

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

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

\*shared first authors, †Shared last authors

<sup>1</sup>Department of Biomedicine/Human Genetics, Aarhus University, Aarhus, Denmark

<sup>2</sup>The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark

<sup>3</sup>Center for Genomics and Personalized Medicine, Aarhus, Denmark

<sup>4</sup>deCODE genetics / Amgen, Reykjavik, Iceland

<sup>5</sup>Mental Health Centre Sct. Hans, Capital Region of Denmark, Institute of Biological Psychiatry, Copenhagen University Hospital, Copenhagen, Denmark

<sup>6</sup>Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

<sup>7</sup>Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

<sup>8</sup>Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark

<sup>9</sup>BiRC, Bioinformatics Research Centre, Aarhus University, Denmark

<sup>10</sup>NCCR - National Centre for Register-Based Research, Business and Social Sciences, Aarhus University, Aarhus, Denmark

<sup>11</sup>Centre for Integrated Register-based Research, CIRRAU, Aarhus University, Aarhus, Denmark

<sup>12</sup>Psychosis Research Unit, Aarhus University Hospital-Psychiatry, Denmark

<sup>13</sup>Mental Health Centre Copenhagen, Capital Region of Denmark, Copenhagen University Hospital, Copenhagen, Denmark

<sup>14</sup>Center for Disease Neurogenomics, <sup>15</sup>Icahn Institute for Data Science and Genomic Technology,

<sup>16</sup>Friedman Brain Institute, <sup>17</sup>Department of Psychiatry, <sup>18</sup>Department of Genetics and Genomic Science, Icahn School of Medicine at Mount Sinai, New York, USA

<sup>19</sup>Mental Illness Research Education and Clinical Center (MIRECC), JJ Peters VA Medical Center, Bronx, New York, USA

<sup>20</sup>Department of Human Genetics and Psychiatry, Radboud University Medical Center, Nijmegen, The Netherlands

<sup>21</sup>Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands.

<sup>22</sup>Department of Medicine, Harvard Medical School, Boston, MA, USA.

<sup>23</sup>Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland

<sup>24</sup>Faculty of Medicine, University of Iceland, Reykjavik, Iceland.

<sup>25</sup>Department of Child and Adolescent Psychiatry, National University Hospital, Reykjavik, Iceland.

<sup>26</sup>The Centre for Child Development and Behaviour, Capital Area Primary Health Care, Reykjavik, Iceland.

<sup>27</sup>Departments of Psychiatry and of Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, NY

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

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

**Correspond with:** Ditte Demontis, email: [ditte@biomed.au.dk](mailto:ditte@biomed.au.dk), Anders D. Børglum, email: [anders@biomed.au.dk](mailto:anders@biomed.au.dk)

## Abstract

Attention deficit hyperactivity disorder (ADHD) is a prevalent childhood psychiatric 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 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 90% of ADHD influencing variants are shared with other psychiatric disorders (autism, schizophrenia and depression) and 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 childhood psychiatric disorders affecting around 5% of children and persists into adulthood in two-thirds of cases<sup>1,2</sup>. It is characterized by extensive hyperactive, impulsive and/or inattentive behaviors that impair functioning. The disorder is associated with multiple adverse outcomes such as injuries<sup>3</sup>, accidents<sup>4</sup>, depression<sup>5</sup>, substances use disorders<sup>6</sup>, aggression<sup>7</sup>, premature death<sup>8</sup>, high rate of unemployment<sup>9</sup>, and has large societal costs<sup>10-12</sup>.

ADHD has a major genetic component with an estimated twin heritability of 0.74<sup>13</sup>. Despite the large influence of genetics on ADHD, the 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<sup>14</sup> 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<sup>15</sup> 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. 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<sup>16</sup>. 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 cases with autism spectrum disorder.

To better understand the biological mechanisms underlying ADHD, it is fundamental to conduct large genetic studies as demonstrated for other psychiatric disorders<sup>17-19</sup>. 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 modelling 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 cases and 186,843 controls (effective sample size = 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 (Ref.<sup>20</sup>). They were identified in the Danish Psychiatric Central Research Register<sup>21</sup> and the National Patient Register<sup>22</sup> 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 details previously<sup>14</sup>.

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<sup>23</sup>. The meta-analysis identified 32 lead variants ( $r^2 < 0.1$ ) located in 27 genome-wide significant loci (Figure 1; Table 1, locus plots in Supplementary Data 1), including 21 novel loci. No statistically significant heterogeneity was observed between cohorts (Supplementary Figure 1). The three strongest associated loci ( $P < 5 \times 10^{-14}$ ) were located on chromosome 1 (in and around *PTPRF*), chromosome 5 (downstream of *MF2C*) 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<sup>14</sup> were significant in the present study (Table 1), and the remaining six loci demonstrated P-values  $< 8 \times 10^{-4}$  (Supplementary Table 3). Overall, the direction of association of the top loci (726 loci with  $P < 1 \times 10^{-4}$ ) 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 ( $h^2_{\text{SNP}}$ ) was estimated to 0.14 (s.e. = 0.01), which is lower than the previously reported  $h^2_{\text{SNP}}$  of 0.22<sup>14</sup>. The  $h^2_{\text{SNP}}$  for iPSYCH ( $h^2_{\text{SNP}} = 0.23$ ; s.e. = 0.01) was in line with the previous finding, but lower  $h^2_{\text{SNP}}$  was observed for PGC ( $h^2_{\text{SNP}} = 0.12$ ; s.e. = 0.03) and deCODE ( $h^2_{\text{SNP}} = 0.081$ ; s.e. = 0.014). Between-cohort heterogeneity in  $h^2_{\text{SNP}}$  is not unusual and has been observed for other disorders like e.g. MDD<sup>24</sup>.

### **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<sup>25</sup> (datasets selected in FUMA are listed in the Supplementary Information). Seventy-six (76) plausible ADHD risk genes were identified (Supplementary Table 7). We found that this set of genes is significantly enriched in genes upregulated during early embryonic brain development (19<sup>th</sup> post-conceptual week;  $P=0.0002$ ; 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 were evaluated using SynGO data<sup>26</sup>. Nine genes mapped to SynGO annotations, and genes encoding integral components of the postsynaptic density membrane were borderline significantly enriched ( $5.43 \times 10^{-3}$ ;  $q$ -value 0.022; genes *PTPRF*, *SORCS3*, *DCC*; Supplementary Figure 5; Supplementary Table 8). Additionally, enrichment of the 76 genes in biological pathways was tested using data from 26 databases, but no significant pathways showed significant enrichment after Bonferroni correction.

### **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<sup>27</sup> and expression data from the PsychENCODE Consortium<sup>28</sup> 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 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 risk 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 ADHD risk. The *B4GALT2-205* isoform located in the genome-wide significant locus on chromosome 1 showed the strongest association ( $P = 7 \times 10^{-11}$ ), 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* ( $P = 1.4 \times 10^{-8}$ ), 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<sup>29</sup> identified 45 exome-wide significant genes ( $P < 2.72 \times 10^{-6}$  (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<sup>30</sup> 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<sup>31</sup> as well as brain-specific cell types<sup>32</sup>. 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<sup>33</sup> 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<sup>34</sup>; 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 ADHD plausible 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 was significantly enriched in cases compared to controls ( $P = 0.015$ ). For comparison, there was no enrichment in rare synonymous variants in the latter 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 of ADHD with other phenotypes was estimated using published GWASs (258 phenotypes) and GWASs of UK Biobank data (514 phenotypes), available in LDhub<sup>35</sup>. ADHD was significantly correlated ( $P < 2 \times 10^{-4}$ ) 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 psychiatric disorders, autism spectrum disorder (ASD) ( $r_g = 0.42$ , s.e. = 0.05), schizophrenia (SZ) ( $r_g = 0.17$ , s.e. = 0.03), major depressive disorder (MDD) ( $r_g = 0.31$ , s.e. = 0.07), and cannabis use disorder (CUD) ( $r_g = 0.61$ , 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<sup>36</sup> 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 same and opposite effects). Approximately 7,000 (s.e. = 285) common variants were found to influence ADHD, which is less than for SZ (8,5K; s.e. = 240), MDD (17.9K; s.e. = 1,700) and ASD (9.9K; s.e. = 1,050), and less than observed previously for bipolar disorder (BD) (8,6K, s.e. = 200)<sup>17</sup>. Remarkably, 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 (93%-94%; Figure 2; Supplementary Table 17). While the fraction of concordant variants (within the shared part) with MDD was high (96%), it was considerably lower for SZ (56%) and in-between for ASD (69%). When considering other phenotypes, insomnia demonstrated the smallest overlap with ADHD in terms of actual number of variants (4.4K, s.e. = 1,296; 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, 84% and 80% of shared variants have concordant

directions, respectively, while only 22% and 19% 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**

In order to evaluate the impact of ADHD risk variants on cognitive domains, we assessed the association of ADHD polygenic scores (PGS) with 15 cognitive measures in the Philadelphia Neurodevelopmental Cohort (PNC)<sup>37,38</sup>. This cohort is from the greater Philadelphia area, and include individuals in the age of 8-21 years 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<sup>39</sup> 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)<sup>40</sup> was used as a measure to get a proxy for overall IQ<sup>38</sup>.

ADHD-PGS was negatively associated with seven neurocognitive domains (Figure 3) with the strongest association for the WRAT-4 test (beta = -0.09,  $P = 4.35 \times 10^{-10}$ ). Besides that, ADHD-PGS was associated with measures of executive control (attention: beta = -0.07,  $P = 3.25 \times 10^{-7}$ ; working memory: beta = -0.05,  $P = 2.45 \times 10^{-3}$ ), complex cognition (verbal reasoning: beta = -0.08,  $P = 4.74 \times 10^{-12}$ ; non-verbal reasoning: beta = -0.06,  $P = 6.28 \times 10^{-4}$ ; spatial reasoning: beta = -0.06,  $P = 5.15 \times 10^{-5}$ ) and one measure of episodic memory (facial memory: beta = -0.05,  $P = 3.23 \times 10^{-3}$ ) (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)<sup>14</sup> and more than doubled the number of associated loci, indicating that we have passed the inflection point<sup>41</sup> 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. The observation that previously identified loci may not reach genome-wide significance in a subsequent larger GWAS, has also been seen for other psychiatric disorders, e.g. bipolar disorder, where eight out of 19 loci were significant in a subsequent larger study<sup>42</sup>.

We report a lower  $h^2_{\text{SNP}}$  for ADHD ( $h^2_{\text{SNP}}=0.14$ ) than estimated previously ( $h^2_{\text{SNP}}=0.22$ ). This is driven by a lower  $h^2_{\text{SNP}}$  in the PGC and deCODE cohorts compared to iPSYCH. Different ascertainment strategies and designs among PGC cohorts could decrease the  $h^2_{\text{SNP}}$ , while lower effective sample size<sup>43</sup> on Iceland, and thus fewer recent variants might bias  $h^2_{\text{SNP}}$  downwards in the deCODE cohort<sup>44</sup>.

Our analyses highlighted that ADHD is a very polygenic disorder by estimating that around 7,000 common variants can explain 90% of the  $h^2_{\text{SNP}}$ . Interestingly, the estimated number of ADHD risk variants was lower than observed for three genetically correlated psychiatric disorders (SZ, MDD, ASD). It could be hypothesized that a relatively larger phenotypic/genetic heterogeneity within these three disorders (as reported in e.g. <sup>45,46</sup>) could explain a part of the larger number of common risk variants influencing these phenotypes. Surprisingly, when assessing both concordant and discordant allelic directions, we found that over 90% of ADHD

risk variants also influence the other investigated psychiatric disorders. This extensive sharing with ASD, SZ and MDD is at the same level as observed for SZ and bipolar disorder<sup>36</sup>, which are among the most genetically correlated mental disorders<sup>47</sup>. Notably, for MDD almost all 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 respects to their common genetic architecture than previously thought based on their overall genetic correlations<sup>36,47</sup>. When considering 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 all variants that influence ADHD overlap with educational attainment<sup>48</sup> and that around 22% of ADHD risk variants are associated with increased educational attainment while the vast majority (78%) are associated with decreased educational attainment. This is consistent with the overall negative genetic correlation with educational attainment and the positive genetic correlation of ADHD with e.g. ASD that has a positive genetic correlation with educational attainment<sup>45,49</sup>. Thus, shared ADHD/ASD variants might be a part of the 22% of variants concordant between ADHD and educational attainment. Likewise, risk variants that differentiate ADHD from ASD might primarily be found among those that overlap with decreased educational attainment, which is consistent with results recently been reported in ADHD-ASD differentiating genetic analyses<sup>50</sup>.

Fine-mapping of the 27 loci identified credible variants that in general had low posterior probabilities. Only four variants had 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 likely risk genes, which were enriched among genes upregulated during early

embryonic development and involved in cognitive abilities identified by GWAS of cognitive phenotypes. Among these genes were *PPP1R16A* and *B4GALT2* (mapped by psychENCODE eQTLs; Supplementary Figure 11.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<sup>48</sup> with increased predicted expression in cortical brain regions associated with increased measures of cognitive ability (assessed in gene2pheno/PhenomeXcan\_SingleTissue<sup>51</sup>, *PPP1R16A* (cortex): educational years<sup>48</sup>  $P = 2.7 \times 10^{-6}$ ; UK Biobank college or university degree  $P = 1.05 \times 10^{-7}$ ; *B4GALT2* (frontal cortex)  $P = 1.98 \times 10^{-7}$ ). 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<sup>52</sup>) and also for other psychiatric disorders<sup>53</sup>. We would also like to highlight *FOXP1* and *FOXP2*, which were among the 76 risk genes. The genome-wide significant signal was located within the transcribed regions of both genes, and were additionally implicated in ADHD by genome-wide significant variants being eQTLs (*FOXP2*, Supplementary Figure 11.C) or located in chromatin interacting regions (*FOXP1*, Supplementary Figure 11.D) in brain tissue. *FOXP2* was identified in the ADHD2019 study<sup>14</sup>, and the association was even stronger in the current study. This gene was also recently identified as a risk gene for cannabis use disorder<sup>54</sup>. *FOXP1* is a new ADHD risk gene, which has previously been associated with schizophrenia<sup>19</sup>. Both *FOXP1* and *FOXP2* encode transcription factors that can heterodimerize to regulate transcription in brain tissues<sup>55,56</sup>. Highly penetrant rare variants have implicated the genes in language and speech disorders and intellectual disability<sup>57</sup>. The relative levels and the specific combination of *FOXP1* and *FOXP2* hetero/homo-dimers seem to regulate transcription of target genes differently<sup>58</sup>, and thus it

could be speculated that deviations in the relative levels of the proteins (altered by the regulatory common risk variants) might affect brain development.

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. Notably, in particular *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<sup>47</sup>, 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<sup>53,59,60,61</sup>.

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 e.g. attention and working memory<sup>62</sup>, and midbrain dopaminergic neurons are essential for controlling key functions, such as voluntary movement<sup>63</sup> and reward processing<sup>64</sup>. 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<sup>65-68</sup>. 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*<sup>15,20</sup> 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, SZ, 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<sup>21</sup> and the Danish National Patient register<sup>22</sup>. 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<sup>15</sup>. iPSYCH1 samples were genotyped using Illumina's PsychChip array and iPSYCH2 samples using the Illumina's global screening array v.2 (Illumina, CA, San Diego, USA). iPSYCH1 genotypes were called using GenCall and Birdseed and iPSYCH 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"<sup>69</sup>. 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 ( $|F_{het}| < 0.2$ ), 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<sup>70</sup> and Minimac<sup>71</sup>,

using the Haplotype Reference Consortium<sup>72</sup> 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 on 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.<sup>73</sup> excluded) with minor allele frequency  $> 0.05$  and no deviation from Hardy-Weinberg equilibrium (HWE  $P > 1 \times 10^{-4}$ ) 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<sup>74</sup>. 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 register information about birth country of the individuals and their parents which were required 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 ADHD cases with a clinical diagnosis of ADHD 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). The

control sample did not contain individuals with a diagnosis of SZ, 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<sup>75</sup>. 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<sup>76</sup>.

### *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<sup>14</sup>.

### **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<sup>54</sup>. Findings from analyses of the genetic structure of the Icelandic population by Price et al.<sup>77</sup> 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<sup>14</sup>.

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 using an fixed effects standard error weighted meta-analysis using METAL (version 2011-03-25)<sup>23</sup>. 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 with associations in the ADHD2019 data were evaluated by a sign-test at different p-value thresholds in the present GWAS (see thresholds in Supplementary Table 4).

### **Conditional analysis**

We identified potential 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 secondary independent lead variants, we performed association analyses of the secondary lead variants while conditioning of the index variant in the locus using COJO as implemented in GCTA<sup>78</sup>.

### **Identification of sets of credible variants**

In order 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. <sup>79</sup>), PAINTOR v.3.0 (Ref.<sup>80</sup>) and CAVIARBF v.0.2.1 (Ref.<sup>81</sup>), using CAUSALdb-finemapping-pip downloaded from <https://github.com/mulinlab/CAUSALdb-finemapping-pip><sup>82</sup>. 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 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 ( $h^2_{\text{SNP}}$ ) and pair-wise genetic correlation among the cohorts were calculated using LD score regression<sup>83</sup> 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  $h^2_{\text{SNP}}$  estimates from observed scale to the liability scale was done using a population prevalence of 5%. Test for significant differences in  $h^2_{\text{SNP}}$  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<sup>25</sup>. Protein coding genes were mapped if they were located with a distance of 10Kb up- or downstream 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). No additional variant filtering by functional annotation was applied in the eQTL and chromatin interaction mapping. The mapping linked credible variants to 76 ADHD risk genes. These genes were used in gene-set 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.

SynGO<sup>26</sup> (dataset version: 20210225) was used to test for enrichment among the 76 risk genes for genes involved in synaptic processes and locations. We analysed 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 gene were included in the enrichment analysis.

The set of 76 mapped ADHD risk genes were also used to test for enrichment in pathways/gene sets using Enrichr<sup>84,85</sup> and its implemented databases (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. Additionally, we tested for enrichment among the 76 genes of genes reported from the GWAS catalog (2019) and UK biobank GWASs (v1), and used [https://appytters.maayanlab.cloud/Enrichr\\_Manhattan\\_Plot/](https://appytters.maayanlab.cloud/Enrichr_Manhattan_Plot/) to visualize the results.

## **Transcriptomic imputation model construction and TWAS**

Transcriptomic imputation models were constructed as previously described<sup>27</sup> for dorso-lateral prefrontal cortex (DLPFC) transcript levels<sup>86</sup>. 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<sup>27</sup>. Genotypes were imputed using the University of Michigan server<sup>87</sup> with the Haplotype Reference Consortium (HRC) reference panel<sup>88</sup>. 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<sup>86</sup>. For the construction of the transcriptomic imputation models we used EpiXcan<sup>27</sup>, an elastic net based method, which weighs SNPs based on available epigenetic annotation information<sup>89</sup>. We performed the transcript-trait association analysis for ADHD as previously described<sup>27</sup>. Briefly, we applied the S-PrediXcan method<sup>27</sup> 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.<sup>25</sup>) 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 ( $P = 0.05/18381 = 2.72 \times 10^{-6}$ ).

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 age) available in FUMA (See Supplementary Information for data sets selected in FUMA).

Enrichment in  $h^2_{\text{SNP}}$  of ADHD associated variants located in or close to genes expressed in specific brain regions was estimated using LDSC-SEG<sup>30</sup>. Annotations indicating specific expression in 13 brain regions from the GTEx gene-expression database was downloaded from: [https://alkesgroup.broadinstitute.org/LDSCORE/LDSC\\_SEG\\_ldscores/](https://alkesgroup.broadinstitute.org/LDSCORE/LDSC_SEG_ldscores/).

### **Cell type-specific expression of ADHD risk genes**

We tested for enrichment in the ADHD  $h^2_{\text{SNP}}$  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<sup>31</sup> and an scATAC-seq study (single-cell assay for transposase accessible chromatin)<sup>32</sup> using LD-score partitioned heritability approach<sup>90</sup>. 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.

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 details in Watanabe et al.<sup>33</sup>. 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 corrects for all tested cell types across datasets. In a second step systematical 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 has also been whole exome-sequenced. DNA was extracted from dried blood spot samples of the study subjects and whole genome amplified in triplicates<sup>91,92</sup>, the coding regions of the genome were extracted using the Illumina 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) were also included in the recent study by Satterstrom et al.<sup>16</sup>, and the same quality control procedure was applied here. In short, the sequencing data were aligned to the reference genome using the BWA<sup>93</sup> (Hg19) and genotype

calling was done using the best practice recommended by the Genome Analysis Toolkit<sup>94</sup> (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<sup>95</sup> 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 being genetic outliers identified by principal component analysis (using a set of common variants and the software Eigensoft<sup>74</sup>), 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. Heterozygote 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  $1 \times 10^{-9}$ . 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  $\geq 0.2$ . After QC, the number of individuals were 8,895 ADHD cases and 9,001 controls.

The QCed variants were annotated using SnpEff<sup>96</sup> version 4.3t. The variants were also annotated with information about allele counts in the gnomAD<sup>97</sup> exomes r2.1.1 database using SnpSift<sup>96</sup> 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  $10 \times$  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 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 18 genes identified based on credible variants was found. We therefore looked specifically into these genes in order 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-Finish European individuals), done as described previously<sup>16</sup>. 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<sup>35</sup> (published GWASs: 255 phenotypes; UK Biobank GWASs: 514 phenotypes). Additionally, genetic correlations with three phenotypes not available in LDhub (cannabis use disorder<sup>54</sup>, smoking initiation<sup>98</sup> and education attainment<sup>48</sup>) were estimated locally using LD score regression<sup>83</sup>.

We applied MiXeR<sup>36</sup> to our ADHD GWAS summary statistics and a selection of complex traits (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. 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 produce estimates of “polygenicity” (i.e. the proportion of non-null SNPs) and “discoverability” (i.e. the variance of effect sizes of non-null SNPs). 2) We then used the variance estimates from the previous univariate step to run the 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) SNPs with a specific effect on the first or on the second trait; and (iii) SNPs with a nonzero effect on both traits. These four parameters can be mathematically combined in the final estimates of a polygenic overlap and genetic correlation (for details on the method see also<sup>17</sup>). The estimates reported explain 90% of  $h^2_{\text{SNP}}$ .

## 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  $\pm$  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<sup>99</sup> 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/PRS-Cs>) was utilized. The following default settings were used for PRS-CS: parameter a in the  $\gamma$ - $\gamma$  prior = 1, parameter b in the  $\gamma$ - $\gamma$  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<sup>100</sup> was then used to calculate individual-level ADHD PGS. Linear regression was used to test the association between ADHD PGS and neurocognitive 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 is

described in details elsewhere (Ref<sup>39</sup>). Additionally, association of ADHD-PGS with results from the Wide Range Achievement Test (WRAT-4)<sup>40</sup> was 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](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).

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 serve 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](mailto:ditte@biomed.au.dk), Anders D. Børghlum: [anders@biomed.au.dk](mailto: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 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).
- 17 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).
- 18 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).
- 19 Pardini, 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).
- 20 Bybjerg-Grauholm, J. *et al.* The iPSYCH2015 Case-Cohort sample: updated directions for unravelling genetic and environmental architectures of severe mental disorders. *medRxiv* (2020).

- 21 Mors, O., Pertou, G. P. & Mortensen, P. B. The Danish Psychiatric Central Research Register. *Scand J Public Health* **39**, 54-57, doi:10.1177/1403494810395825 (2011).
- 22 Lynge, E., Sandegaard, J. L. & Rebolj, M. The Danish National Patient Register. *Scand J Public Health* **39**, 30-33, doi:10.1177/1403494811401482 (2011).
- 23 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).
- 24 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).
- 25 Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature communications* **8**, 1826, doi:10.1038/s41467-017-01261-5 (2017).
- 26 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).
- 27 Zhang, W. *et al.* Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. *Nat Commun* **10**, 3834, doi:10.1038/s41467-019-11874-7 (2019).
- 28 Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, doi:10.1126/science.aat8464 (2018).
- 29 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).
- 30 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).
- 31 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).
- 32 Corces, M. R. *et al.* Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat Genet* **52**, 1158-1168, doi:10.1038/s41588-020-00721-x (2020).
- 33 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).
- 34 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).
- 35 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).
- 36 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).
- 37 Satterthwaite, T. D. *et al.* Neuroimaging of the Philadelphia neurodevelopmental cohort. *Neuroimage* **86**, 544-553, doi:10.1016/j.neuroimage.2013.07.064 (2014).
- 38 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).

- 39 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).
- 40 Wilkinson, G. S. & Robertson, G. J. Wide range achievement test (WRAT4). *Lutz, FL: Psychological Assessment Resources* (2006).
- 41 Levinson, D. F. *et al.* Genetic studies of major depressive disorder: why are there no genome-wide association study findings and what can we do about it? *Biological psychiatry* **76**, 510-512, doi:10.1016/j.biopsych.2014.07.029 (2014).
- 42 Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature genetics* **51**, 793-803, doi:10.1038/s41588-019-0397-8 (2019).
- 43 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).
- 44 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).
- 45 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).
- 46 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).
- 47 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).
- 48 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).
- 49 Karlsson Linner, R. *et al.* Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature genetics* **51**, 245-257, doi:10.1038/s41588-018-0309-3 (2019).
- 50 Mattheisen, M. *et al.* Identification of shared and differentiating genetic risk for autism spectrum disorder, attention deficit hyperactivity disorder and case subgroups. *medRxiv* (2021).
- 51 Pividori, M. *et al.* PhenomeXcan: Mapping the genome to the phenome through the transcriptome. *Sci Adv* **6**, doi:10.1126/sciadv.aba2083 (2020).
- 52 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).
- 53 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).
- 54 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).
- 55 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).



- 56 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).
- 57 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).
- 58 Sin, C., Li, H. & Crawford, D. A. Transcriptional regulation by FOXP1, FOXP2, and FOXP4 dimerization. *J Mol Neurosci* **55**, 437-448, doi:10.1007/s12031-014-0359-7 (2015).
- 59 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).
- 60 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).
- 61 Sazonovs, A. *et al.* Sequencing of over 100,000 individuals identifies multiple genes and rare variants associated with Crohns disease susceptibility. *medRxiv* (2021).
- 62 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).
- 63 Sonne, J., Reddy, V. & Beato, M. R. in *StatPearls* (2021).
- 64 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).
- 65 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).
- 66 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).
- 67 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).
- 68 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).
- 69 Lam, M. *et al.* RICOPIILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics*, doi:10.1093/bioinformatics/btz633 (2019).
- 70 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).
- 71 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).
- 72 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).
- 73 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).

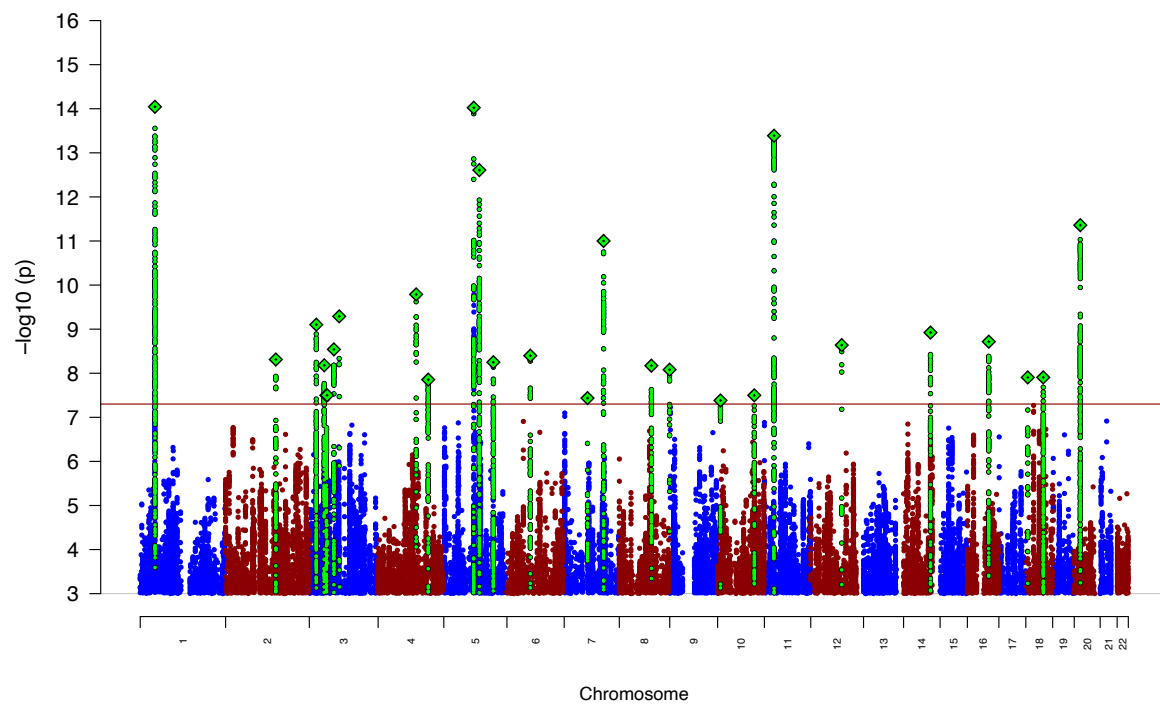
- 74 Wu, C., DeWan, A., Hoh, J. & Wang, Z. A comparison of association methods correcting for population stratification in case-control studies. *Ann Hum Genet* **75**, 418-427, doi:10.1111/j.1469-1809.2010.00639.x (2012).
- 75 Eggertsson, H. P. *et al.* GraphTyper2 enables population-scale genotyping of structural variation using pangenome graphs. *Nature communications* **10**, 5402, doi:10.1038/s41467-019-13341-9 (2019).
- 76 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).
- 77 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).
- 78 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).
- 79 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).
- 80 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).
- 81 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).
- 82 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).
- 83 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).
- 84 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).
- 85 Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Curr Protoc* **1**, e90, doi:10.1002/cpz1.90 (2021).
- 86 Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, doi:10.1126/science.aat8127 (2018).
- 87 Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287, doi:10.1038/ng.3656 (2016).
- 88 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).
- 89 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).
- 90 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).
- 91 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).
- 92 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).

- 93 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).
- 94 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 / editorial board, Andreas D. Baxevanis ... [et al.]* **11**, 11 10 11-11 10 33, doi:10.1002/0471250953.bi1110s43 (2013).
- 95 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).
- 96 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).
- 97 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).
- 98 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).
- 99 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).
- 100 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).

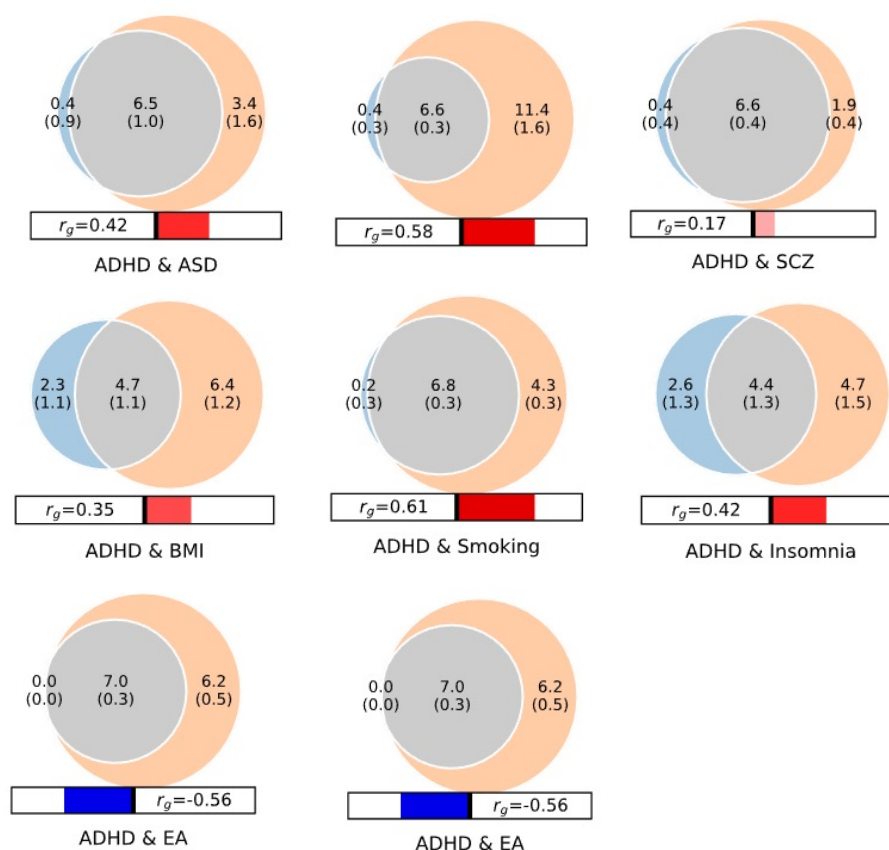


**Table 1.** Results for the 27 genome-wide significant index variants identified in the GWAS meta-analysis of 38,691 individuals with ADHD and 186,843 controls. The location (chromosome (chr)) base position (bp) in hg19), alleles (A1 and A2), odds ratio (OR) of the effect with respect to A1, standard 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<sup>14</sup>). 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).

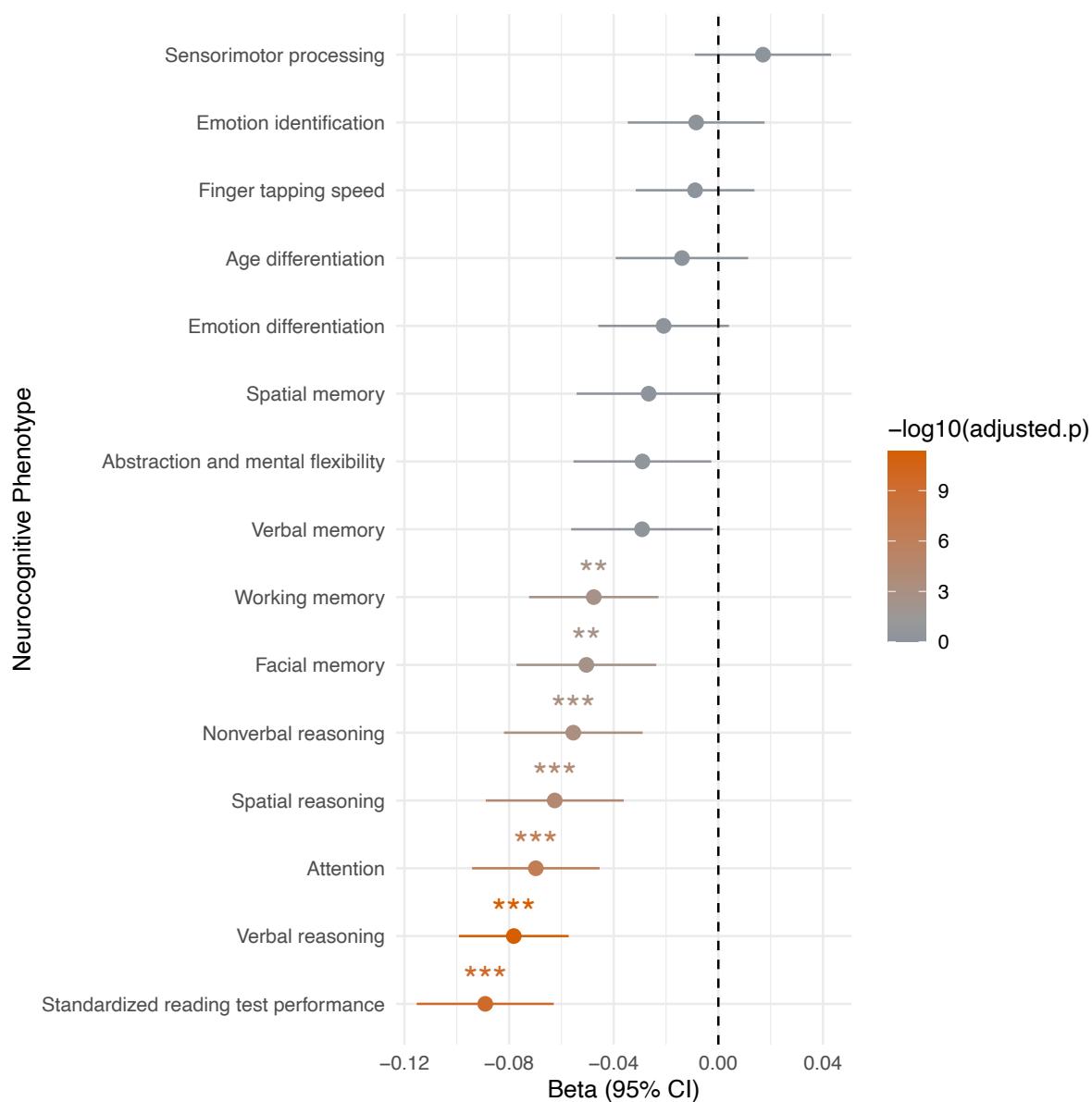
Genomic locus	chr	bp	rs ID	A1	A2	Nearby genes	Frequency cases	Frequency controls	OR	SE	P-value	Novel
1	1	44076469	rs549845	G	A	<i>PTPRF, KDM4A</i>	0.321	0.326	1.082	0.01	9.03E-15	no
2	2	145714354	rs1438898	A	C		0.762	0.769	1.065	0.01	4.88E-09	yes
3	3	20724204	rs2886697	G	A		0.634	0.643	1.061	0.01	7.90E-10	no
4	3	43691501	rs9877066	G	A	<i>SNRK, ANO10, ABHD5, TRAIIP, CAMKV, MST1R, CTD-</i>	0.944	0.951	0.888	0.02	6.60E-09	yes
5	3	49916710	rs7613360	C	T	<i>2330K9.3, MON1A, IQCF3, IQCF2, IQCF5,</i>	0.598	0.614	0.948	0.01	3.18E-08	yes
6	3	51884072	rs2311059	G	A	<i>IQCF1</i>	0.314	0.308	0.944	0.01	3.16E-08	yes
7	3	71499401	rs17718444	C	T	<i>FOXP1</i>	0.695	0.660	1.063	0.01	2.87E-09	yes
8	3	87015142	rs114142727	C	G	<i>VGLL3</i>	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	<i>LSM6, RP11-6L6.2, SLC10A7</i>	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	<i>COL19A1</i>	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	<i>FOXP2</i>	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	<i>C8orf82, ARHGAP39</i>	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	<i>SORCS3</i>	0.597	0.569	1.054	0.01	3.17E-08	no
21	11	28602173	rs2582895	C	A	<i>METTL15</i>	0.634	0.618	1.075	0.01	4.09E-14	yes
22	12	89771903	rs704061	T	C	<i>DUSP6, POC1B</i>	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	<i>CDH8</i>	0.630	0.606	1.063	0.01	1.92E-09	yes
25	18	5871800	rs76857496	C	A	<i>TMEM200C</i>	0.870	0.859	1.083	0.01	1.24E-08	yes
26	18	50625779	rs7506904	G	A	<i>DCC</i>	0.343	0.372	0.946	0.01	1.24E-08	yes
27	20	21250843	rs6082363	T	C	<i>XRN2, NKX2-4</i>	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 appears as one diamond in the plot. The red horizontal line represents the threshold for genome-wide significant association ( $P = 5 \times 10^{-8}$ ).



**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, insomnia, educational attainment (EA) and age at first birth 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 indicate the  $-\log_{10}(\text{Bonferroni adjusted } P\text{-value})$  and significant results are indicated by stars (\*  $P < 0.05$ ; \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ ).