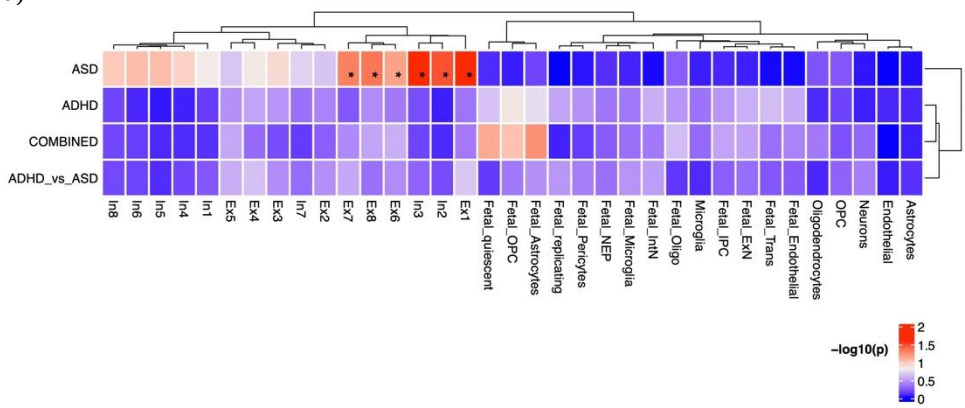
b)



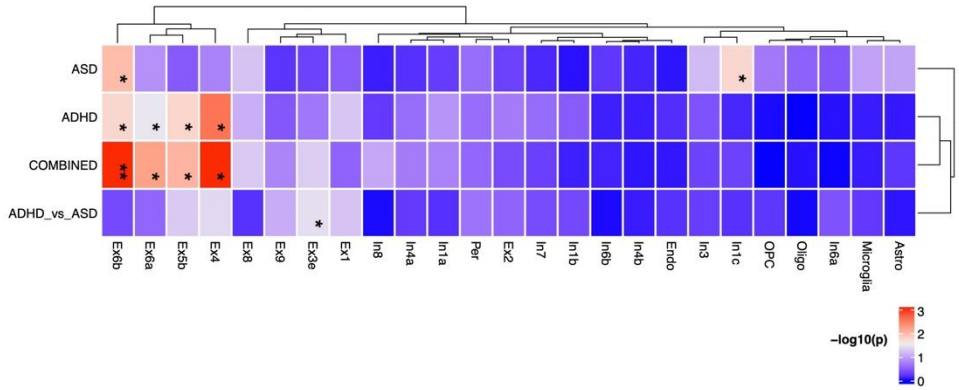
Results of gene-property analysis in MAGMA [7] as implemented in FUMA [3]. Tissue specific data is obtained from the GTEx v8 dataset (https://www.gtexportal.org) [12]. Tissues with red bar surpass experiment-wide significance. **(a)** General tissue type analysis with 30 tissues. **(b)** Individual tissue type analyses with 54 tissue types.

Supplemental Figure S9 FUMA single-cell analyses for ASD, ADHD, combined, and differentiating GWAS.

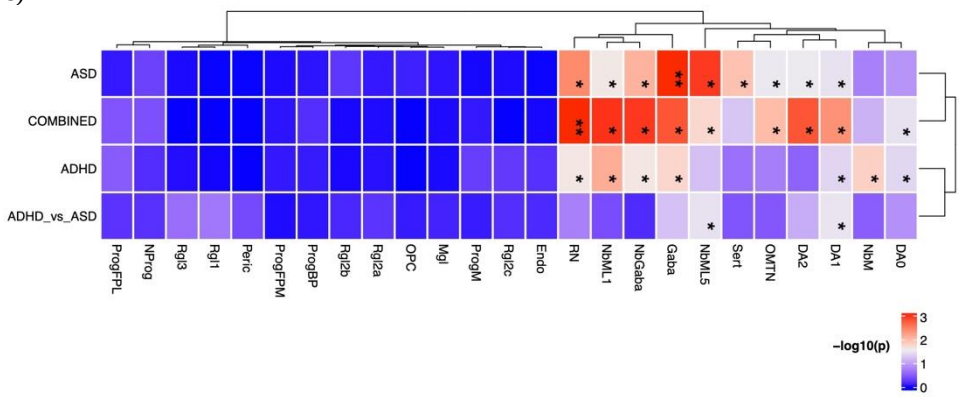
a)



b)



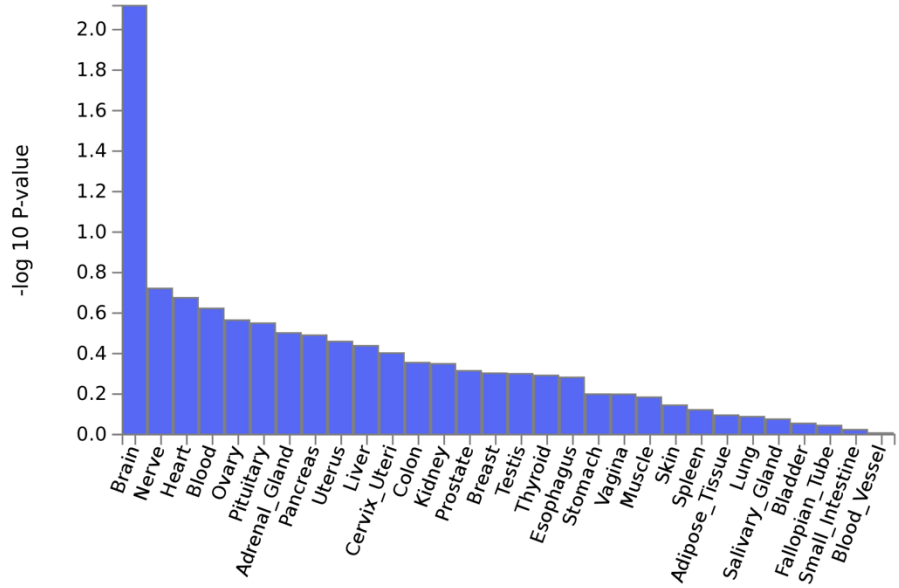
c)



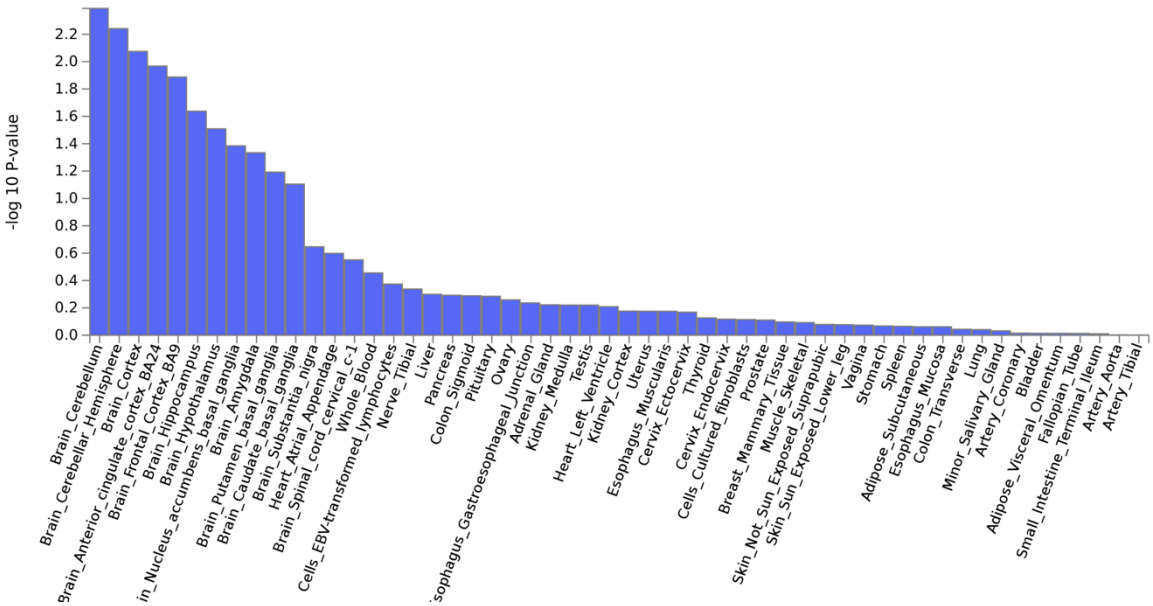
Results are shown for the psychENCODE datasets [4] with **(a)** human developmental and **(b)** human adult brain samples as well as the human midbrain cell types (ventral midbrain from 6-11 weeks embryos) from La Manno et al. [5] **(c**, GSE76381). For psychENCODE developmental dataset, 4,249 cells were available, for the psychENCODE adult dataset, 27,380 cells were included in the analysis. Mapping to unique ENSG IDs was available for 15,019 and 16,243 genes, respectively. For the human midbrain samples 1,695 cells were used. Mapping to unique ENSG ID was available for 16,885 genes. Cell type results highlighted with one (*) asterisk achieve a p-value < 0.05, those with two asterisks (**) a p-value < 0.001. Only one cell-type (RN in the GSE76381 dataset) survives correction for multiple testing across all tested celltypes ($p = 1.29 \times 10^{-4}$ for the combined GWAS). For naming conventions on different cell types please see the original publication [PMID 30545857 27716510]. In brief for the ***PsychENCODE*** data: *Ex1 to Ex9* and *In1 to In8* - excitatory and inhibitory neurons; *OPC* - oligodendrocyte progenitor cells, *IPC* - *intermediate progenitor cells; NEP* - neuroepithelial cells; *trans* - transient cell type. For ***GSE76381***: *DA0-2* - dopaminergic neurons; *Endo* - endothelial cells; *Gaba* - GABAergic neurons; *Mgl* - microglia; *NProg* - neuronal progenitor; NbGaba - neuroblast gabaergic; *NbM* - medial neuroblast; *NbML1+5* - mediolateral neuroblasts; *OMTN* - oculomotor and trochlear nucleus; *OPC* - oligodendrocyte precursor cells. *Peric* - pericytes; *Prog* - progenitor medial floorplate (FPM), lateral floorplate (FPL), midline (M), basal plate (BP); *RN* - red nucleus; *Rgl1-3* - radial glia-like cells; *Sert* – serotonergic.

Supplemental Figure S10 MAGMA tissue expression analysis for ADHD vs ASD GWAS.

a)

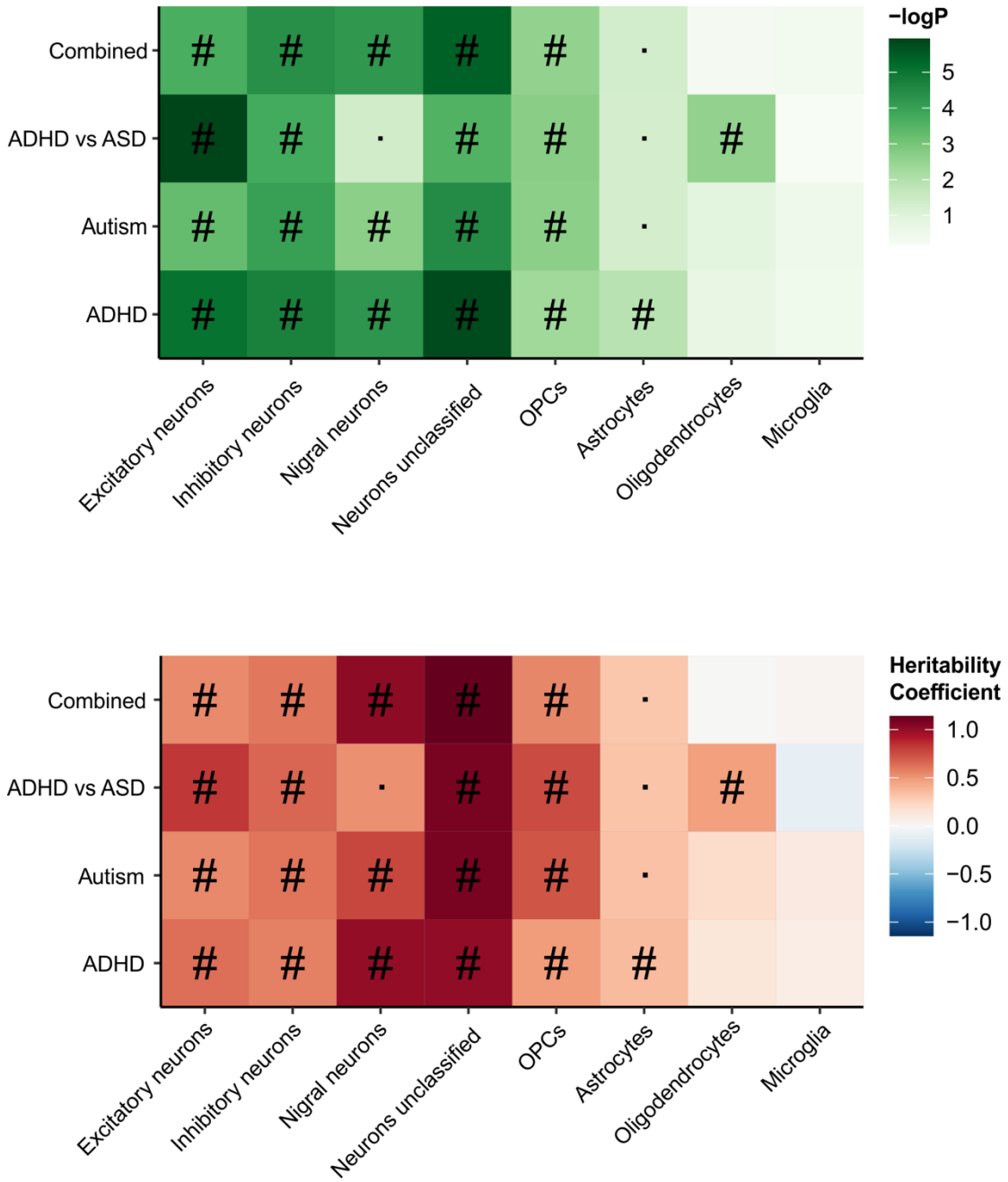


b)



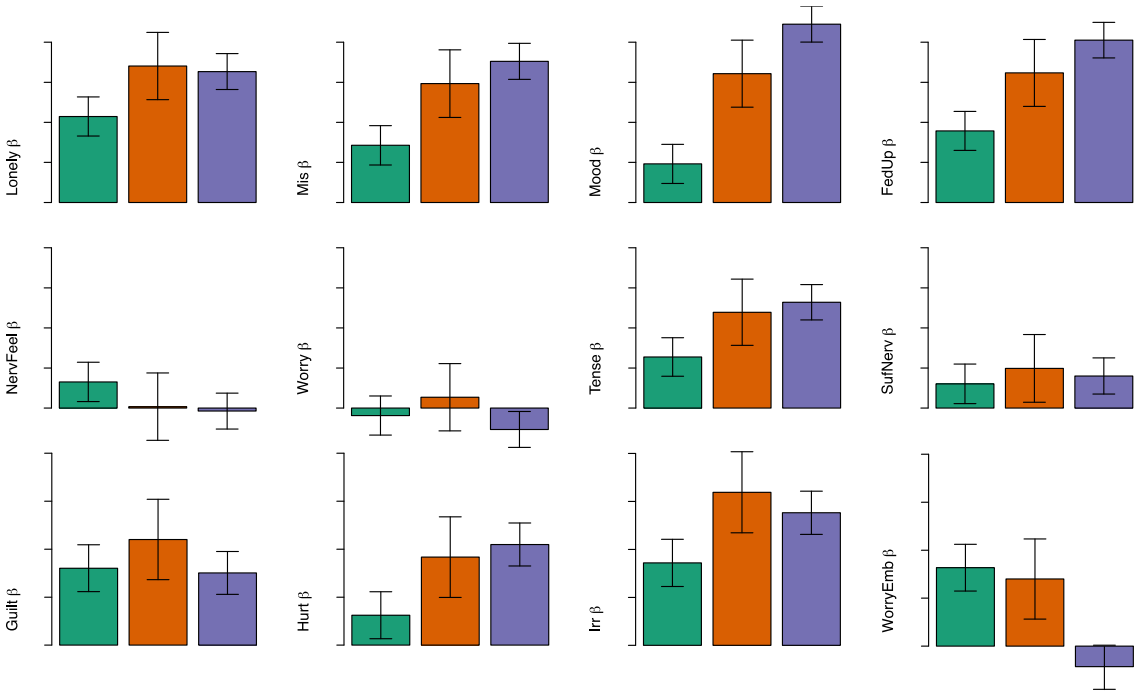
Results of gene-property analysis in MAGMA [7] as implemented in FUMA [3]. Tissue specific data is obtained from the GTEx v8 dataset (https://www.gtexportal.org) [12]. Tissues with red bar surpass experiment-wide significance. **(a)** General tissue type analysis with 30 tissues. **(b)** Individual tissue type analyses with 54 tissue types.

Supplemental Figure S11: Single cell enrichment analysis for epigenomic peaks.



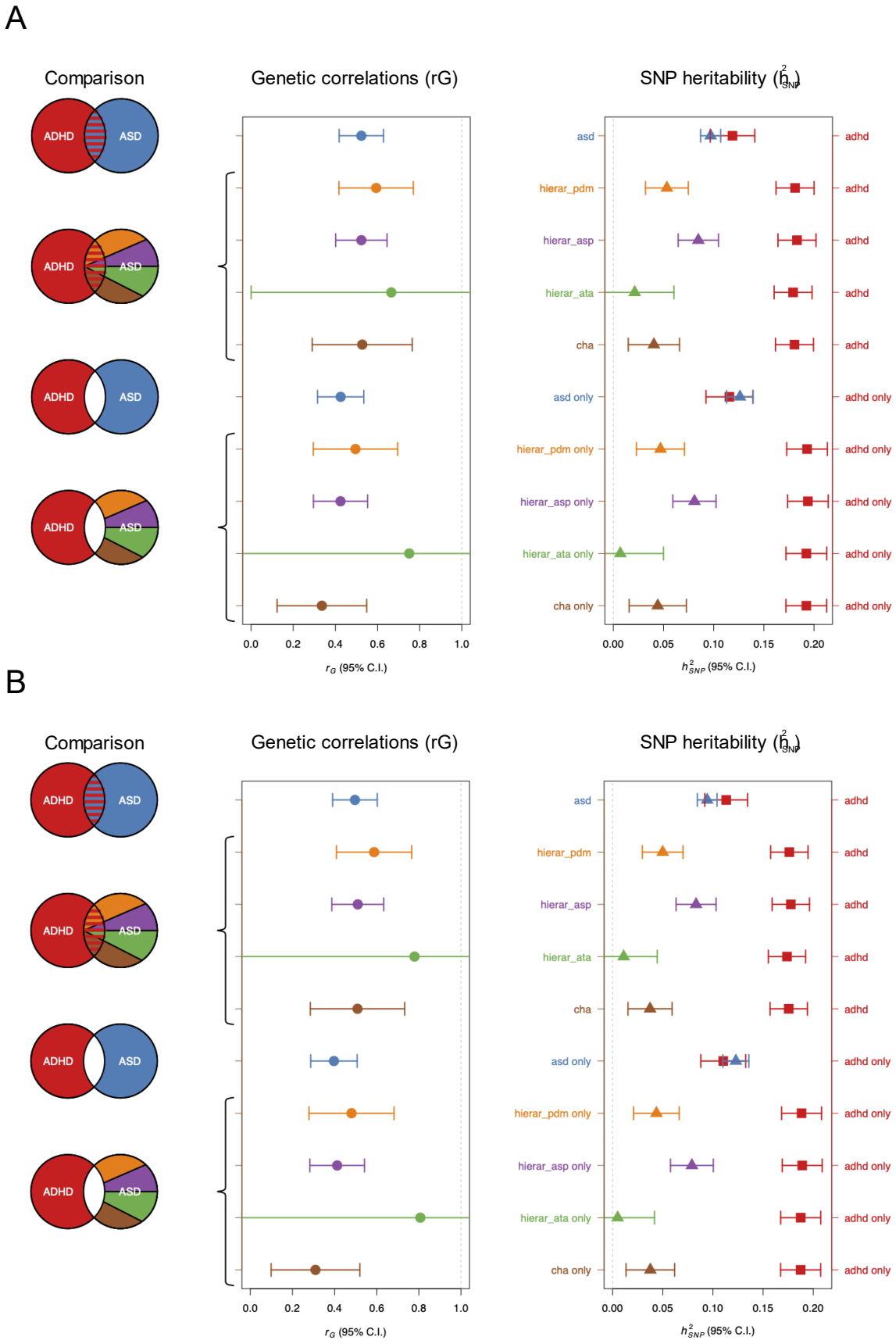
Enrichment of heritability within cell-specific open chromatin identified by scATAC-seq assay (single-cell assay for transposase accessible chromatin) calculated using LD-score partitioned heritability. Top part of figure: # - Test wide significant at FDR < 0.05; · - Nominally significant at p < 0.05. Bottom part of figure: The heritability coefficient is the regression coefficient normalized by the per-SNP heritability.

Supplemental Figure S12: Multivariate PRS analyses for Neuroticism subitems



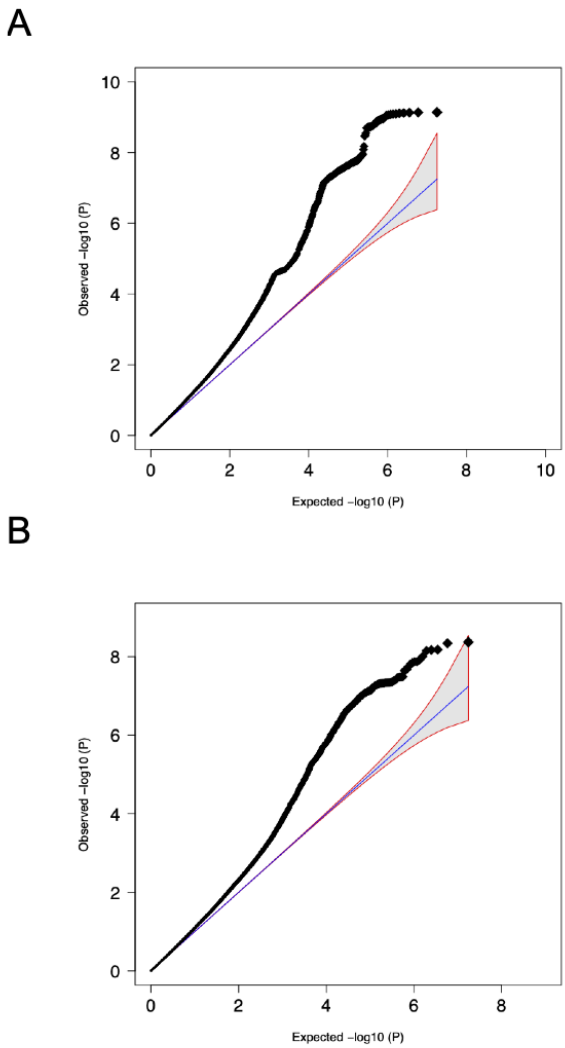
Comparison of PRS profiles across ADHD/ASD subtypes for 12 neuroticism subitems. Green bars represent ASD only cases, orange bars depict comorbid samples, and purple bars show average PRS for ADHD only cases. **Lonely** - Do you often feel lonely? (yes/no); **Mis** - Do you ever feel 'just miserable' for no reason? (yes/no); **Mood** - Does your mood often go up and down? (yes/no); **FedUp** - Do you often feel 'fed-up'? (yes/no); ***NervFeel*** - Would you call yourself a nervous person? (yes/no); ***Worry*** - Are you a worrier? (yes/no); ***Tense*** - Would you call yourself tense or 'highly strung'? (yes/no); ***SufNerv*** - Do you suffer from 'nerves'? (yes/no); **Guilt** - Are you often troubled by feelings of guilt? (yes/no); **Hurt** - Are your feelings easily hurt? (yes/no); **Irr** - Are you an irritable person? (yes/no); **WorryEmb** - Do you worry too long after an embarrassing experience? (yes/no). In the text above an underlined item (first row of traits in figure) belongs to the depressed affect cluster in Nagel et al [13] while an item in italic (second row in figure) belongs to the worry cluster. Items that are neither underlined nor italic (last row in figure) do not belong to the two clusters.

Supplemental Figure S13: GCTA-based heritability estimates and genetic correlation for ASD (with subtypes) and ADHD



All analyses used the GCTA framework. For analyses datasets were split to allow for comparisons of independent datasets. Sample split was kept the same across analyses (with control samples using intra case ratios for splitting). For ASD subtypes a hierarchical approach was taken in the reverse order the comparison groups appear in the plot (i.e. first all individuals with childhood autism (cha), then those with atypical autism (ata) and no comorbid childhood autism, then those with Asperger's syndrome (asp) and no comorbid childhood autism or atypical, and finally the remaining individuals in the pervasive disorders group). Comparisons include **(A)** a base dataset that *excludes* individuals with mild and moderate intellectual disability and **(B)** a base dataset that *includes* individuals with mild and moderate intellectual disability. For all comparisons the following color coding applies: **red** – ADHD (i.e. without individuals with a comorbid ASD), **blue** – ASD (i.e. without individuals with comorbid ADHD), **brown** – childhood autism (*cha*, ICD10 F84.0), **green** - atypical autism (*ata*, ICD10 F84.1), **purple** – Asperger's syndrome (*asp*, ICD10 F84.5), and **orange** – pervasive disorders, unspecified and others (*pdm*, ICD10 F84.8+9). If a "only" follows the name of the group only non-comorbid cases between ADHD and ASD are includes (e.g. all ADHD cases that are not comorbid ASD cases). Please also see **Supplemental Table S7** at the end of this document.

Supplemental Figure S14: QQ plots for combined and ADHD vs ASD GWASs



QQ-plot for **(A)** combined GWAS (34,462 cases and 41,201 controls) and for **(B)** ADHD vs ASD GWAS (11,964 ADHD only cases and 9,315 ASD only cases). The expected -log(10) under the null is plotted against the observed -log10(P) of the two aforementioned GWASs. The shading indicates 95%-confidence region under the null. The genomic inflation factor is 1.134 (with an intersect of 1.0134 in the LD score analysis) and 1.089 (intersect 0.9863) for the combined and ADHD vs ASD GWASs, respectively.

References

1. Bipolar, D., Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic address, d.r.v.e., Bipolar, D. & Schizophrenia Working Group of the Psychiatric Genomics, C. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell* **173**, 1705-1715 e16 (2018).
2. 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).
3. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
4. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**(2018).
5. La Manno, G. *et al.* Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell* **167**, 566-580 e19 (2016).
6. Finucane, H.K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50**, 621-629 (2018).
7. 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).
8. 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).
9. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* **51**, 1339-1348 (2019).
10. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* **51**, 431-444 (2019).
11. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* **51**, 63-75 (2019).
12. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
13. 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).