Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains

Ditte Demontis[1,2,3*], G. Bragi Walters[4,24*], Georgios Athanasiadis[2,5,39*], Raymond Walters[6,7*], Karen Therrien[14,15,16,17,18,40], Leila Farajzadeh[1,2,3], Georgios Voloudakis[14,15,16,17,18], Jaroslav Bendl[14,15,16,18], Biao Zeng[14,15,16,18], Wen Zhang[14,15,16,18],, Jakob Grove[1,2,9], Thomas D. Als[1,2,3], Jinjie Duan[1,2,3], F. Kyle Satterstrom[6,7], Jonas Bybjerg-Grauholm[2,8], Marie Bækved-Hansen[2,8], Olafur O. Gudmundsson[4,24], Sigurdur H. Magnusson[4], Gisli Baldursson[25], Katrin Davidsdottir[26], Gyda S. Haraldsdottir[26], Trine Tollerup Nielsen[1], Esben Agerbo[2,10,11], Gabriel E. Hoffman[14,15,16,18], Søren Dalsgaard[10,33,34], Joanna Martin[28], Marta Ribasés[29,30,31,32], Dorret I. Boomsma[37,38], Maria Soler Artigas[29,30,31,32], Nina Roth Mota[20,21], Daniel Howrigan[6,7], Sarah E. Medland[35], Tetyana Zayats[6,7,36], Veera Manikandan[1,2,3], ADHD Working Group of the Psychiatric Genomics Consortium, iPSYCH-Broad Consortium, Merete Nordentoft[2,13], Ole Mors[2,12], David M. Hougaard[2,8], Preben Bo Mortensen[2,10,11], Mark J. Daly[6,7,22,23], Stephen V. Faraone[27], Hreinn Stefansson[4], Panos Roussos[14,15,16,17,18,19], Barbara Franke[20,21,41,†], Thomas Werge[2,5,†], Benjamin M. Neale[6,7,†], Kari Stefansson[4,24†], Anders D. Børglum[1,2,3,†]

*shared first authors, †Shared last authors

[1]Department of Biomedicine/Human Genetics, Aarhus University, Aarhus, Denmark
[2]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark
[3]Center for Genomics and Personalized Medicine, Aarhus, Denmark
[4]deCODE genetics / Amgen, Reykjavik, Iceland
[5]Mental Health Centre Sct. Hans, Capital Region of Denmark, Institute of Biological Psychiatry, Copenhagen University Hospital, Copenhagen, Denmark
[6]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
[7]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[8]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark
[9]BiRC, Bioinformatics Research Centre, Aarhus University, Denmark
[10]NCRR - National Centre for Register-Based Research, Business and Social Sciences, Aarhus University, Aarhus, Denmark
[11]Centre for Integrated Register-based Research, CIRRAU, Aarhus University, Aarhus, Denmark
[12]Psychosis Research Unit, Aarhus University Hospital-Psychiatry, Denmark
[13]Mental Health Centre Copenhagen, Capital Region of Denmark, Copenhagen University Hospital, Copenhagen, Denmark
[14]Center for Disease Neurogenomics, [15]Icahn Institute for Data Science and Genomic Technology, [16]Friedman Brain Institute, [17]Department of Psychiatry, [18]Department of Genetics and Genomic Science, Icahn School of Medicine at Mount Sinai, New York, USA
[19]Mental Illness Research Education and Clinical Center (MIRECC), JJ Peters VA Medical Center, Bronx, New York, USA
[20]Department of Human Genetics , Radboud University Medical Center, Nijmegen, The Netherlands
[21]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands.
[22]Department of Medicine, Harvard Medical School, Boston, MA, USA.
[23]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
[24]Faculty of Medicine, University of Iceland, Reykjavík, Iceland.
[25]Department of Child and Adolescent Psychiatry, National University Hospital, Reykjavik, Iceland.
[26]The Centre for Child Development and Behaviour, Capital Area Primary Health Care, Reykjavik, Iceland.
[27]Departments of Psychiatry and of Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, NY

[28]Psychological Medicine and Clinical Neurosciences, Cardiff University, UK
[29] Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain
[30]Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain
[31]Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain.
[32]Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain
[33]Mental Health Services of the Capital Region, Child and Adolescent Psychiatric Center, Glostrup, Denmark
[34]Institute of Clinical Medicine, University of Copenhagen, Denmark
[35]Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia
[36]PROMENTA, Department of Psychology, University of Oslo, Oslo, Norway
[37]Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[38]Amsterdam Public Health Research Institute, Amsterdam, The Netherlands
[39]Department of Evolutionary Biology, Ecology and Environmental Sciences, University of Barcelona, Barcelona, Spain
[40]Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, USA
41Department of Psychiatry, Radboud University Medical Center, Nijmegen, The Netherlands

Correspond with: Ditte Demontis, email: ditte@biomed.au.dk, Anders D. Børglum, email: anders@biomed.au.dk

Abstract

Attention deficit hyperactivity disorder (ADHD) is a prevalent childhood neurodevelopmental disorder, with a major genetic component. Here we present a GWAS meta-analysis of ADHD comprising 38,691 individuals with ADHD and 186,843 controls. We identified 27 genome-wide significant loci, which is more than twice the number previously reported. Fine-mapping risk loci highlighted 76 potential risk genes enriched in genes expressed in brain, particularly the frontal cortex, and in early brain development. Overall, ADHD genetic risk was associated with several brain specific neuronal sub-types and especially midbrain dopaminergic neurons. In a subsample of 17,896 exome-sequenced individuals, we identified increased load of rare protein-truncating variants in cases for a set of risk genes enriched with likely causal common variants, suggesting implication of *SORCS3* in ADHD by both common and rare variants. We found ADHD to be highly polygenic, with around seven thousand variants explaining 90% of the SNP heritability. Bivariate gaussian mixture modeling estimated that more than 84% of ADHD influencing variants are shared with other psychiatric disorders (autism, schizophrenia and depression) and related phenotypes (e.g., educational attainment) when both concordant and discordant variants are considered. Additionally, we demonstrated that common variant ADHD risk was associated with impaired complex cognition such as verbal reasoning and a range of executive functions including attention.

Introduction

Attention deficit hyperactivity disorder (ADHD) is one of the most prevalent neurodevelopmental disorders affecting around 5% of children and persists into adulthood in two-thirds of cases[1,2]. It is characterized by extensive hyperactive, impulsive and/or inattentive behaviors that impair daily functioning. The disorder is associated with multiple adverse outcomes such as injuries[3], accidents[4], depression[5], substance use disorders[6], aggression[7], premature death[8], high rate of unemployment[9], and has large societal costs[10-12].

ADHD has a major genetic component with an estimated twin heritability of 0.74[13]. Despite this, ADHD's complex polygenic architecture makes it difficult to unravel the underlying biological causes of the disorder. Recently, we discovered the first 12 genome-wide significant loci for ADHD[14] in a GWAS of 20,183 cases and 35,191 controls (here referred to as ADHD2019) that combined the first wave of data from large Danish iPSYCH[15] cohort (iPSYCH1) with 11 ADHD cohorts collected by the Psychiatric Genomics Consortium (PGC). The results implicated brain-expressed genes and demonstrated considerable genetic overlap of ADHD with a range of phenotypes, including phenotypes within psychiatric, cognitive and metabolic domains. Related to this, a recent cross-disorder GWAS of ADHD and autism[16] has identified both shared and differentiating loci and showed that individuals diagnosed with both ADHD and autism have distinctive patterns of genetic association with other traits compared to those with only a single diagnosis, highlighting that further mapping of the shared genetic risk component with other psychiatric disorders is important for understanding the complexity of the genetics underlying ADHD. Additionally, we established the role of common variants in ADHD, explaining around 22% of the variance in the phenotype. Besides common risk variants, analyses of whole-exome sequencing data from a subset of the iPSYCH cohort have recently shown that rare variants also contribute to the risk for ADHD[17]. The burden of rare deleterious variants in evolutionary conserved genes in ADHD cases was increased compared to controls at a level comparable to what is found in autism spectrum disorder cases.

To better understand the biological mechanisms underlying ADHD, it is fundamental to conduct large genetic studies as demonstrated for other psychiatric disorders[18-20]. Here we present results from an updated GWAS meta-analysis of ADHD combining data from the extended Danish iPSYCH cohort (iPSYCH1 plus new iPSYCH2 data), the Icelandic deCODE cohort and the PGC, almost doubling the number of cases compared with ADHD2019. We fine-map identified risk loci and integrate our results with functional genomics data to pinpoint potential causal genes and evaluate the burden of rare deleterious variants in top-associated

genes. We characterize the polygenic architecture of ADHD and its overlap with other phenotypes by e.g. bivariate causal mixture modeling and perform polygenic score (PGS) analyses in order to test for association of ADHD-PGS with neurocognitive measures in the Philadelphia Neurodevelopmental Cohort (PNC).

RESULTS

Identification of new ADHD risk loci by GWAS meta-analysis

We conducted a GWAS meta-analysis based on expanded data from iPSYCH (25,895 cases; 37,148 controls), deCODE genetics (8,281 cases; 137,993 controls) and previously published data from 10 ADHD cohorts with European ancestry collected by the PGC (4,515 cases; 11,702 controls), which resulted in a total sample size in the meta-analysis of 38,691 individuals with ADHD and 186,843 controls (effective sample size (Neff_half) = 51,568; cohorts are listed in Supplementary Table 1). The iPSYCH cases comprise all individuals born in Denmark between 1981 and 2008 diagnosed with ADHD up to 2016[15,21]. They were identified in the Danish Psychiatric Central Research Register[22] and the National Patient Register[23] based on ICD10 diagnosis codes (Online Methods). Controls are population-based individuals without a diagnosis of ADHD. deCODE cases were also clinically diagnosed according to ICD10 or identified based on medication prescribed specifically for ADHD (mostly methylphenidate). deCODE controls were not diagnosed with ADHD or other major psychiatric disorders. Ascertainment and diagnosis criteria for the PGC cohorts have been described in detail previously[14].

Quality control and imputation was done separately for each cohort (Online Methods), and GWAS results from logistic regression (using relevant covariates) in the single cohorts were combined into a GWAS meta-analysis using an inverse-variance-weighted fixed effects model[24]. The meta-analysis identified 32 independent lead variants (i.e., with a squared correlation (r^2) < 0.1 between variants) located in 27 genome-wide significant loci (Figure

1; Table 1, locus plots in Extended data figures 1, forest plots in Extended data figures 2), including 21 novel loci. No statistically significant heterogeneity was observed between cohorts (Supplementary Figure 1). The three strongest associated loci ($P < 5x10^{-14}$) were located on chromosome 1 (in and around *PTPRF*), chromosome 5 (downstream of *MEF2C*) and chromosome 11 (downstream of *METTL15*). The latter locus on chromosome 11 is a novel ADHD risk locus. Four loci on chromosomes 1, 5, 11 and 20 had secondary genome-wide significant lead variants ($r^2 < 0.1$ between the index variant and the secondary lead variant within a region of 0.5 MB), but none of these remained genome-wide significant in analyses conditioning on the index variant using COJO (Supplementary Table 2).

Six of the previously identified 12 loci in the ADHD2019 study[14] were significant in the present study (Table 1), and the remaining six loci demonstrated P-values < $8x10^{-4}$ (Supplementary Table 3). Overall, the direction of association of the top loci (726 variants with $P < 1x10^{-4}$) was consistent with the direction of association in ADHD2019 for all loci, except for one (Supplementary Table 4).

Genetic correlations among cohorts and SNP-heritability

Genetic correlation analyses supported a high consistency in the phenotype across cohorts, including a high genetic correlation (r_g) between iPSYCH1 and iPSYCH2 ($r_g = 0.97$; s.e. = 0.06; Supplementary Table 5). The r_g was 0.93 (s.e.= 0.21) between deCODE and PGC, and 0.82 (s.e. = 0.08) between iPSYCH and deCODE; none of the genetic correlations were significantly different from 1. LD score regression analysis found an intercept of 1.04 (s.e.=0.009) and ratio of 0.092 (s.e. = 0.02), the latter indicating that around 90% of the deviation from null, in the distribution of the test statistics, reflects polygenicity (QQ-plot shown in Supplementary Figure 2). The SNP heritability (h^2_{SNP}) was estimated to 0.14 (s.e. = 0.01), which is lower than the previously reported h^2_{SNP} of 0.22[14]. The h^2_{SNP} for iPSYCH (h^2_{SNP} = 0.23; s.e. = 0.01) was in line with the previous finding, but lower h^2_{SNP} was observed for

PGC (h^2_{SNP} = 0.12; s.e. = 0.03) and deCODE (h^2_{SNP} = 0.081; s.e. = 0.014). The difference in SNP heritability was not caused by different sex distributions across cohorts as there were no significant differences in h^2_{SNP} between males and females in the iPSYCH and deCODE cohorts (Supplementary Table 5). Between-cohort heterogeneity in h^2_{SNP} is not unusual and has been observed for other disorders like e.g. MDD[25].

Mapping risk variants to genes and enrichment analyses

In order to link identified risk variants to genes by incorporating functional genomics information, we first identified sets of Bayesian credible variants for each risk locus, which most likely (probability > 95%) include a causal variant (Supplementary Table 6). The sets of credible variants were linked to genes based on genomic position, and information about expression quantitative trait loci (eQTLs) and chromatin interaction mapping was derived from human brain tissue datasets implemented in FUMA[26] (datasets selected in FUMA are listed in the Supplementary Information). Seventy-six (76) plausible ADHD risk genes were identified (Supplementary Table 7); four of the 76 were mapped by position alone. We found that this set of genes is significantly enriched in genes upregulated during early embryonic brain development (19th post-conceptual week; P_{one_sided} =0.0008; Supplementary Figure 3) and highly enriched for genes identified in GWASs of cognition-related phenotypes and reproduction (Supplementary Figure 4). Assessment of the role of the genes in synapses was evaluated using SynGO data[27]. Nine genes mapped to SynGO annotations, and genes encoding integral components of the postsynaptic density membrane were borderline significantly enriched ($5.43x10^{-3}$; q-value 0.022; genes *PTPRF*, *SORCS3*, *DCC*; Supplementary Figure 5; Supplementary Table 8). One SynGO mapped gene was also a part of the upregulated genes during early embryonic brain development (the Rho GTPase Activating Protein 39, *ARHGAP39*). Additionally, enrichment of the 76 genes in biological pathways was tested using data from 26 databases implemented in Enrichr[28,29], but no pathways showed significant

enrichment after Bonferroni correction (database significant findings can be found in Supplementary Table 8.2). Finally, MAGMA[30] pathway/gene-set analysis using gene-based P-values derived from the full GWAS summary statistics (i.e., no preselection of specific genes) did not reveal any significant findings (top gene-sets can be found in Supplementary Table 8.3).

Transcriptome-wide association analysis (TWAS) of the genetically regulated gene expression

To identify and prioritize ADHD risk genes we also performed a transcriptome-wide association study (TWAS) of the genetically regulated gene expression using EpiXcan[31] and expression data from the PsychENCODE Consortium[32] on genes as well as isoforms detected in 924 samples from the dorsolateral prefrontal cortex (DLPFC). The TWAS identified 15 genes (Supplementary Table 9) and 18 isoforms (Supplementary Table 10), which together identified 23 distinct genes (Supplementary Figure 6) with significantly different predicted gene expression levels in ADHD cases compared to controls (after Bonferroni correction correcting for all the 34,646 genes and isoforms tested; Supplementary Figure 6). Eight of the genes were among the 76 genes mapped by credible variants in FUMA. If we instead applied a less stringent correction using a false discovery rate < 5% we identified 237 genes with different predicted expression among cases and controls, of which 19 genes were also among the 76 prioritized risk genes. The *B4GALT2-205* isoform located in the genome-wide significant locus on chromosome 1 showed the strongest association ($P = 7x10^{-11}$), with lower predicted expression in ADHD compared to controls (Supplementary Figure 7.A). The expression model for *B4GALT2-205* implicated four genome-wide significant variants. The second top gene was *PPP1R16A* ($P = 1.4x10^{-8}$), which showed a predicted under-expression in cases compared to controls. The expression model for this gene implicated one genome-wide significant variant (Supplementary Figure 7.B).

Gene-based association, tissue and cell-type specific expression of ADHD risk genes

Gene-based association analysis using MAGMA[30] identified 45 exome-wide significant genes ($P < 2.72x10^{-6}$ (0.05/18381 genes)) associated with ADHD (Supplementary Table 11). Gene association results across the entire genome were tested for a relationship with tissue specific gene expression. This showed that brain-expressed genes, and in particular genes expressed in the cortex, are associated with ADHD (Supplementary Figure 8). This result was supported by LDSC-SEG[33] analysis, showing a significant enrichment in the heritability by variants located in genes specifically expressed in the frontal cortex (Supplementary Table 12).

Next, we examined neuronal cell-type specific gene expression in ADHD using two approaches. First, we tested for enrichment of variants located in cell-specific epigenomic peaks by intersecting our genetic associations with data from two recent catalogs of the human epigenome that profile major human body cell types[34] as well as brain-specific cell types[35]. Here we found enrichment for genes expressed in major brain neuronal cell types including both excitatory and inhibitory neurons (Supplementary Figure 9). Second, we performed cell-type specific analyses in FUMA[36] based on single cell RNA-sequencing data. This revealed a significant association ($P = 0.005$) between ADHD-associated genes and genes expressed in dopaminergic midbrain neurons (Linnarsson midbrain data[37]; Supplementary Figure 10; Supplementary Table 13).

Convergence of common and rare variant risk

In order to test for convergence of risk conferred by common variants and rare protein-truncating variants (rPTVs), we analyzed whole-exome sequencing data from a subset of the iPSYCH cohort consisting of 8,895 ADHD cases and 9,001 controls. We tested three gene-sets: 1) the 76 prioritized risk genes identified by positional and functional annotation, 2) the 45 significant genes in the MAGMA analysis, and 3) 18 genes with at least five credible

variants located in the coding region (Supplementary Table 14). While there was no indication of increased burden of rPTVs in the first gene-set (P = 0.39), the second gene-set showed borderline nominal significant enrichment (P = 0.05) and the set of genes identified based on credible variants had a significant increased burden of rPTVs in individuals with ADHD compared to controls (P = 0.015). For comparison, there was no enrichment in rare synonymous variants in the third gene-set (P = 0.59). When evaluating the 18 genes from the "credible gene-set" individually, *SORCS3* was nominally significantly (P = 0.008; Supplementary Table 14) enriched in rare rPTVs in ADHD cases when compared to a combined group of iPSYCH controls and gnomAD individuals (non-psychiatric non-Finnish Europeans; N=58,121), suggesting that *SORCS3* might be implicated in ADHD both by common and rare deleterious variants (Supplementary Table 14).

Genetic overlap of ADHD with other phenotypes

The genome-wide genetic correlation (r_g) of ADHD with other phenotypes was estimated using published GWASs (258 phenotypes) and GWASs of UK Biobank data (514 phenotypes), available in LDhub[38]. ADHD showed significant genetic correlation ($P < 2x10^{-4}$) with 56 phenotypes representing domains previously found to have significant genetic correlations with ADHD: cognition (e.g. educational attainment r_g = -0.55, s.e. = 0.021), weight/obesity (e.g. body mass index r_g = 0.27, s.e. = 0.03), smoking (e.g. smoking initiation r_g = 0.48; s.e = 0.07), sleep (e.g. insomnia r_g = 0.46, s.e. = 0.05), reproduction (e.g age at first birth r_g = -0.65, s.e. = 0.03), and longevity (e.g. mother´s age at death r_g = -0.42, s.e. = 0.07). When considering other neurodevelopmental and psychiatric disorders, autism spectrum disorder (ASD) (r_g = 0.42, s.e. = 0.05), schizophrenia (SCZ) (r_g = 0.17, s.e. = 0.03), major depressive disorder (MDD) (r_g = 0.31, s.e. = 0.07), and cannabis use disorder (CUD) (rg = 0.61, s.e. = 0.04) were significantly correlated with ADHD (Supplementary Table 15). In UK Biobank data, ADHD

demonstrated the strongest genetic correlation with a low overall health rating (r_g = 0.60, s.e. = 0.2; Supplementary Table 16).

Furthermore, we applied MiXeR[39] which uses uni- and bivariate gaussian mixture modeling to quantify the actual number of variants that: 1) explain 90% of the SNP heritability of ADHD and 2) overlap between ADHD and other phenotypes representing domains with high genetic correlation with ADHD (psychiatric disorders, smoking behavior, weight, reproduction, and sleep were evaluated). MiXeR considers all variants irrespective of the direction of genetic correlation (i.e., both variants with the same and opposite effects). Approximately 7,2 K (standard deviation (std.) = 324) common variants were found to influence ADHD, which is less than for SCZ (9,6 K; s.e. = 199), MDD (11, 7 K; std. = 345) and ASD (10 , 3 K; std. = 1, 011), and less than previously reported for bipolar disorder (BD) (8,6K, std. = 200)[18].

When considering the number of shared loci as a proportion of the total polygenicity of ADHD, the vast majority of variants influencing ADHD were also estimated to influence the other investigated psychiatric disorders (84%- 98%; Figure 2; Supplementary Figure 11, Supplementary Table 17). While the fraction of concordant variants (within the shared part) with ASD and MDD was at the high end (75-76%), it was lower for SCZ (59%) . When considering other phenotypes, insomnia demonstrated the smallest overlap with ADHD in terms of actual number of variants (4.5 K, std. = 1,281 ; 62% of ADHD variants shared) while almost all variants influencing ADHD also influence educational attainment, age at first birth and smoking (Figure 2; Supplementary Table 17). For insomnia and smoking, 83% and 79% of shared variants have concordant directions, respectively, while only 21% and 20% of ADHD risk variants were concordant with educational attainment and age at first birth associated variants, respectively (Supplementary Table 17).

Impact of ADHD polygenic scores on cognitive domains

Educational attainment is one of the phenotypes with the strongest negative genetic correlation with ADHD, as demonstrated above, and cognitive impairments in ADHD are well described[40]. In order to further explore how ADHD risk variants affect specific cognitive domains , we assessed the association of ADHD polygenic scores (PGS) with 15 cognitive measures in the Philadelphia Neurodevelopmental Cohort (PNC)[41,42]. This cohort is from the greater Philadelphia area and include individuals, 8-21 years of age, who received medical care at the Children's Hospital of Philadelphia Network. The PNC cohort (v1 release) is a sample of 8,722 genotyped individuals. The subsample of 4,973 individuals with European descent was utilized in this study. The Computerized Neurocognitive Battery[43] was used to assess cognitive performance in the subjects. The battery consists of 14 tests in 5 domains: executive-control, episodic memory, complex cognitive processing, social cognition, and sensorimotor speed. Additionally, the Wide Range Achievement Test (WRAT-4)[44] was used as a proxy measure for overall IQ[42].

ADHD-PGS was negatively associated with seven neurocognitive domains (Figure 3) with the strongest association for the WRAT-4 test (beta = -0.09, $P = 4.35x10^{-10}$). Besides that, ADHD-PGS was associated with measures of executive control (attention: beta = -0.07, $P = 3.25x10^{-7}$; working memory: beta = -0.05, $P = 2.45x10^{-3}$), complex cognition (verbal reasoning: beta = -0.08, $P = 4.74x10^{-12}$; non-verbal reasoning: beta = -0.06, $P = 6.28x10^{-4}$; spatial reasoning: beta = -0.06, $P = 5.15x10^{-5}$) and one measure of episodic memory (facial memory: beta = -0.05, $P = 3.23x10^{-3}$)(Supplementary Table 18). The negative association of ADHD risk variants with executive functions, especially attention, is in line with the inattention problems often observed in individuals with ADHD.

Discussion

The present study identified 27 genome-wide significant loci in the largest GWAS of ADHD to-date. We analyzed around twice as many ADHD cases as in the previous GWAS meta-

analysis (ADHD2019)[14] and more than doubled the number of associated loci, indicating that we have passed the inflection point for ADHD with respect to the rate of risk loci discovery. Six of the 12 previously identified loci were also significant in this study. Even though some previously identified loci demonstrated less association here, their associations were still strong and there was almost complete concordance in the direction of association between top-associated variants in this study and ADHD2019. It is not seldomly observed in GWAS of complex disorders that some loci may fluctuate around the significance threshold with increasing sample sizes until they eventually achieve stable significance, which can often be attributed to the *winner's curse* phenomenon where effect size estimates close to the discovery threshold tend to be overestimated in initial GWAS[45] .

We report a lower h^2_{SNP} for ADHD (h^2_{SNP}=0.14) than estimated previously (h^2_{SNP}=0.22). This is driven by a lower h^2_{SNP} in the PGC and deCODE cohorts compared to iPSYCH. Different ascertainment and diagnostic strategies and designs among PGC cohorts could decrease the h^2_{SNP}, while lower effective sample size[46] in Iceland, and thus fewer recent variants might bias h^2_{SNP} downwards in the deCODE cohort[47].

We refined ADHD´s genetic architecture by estimating that around 7,2 K (std. = 324) common variants can explain 90% of the $h^2_{SNP.}$ This is a higher estimate than reported based on the 2019 ADHD GWAS (5,6K, std. = 400)[48]. Interestingly, the estimated number of ADHD risk variants was lower than observed for three genetically correlated psychiatric disorders (SCZ, MDD, ASD). It could be hypothesized that a relatively larger phenotypic/genetic heterogeneity within these three disorders (as reported in e.g., (REF[49,50])) could explain a part of the larger number of common risk variants influencing these phenotypes. ADHD is often comorbid with other psychiatric disorders[51] with e.g. 12-16% also diagnosed with autism[16,52,53] and around 40% with depression[54], which is also reflected in the genetic correlations reported

here and previously[14]. Strikingly, when assessing both concordant and discordant allelic directions, over 90% of ADHD risk variants also seem to influence SCZ and MDD and 84% influence ASD . This extensive sharing with SCZ, MDD and ASD is at the same level as observed for SCZ and bipolar disorder[39], which are among the most genetically correlated mental disorders[55]. Notably, for both MDD and ASD around 75% of the variants shared with ADHD demonstrated concordant direction of association. The large sharing of variants influencing ADHD and other psychiatric disorders, when assessing both concordant and discordant allelic directions, suggest that the disorders are even more intermingled with respect to their common genetic architecture than previously thought based on their overall genetic correlations[39,55].

For common variants, the developmental trajectory towards ADHD might therefore be influenced by variants involved in several psychiatric disorders but with disorder-specific effect sizes rather than actual ADHD-specific risk variants. We also note that almost all variants that influence ADHD overlap with educational attainment[56] and that around 21% of ADHD risk variants are associated with increased educational attainment while the vast majority (79%) are associated with decreased educational attainment. This is consistent with the overall negative genetic correlation with educational attainment. For the models indicating a high number of shared variants between phenotypes (ADHD vs MDD, SCZ, BMI, educational attainment, age at first birth, and smoking) we found support (evaluated using the Akaike Information Criterion[57]) for the best fitting MiXeR models above the "minimal model", which indicate that the data support the existence of a polygenic overlap, beyond the minimal level needed to explain the observed genetic correlations. For ADHD vs ASD, the model had limited support and the results should therefore be interpreted with caution.

Fine-mapping of the 27 loci identified credible variants but in general, few variants had high posterior probabilities, with only four variants having posterior probabilities greater than 0.5 in all three fine-mapping methods, and none were linked to specific genes based on our

functional annotation analyses. Linking the credible variants to genes by integration with functional genomics data identified 76 prioritized risk genes, which were enriched among genes upregulated during early embryonic development and involved in cognitive abilities identified by GWAS of cognitive phenotypes. Among the 76 genes were *PPP1R16A* and *B4GALT2* (mapped by psychENCODE eQTLs; Supplementary Figure 12.A and B), which were also the top-ranking genes in our TWAS of DLPFC expression, both showing a predicted decreased expression in cases compared to controls. These genes have not previously been linked to psychiatric disorders, but both have been linked to educational attainment[56]. The set of risk genes also included *PTPRF*, *SORCS3* and *DCC* which encode integral components of the postsynaptic density membrane. Involvement of postsynaptic components in the pathology of ADHD has been reported previously (identified by MTAG analyses of ADHD and related psychiatric disorders[58]) and also for schizophrenia[59]. We would also like to highlight *FOXP1* and *FOXP2*. The genome-wide significant signals were located within the transcribed regions of both genes and were additionally implicated in ADHD by genome-wide significant variants being eQTLs (*FOXP2*, Supplementary Figure 12.C) or located in chromatin interacting regions (*FOXP1*, Supplementary Figure 11.D) in brain tissue. *FOXP2* was identified in the ADHD2019 study[14], and was recently suggested as a risk gene for cannabis use disorder[60]. *FOXP1* is a new ADHD locus, which has previously been associated with schizophrenia[20]. Both *FOXP1* and *FOXP2* encode transcription factors that can heterodimerize to regulate transcription in brain tissues[61,62] and have been implicated in speech disorders and intellectual disability[63] by highly penetrant rare variants.

Overall, less than half of the TWAS Bonferroni significant genes overlapped with the 76 candidate risk genes (40% of “TWAS transcript genes”; and 47% of “TWAS genes”; Supplementary Figure 13). This was not unexpected and could be due to several factors, including noise in the data and that TWAS results are based on expression in adult brains whereas a large proportion of individuals in the GWAS are children. Additionally, eQTLs used

to derive TWAS models might not overlap GWAS identified variants as the two types of methods systematically are biased toward identification of different types of variants[64]. We expect that future studies based on larger expression data sets (and/or age-appropriate brain tissues) in combination with larger GWAS will converge on a larger number of genes identified by both approaches.

We report convergence of common and rare variants in a set of 18 genes defined by location of credible variants. Thirteen of the genes were hit by rPTVs and for eight of them there was a higher load in cases compared to controls. The signal was not driven by a few genes but by several genes with an increased burden of rPTVs. Of particular note *SORCS3* seems to be implicated in ADHD by both common and rare variants. Common variants in *SORCS3* show strong pleiotropic effects across several major psychiatric disorders[55], but to our knowledge, rare variant analyses have not implicated *SORCS3* in psychiatric disorders before. Our results add to the emerging picture of overlap between genes and pathways affected by common and rare variants in psychiatric disorders[59,65,66,67].

We found that ADHD risk was associated with common variants located in genes significantly expressed in the brain, especially the frontal cortex. We also observed an enrichment of ADHD risk variants in genes expressed in major cell types of the brain including both excitatory and inhibitory neurons and in midbrain dopaminergic neurons. The findings for frontal cortex and dopamine neurons fit well with the motor, reward and executive function deficits associated with ADHD; the frontal cortex is involved in executive functions through attention and working memory[68], and midbrain dopaminergic neurons are essential for controlling key functions, such as voluntary movement[69] and reward processing[70]. This interpretation is further supported by our ADHD-PGS analyses in PNC which revealed that common ADHD risk variants impair several domains of cognitive abilities, including attention and working memory.

The PGS analyses in PNC identified strong association of polygenic ADHD risk with decreased overall IQ (approximated by the WRAT test scores), which is in line with the high negative genetic correlation of ADHD with educational attainment and the observation that 78% of all ADHD risk variants are associated with decreased educational attainment. Interestingly, we found that ADHD-PGS associates with decreased attention, which is a key ADHD symptom, and with impairments in measures of other cognitive traits such as working memory. Smaller studies have previously analyzed the impact of ADHD-PGS on executive functions with mixed results[71-74]. This study robustly identifies specific cognitive domains impacted by ADHD-PGS and our results support ADHD-PGS to be negatively associated with neurocognitive performance.

In summary, we identified new ADHD risk loci, highlighted candidate causal genes and implicated genes expressed in frontal cortex and several brain specific neuronal sub-types in ADHD. Our analyses revealed ADHD to be highly polygenic, influenced by thousands of variants, of which the vast majority also influence other psychiatric disorders with concordant or discordant effects. Additionally, we demonstrated that common variant ADHD risk has an impairing impact on a range of executive functions. Overall, the results advance our understanding of the underlying biology of ADHD, and reveal novel aspects of ADHD´s polygenic architecture, its relationship with other phenotypes and its impact on cognitive domains.

METHODS

Samples, quality control and imputation

iPSYCH

The iPSYCH[15,21] cohort consists of 129,950 genotyped individuals, among which 85,891 are cases diagnosed with at least one of six mental disorders (i.e. ADHD, SCZ, BD, MDD, ASD, post-partum disorder) and the remaining are population-based controls. Samples were selected from a baseline birth cohort comprising all singletons born in Denmark between May 1, 1981, and December 31, 2008, who were residents in Denmark on their first birthday and who have a known mother (N = 1,657,449). ADHD cases were diagnosed by psychiatrists at in- or out-patient clinics according to the ICD10 criteria (F90.0, F90.1, F98.8 diagnosis codes) identified using the Danish Psychiatric Central Research Register[22] and the Danish National Patient register[23]. Diagnoses were given in 2016 or earlier for individuals at least 1 year old. Controls were randomly selected from the same nationwide birth cohort and not diagnosed with ADHD. The study was approved by the Danish Data Protection Agency and the Scientific Ethics Committee in Denmark.

The samples were genotyped in two genotyping rounds referred to as iPSYCH1 and iPSYCH2. DNA extraction and subsequent whole-genome amplification was performed as previously described[15]. iPSYCH1 samples were genotyped using Illumina's PsychChip array and iPSYCH2 samples using Illumina´s global screening array v.2 (Illumina, CA, San Diego, USA). iPSYCH1 genotypes were called using GenCall and Birdseed and iPSYCH2 genotypes were called using GenTrain V3.

Pre-imputation quality control and imputation was performed on genotypes from the full set of genotyped individuals for iPSYCH1 and iPSYCH2 separately. Quality control, imputation and primary association analyses were done using the bioinformatics pipeline "Ricopili"[75]. Subjects and variants were included in the imputation based on the following quality control parameters: variant call rate > 0.95 (before sample removal), subject call rate > 0.95, autosomal heterozygosity deviation (| Fhet | < 0.2), variant call rate > 0.98 (after sample removal), difference in variant missingness between cases and controls < 0.02, and SNP Hardy-Weinberg equilibrium (HWE) ($P > 10^{-6}$ in controls or $P > 10^{-10}$ in cases). The iPSYCH1 samples were

genotyped in 23 genotyping waves and thus additional steps were taken in order to eliminate potential batch effects. Only variants present in more than 20 waves and with no significant association with wave status, were retained. Imputation was done using the prephasing/imputation stepwise approach implemented in EAGLE v2.3.5[76] and Minimac[77], using the Haplotype Reference Consortium[78] panel v1.0. iPSYCH1 comprised imputed genotypes from 20,175 ADHD cases and 25,836 population-based controls without ADHD, and iPSYCH2 contained imputed genotypes from 10,624 ADHD cases and 18,255 controls. Best guess genotypes from iPSYCH1 and iPSYCH2 were merged in order to identify potential duplicated samples and related individuals within and across the entire sample. Related (duplicated samples) were identified by "identity by state" analysis in plink v1.9, and one individual was excluded from pairs of subjects with pi_hat > 0.2. For this a set pruned best guess genotypes (imputation INFO score > 0.8; $r^2 < 0.075$; markers located in long range LD regions defined by Price et al.[79] excluded) with minor allele frequency > 0.05 and no deviation from Hardy-Weinberg equilibrium (HWE P > $1x\ 10^{-4}$) was used. This step removed 5,326 individuals.

Genetic outliers were identified by principal component analysis (PCA) which was performed separately for iPSYCH1 and iPSYCH2 using high quality variants as described above and the software Eigensoft[80]. Non-European individuals were excluded if their principal component (PC) values for PC1 and PC2 were greater than six standard deviations from the centre of an ellipsoid where the centre was based on the mean values of PC1 and PC2 of a sub-sample of Danish individuals. The subsample of Danes was defined using registry information requiring the individuals and their parents' birth country to be Denmark. After exclusion of non-European samples PCAs were re-run to exclude remaining population stratification, which was done by visual inspection of PCA plots. After QC the iPSYCH1 ADHD sample included 38,899 individuals and iPSYCH2 included 24,144 individuals.

deCODE

The deCODE cohort consisted of 8,281 individuals with ADHD. These were either individuals with a clinical diagnosis of ADHD (N=5,583) according to the ICD10 criteria (ICD10-F90, F90.1, F98.8) or individuals that have been prescribed medication specific for ADHD symptoms (ATC-NA06BA, mostly methylphenidate) (N=2,698). The control sample did not contain individuals with a diagnosis of SCZ, BD, ASD or self-reported ADHD symptoms or diagnosis. The study was approved by the National Bioethics Committee of Iceland (VSN 15-047) and all participants who donated samples gave informed consent. Samples were assayed with several Illumina arrays at deCODE genetics and genotypes called using GraphTyper2[81]. SNPs with low call rate (<95%), significant deviation from Hardy-Weinberg equilibrium (P<0.001), and excessive inheritance error rates (>0.001) were excluded. Variant imputation was performed based on the IMPUTE HMM model and long-range phasing, as described previously[82]. To rule out genetic heterogeneity between individuals with ADHD identified based on diagnosis codes and individuals identified based on medication prescription, we performed a separate GWAS for each group using non-overlapping controls and the same GWAS procedure as described below (diagnosed ADHD: 5,583 cases, 68,280controls; medication identified ADHD: 2,698 cases and 69,405controls). Subsequently LD score regression was used to estimate h^2_{SNP} and the genetic correlation between the two groups. The genetic correlation was practically one (r_g = 0.98, s.e. = 0.27) and the h^2_{SNP} was similar in the two groups (Supplementary Table 5). These findings support that the genetic architecture underlying ADHD in the two groups is similar, and that pooling individuals together in the GWAS will not introduce significant heterogeneity.

PGC cohorts

We used summary statistics from the 10 PGC cohorts with European ancestry generated as a part of our previous GWAS meta-analysis of ADHD. Detailed information about cohort design, genotyping, QC and imputation can be found in Demontis and Walters et al[14].

GWAS meta-analysis of ADHD

GWASs were performed separately for iPSYCH1 (17,019 cases and 21,880 controls) and iPSYCH2 (8,876 cases and 15,268 controls) using dosages for imputed genotypes and additive logistic regression with the first 10 PCs (from the final PCAs) as covariates using PLINK v1.9. GWAS of deCODE samples (8,281 ADHD cases; 137,993 controls) was done using dosage data and logistic regression with sex, age, and county of origin as covariates. To account for inflation due to population stratification and cryptic relatedness, test statistics were divided by an inflation factor (lambda = 1.23) estimated from LD score regression as done previously[60]. Findings from analyses of the genetic structure of the Icelandic population by Price et al.[83] support that lambda correction will ensure proper correction without false positives. Subsequently alleles were converted to match HRC alleles.

For the PGC cohorts we used GWAS summary statistics for each of the 10 European PGC cohorts generated as a part of our previous GWAS meta-analysis[14].

Sex was used as covariate in the GWAS of deCODE samples but not in the other cohorts. In order to rule out potential biases created by differentiating polygenic architecture in males and females we estimated sex-specific h^2_{SNP} and the genetic correlation between sexes using LD score regression[84] and summary statistics from sex-specific GWAS (iPSYCH1+2: 7,960 females with ADHD and 18,425 controls, 17,929 males with ADHD and 18,716 controls; deCODE: 3,896 females with ADHD and 75,369 controls, 4,385 males with ADHD and 62,624 controls). No difference in h^2_{SNP} between males and females was observed and the genetic correlation between sexes was practically 1 (Supplementary Table 5), supporting previous findings of similar polygenic architecture in males and females[85,86]. Thus, we found no strong

argument for including sex as a covariate. Additionally, we evaluated the impact of using age as a covariate in the GWAS of iPSYCH samples by estimating the correlation of the -log10(P-values) from association results generated with and without age as covariate. The correlation between the -log10(P-values) for the 27 genome-wide significant variants was 1 and very high when considering all variants ($r = 0.95$, $P < 2.2x10^{-16}$). Visual inspection of the correlation plot of the -log10(P-values) revealed no strong impact on highly associated variants and no systematic bias among variants with low association (Supplementary Figure 14).

Summary statistics from GWAS of the individual cohorts, containing variants with imputation quality (INFO score) > 0.8 and minor allele frequency > 0.01, were meta-analyzed with a fixed effects standard error weighted meta-analysis using METAL (version 2011-03-25)[24]. Only variants supported by an effective sample size greater than 60% were retained in the final summary statistics (6,774,228 variants).

Concordance in the direction of associations in the present GWAS with associations in the ADHD2019 data were evaluated by a sign-test at different p-value thresholds (see thresholds in Supplementary Table 4).

Conditional analysis

We identified potentially independent genome-wide significant lead variants for four loci located on chromosome 1 (two secondary lead variants), 5, 11 and 20. In order to evaluate if these variants were independent from the lead variants, we performed association analyses of the secondary variants while conditioning on the index variant in the locus using COJO as implemented in GCTA[87].

Identification of sets of credible variants

To identify sets of causal variants we fine-mapped each of the 27 genome-wide loci using three fine-mapping tools, FINEMAP v. 1.3.1 (Ref.[88]), PAINTOR v.3.0 (Ref.[89]) and

CAVIARBF v.0.2.1 (Ref.[90]), using CAUSALdb-finemapping-pip downloaded from https://github.com/mulinlab/CAUSALdb-finemapping-pip[91]. Since no secondary lead variants remained genome-wide significant after conditional analyses, one causal variant was assumed per locus. Variants located in a region of 1MB around index variants were included in the analyses. We used a threshold of 95% for the total posterior probability of the variants included in the credible sets and only variants claimed to be within the set by all three methods were included in the final credible set for each locus.

Genetic correlations among ADHD cohorts and SNP heritability

SNP heritability (h^2_{SNP}) and pair-wise genetic correlation among the cohorts were calculated using LD score regression[84] analysis of summary statistics from GWAS of deCODE samples, meta-analysis of iPSYCH1+iPSYCH2 and meta-analysis of the 10 PGC cohorts (applying the same approach as described for the meta-analysis of all cohorts). Conversion of h^2_{SNP} estimates from observed scale to the liability scale was done using a population prevalence of 5%. Test for significant differences in h^2_{SNP} between cohorts was done using a Z-test.

Mapping of risk genes, enrichment and pathway analyses

To link identified risk variants to genes, we used the set of credible variants (identified as described above) for each locus and linked variants to genes based on genomic position and functional annotations in FUMA[26]. Protein coding genes were mapped if they were located with a distance of 10Kb up- or downstream of the index variants or if a credible variant was annotated to the gene based on eQTL data or chromatin interaction data from human brain (data sets used in the mapping can be found in the Supplementary Note). The mapping linked credible variants to 76 ADHD prioritized risk genes. These genes were used in gene-set enrichment analyses in order to evaluate if the candidate genes were enriched among 1) genes differentially expressed in specific brain tissues, 2) genes differentially expressed at specific

brain developmental stages, 3) genes encoding proteins involved in synapses and 4) genes encoding proteins in specific biological pathways. We corrected for multiple testing separately for each of these hypotheses. The first two aims were addressed by performing enrichment analyses in the GENE2FUNC module in FUMA. Enrichment of ADHD risk genes among predefined sets of differently expressed genes in GTEx (54 tissue types) and Brainspan (29 different ages of samples and 11 general developmental stages) data using hypergeometric test and protein coding genes were chosen as background genes.

The third aim was addressed using SynGO[27] (dataset version: 20210225) test for enrichment among the 76 risk genes for genes involved in synaptic processes and locations. We analyzed for enrichment in two subsets; "biological process" (201 gene sets) and "cellular component" (92 gene sets). We controlled using a background set of "brain expressed" genes provided by the SynGo platform (defined as 'expressed in any GTEx v7 brain tissues) containing 18,035 unique genes of which 1,225 overlap with SynGO annotated genes. For each ontology term, a one-sided Fisher exact test was performed to compare the list of ADHD risk genes and the selected background set. To find enriched terms within the entire SynGO ontology, the most specific term is selected where each 'gene cluster' (unique set of genes) is found and then multiple testing correction is applied using False Discovery Rate (FDR) on the subset of terms that contain these 'gene clusters'. Only ontology terms with gene sets with a minimum of three genes were included in the enrichment analysis.

The fourth aim was addressed by testing if the 76 genes were enriched in pathways/gene sets using Enrichr[28,29] and its implemented data bases (26 databases). Only pathways enriched with more than two genes were considered. We made a conservative approach and only considered pathways to be significant if the within database adjusted P-value was smaller than 0.002 (0.05/26 databases evaluated). After correction for the number of data bases no significantly enriched pathways were identified.

Finally, we tested for enrichment among the 76 genes of genes reported from the GWAS catalog (2019) and UK biobank GWASs (v1), and used https://appyters.maayanlab.cloud/Enrichr_Manhattan_Plot/ to visualize the results.

We also conducted pathway enrichment analysis using results from the full GWAS meta-analysis (i.e. no preselection of genes) by performing MAGMA[30] gene-set analysis in FUMA. We tested 15,496 gene sets from MsigDB v7.0[92] (Curated gene sets: 5500, GO terms: 9996). MAGMA uses gene-based P-values to test for enrichment in association signals in genes belonging to specific biological pathways or processes and applies a competitive test to analyze if the genes of a gene-set are more strongly associated with the phenotype than other genes, while correcting for a series of confounding effects such as gene length and size of the gene-set. Genes located in the MHC region (hg19:chr6:25-35M) were excluded and a window size of zero bases around genes was used. Correction for multiple testing was done using Bonferroni correction.

Transcriptomic imputation model construction and TWAS

Transcriptomic imputation models were constructed as previously described[31] for dorso-lateral prefrontal cortex (DLPFC) transcript levels[93]. The genetic dataset of the PsychENCODE cohort was uniformly processed for quality control (QC) steps before genotype imputation. The analysis was restricted to samples with European ancestry as previously described[31]. Genotypes were imputed using the University of Michigan server[94] with the Haplotype Reference Consortium (HRC) reference panel[95]. Gene expression information (both at the level of gene and transcript) was derived from RNA-seq counts which were adjusted for known and hidden confounds, followed by quantile normalization[93]. For the construction of the transcriptomic imputation models we used EpiXcan[31], an elastic net based method, which weighs SNPs based on available epigenetic annotation information[96]. We performed the transcript-trait association analysis for ADHD as previously described[31]. Briefly, we applied the S-PrediXcan method[31]

to integrate the ADHD GWAS meta-analysis summary statistics and the transcriptomic imputation models constructed above to obtain association results at both the level of genes and transcripts.

Gene-based association and tissue-specific expression of ADHD risk genes

We used MAGMA v1.08 implemented in FUMA v1.3.6a (Ref.[26]) to perform gene-based association analysis using the full summary statistics from the GWAS meta-analysis. Genome-wide significance was assessed through Bonferroni correction for the number of genes tested ($P = 0.05/18381 = 2.72\times10^{-6}$).

The relationships between tissue specific gene expression profiles and ADHD-gene associations was tested using MAGMA gene-property analysis of expression data from GTEx (54 tissue types) and BrainSpan (29 brain samples at different ages) available in FUMA (See Supplementary Information for data sets selected in FUMA).

Enrichment in h^2_{SNP} of ADHD associated variants located in or close to genes expressed in specific brain regions was estimated using LDSC-SEG[33]. Annotations indicating specific expression in 13 brain regions from the GTEx gene-expression database were downloaded from: https://alkesgroup.broadinstitute.org/LDSCORE/LDSC_SEG_ldscores/.

Cell type-specific expression of ADHD risk genes

We tested for enrichment in the ADHD h^2_{SNP} heritability of variants located in cell type specific epigenetic peaks by examining the overlap of common genetic risk variants with open chromatin from a DHS study (DNase I hypersensitive sites) profiling major human cell types[34] and an scATAC-seq study (single-cell assay for transposase accessible chromatin)[35] using LD-score partitioned heritability approach[97]. 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. We applied Bonferroni correction (correcting for 23 cell types) and results below P = 0.0022 were considered significant.

Additionally, we performed cell-type specific analyses implemented in FUMA, using data from 13 single-cell RNA sequencing data sets from human brain (data sets listed in the Supplementary Information). The method is described in detail in Watanabe et al.[36]. In short, the method uses MAGMA gene-property analysis to test for association between cell specific gene expression and ADHD-gene association and correction for multiple testing adjusts for all tested cell types across datasets. In a second step, systematic step-wise conditional analysis per dataset is performed in order to correct for false positives when there is high correlation in expression profiles among cell-types; only cell-type specific expression of ADHD risk genes in DA1 neurons remained significant after this step.

Overlap of common ADHD risk variants with rPTVs

We analyzed the overlap of common variants with rPTVs in a subset of iPSYCH samples that have also been whole exome sequenced. DNA was extracted from dried blood spot samples of the study subjects and whole genome amplified in triplicates[98,99], the coding regions of the genome were extracted using the llumina Nextera capture kit and sequencing was performed in multiple waves (Pilot 1, Wave 1, Wave 2 and Wave 3) using the Illumina HiSeq platform at the Broad Institute.

A major part of the data (Pilot 1, Wave 1, Wave 2) was also included in the recent study by Satterstrom et al.[17], and the same quality control procedure was applied here. In short, the sequencing data were aligned to the reference genome using the BWA[100] (Hg19) and genotype calling was done using the best practice recommended by the Genome Analysis Toolkit[101] (GATK) v.3.4, and additional QC steps performed using Hail (Hail Team. Hail 0.2. https://github.com/hail-is/hail). All variants annotated to ACMG[102] genes were removed due to Danish regulations. Samples were removed if they lacked complete phenotype information,

inconsistencies of the imputed sex with the reported sex, if they were duplicates or genetic outliers identified by principal component analysis (using a set of common variants and the software Eigensoft[80]), if they had an estimated level of contamination > 5% or if they had an estimated level of chimeric reads > 5%.

Only autosomal genotypes were included in our analyses. Genotypes were removed if they did not pass GATK variant quality score recalibration (VQSR) and had read depths < 10 or > 1,000. Homozygous alleles were removed if they had reference calls with genotype quality < 25, homozygous alternate alleles with PL(HomRef) less than 25 or < 90% reads supporting alternate allele. Heterozygous alleles were removed if they had PL(HomRef) < 25 or < 25% reads supporting the alternate allele, less than < 90% informative reads, or a probability of the allele balance calculated from a binomial distribution centered on 0.5 less than 1×10^{-9}. After these genotype filters, variants with a call rate < 90% were removed, then samples with a call rate < 95% were removed and then additional removal of variants with a call rate < 95%. Additionally, one of each pair of related samples was removed from pairs with pi-hat values ≥ 0.2. After QC, the number of individuals were 8,895 ADHD cases and 9,001 controls.

The QCed variants were annotated using SnpEff[103] version 4.3t. The variants were also annotated with information about allele counts in the gnomAD[104] exomes r2.1.1 database using SnpSift[103] version 4.3t. Variants were only included if they were located in consensus high-confident regions with high read depth in both iPSYCH and gnomAD (80% of the samples in both datasets had at least 10× sequencing coverage in the region). Variants were defined as rPTVs if they were annotated as having large effects on gene function (nonsense variant, frameshift, splice site). We defined a variant as being rare if it had an allele count of five or less across the combination of the full iPSYCH exome-sequencing dataset (n=28,448) and non-Finnish Europeans in the nonpsychiatric gnomAD exome database (n = 44,779).

We tested for increased burden of rPTVs in ADHD compared to controls in three gene-sets (1) the 76 genes linked to credible variants based on position and functional genomic data, (2) the 45 exome-wide significant genes identified in MAGMA analysis, (3) Genes with at least five credible variants within the coding regions. The requirement of five credible variants was chosen in order to prioritize the most likely causal genes. This threshold excluded eight genes located in the same locus covering a broad LD region on chromosome 3 (Supplementary Data 1; page 25). Additionally, two other genes with less than five credible variants were excluded located in two other loci on chromosome 3.

The burden of rPTVs and rare synonymous (rSYNs) in cases compared to controls was tested for the three gene-sets with logistic regression corrected using the following covariates: birth year, sex, first ten principal components, number of rSYN, percentage of target with coverage > 20x, mean read depth at sites within the exome target passing VQSR, total number of variants, sequencing wave.

Only significant enrichment in the set of 18 genes identified based on credible variants was found. We therefore looked specifically into these genes to identify if the signal was driven by specific genes. rPTVs were found in 13 of the genes and out of these eight genes had more rPTVs in cases compared to controls when looking at the raw counts (Supplementary Table 14). We performed gene-based burden test using EPACTS (https://genome.sph.umich.edu/wiki/EPACTS) and logistic Wald test (correcting using the covariates as described above). Additionally, in order to increase power to detect increased burden of rPTVs at the gene-level in ADHD cases, we combined iPSYCH controls with information about rPTVs in gnomAD (non-Finnish European individuals), done as described previously[17]. We performed gene-based test using Fisher´s exact test and only genes with higher number of rPTVs in cases compared to controls in the iPSYCH data were considered.

Genetic overlap with other phenotypes

We estimated genetic correlations of ADHD with other phenotypes in LDhub[38] (published GWASs: 255 phenotypes; UK Biobank GWASs: 514 phenotypes). Additionally, genetic correlations with three phenotypes not available in LDhub (cannabis use disorder[60], smoking initiation[105] and education attainment[56]) were estimated locally using LD score regression[84].

We applied MiXeR[39] to our ADHD GWAS summary statistics and GWAS from a selection of complex traits showing high genetic correlation with ADHD: ASD[49], SCZ[106], BMI[107], educational attainmet[108], Age at first birth[109], smoking initiation[105], insomnia[110] and an unpublished new GWAS meta-analysis of major depressive disorder including 371,184 MD cases and 978,703 controls (Supplementary Table 17) to quantify (i) the number of variants influencing each trait and (ii) the genetic overlap between ADHD and each of the other traits. Before MiXeR was run, summary statistics were prepared by removing variants in the major histocompatibility complex (MHC) region, chromosome 6, base position 26000000-34000000 and by running the munge_sumstats.py function from LDSC, where case/control studies were balanced with the effective sample size (Neff=4/(1/ncase+1/ncontrol)). We used MiXeR with default settings (https://github.com/precimed/mixer) in a two-step process: 1) We ran a univariate model for each trait to estimate the number of common variants having a non-zero genetic additive impact on the phenotype. The univariate model generates estimates of "polygenicity" (i.e., the proportion of non-null variants) and "discoverability " (i.e., the variance of effect sizes of non-null SNPs). In this analysis MiXeR incorporates LD information and allele frequencies for 9,997,231 variants extracted from 1000 Genomes phase 3 data. Model fit was based on likelihood maximization of the signed test statistics (GWAS z-scores). Estimates and standard errors were calculated by performing 20 iterations using 2 million randomly selected variants (with MAF > 0.05) for each iteration, followed by random pruning at a linkage disequilibrium threshold of r^2=0.8. The estimates of the total number of phenotype-influencing variants reported explain 90% of h^2_{SNP}.

2) The variance estimates from the previous univariate step were used to run a bivariate model in a pairwise fashion (i.e. ADHD vs. each of the other traits) which produced estimates of four components representing (i) null SNPs in both traits; (ii-iii) SNPs with a specific effect on the first or on the second trait; and (iii) SNPs with a non-zero effect on both traits (for details on the method see also[18]). The models were evaluated by the Akaike Information Criterion[57] (AIC) and illustrated with modeled versus observed conditional quantile-quantile (Q-Q) plots (Supplementary Figure 11). The AIC values can be found in Supplementary Table 17.

PGS analysis of cognitive measures in PNC

PGS analysis was performed on 4,973 individuals of European ancestry from the Philadelphia Neurodevelopmental Cohort (PNC), ages 8-21. Genotypes used for PGS generation were from the first PNC release (dbGaP phs000607.v1.p1), while the neurocognitive phenotypes used were from the third release (dbGaP phs000607.v3.p2). Pre-imputation quality control steps included removing individuals whose genotypically inferred and phenotypically reported sex did not align, those with heterozygosity rates ± three standard deviations from the mean, those who did not meet the individual-level missingness filter of 0.05, and those who did not meet the identity by descent (IBD) filter (PI-HAT > 0.185). Additionally, SNPs were removed if they had high missingness rate (> 0.05), high deviation from Hardy-Weinberg equilibrium (HWE) ($P < 0.00001$), and low minor allele frequency (MAF) filters (MAF < 0.01). Genotype imputation was performed on the Michigan Imputation Server (https://imputationserver.sph.umich.edu/index.html#!), using the reference panel HRC r1.1 2016 and selecting the "Mixed population" option. Post-imputation processing included removing SNPs with an imputation $R^2 < 0.03$ and filtering based on the HWE, MAF, and missingness thresholds stated above. Outlier individuals were identified and removed by plotting the first two Multi-Dimensional Scaling (MDS) dimensions. Individuals of European ancestry were identified using the GemTools package

(http://www.compgen.pitt.edu/GemTools/GEM%20Documentation.pdf) and Ward's hierarchical clustering in R.

The software PRS-CS[111] was used to process ADHD GWAS summary statistics and assign per-allele posterior SNP effect sizes. A European LD reference panel generated from the 1000 Genomes Project data (can be downloaded here: https://github.com/getian107/PRScs) was utilized. The following default settings were used for PRS-CS: parameter a in the γ-γ prior = 1, parameter b in the γ-γ prior = 0.5, MCMC iterations = 1000, number of burn-in iterations = 500, and thinning of the Markov chain factor = 5. Additionally, the global shrinkage parameter phi was determined using a fully Bayesian method. Plink v2.0[112] was then used to calculate individual-level ADHD PGS. Linear regression was used to test the association between ADHD PGS and neurocognitive phenotypes measured in the PNC. The neurocognitive measures were obtained using the Computerized Neurocognitive Battery, which consists of 14 tests in 5 domains: executive-control, episodic memory, complex cognitive processing, social cognition, and sensorimotor speed. The battery has been described in detail elsewhere (Ref. [43]). Additionally, association of ADHD-PGS with results from the Wide Range Achievement Test (WRAT-4)[44] were analyzed.

Neurocognitive phenotype metrics and the ADHD-PGS values were each scaled such that the mean value is 0 and the standard deviation is 1 (using the function scale() in base R). Age (at time of neurocognitive testing), age squared, genotyping batch, sex, and the first 10 MDS dimensions were used as covariates. The total variance explained by ADHD-PGS and model covariates for each neurocognitive phenotype was reported using Adjusted R^2. Additionally, the variance explained by ADHD-PGS and each covariate individually was calculated in R using a variance partitioning tool (https://github.com/GabrielHoffman/misc_vp/blob/master/calcVarPart.R). Reported *P*-values were Bonferroni adjusted to account for the number of independent tests performed.

Acknowledgements
D.D. was supported by the Novo Nordisk Foundation (NNF20OC0065561), the Lundbeck Foundation (R344-2020-1060) the European Union's Horizon 2020 research and innovation programme under grant agreement No. 965381(TIMESPAN). The iPSYCH team was supported by grants from the Lundbeck Foundation (R102-A9118, R155-2014-1724, and R248-2017-2003), NIH/NIMH (1U01MH109514-01 and 1R01MH124851-01 to A.D.B.) and the Universities and University Hospitals of Aarhus and Copenhagen. The Danish National Biobank resource was supported by the Novo Nordisk Foundation. High-performance computer capacity for handling and statistical analysis of iPSYCH data on the GenomeDK HPC facility was provided by the Center for Genomics and Personalized Medicine and the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark (grant to A.D.B.).
Research reported in this publication was supported by the National Institute Of Mental Health of the National Institutes of Health under Award Number R01MH124851. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The work was also supported by the European College of Neuropsychopharmacology (ECNP) Network "ADHD Across the Lifespan". B.F. was also supported by funding from the European Community's Horizon 2020 Programme (H2020/2014 – 2020) under grant agreements No. 728018 (Eat2beNICE) and No. 847879 (PRIME). B.F. also received relevant funding from the Netherlands Organization for Scientific Research (NWO) for the Dutch National Science Agenda NeurolabNL project (grant 400-17-602).
S.E.M. was funded by NHMRC grants APP1172917, APP1158125 and APP1103623.
This work was supported by the Instituto de Salud Carlos III (PI19/01224, PI20/0004), by the Pla Estratègic de Recerca i Innovació en Salut (PERIS), Generalitat de Catalunya (METAL-Cat; SLT006/17/287); by the Agència de Gestió d'Ajuts Universitaris i de Recerca AGAUR, Generalitat de Catalunya (2017SGR1461), Ministry of Science, Innovation and Universities (IJC2018-035346-I to M.S.A); by the European Regional Development Fund (ERDF) and by "la Marató de TV3" (092330/31) and the ECNP Network 'ADHD across the Lifespan' (https://www.ecnp.eu/researchinnovation/ECNP-networks/List-ECNP-Networks/).
T.Z. is funded by NIH, Grant No. R37MH107649-07S1 and by Research Council of Norway, NRC, Grant No. 288083.
This study was also supported by the National Institutes of Health (NIH), Bethesda, MD under award numbers T32MH087004 (to K.T.), K08MH122911 (to G.V.), R01MH125246 (to P.R.) and U01MH116442 (to P.R.).

Conflicts of interest

B.M.N. currently serves as a member of the scientific advisory board at Deep Genomics and Neumora (previously RBNC) and consultant for Camp4 Therapeutics, Takeda Pharmaceutical, and Biogen.

Author contributions

Analysis: D.D., G.B.W., G.A., R.W., K.T., L.F., G.V., J.B., B.Z., W.Z., J.D., S.H.M. and T.T.N. **Sample and/or data provider and processing:** D.D, G.B.W, J.G., T.D.A., J.D., F.K.S, J.B.G., M.B.H., O.O.G., G.B., K.D., G.S.H., ADHD Working Group of the Psychiatric Genomics Consortium (PGC), iPSYCH-Broad Consortium, E.A., G.E.H., M.N., O.M., D.M.H., P.B.M., M.J.D., H.S., T.W., B.M.N., K.S. and A.D.B. **Writing:** D.D., G.B.W., K.T., G.A., G.V., J.B., H.S. and A.D.B. **Core revision:** D.D., G.B.W., G.A., R.W., K.T., G.V., J.B., S.D., J.M., M.R., F.K.S, D.I.B., M.S.A., N.R.M., D.H., S.E.M., T.Z., V.M., S.V.F., H.S., P.R., B.F., B.M.N., K.S. and A.D.B. **Study direction:** D.D. and

A.D.B. **Study supervision:** D.D., G.B.W., H.S., P.R., B.F., T.W., B.M.N., K.S. and A.D.B. All authors contributed with critical revision of the manuscript.

Data availability

Summary statistics from the ADHD GWAS meta-analysis is available for download at the PGC website (https://www.med.unc.edu/pgc/download-results/).

Correspondence and requests for materials should be addressed to:

Ditte Demontis: ditte@biomed.au.dk, Anders D. Børglum: anders@biomed.au.dk

Code availability

No previously unreported custom computer code or algorithm were used to generate results.

References

1 Faraone, S. V. *et al.* Attention-deficit/hyperactivity disorder. *Nature Reviews Disease Primers*, 15020, doi:10.1038/nrdp.2015.20 (2015).
2 Franke, B. *et al.* The genetics of attention deficit/hyperactivity disorder in adults, a review. *Molecular psychiatry* **17**, 960-987, doi:10.1038/mp.2011.138 (2012).
3 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, doi:10.1016/S2215-0366(15)00271-0 (2015).
4 Chang, Z., Lichtenstein, P., D'Onofrio, B. M., Sjolander, A. & Larsson, H. Serious transport accidents in adults with attention-deficit/hyperactivity disorder and the effect of medication: a population-based study. *JAMA Psychiatry* **71**, 319-325, doi:10.1001/jamapsychiatry.2013.4174 (2014).
5 Babinski, D. E., Neely, K. A., Ba, D. M. & Liu, G. Depression and Suicidal Behavior in Young Adult Men and Women With ADHD: Evidence From Claims Data. *J Clin Psychiatry* **81**, doi:10.4088/JCP.19m13130 (2020).
6 Capusan, A. J., Bendtsen, P., Marteinsdottir, I. & Larsson, H. Comorbidity of Adult ADHD and Its Subtypes With Substance Use Disorder in a Large Population-Based Epidemiological Study. *J Atten Disord* **23**, 1416-1426, doi:10.1177/1087054715626511 (2019).
7 Boomsma, D. I., van Beijsterveldt, T., Odintsova, V. V., Neale, M. C. & Dolan, C. V. Genetically Informed Regression Analysis: Application to Aggression Prediction by Inattention and Hyperactivity in Children and Adults. *Behav Genet* **51**, 250-263, doi:10.1007/s10519-020-10025-9 (2021).
8 Dalsgaard, S., Ostergaard, S. D., Leckman, J. F., Mortensen, P. B. & Pedersen, M. G. Mortality in children, adolescents, and adults with attention deficit hyperactivity disorder: a nationwide cohort study. *Lancet* **385**, 2190-2196, doi:10.1016/S0140-6736(14)61684-6 (2015).
9 Jangmo, A. *et al.* Attention-deficit/hyperactivity disorder and occupational outcomes: The role of educational attainment, comorbid developmental disorders, and intellectual disability. *PloS one* **16**, e0247724, doi:10.1371/journal.pone.0247724 (2021).

10 Zhao, X. *et al.* Family Burden of Raising a Child with ADHD. *Journal of abnormal child psychology* **47**, 1327-1338, doi:10.1007/s10802-019-00518-5 (2019).
11 Le, H. H. *et al.* Economic impact of childhood/adolescent ADHD in a European setting: the Netherlands as a reference case. *European child & adolescent psychiatry* **23**, 587-598, doi:10.1007/s00787-013-0477-8 (2014).
12 Libutzki, B. *et al.* Direct medical costs of ADHD and its comorbid conditions on basis of a claims data analysis. *Eur Psychiatry* **58**, 38-44, doi:10.1016/j.eurpsy.2019.01.019 (2019).
13 Faraone, S. V. & Larsson, H. Genetics of attention deficit hyperactivity disorder. *Molecular psychiatry*, doi:10.1038/s41380-018-0070-0 (2018).
14 Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature genetics* **51**, 63-75, doi:10.1038/s41588-018-0269-7 (2019).
15 Pedersen, C. B. *et al.* The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Molecular psychiatry*, doi:10.1038/mp.2017.196 (2017).
16 Mattheisen, M. *et al.* Identification of shared and differentiating genetic risk for autism spectrum disorder, attention deficit hyperactivity disorder and case subgroups. *medRxiv* (2021).
17 Satterstrom, F. K. *et al.* Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. *Nature neuroscience* **22**, 1961-1965, doi:10.1038/s41593-019-0527-8 (2019).
18 Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature genetics* **53**, 817-829, doi:10.1038/s41588-021-00857-4 (2021).
19 Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature neuroscience* **22**, 343-352, doi:10.1038/s41593-018-0326-7 (2019).
20 Pardinas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature genetics* **50**, 381-389, doi:10.1038/s41588-018-0059-2 (2018).
21 Bybjerg-Grauholm, J. *et al.* The iPSYCH2015 Case-Cohort sample: updated directions for unravelling genetic and environmental architectures of severe mental disorders. *medRxiv* (2020).
22 Mors, O., Perto, G. P. & Mortensen, P. B. The Danish Psychiatric Central Research Register. *Scand J Public Health* **39**, 54-57, doi:10.1177/1403494810395825 (2011).
23 Lynge, E., Sandegaard, J. L. & Rebolj, M. The Danish National Patient Register. *Scand J Public Health* **39**, 30-33, doi:10.1177/1403494811401482 (2011).
24 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).
25 Trzaskowski, M. *et al.* Quantifying between-cohort and between-sex genetic heterogeneity in major depressive disorder. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **180**, 439-447, doi:10.1002/ajmg.b.32713 (2019).
26 Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826, doi:10.1038/s41467-017-01261-5 (2017).

27 Koopmans, F. *et al.* SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron* **103**, 217-234 e214, doi:10.1016/j.neuron.2019.05.002 (2019).
28 Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Curr Protoc* **1**, e90, doi:10.1002/cpz1.90 (2021).
29 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research* **44**, W90-97, doi:10.1093/nar/gkw377 (2016).
30 de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219, doi:10.1371/journal.pcbi.1004219 (2015).
31 Zhang, W. *et al.* Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. *Nature communications* **10**, 3834, doi:10.1038/s41467-019-11874-7 (2019).
32 Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, doi:10.1126/science.aat8464 (2018).
33 Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature genetics* **50**, 621-629, doi:10.1038/s41588-018-0081-4 (2018).
34 Meuleman, W. *et al.* Index and biological spectrum of human DNase I hypersensitive sites. *Nature* **584**, 244-251, doi:10.1038/s41586-020-2559-3 (2020).
35 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, doi:10.1038/s41588-020-00721-x (2020).
36 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, doi:10.1038/s41467-019-11181-1 (2019).
37 La Manno, G. *et al.* Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell* **167**, 566-580 e519, doi:10.1016/j.cell.2016.09.027 (2016).
38 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, doi:10.1093/bioinformatics/btw613 (2017).
39 Frei, O. *et al.* Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nature communications* **10**, 2417, doi:10.1038/s41467-019-10310-0 (2019).
40 Franke, B. *et al.* Live fast, die young? A review on the developmental trajectories of ADHD across the lifespan. *Eur Neuropsychopharmacol* **28**, 1059-1088, doi:10.1016/j.euroneuro.2018.08.001 (2018).
41 Satterthwaite, T. D. *et al.* Neuroimaging of the Philadelphia neurodevelopmental cohort. *Neuroimage* **86**, 544-553, doi:10.1016/j.neuroimage.2013.07.064 (2014).
42 Calkins, M. E. *et al.* The Philadelphia Neurodevelopmental Cohort: constructing a deep phenotyping collaborative. *J Child Psychol Psychiatry* **56**, 1356-1369, doi:10.1111/jcpp.12416 (2015).
43 Gur, R. C. *et al.* Age group and sex differences in performance on a computerized neurocognitive battery in children age 8-21. *Neuropsychology* **26**, 251-265, doi:10.1037/a0026712 (2012).
44 Wilkinson, G. S. & Robertson, G. J. Wide range achievement test (WRAT4). *Lutz, FL: Psychological Assessment Resources* (2006).

45 Uffelmann, E. *et al.* Genome-wide association studies. *Nature Reviews Methods Primers* **1**, 59, doi:10.1038/s43586-021-00056-9 (2021).
46 Bataillon, T. *et al.* The effective size of the Icelandic population and the prospects for LD mapping: inference from unphased microsatellite markers. *Eur J Hum Genet* **14**, 1044-1053, doi:10.1038/sj.ejhg.5201669 (2006).
47 Gazal, S. *et al.* Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nature genetics* **49**, 1421-1427, doi:10.1038/ng.3954 (2017).
48 Hindley, G. *et al.* The shared genetic basis of mood instability and psychiatric disorders: A cross-trait genome-wide association analysis. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **189**, 207-218, doi:10.1002/ajmg.b.32907 (2022).
49 Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics* **51**, 431-444, doi:10.1038/s41588-019-0344-8 (2019).
50 Bansal, V. *et al.* Genome-wide association study results for educational attainment aid in identifying genetic heterogeneity of schizophrenia. *Nature communications* **9**, 3078, doi:10.1038/s41467-018-05510-z (2018).
51 Plana-Ripoll, O. *et al.* Exploring Comorbidity Within Mental Disorders Among a Danish National Population. *JAMA Psychiatry* **76**, 259-270, doi:10.1001/jamapsychiatry.2018.3658 (2019).
52 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, doi:10.1177/1087054717713638 (2020).
53 Jensen, C. M. & Steinhausen, H. C. Comorbid mental disorders in children and adolescents with attention-deficit/hyperactivity disorder in a large nationwide study. *Atten Defic Hyperact Disord* **7**, 27-38, doi:10.1007/s12402-014-0142-1 (2015).
54 Chen, Q. *et al.* Common psychiatric and metabolic comorbidity of adult attention-deficit/hyperactivity disorder: A population-based cross-sectional study. *PloS one* **13**, e0204516, doi:10.1371/journal.pone.0204516 (2018).
55 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 e1411, doi:10.1016/j.cell.2019.11.020 (2019).
56 Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics* **50**, 1112-1121, doi:10.1038/s41588-018-0147-3 (2018).
57 Akaike, H. A new look at the statistical model identification. *IEE Transactions on Automatic Controls* **19**, 716-723 (1974).
58 Yao, X. *et al.* Integrative analysis of genome-wide association studies identifies novel loci associated with neuropsychiatric disorders. *Transl Psychiatry* **11**, 69, doi:10.1038/s41398-020-01195-5 (2021).
59 Consortium, T. S. W. G. o. t. P. G., Ripke, S., Walters, J. T. & O'Donovan, M. C. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *medRxiv* (2020).
60 Johnson, E. C. *et al.* A large-scale genome-wide association study meta-analysis of cannabis use disorder. *Lancet Psychiatry*, doi:10.1016/S2215-0366(20)30339-4 (2020).

61 Araujo, D. J. *et al.* FoxP1 orchestration of ASD-relevant signaling pathways in the striatum. *Genes Dev* **29**, 2081-2096, doi:10.1101/gad.267989.115 (2015).
62 Fong, W. L., Kuo, H. Y., Wu, H. L., Chen, S. Y. & Liu, F. C. Differential and Overlapping Pattern of Foxp1 and Foxp2 Expression in the Striatum of Adult Mouse Brain. *Neuroscience* **388**, 214-223, doi:10.1016/j.neuroscience.2018.07.017 (2018).
63 Sollis, E. *et al.* Equivalent missense variant in the FOXP2 and FOXP1 transcription factors causes distinct neurodevelopmental disorders. *Hum Mutat* **38**, 1542-1554, doi:10.1002/humu.23303 (2017).
64 Mostafavi, H., Spence, J. P., Naqvi, S. & Pritchard, J. K. Limited overlap of eQTLs and GWAS hits due to systematic differences in discovery. *bioRxiv* (2022).
65 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 e523, doi:10.1016/j.cell.2019.12.036 (2020).
66 Singh, T., Neale, B. M., Daly, M. J. & Consortium, o. b. o. t. S. E. M.-A. S. Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. *medRxiv* (2020).
67 Sazonovs, A. *et al.* Sequencing of over 100,000 individuals identifies multiple genes and rare variants associated with Crohns disease susceptibility. *medRxiv* (2021).
68 Bahmani, Z. *et al.* Prefrontal Contributions to Attention and Working Memory. *Current topics in behavioral neurosciences* **41**, 129-153, doi:10.1007/7854_2018_74 (2019).
69 Sonne, J., Reddy, V. & Beato, M. R. in *StatPearls* (2021).
70 Morales, M. & Margolis, E. B. Ventral tegmental area: cellular heterogeneity, connectivity and behaviour. *Nature reviews. Neuroscience* **18**, 73-85, doi:10.1038/nrn.2016.165 (2017).
71 Chang, S., Yang, L., Wang, Y. & Faraone, S. V. Shared polygenic risk for ADHD, executive dysfunction and other psychiatric disorders. *Transl Psychiatry* **10**, 182, doi:10.1038/s41398-020-00872-9 (2020).
72 Nigg, J. T. *et al.* Working Memory and Vigilance as Multivariate Endophenotypes Related to Common Genetic Risk for Attention-Deficit/Hyperactivity Disorder. *Journal of the American Academy of Child and Adolescent Psychiatry* **57**, 175-182, doi:10.1016/j.jaac.2017.12.013 (2018).
73 Aguilar-Lacasana, S. *et al.* Polygenic risk for ADHD and ASD and their relation with cognitive measures in school children. *Psychological medicine*, 1-9, doi:10.1017/S0033291720003189 (2020).
74 Martin, J., Hamshere, M. L., Stergiakouli, E., O'Donovan, M. C. & Thapar, A. Neurocognitive abilities in the general population and composite genetic risk scores for attention-deficit hyperactivity disorder. *J Child Psychol Psychiatry* **56**, 648-656, doi:10.1111/jcpp.12336 (2015).
75 Lam, M. *et al.* RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics*, doi:10.1093/bioinformatics/btz633 (2019).
76 Loh, P. R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nature genetics* **48**, 1443-1448, doi:10.1038/ng.3679 (2016).
77 Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature genetics* **44**, 955-959, doi:10.1038/ng.2354 (2012).

78 Iglesias, A. I. *et al.* Haplotype reference consortium panel: Practical implications of imputations with large reference panels. *Hum Mutat* **38**, 1025-1032, doi:10.1002/humu.23247 (2017).
79 Price, A. L. *et al.* Long-range LD can confound genome scans in admixed populations. *American journal of human genetics* **83**, 132-135; author reply 135-139, doi:10.1016/j.ajhg.2008.06.005 (2008).
80 Wu, C., DeWan, A., Hoh, J. & Wang, Z. A comparison of association methods correcting for population stratification in case-control studies. *Ann Hum Genet* **75**, 418-427, doi:10.1111/j.1469-1809.2010.00639.x (2012).
81 Eggertsson, H. P. *et al.* GraphTyper2 enables population-scale genotyping of structural variation using pangenome graphs. *Nature communications* **10**, 5402, doi:10.1038/s41467-019-13341-9 (2019).
82 Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nature genetics* **47**, 435-444, doi:10.1038/ng.3247 (2015).
83 Price, A. L. *et al.* The impact of divergence time on the nature of population structure: an example from Iceland. *PLoS genetics* **5**, e1000505, doi:10.1371/journal.pgen.1000505 (2009).
84 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics* **47**, 291-295, doi:10.1038/ng.3211 (2015).
85 Martin, J. *et al.* A Genetic Investigation of Sex Bias in the Prevalence of Attention-Deficit/Hyperactivity Disorder. *Biological psychiatry* **83**, 1044-1053, doi:10.1016/j.biopsych.2017.11.026 (2018).
86 Martin, J. *et al.* Examining Sex-Differentiated Genetic Effects Across Neuropsychiatric and Behavioral Traits. *Biological psychiatry* **89**, 1127-1137, doi:10.1016/j.biopsych.2020.12.024 (2021).
87 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics* **88**, 76-82, doi:10.1016/j.ajhg.2010.11.011 (2011).
88 Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493-1501, doi:10.1093/bioinformatics/btw018 (2016).
89 Greenbaum, J. & Deng, H. W. A Statistical Approach to Fine Mapping for the Identification of Potential Causal Variants Related to Bone Mineral Density. *J Bone Miner Res* **32**, 1651-1658, doi:10.1002/jbmr.3154 (2017).
90 Chen, W. *et al.* Fine Mapping Causal Variants with an Approximate Bayesian Method Using Marginal Test Statistics. *Genetics* **200**, 719-736, doi:10.1534/genetics.115.176107 (2015).
91 Wang, J. *et al.* CAUSALdb: a database for disease/trait causal variants identified using summary statistics of genome-wide association studies. *Nucleic acids research* **48**, D807-D816, doi:10.1093/nar/gkz1026 (2020).
92 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425, doi:10.1016/j.cels.2015.12.004 (2015).
93 Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, doi:10.1126/science.aat8127 (2018).
94 Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287, doi:10.1038/ng.3656 (2016).

95 McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-1283, doi:10.1038/ng.3643 (2016).
96 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).
97 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics* **47**, 1228-1235, doi:10.1038/ng.3404 (2015).
98 Borglum, A. D. *et al.* Genome-wide study of association and interaction with maternal cytomegalovirus infection suggests new schizophrenia loci. *Molecular psychiatry* **19**, 325-333, doi:10.1038/mp.2013.2 (2014).
99 Hollegaard, M. V. *et al.* Robustness of genome-wide scanning using archived dried blood spot samples as a DNA source. *BMC Genet* **12**, 58, doi:10.1186/1471-2156-12-58 (2011).
100 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:btp324.10.1093/bioinformatics/btp324 (2009).
101 Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **11**, 11 10 11-11 10 33, doi:10.1002/0471250953.bi1110s43 (2013).
102 Kalia, S. S. *et al.* Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med* **19**, 249-255, doi:10.1038/gim.2016.190 (2017).
103 Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80-92, doi:10.4161/fly.19695 (2012).
104 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443, doi:10.1038/s41586-020-2308-7 (2020).
105 Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature genetics*, doi:10.1038/s41588-018-0307-5 (2019).
106 Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502-508, doi:10.1038/s41586-022-04434-5 (2022).
107 Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet* **27**, 3641-3649, doi:10.1093/hmg/ddy271 (2018).
108 Okbay, A. *et al.* Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature genetics* **54**, 437-449, doi:10.1038/s41588-022-01016-z (2022).
109 Mills, M. C. *et al.* Identification of 371 genetic variants for age at first sex and birth linked to externalising behaviour. *Nat Hum Behav* **5**, 1717-1730, doi:10.1038/s41562-021-01135-3 (2021).
110 Watanabe, K. *et al.* Genome-wide meta-analysis of insomnia in over 2.3 million individuals implicates involvement of specific biological pathways through gene-prioritization. *medRxiv* (2020).

111 Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature communications* **10**, 1776, doi:10.1038/s41467-019-09718-5 (2019).
112 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).

Table 1. Results for the 27 genome-wide significant index variants identified in the GWAS meta-analysis of 38,691 individuals with ADHD and 186,843 controls. The location (chromosome (chr)) base position (bp) in hg19), alleles (A1 and A2), odds ratio (OR) of the effect with respect to A1, standard error (SE) and association *P*-values from inverse-variance weighted fixed effects model of the index variants are given. "Novel" indicates if the locus is a new ADHD risk locus i.e., not identified in ADHD2019 (Ref. [14]). Nearby genes located within 50 kb from index variants are listed (for a list of mapped genes based on other criteria see Supplementary Table 8).

Genomic locus	chr	bp	rs ID	A1	A2	Nearby genes	Frequency cases	Frequency controls	OR	SE	P-value	Novel
1	1	44076469	rs549845	G	A	*PTPRF, KDM4A*	0.321	0.326	1.082	0.01	9.03E-15	no
2	2	145714354	rs1438898	A	C		0.762	0.769	1.065	0.01	4.88E-09	yes
3	3	20724204	rs2886697	G	A		0.634	0.643	1.061	0.01	7.90E-10	no
4	3	43691501	rs9877066	G	A	*SNRK, ANO10, ABHD5*	0.944	0.951	0.888	0.02	6.60E-09	yes
5	3	49916710	rs7613360	C	T	*TRAIP, CAMKV, MST1R, CTD-2330K9.3, MON1A*	0.598	0.614	0.948	0.01	3.18E-08	yes
6	3	51884072	rs2311059	G	A	*IQCF3, IQCF2, IQCF5, IQCF1*	0.314	0.308	0.944	0.01	3.16E-08	yes
7	3	71499401	rs17718444	C	T	*FOXP1*	0.695	0.660	1.063	0.01	2.87E-09	yes
8	3	87015142	rs114142727	C	G	*VGLL3*	0.988	0.988	1.285	0.04	5.13E-10	yes
9	4	112217523	rs17576773	C	T		0.888	0.880	1.101	0.02	1.63E-10	yes
10	4	147099654	rs6537401	G	A	*LSM6, RP11-6L6.2, SLC10A7*	0.660	0.655	0.945	0.01	1.40E-08	yes
11	5	87854395	rs4916723	A	C		0.553	0.573	0.918	0.01	9.48E-15	no
12	5	103964585	rs77960	G	A		0.665	0.682	0.929	0.01	2.46E-13	yes
13	5	144474779	rs10875612	C	T		0.483	0.470	0.947	0.01	5.62E-09	yes
14	6	70858701	rs2025286	A	C	*COL19A1*	0.553	0.550	0.947	0.01	4.00E-09	yes
15	7	67685754	rs73145587	A	T		0.910	0.901	1.107	0.02	3.67E-08	yes
16	7	114158954	rs9969232	G	A	*FOXP2*	0.344	0.382	0.934	0.01	9.98E-12	no
17	8	93277087	rs7844069	T	G		0.428	0.399	1.057	0.01	6.74E-09	yes
18	8	145802447	rs4925811	T	G	*C8orf82, ARHGAP39*	0.515	0.531	0.944	0.01	8.30E-09	yes
19	10	8784773	rs11255890	C	A		0.389	0.401	1.054	0.01	4.14E-08	yes
20	10	106453832	rs11596214	G	A	*SORCS3*	0.597	0.569	1.054	0.01	3.17E-08	no
21	11	28602173	rs2582895	C	A	*METTL15*	0.634	0.618	1.075	0.01	4.09E-14	yes
22	12	89771903	rs704061	T	C	*DUSP6, POC1B*	0.554	0.560	0.946	0.01	2.30E-09	no
23	14	98690923	rs76284431	T	A		0.847	0.842	0.922	0.01	1.19E-09	yes
24	16	61966703	rs1162202	C	T	*CDH8*	0.630	0.606	1.063	0.01	1.92E-09	yes
25	18	5871800	rs76857496	C	A	*TMEM200C*	0.870	0.859	1.083	0.01	1.24E-08	yes
26	18	50625779	rs7506904	G	A	*DCC*	0.343	0.372	0.946	0.01	1.24E-08	yes
27	20	21250843	rs6082363	T	C	*XRN2, NKX2-4*	0.296	0.291	1.073	0.01	4.38E-12	yes

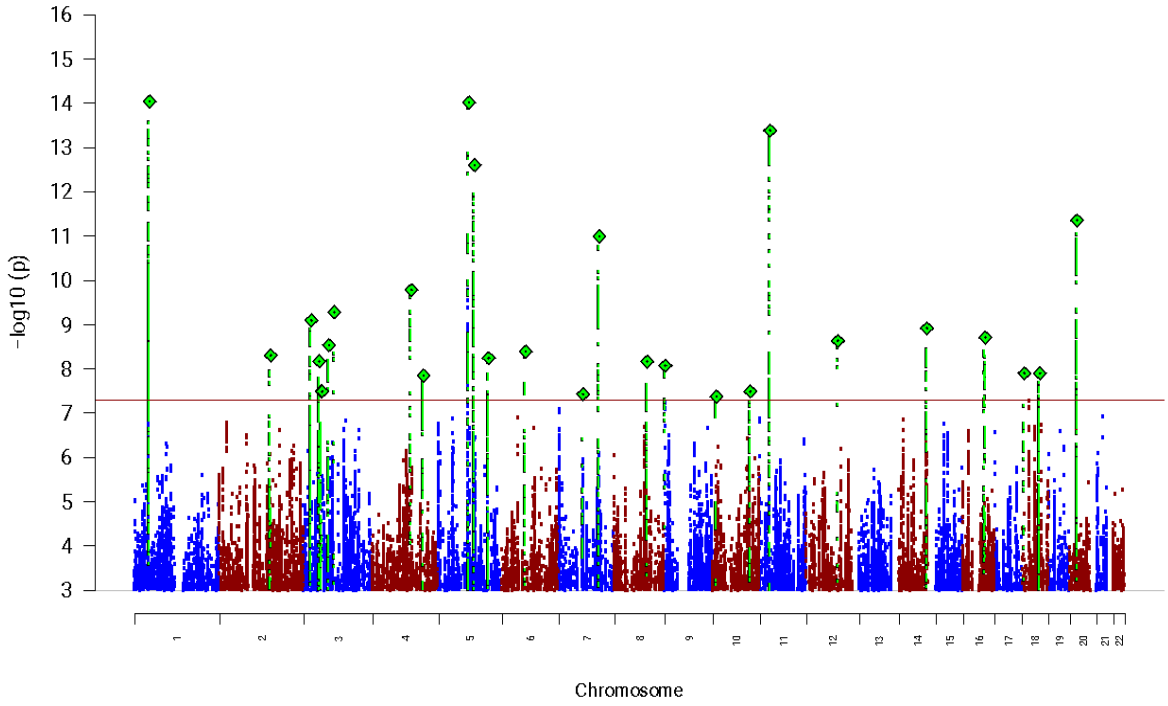

Figure 1. Results from GWAS meta-analysis of iPSYCH, deCODE and PGC cohorts in total including 38,899 cases and 186,843 controls. Two-sided *P*-values from meta-analysis using an inverse-variance weighted fixed effects model. Index variants in each of the genome-wide significant loci are marked as a green diamond (note that two loci on chromosome 3, index variants rs7613360 and rs2311059, are located in close proximity and therefore appear as one diamond in the plot). The red horizontal line represents the threshold for genome-wide significant association ($P = 5x10^{-8}$).

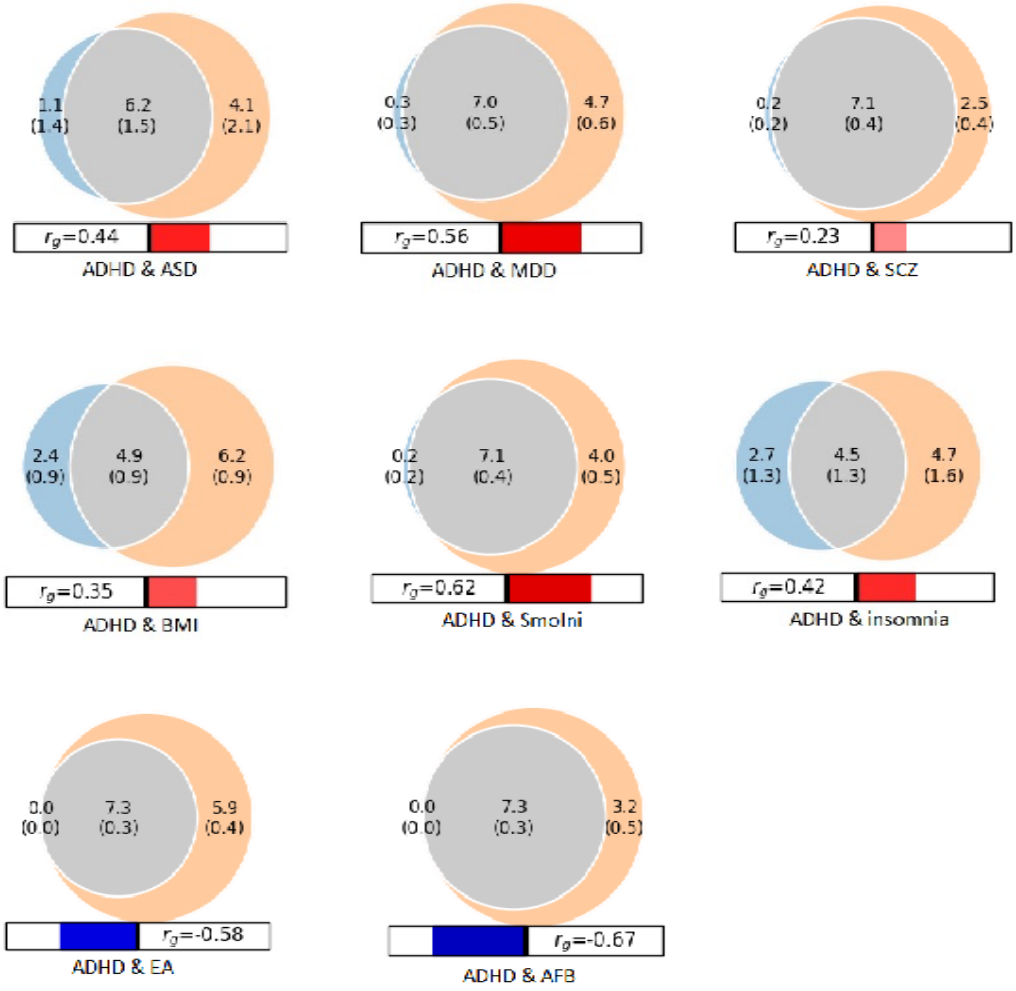

Figure 2. Venn diagrams showing MiXeR results of the estimated number of variants shared between ADHD and psychiatric disorders (with significant genetic correlations with ADHD) and phenotypes representing other domains with high genetic correlation with ADHD. Circles represent shared loci (gray), unique to ADHD (light blue) and unique to the phenotype of interest (orange). The number of shared variants (and standard errors) are shown in thousands. The size of the circles reflects the polygenicity of each phenotype, with larger circles corresponding to greater polygenicity. The estimated genetic correlation (r_g) between ADHD and each phenotype from LDSC is shown below the corresponding Venn diagram, with an accompanying scale (−1 to +1) with blue and red representing negative and positive genetic correlations, respectively. Bivariate results for ADHD, autism spectrum disorder (ASD), major depressive disorder (MDD), schizophrenia (SCZ), body mass index (BMI), smoking initiation (SmoIni), insomnia, educational attainment (EA) and age at first birth (AFB) are shown (see also Supplementary Table 17).

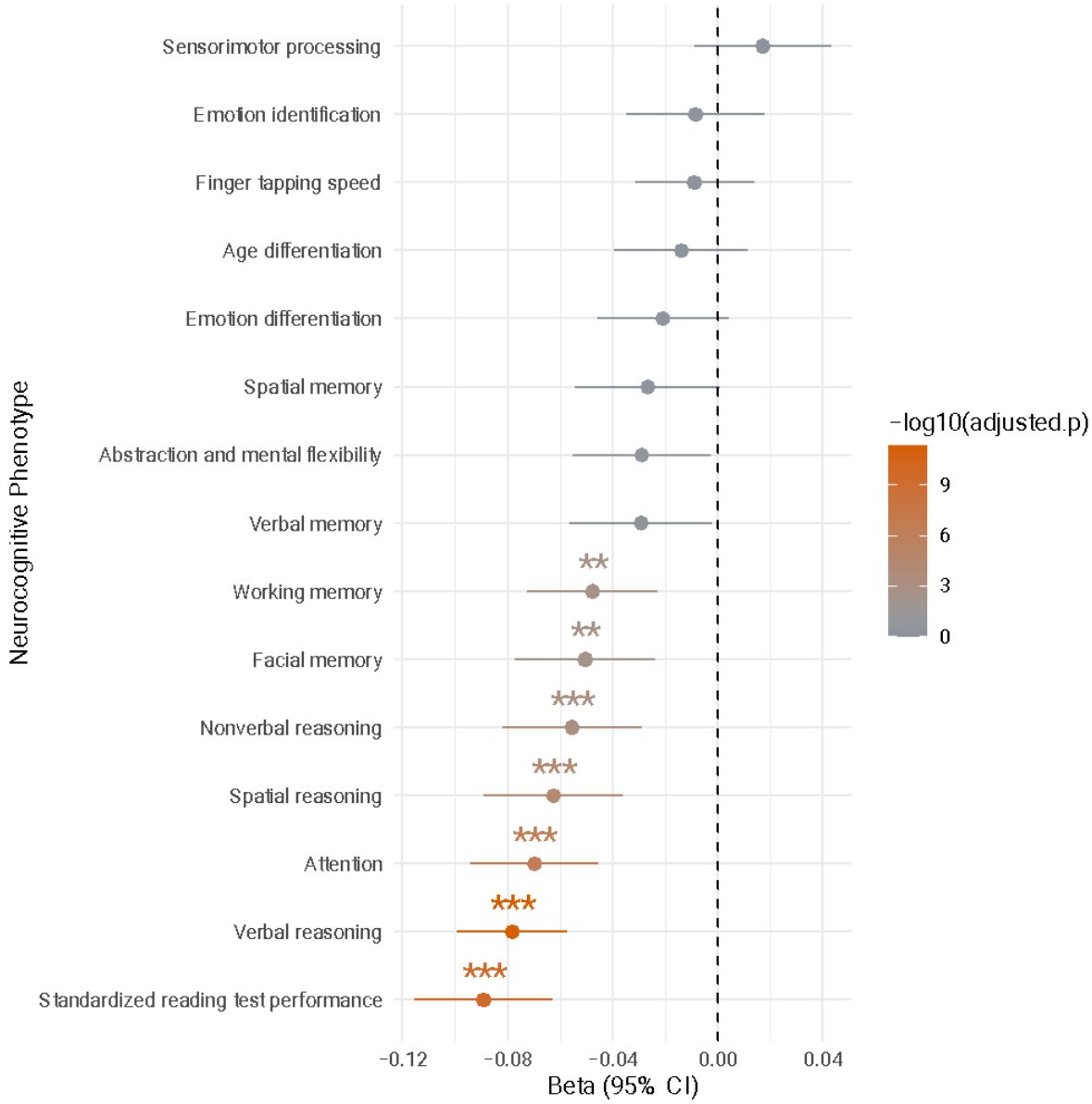

Figure 3. Association of ADHD-PGS with measures of cognitive abilities in the PNC cohort (N=4,973). Beta values (and standard errors indicated as horizontal bars) from linear regression testing for the association of ADHD-PGS with the 15 neurocognitive measures listed on the y-axis. The color bar at the right indicates the -log10(Bonferroni adjusted *P*-value) and significant results are indicated by stars (* $P < 0.05$; ** $P < 0.01$, *** $P < 0.001$).