

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

Received: 28 January 2022

A list of authors and their affiliations appears at the end of the paper

Accepted: 9 December 2022

Published online: 26 January 2023

 Check for updates

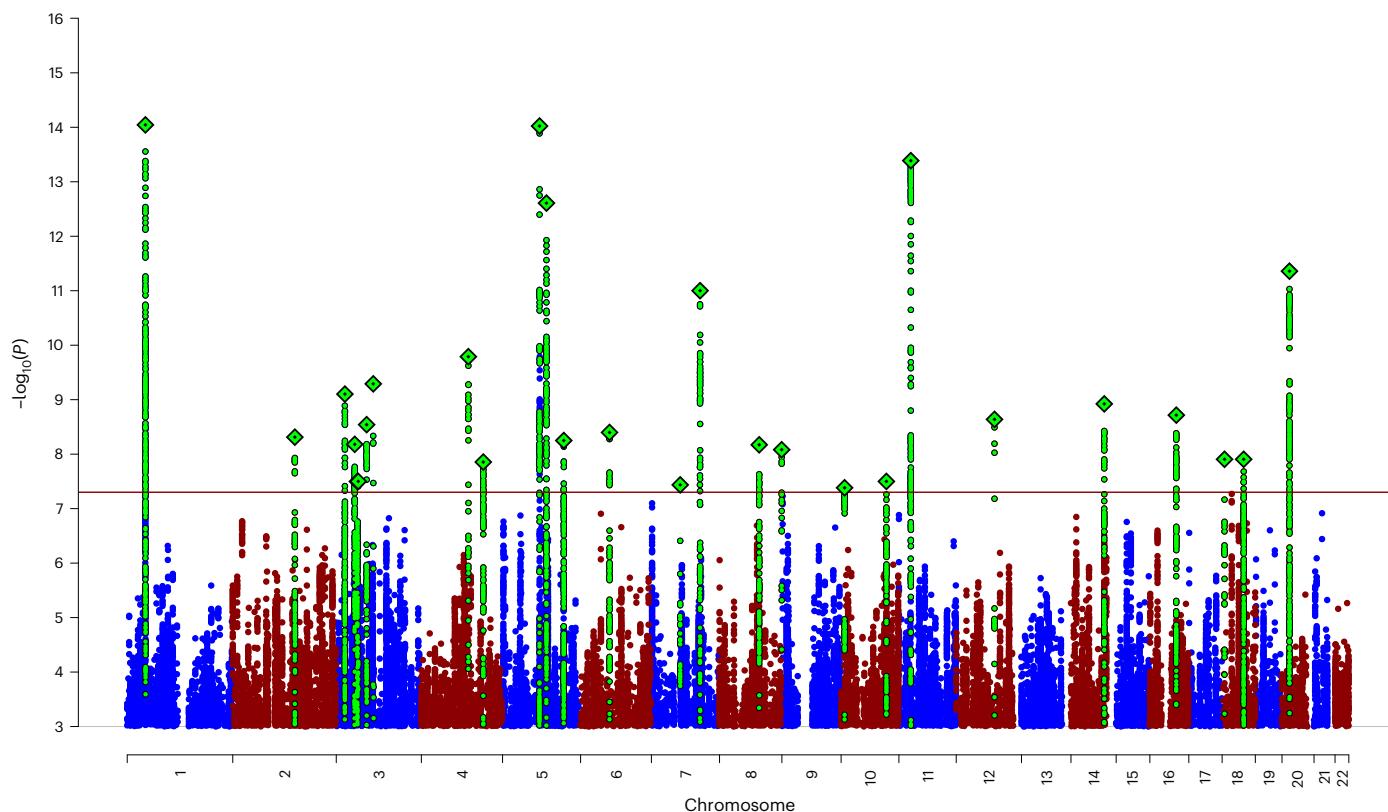
Attention-deficit hyperactivity disorder (ADHD) is a prevalent neurodevelopmental disorder with a major genetic component. Here, we present a genome-wide association study meta-analysis of ADHD comprising 38,691 individuals with ADHD and 186,843 controls. We identified 27 genome-wide significant loci, highlighting 76 potential risk genes enriched among genes expressed particularly in early brain development. Overall, ADHD genetic risk was associated with several brain-specific neuronal subtypes and midbrain dopaminergic neurons. In exome-sequencing data from 17,896 individuals, we identified an increased load of rare protein-truncating variants in ADHD for a set of risk genes enriched with probable causal common variants, potentially implicating *SORCS3* in ADHD by both common and rare variants. Bivariate Gaussian mixture modeling estimated that 84–98% of ADHD-influencing variants are shared with other psychiatric disorders. In addition, common-variant ADHD risk was associated with impaired complex cognition such as verbal reasoning and a range of executive functions, including attention.

ADHD is a prevalent neurodevelopmental disorder, 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 daily functioning. The disorder is associated with multiple adverse outcomes, such as injuries<sup>3</sup>, accidents<sup>4</sup>, depression<sup>5</sup>, substance use disorders<sup>6</sup>, aggression<sup>7</sup>, premature death<sup>8</sup> and high rates of unemployment<sup>9</sup> and has large social costs<sup>10–12</sup>.

ADHD has a major genetic component, with an estimated twin heritability of 0.74 (ref.<sup>13</sup>). Despite this, the complex polygenic architecture of ADHD makes it difficult to unravel its underlying biological causes. Previously, we discovered the first 12 genome-wide significant loci for ADHD<sup>14</sup> in a genome-wide association study (GWAS) of 20,183 cases and 35,191 controls (here referred to as ADHD2019) that combined the first wave of data from the Danish iPSYCH<sup>15</sup> cohort (iPSYCH1) with 11 ADHD cohorts collected by the Psychiatric Genomics Consortium (PGC). We established the role of common variants in ADHD, explaining around 22% of the variance in the phenotype. The results implicated

brain-expressed genes and demonstrated considerable genetic overlap of ADHD with a range of phenotypes, for instance, within psychiatric, cognitive and metabolic domains. In addition, a recent cross-disorder GWAS of ADHD and autism<sup>16</sup> identified shared and differentiating loci and showed that individuals with both ADHD and autism have distinctive patterns of genetic association with other traits compared with those with only a single diagnosis. This highlights that further mapping of the shared genetic risk component with other psychiatric disorders is important for understanding the complexity of the genetics underlying ADHD. Analyses of whole-exome-sequencing data have shown that rare variants also contribute to the risk of ADHD<sup>17</sup>, especially in mutationally constrained genes.

To better understand the biological mechanisms underlying ADHD, it is of fundamental importance to conduct large genetic studies, as has been demonstrated in other psychiatric disorders<sup>18–20</sup>. Here, we present results from an updated GWAS meta-analysis of ADHD, combining data from the newly extended Danish iPSYCH cohort,



**Fig. 1 | Results from GWAS meta-analysis of iPSYCH, deCODE and PGC cohorts including 38,899 cases and 186,843 controls in total.** The y axis represents  $-\log_{10}(P)$  (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 green diamonds (note that two loci on chromosome 3, index variants rs7613360 and rs2311059, are located in close proximity and therefore appear as one diamond in the plot). The red horizontal line represents the threshold for genome-wide significant association ( $P = 5 \times 10^{-8}$ ).

the Icelandic deCODE cohort and the PGC, almost doubling the number of cases compared with ADHD2019. We fine-map identified risk loci and use integration 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 bivariate mixture modeling and perform polygenic score (PGS) analyses to test for associations 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 published data from ten ADHD cohorts with European ancestry collected by the PGC (4,515 cases; 11,702 controls), resulting in a total sample size of 38,691 individuals with ADHD and 186,843 controls (effective sample size ( $n_{\text{eff\_half}}$ ) = 51,568; cohorts listed in Supplementary Table 1).

The GWAS meta-analysis identified 32 independent lead variants (that is, those with a squared correlation ( $r^2$ ) < 0.1 between variants) located in 27 genome-wide significant loci (Fig. 1, Table 1, locus plots in Supplementary Data 1 and forest plots in Supplementary Data 2), including 21 novel loci. No statistically significant heterogeneity was observed between cohorts (Supplementary Fig. 1). The three most strongly associated loci ( $P < 5 \times 10^{-14}$ ) were located on chromosome 1 (in and around *PTPRF*), chromosome 5 (downstream of *MEF2C*) and chromosome 11 (downstream of *METTL15*); the latter is a new 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 remained genome-wide significant in analyses conditioning on the index variant using COJO<sup>21</sup> (Supplementary Table 2).

Six of the previously identified 12 loci in the ADHD2019 study<sup>14</sup> were found to be significant in the present study (Table 1), and the remaining six loci had  $P$  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 but one (Supplementary Table 4).

### Genetic correlations among cohorts and SNP heritability

Genetic correlation analyses supported a high consistency in the phenotype across cohorts (genetic correlation ( $r_g$ ) ranging from 0.82 to 0.93, Supplementary Table 5) and between iPSYCH1 and iPSYCH2 ( $r_g = 0.97$ ; s.e. = 0.06). None of the genetic correlations was significantly different from 1. Linkage disequilibrium (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 (quantile-quantile plot shown in Supplementary Fig. 2). The SNP heritability ( $h^2_{\text{SNP}}$ ) was estimated to be 0.14 (s.e. = 0.01), lower than the previously reported  $h^2_{\text{SNP}}$  of 0.22 (ref. <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  $h^2_{\text{SNP}}$  values were observed for PGC ( $h^2_{\text{SNP}} = 0.12$ ; s.e. = 0.03) and deCODE ( $h^2_{\text{SNP}} = 0.08$ ; s.e. = 0.014). The difference in  $h^2_{\text{SNP}}$  was not caused by different sex distributions across cohorts, as there were no significant differences in  $h^2_{\text{SNP}}$  between males and females in the iPSYCH and deCODE cohorts (Supplementary Table 5). Between-cohort heterogeneity in  $h^2_{\text{SNP}}$  is not unusual and has

**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**

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

The location (chromosome (Chr.) position (bp) in hg19), alleles (A1 and A2), frequency (Freq.) of A1 in cases and controls, OR of the effect with respect to A1, standard error (s.e.) and association P values (two-sided) from inverse variance-weighted fixed-effects model of the index variants are given. ‘New’ indicates whether the locus is a new ADHD risk locus that was not identified in ADHD2019 (ref. <sup>14</sup>). Nearby genes located within 50 kb of index variants are listed (for a list of mapped genes based on other criteria, see Supplementary Table 8).

been observed in other diagnoses such as major depressive disorder (MDD)<sup>22</sup>.

#### Mapping risk variants to genes and enrichment analyses

To link identified risk variants to genes, we first identified sets of Bayesian credible variants for each risk locus, with each set most likely (probability >95%) to include a causal variant (Supplementary Table 6). Credible variants were subsequently linked to genes based on genomic position, information about expression quantitative trait loci (eQTLs) and chromatin interaction mapping in human brain tissue as implemented in FUMA<sup>23</sup> (datasets selected are listed in the Supplementary Note). We identified 76 plausible ADHD risk genes (Supplementary Table 7); four of the 76 were mapped by position alone. We found that this set of genes was significantly enriched among genes upregulated during early embryonic brain development (week 19 postconception;  $P_{\text{one\_sided}} = 0.0008$ ; Supplementary Fig. 3) and highly enriched for genes identified in GWAS of cognition-related phenotypes and reproduction (Supplementary Fig. 4). The role of the genes in synapses was evaluated

using SynGO data<sup>24</sup>; nine genes mapped to SynGO annotations, and genes encoding integral components of the postsynaptic density membrane were borderline significantly enriched ( $P = 5.43 \times 10^{-3}$ ;  $q = 0.022$ ; genes PTPRF, SORCS3 and DCC; Supplementary Fig. 5 and Supplementary Table 8). One SynGO-mapped gene was also among the genes upregulated during early embryonic brain development (ARHGAP39). In addition, enrichment of the 76 genes in biological pathways was tested using data from 26 databases implemented in Enrichr<sup>25,26</sup>. No pathway showed significant enrichment after Bonferroni correction (database significant findings can be found in Supplementary Table 9). Finally, MAGMA<sup>27</sup> gene-set analysis using gene-based P values derived from the full GWAS summary statistics (that is, with no preselection of specific genes) did not reveal any significant finding (top gene sets can be found in Supplementary Table 10).

#### Transcriptome-wide association analysis

To identify and prioritize ADHD risk genes, we also performed a transcriptome-wide association study (TWAS) of genetically regulated

gene expression using EpiXcan<sup>28</sup> and expression data from the PsychENCODE Consortium<sup>29</sup> for both genes and isoforms detected in 924 samples from the dorsolateral prefrontal cortex (DLPFC). The TWAS identified 15 genes (Supplementary Table 11) and 18 isoforms (Supplementary Table 12), which together led to the identification of 23 distinct genes (Supplementary Fig. 6) with significantly different predicted gene expression levels in ADHD cases compared with controls (after Bonferroni correction for all 34,646 genes and isoforms tested; Supplementary Fig. 6). Eight of the genes were among the 76 genes mapped by credible variants in FUMA. When using a less stringent correction (false discovery rate <5%), we identified 237 genes with different predicted expression among cases and controls, of which 19 were also among the 76 prioritized risk genes. The *B4GALT2-205* isoform located in the genome-wide significant locus on chromosome 1 showed the strongest association ( $P = 7 \times 10^{-11}$ ), with lower predicted expression in ADHD compared with controls (Supplementary Fig. 7a). 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 predicted underexpression in cases compared with controls. The expression model for this gene implicated one genome-wide significant variant (Supplementary Fig. 7b).

### Tissue-specific and cell-type-specific expression of ADHD risk genes

Gene-based association analysis using MAGMA<sup>27</sup> identified 45 exome-wide significant genes ( $P < 2.72 \times 10^{-6}$ ; 0.05/18,381 genes) associated with ADHD (Supplementary Table 13). Gene association results across the entire genome were tested for a relationship with tissue-specific gene expression. This showed that brain-expressed genes—in particular, genes expressed in the cortex—were associated with ADHD (Supplementary Fig. 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 14).

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 Fig. 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 Fig. 10 and Supplementary Table 15).

### Convergence of common and rare variant risk

To test for convergence of risk conferred by common variants and rare protein-truncating variants (rPTVs), we analyzed whole-exome-sequencing data from a subset of the iPSYCH cohort consisting of 8,895 ADHD cases and 9,001 controls. We tested three gene sets: (1) the 76 prioritized risk genes identified by positional and functional annotation; (2) the 45 significant genes in the MAGMA analysis; and (3) 18 genes with at least five credible variants located in the coding region (Supplementary Table 16). Although there was no indication of increased burden of rPTVs in the first gene set ( $P = 0.39$ , odds ratio (OR) = 1.30, s.e. = 0.16), the second gene set showed borderline nominally significant enrichment ( $P = 0.05$ , OR = 1.43, s.e. = 0.18), and the set of genes identified based on credible variants had a significantly increased burden of rPTVs in individuals with ADHD compared with controls ( $P = 0.015$ , OR = 2.19, s.e. = 0.32). For comparison, there was no enrichment in rare synonymous variants (rSYNs) in the third gene set ( $P = 0.59$ ). When evaluating the 18 genes from the ‘credible gene

set’ individually, *SORCS3* was nominally significantly ( $P = 0.008$ ; Supplementary Table 16) enriched in rPTVs in ADHD cases when compared with a combined group of iPSYCH controls and gnomAD individuals (nonpsychiatric non-Finnish Europeans;  $n = 58,121$ ); this suggests that *SORCS3* might be implicated in ADHD by both common and rare deleterious variants.

### Genetic overlap of ADHD with other phenotypes

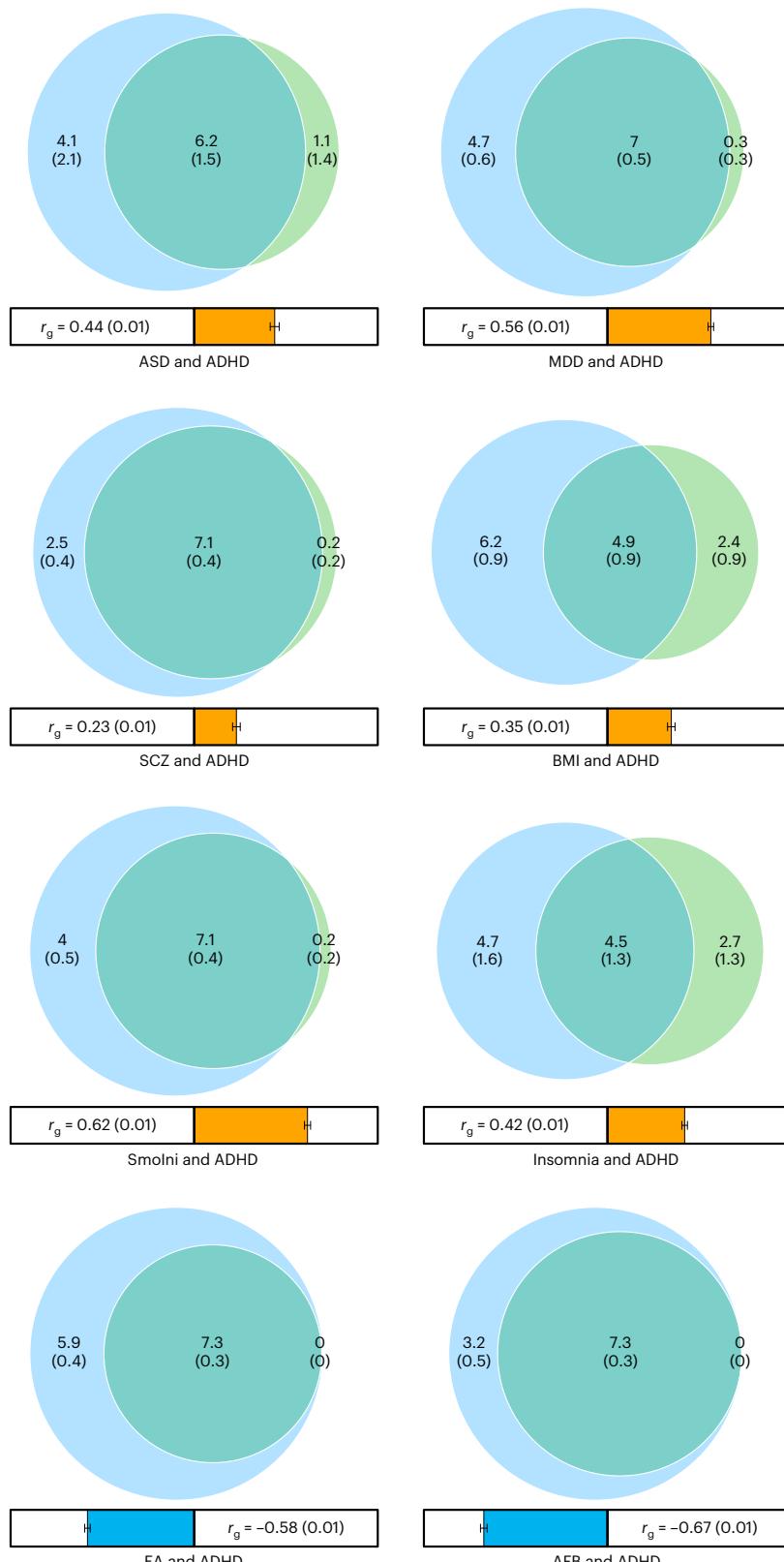
The genome-wide  $r_g$  of ADHD with other phenotypes was estimated using published GWAS (258 phenotypes) and GWAS of UK Biobank data (514 phenotypes) available in LDhub<sup>35</sup>. ADHD showed significant genetic correlation ( $P < 2 \times 10^{-4}$ ) with 56 phenotypes representing domains previously found to have significant genetic correlations with ADHD: cognition (for example, educational attainment,  $r_g = -0.55$ , s.e. = 0.021), weight/obesity (for example, body mass index (BMI),  $r_g = 0.27$ , s.e. = 0.03), smoking (for example, smoking initiation,  $r_g = 0.48$ ; s.e. = 0.07), sleep (for example, insomnia,  $r_g = 0.46$ , s.e. = 0.05), reproduction (for example, age at first birth,  $r_g = -0.65$ , s.e. = 0.03) and longevity (for example, mother’s age at death,  $r_g = -0.42$ , s.e. = 0.07). When considering other neurodevelopmental and psychiatric disorders, autism spectrum disorder (ASD) ( $r_g = 0.42$ , s.e. = 0.05), schizophrenia (SCZ) ( $r_g = 0.17$ , s.e. = 0.03), MDD ( $r_g = 0.31$ , s.e. = 0.07) and cannabis use disorder ( $r_g = 0.61$ , s.e. = 0.04) were significantly correlated with ADHD (Supplementary Table 17). 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 18).

Furthermore, we applied MiXeR<sup>36</sup>, which uses univariate and bivariate Gaussian mixture modeling, to quantify the actual number of variants that (1) explain 90% of the  $h^2_{SNP}$  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, that is, variants with the same and opposite directions of effects. Approximately 7.3K (s.d. = 324) common variants were found to influence ADHD, less than our estimates for SCZ (9.6K; s.d. = 199), MDD (11.7K; s.d. = 345) and ASD (10.3K; s.d. = 1,011) and less than previously reported for bipolar disorder (BD) (8.6K; s.d. = 200)<sup>18</sup>.

When considering the number of shared loci as a proportion of the total polygenicity of ADHD, the vast majority of variants influencing ADHD were also estimated to influence the other investigated psychiatric disorders (84–98%; Fig. 2, Supplementary Fig. 11 and Supplementary Table 19). Although the fraction of concordant variants (within the shared part) with ASD and MDD was at the high end (75–76%), it was lower for SCZ (59%). When considering other phenotypes, insomnia demonstrated the smallest overlap with ADHD in terms of actual number of variants (4.5K, s.d. = 1,281; 62% of ADHD variants shared), whereas almost all variants influencing ADHD also influenced educational attainment, age at first birth and smoking (Fig. 2 and Supplementary Table 19). For insomnia and smoking, 83% and 79% of shared variants had concordant directions, respectively, whereas only 21% and 20% of ADHD risk variants were concordant with variants associated with educational attainment and age at first birth, respectively (Supplementary Table 19).

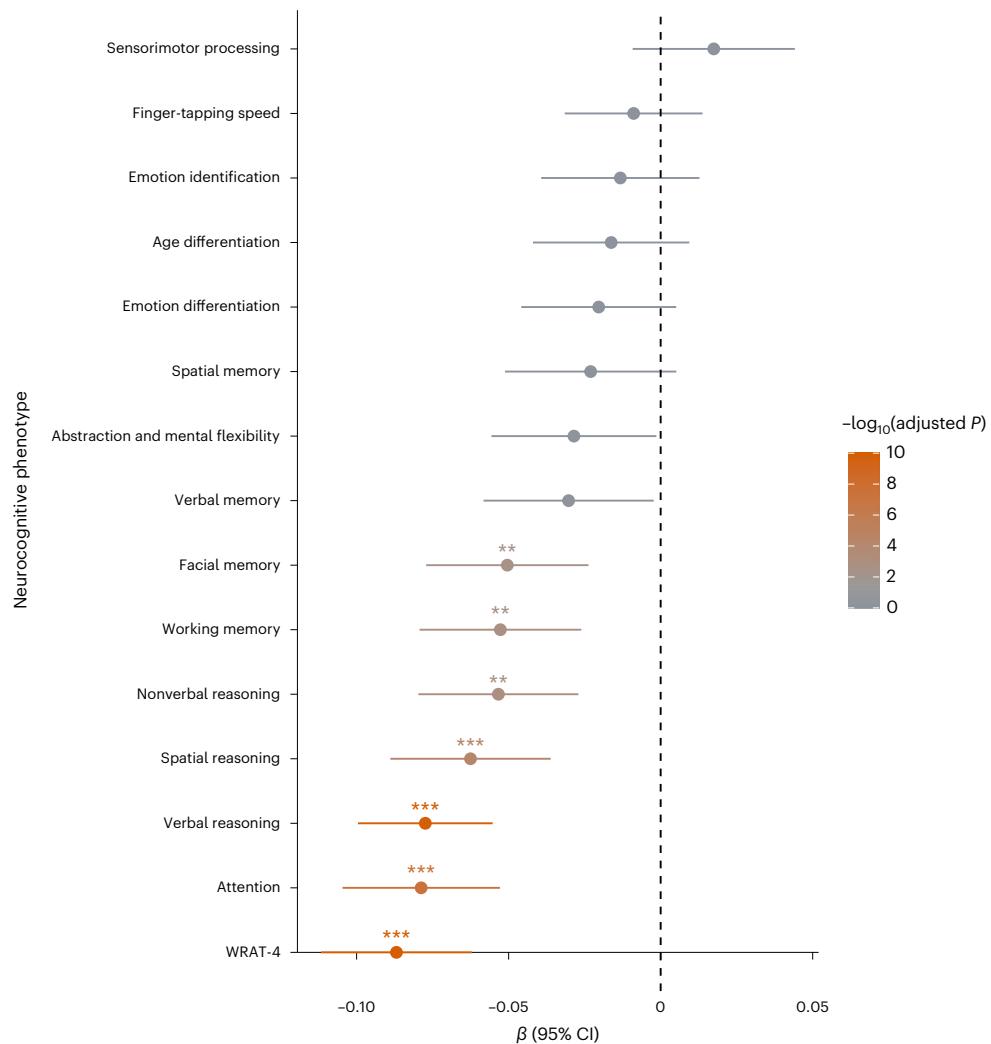
### Impact of ADHD-PGS on cognitive domains

Educational attainment is among the phenotypes with the strongest negative genetic correlations with ADHD, as demonstrated above, and cognitive impairments in ADHD have been well described<sup>37</sup>. To further explore how ADHD risk variants affect specific cognitive domains, we assessed the association of ADHD-PGS with 15 cognitive measures in the Philadelphia Neurodevelopmental Cohort (PNC)<sup>38,39</sup>. This cohort is from the greater Philadelphia area and includes individuals of 8–21 years of age who received medical care at the Children’s Hospital of Philadelphia Network. A subsample of the PNC cohort (v.1 release) of



**Fig. 2 | Venn diagrams showing MiXeR results for 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 variants (dark green), variants unique to ADHD (light green) and variants unique to the other phenotype of interest (light blue). The number of shared variants (and standard deviation) is shown in thousands. The size of the circles reflects**

the polygenicity of each phenotype, with larger circles corresponding to greater polygenicity. The estimated  $r_g$  between ADHD and each phenotype from LDSC is shown below the corresponding Venn diagram, with an accompanying scale ( $-1$  to  $+1$ ), with blue and orange representing negative and positive genetic correlations, respectively. Bivariate results for the overlap of ADHD with ASD, MDD, SCZ, BMI, smoking initiation (Smolni), insomnia, educational attainment (EA) and age at first birth (AFB) are shown (see also Supplementary Table 17).



**Fig. 3 | Association of ADHD-PGS with measures of cognitive abilities in the PNC cohort ( $n = 4,973$ ).**  $\beta$  values (represented as dots, with standard errors indicated as horizontal bars) from linear regression testing for the association of ADHD-PGS with the 15 neurocognitive measures listed on the yaxis (Wide

Range Achievement Test, WRAT-4). The color bar at the right indicates the  $-\log_{10}$  Bonferroni-adjusted two-sided  $P$  value, and  $P$ -value thresholds are indicated by asterisks (\* $P = 0.05$ ; \*\* $P = 0.01$ , \*\*\* $P = 0.001$ ).

4,973 individuals with European descent was used in this study. The Computerized Neurocognitive Battery (CNB)<sup>40</sup> was used to assess cognitive performance in the study participants. The battery consists of 14 tests in five domains: executive control, episodic memory, complex cognitive processing, social cognition and sensorimotor speed. In addition, the Wide Range Achievement Test (WRAT-4)<sup>41</sup> was used as a proxy measure for overall IQ<sup>39</sup>.

ADHD-PGS was negatively associated with seven neurocognitive measures (Fig. 3), with the strongest association for the WRAT-4 test ( $\beta = -0.09, P = 1.09 \times 10^{-10}$ ). ADHD-PGS was associated with measures of executive control (attention:  $\beta = -0.08, P = 3.94 \times 10^{-8}$ ; working memory:  $\beta = -0.05, P = 1.56 \times 10^{-3}$ ), complex cognition (verbal reasoning:  $\beta = -0.08, P = 1.31 \times 10^{-10}$ ; nonverbal reasoning:  $\beta = -0.05, P = 1.08 \times 10^{-3}$ ; 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 20). 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 compared with the ADHD2019 (ref. <sup>14</sup>) study and more than doubled the number of associated loci, indicating that we have passed the inflection point for ADHD with respect to the rate of risk loci discovery.

Six of the 12 previously identified loci were also found to be significant in this study. Although some previously identified loci demonstrated weaker associations here, their associations remained strong, and there was almost complete concordance in the direction of association between top-associated variants in this study and in ADHD2019. In GWAS of complex disorders, it is not uncommon for some loci to fluctuate around the significance threshold with increasing sample sizes until they eventually achieve stable significance; this can often be attributed to the ‘winner’s curse’ phenomenon, where effect-size estimates close to the discovery threshold tend to be overestimated in initial GWAS<sup>42</sup>.

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

We refined the genetic architecture of ADHD by estimating that around 7.3K (s.d. = 324) common variants can explain 90% of the  $h^2_{\text{SNP}}$ . This is a higher estimate than that reported based on the 2019 ADHD GWAS (5.6K, s.d. = 400)<sup>45</sup>, but the current estimate is based on a better fit to the causal mixture model (Akaike Information Criterion (AIC) = 80 versus AIC = 31 in Hindley et al.<sup>45</sup>). ADHD is often comorbid with other psychiatric disorders<sup>46</sup>, with 12–16% of individuals with ADHD also diagnosed with ASD<sup>16,47,48</sup> and around 40% with depression<sup>49</sup>, which is also reflected in the genetic correlations reported here and previously<sup>14</sup>. Strikingly, when both concordant and discordant allelic directions were assessed, more than 90% of ADHD risk variants also seemed to influence SCZ and MDD, and 84% influenced ASD. This extensive sharing with SCZ, MDD and ASD is at the same level as that observed for SCZ and BD<sup>36</sup>, which are among the most genetically correlated mental disorders<sup>50</sup>. Notably, for both MDD and ASD, around 75% of the variants shared with ADHD demonstrated concordant direction of association. The high level of sharing of variants influencing ADHD and other psychiatric disorders, when assessing both concordant and discordant allelic directions, suggests that the disorders are even more intermingled with respect to their common genetic architecture than was previously thought based on their overall genetic correlations<sup>36,50</sup>. For common variants, the developmental trajectory toward ADHD might therefore be influenced by variants involved in several psychiatric disorders but with disorder-specific allelic directions and effect sizes rather than actual ADHD-specific loci.

We also note that almost all variants that influence ADHD overlap with educational attainment<sup>51</sup>, and that the vast majority (79%) are associated with decreased educational attainment, consistent with the overall negative genetic correlation. For the models indicating a high number of shared variants (ADHD versus MDD, SCZ, BMI, educational attainment, age at first birth and smoking), we found support (evaluated using the AIC<sup>52</sup>) for the best fitting MiXeR models above the ‘minimal model’, indicating that the data support the existence of a polygenic overlap beyond the minimal level needed to explain the observed genetic correlations. For ADHD versus ASD, the model had limited support, and the results should therefore be interpreted with caution.

Fine-mapping of the 27 loci identified credible variants, but only four variants had posterior probabilities greater than 0.5 in all three fine-mapping methods, and none was linked to specific genes based on our functional annotation analyses. Linking the credible variants to genes by integration with functional genomics data identified 76 prioritized risk genes, which were enriched among genes upregulated during early embryonic development and involved in cognitive abilities as identified by GWAS of cognitive phenotypes. Among the 76 genes were *PPPIR16A* and *B4GALT2* (mapped by psychENCODE eQTLs; Supplementary Fig. 12a,b), which were also the top-ranking genes in our TWAS of DLPFC expression, both showing a predicted decreased expression in cases compared with controls. These genes have not previously been linked to psychiatric disorders, but both have been linked to educational attainment<sup>51</sup>. 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<sup>53</sup> and also for SCZ<sup>54</sup>. We also highlight *FOXP1* and *FOXP2*. The association signals were located within the transcribed regions of both genes and had credible variants that were eQTLs (*FOXP2*, Supplementary Fig. 12c) or located in chromatin-interacting regions (*FOXP1*, Supplementary Fig. 12d) in brain tissue. *FOXP2* was identified in the ADHD2019 study<sup>14</sup> and is also a risk gene for cannabis use disorder<sup>55</sup>, whereas *FOXP1* is a new ADHD locus. Both *FOXP1* and *FOXP2* encode transcription factors that can heterodimerize to regulate transcription in brain tissues<sup>56,57</sup> and have been implicated in speech disorders and intellectual disability<sup>58</sup> by highly penetrant rare variants.

Overall, fewer than half of the TWAS Bonferroni-significant genes overlapped with the 76 candidate risk genes (40% of ‘TWAS

transcript genes’; and 47% of ‘TWAS genes’; Supplementary Fig. 13). This was not unexpected and could have been due to noise in the data and/or the fact that the TWAS models were based on expression in adult brains whereas a large proportion of individuals in the GWAS were children. In addition, eQTLs used to derive TWAS models might not overlap with GWAS-identified variants as the two methods are systematically biased toward identification of different types of variants<sup>59</sup>.

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 eight had a higher load in cases compared with controls; thus, the signal was not driven by a few genes but by several genes with an increased burden of rPTVs. Of particular note, *SORCS3* seems to be implicated in ADHD by both common and rare variants. Common variants in *SORCS3* show strong pleiotropic effects across several major psychiatric disorders<sup>50</sup>, but to our knowledge, rare variant analyses have not previously implicated *SORCS3* in psychiatric disorders. Our results add to the emerging picture of overlap between genes and pathways affected by common and rare variants in psychiatric disorders<sup>54,60–62</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 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 including attention and working memory<sup>63</sup>, and midbrain dopaminergic neurons are essential for controlling key functions such as voluntary movement<sup>64</sup> and reward-processing<sup>65</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.

The PGS analyses in PNC identified strong association of polygenic ADHD risk with decreased overall IQ (approximated by WRAT test scores), in line with the high negative genetic correlation of ADHD with educational attainment and the observation that 79% of all ADHD risk variants are associated with decreased educational attainment. We also found that ADHD-PGS was associated with decreased attention, which is a key ADHD symptom, and with impairments in other cognitive traits such as working memory. Smaller studies have analyzed the impact of ADHD-PGS on executive functions, with mixed results<sup>66–69</sup>. This study robustly identifies specific cognitive domains affected by ADHD-PGS, and our results support a negative association of ADHD-PGS with neurocognitive performance.

In summary, we identified new ADHD risk loci, highlighted candidate causal genes, and implicated genes expressed in the frontal cortex and several brain-specific neuronal subtypes 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. In addition, 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 new aspects of the polygenic architecture of ADHD, its relationship with other phenotypes and its impact on cognitive domains.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-022-01285-8>.

## References

1. Faraone, S. V. et al. Attention-deficit/hyperactivity disorder. *Nat. Rev. Dis. Prim.* **1**, 15020 (2015).
2. Franke, B. et al. The genetics of attention deficit/hyperactivity disorder in adults, a review. *Mol. Psychiatry* **17**, 960–987 (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 (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 (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**, 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 (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 (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 (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 (2021).
10. Zhao, X. et al. Family burden of raising a child with ADHD. *J. Abnorm. Child Psychol.* **47**, 1327–1338 (2019).
11. Le, H. H. et al. Economic impact of childhood/adolescent ADHD in a European setting: the Netherlands as a reference case. *Eur. Child Adolesc. Psychiatry* **23**, 587–598 (2014).
12. Libutski, B. et al. Direct medical costs of ADHD and its comorbid conditions on basis of a claims data analysis. *Eur. Psychiatry* **58**, 38–44 (2019).
13. Faraone, S. V. & Larsson, H. Genetics of attention deficit hyperactivity disorder. *Mol. Psychiatry* **24**, 562–575 (2019).
14. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
15. Pedersen, C. B. et al. The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol. Psychiatry* **23**, 6–14 (2018).
16. Mattheisen, M. et al. Identification of shared and differentiating genetic architecture for autism spectrum disorder, attention-deficit hyperactivity disorder and case subgroups. *Nat. Genet.* **54**, 1470–1478 (2022).
17. Satterstrom, F. K. et al. Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. *Nat. Neurosci.* **22**, 1961–1965 (2019).
18. Mullins, N. et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).
19. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
20. Pardinas, A. F. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).
21. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
22. Trzaskowski, M. et al. Quantifying between-cohort and between-sex genetic heterogeneity in major depressive disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **180**, 439–447 (2019).
23. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
24. Koopmans, F. et al. SynGO: an evidence-based, expert-curated knowledge base for the synapse. *Neuron* **103**, 217–234.e4 (2019).
25. Xie, Z. et al. Gene set knowledge discovery with enrichr. *Curr. Protoc.* **1**, e90 (2021).
26. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
27. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
28. Zhang, W. et al. Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. *Nat. Commun.* **10**, 3834 (2019).
29. Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).
30. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
31. Meuleman, W. et al. Index and biological spectrum of human DNase I hypersensitive sites. *Nature* **584**, 244–251 (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 (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 (2019).
34. La Manno, G. et al. Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell* **167**, 566–580.e19 (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 (2017).
36. Frei, O. et al. Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nat. Commun.* **10**, 2417 (2019).
37. Franke, B. et al. Live fast, die young? A review on the developmental trajectories of ADHD across the lifespan. *Eur. Neuropsychopharmacol.* **28**, 1059–1088 (2018).
38. Satterthwaite, T. D. et al. Neuroimaging of the Philadelphia Neurodevelopmental Cohort. *Neuroimage* **86**, 544–553 (2014).
39. Calkins, M. E. et al. The Philadelphia Neurodevelopmental Cohort: constructing a deep phenotyping collaborative. *J. Child Psychol. Psychiatry* **56**, 1356–1369 (2015).
40. 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 (2012).
41. Wilkinson, G. S. & Robertson, G. J. *Wide Range Achievement Test (WRAT4)* (Psychological Assessment Resources, 2006).
42. Uffelmann, E. et al. Genome-wide association studies. *Nat. Rev. Methods Prim.* **1**, 59 (2021).

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 (2006).
44. Gazal, S. et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49**, 1421–1427 (2017).
45. Hindley, G. et al. The shared genetic basis of mood instability and psychiatric disorders: a cross-trait genome-wide association analysis. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **189**, 207–218 (2022).
46. Plana-Ripoll, O. et al. Exploring comorbidity within mental disorders among a Danish national population. *JAMA Psychiatry* **76**, 259–270 (2019).
47. Zablotsky, B., Bramlett, M. D. & Blumberg, S. J. The co-occurrence of autism spectrum disorder in children with ADHD. *J. Atten. Disord.* **24**, 94–103 (2020).
48. Jensen, C. M. & Steinhausen, H. C. Comorbid mental disorders in children and adolescents with attention-deficit/hyperactivity disorder in a large nationwide study. *Atten. Defic. Hyperact. Disord.* **7**, 27–38 (2015).
49. Chen, Q. et al. Common psychiatric and metabolic comorbidity of adult attention-deficit/hyperactivity disorder: a population-based cross-sectional study. *PLoS ONE* **13**, e0204516 (2018).
50. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell* **179**, 1469–1482.e11 (2019).
51. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
52. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Controls* **19**, 716–723 (1974).
53. Yao, X. et al. Integrative analysis of genome-wide association studies identifies novel loci associated with neuropsychiatric disorders. *Transl. Psychiatry* **11**, 69 (2021).
54. Trubetskov, V. et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
55. Johnson, E. C. et al. A large-scale genome-wide association study meta-analysis of cannabis use disorder. *Lancet Psychiatry* **7**, 1032–1045 (2020).
56. Araujo, D. J. et al. FoxP1 orchestration of ASD-relevant signaling pathways in the striatum. *Genes Dev.* **29**, 2081–2096 (2015).
57. 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 (2018).
58. Sollis, E. et al. Equivalent missense variant in the FOXP2 and FOXP1 transcription factors causes distinct neurodevelopmental disorders. *Hum. Mutat.* **38**, 1542–1554 (2017).
59. Mostafavi, H., Spence, J. P., Naqvi, S. & Pritchard, J. K. Limited overlap of eQTLs and GWAS hits due to systematic differences in discovery. Preprint at bioRxiv <https://doi.org/10.1101/2022.05.07.491045> (2022).
60. Satterstrom, F. K. et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell* **180**, 568–584.e23 (2020).
61. Singh, T. et al. Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature* **604**, 509–516 (2022).
62. Sazonovs, A. et al. Large-scale sequencing identifies multiple genes and rare variants associated with Crohn's disease susceptibility. *Nat. Genet.* **54**, 1275–1283 (2021).
63. Bahmani, Z. et al. Prefrontal contributions to attention and working memory. *Curr. Top. Behav. Neurosci.* **41**, 129–153 (2019).
64. Sonne, J., Reddy, V. & Beato, M. R. Substantia nigra. in *StatPearls* (StatPearls Publishing, 2021).
65. Morales, M. & Margolis, E. B. Ventral tegmental area: cellular heterogeneity, connectivity and behaviour. *Nat. Rev. Neurosci.* **18**, 73–85 (2017).
66. Chang, S., Yang, L., Wang, Y. & Faraone, S. V. Shared polygenic risk for ADHD, executive dysfunction and other psychiatric disorders. *Transl. Psychiatry* **10**, 182 (2020).
67. Nigg, J. T. et al. Working memory and vigilance as multivariate endophenotypes related to common genetic risk for attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **57**, 175–182 (2018).
68. Aguilar-Lacasana, S. et al. Polygenic risk for ADHD and ASD and their relation with cognitive measures in school children. *Psychol. Med.* **52**, 1356–1364 (2022).
69. 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 (2015).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2023, corrected publication 2023

Ditte Demontis  , G. Bragi Walters  , Georgios Athanasiadis  , Raymond Walters  , Karen Therrien  , Trine Tollerup Nielsen  , Leila Farajzadeh  , Georgios Voloudakis  , Jaroslav Bendl  , Biao Zeng  , Wen Zhang  , Jakob Grove  , Thomas D. Als  , Jinjie Duan  , F. Kyle Satterstrom  , Jonas Bybjerg-Grauholt  , Marie Bækved-Hansen  , Olafur O. Gudmundsson  , Sigurdur H. Magnusson  , Gisli Baldursson  , Katrin Davidsdottir  , Gyda S. Haraldsdottir  , Esben Agerbo  , Gabriel E. Hoffman  , Søren Dalsgaard  , Joanna Martin  , Marta Ribasés  , Dorret I. Boomsma  , Maria Soler Artigas  , Nina Roth Mota  , Daniel Howrigan  , Sarah E. Medland  , Tetyana Zayats  , Veera M. Rajagopal  , ADHD Working Group of the Psychiatric Genomics Consortium\*, iPSYCH-Broad Consortium\*, Merete Nordentoft  , Ole Mors  , David M. Hougaard  , Preben Bo Mortensen  , Mark J. Daly  , Stephen V. Faraone  , Hreinn Stefansson  , Panos Roussos  , Barbara Franke  , Thomas Werge  , Benjamin M. Neale  , Kari Stefansson  , Anders D. Børglum  

<sup>1</sup>Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark. <sup>2</sup>The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus, Denmark. <sup>3</sup>Center for Genomics and Personalized Medicine, Aarhus, Denmark. <sup>4</sup>deCODE Genetics/Amgen, Reykjavik, Iceland. <sup>5</sup>Faculty of Medicine, University of Iceland, Reykjavik, Iceland. <sup>6</sup>Mental Health Centre Sct. Hans, Capital Region of Denmark, Institute of Biological Psychiatry, Copenhagen University Hospital, Copenhagen, Denmark. <sup>7</sup>Department of Evolutionary Biology, Ecology and Environmental Sciences, University of Barcelona, Barcelona, Spain. <sup>8</sup>Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. <sup>9</sup>Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>10</sup>Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>11</sup>Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>12</sup>Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>13</sup>Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>14</sup>Department of Genetics and Genomic Science, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>15</sup>Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>16</sup>Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. <sup>17</sup>Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark. <sup>18</sup>Department of Child and Adolescent Psychiatry, National University Hospital, Reykjavik, Iceland. <sup>19</sup>The Centre for Child Development and Behaviour, Capital Area Primary Health Care, Reykjavik, Iceland. <sup>20</sup>National Centre for Register-Based Research, Business and Social Sciences, Aarhus University, Aarhus, Denmark. <sup>21</sup>Centre for Integrated Register-based Research, CIRRAU, Aarhus University, Aarhus, Denmark. <sup>22</sup>Mental Health Services of the Capital Region, Child and Adolescent Psychiatric Center, Glostrup, Denmark. <sup>23</sup>Institute of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. <sup>24</sup>Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK. <sup>25</sup>Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. <sup>26</sup>Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain. <sup>27</sup>Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain. <sup>28</sup>Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain. <sup>29</sup>Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. <sup>30</sup>Amsterdam Public Health Research Institute, Amsterdam, the Netherlands. <sup>31</sup>Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands. <sup>32</sup>Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands. <sup>33</sup>Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. <sup>34</sup>PROMENTA, Department of Psychology, University of Oslo, Oslo, Norway. <sup>35</sup>Mental Health Centre Copenhagen, Capital Region of Denmark, Copenhagen University Hospital, Copenhagen, Denmark. <sup>36</sup>Psychosis Research Unit, Aarhus University Hospital-Psychiatry, Aarhus, Denmark. <sup>37</sup>Department of Medicine, Harvard Medical School, Boston, MA, USA. <sup>38</sup>Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. <sup>39</sup>Departments of Psychiatry and of Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, NY, USA. <sup>40</sup>Mental Illness Research Education and Clinical Center, JJ Peters VA Medical Center, Bronx, NY, USA. <sup>41</sup>Department of Psychiatry, Radboud University Medical Center, Nijmegen, the Netherlands. <sup>42</sup>These authors contributed equally: Ditte Demontis, G. Bragi Walters, Georgios Athanasiadis, Raymond Walters. <sup>43</sup>These authors jointly supervised this work: Barbara Franke, Thomas Werge, Benjamin M. Neale, Kari Stefansson, Anders D. Børglum. \*Lists of authors and their affiliations appear at the end of the paper.  e-mail: ditte@biomed.au.dk; anders@biomed.au.dk

## ADHD Working Group of the Psychiatric Genomics Consortium

Alexandra Havdahl<sup>42,43</sup>, Alysa Doyle<sup>44</sup>, Andreas Reif<sup>45</sup>, Anita Thapar<sup>24</sup>, Barbara Franke<sup>31,32,41,86</sup>, Benjamin M. Neale<sup>8,9,86</sup>, Bru Cormand<sup>46,47,48,49</sup>, Calwing Liao<sup>50</sup>, Christie Burton<sup>51</sup>, Clayton H. D. Bau<sup>52,53</sup>, Diego Luiz Rovaris<sup>54</sup>, Ditte Demontis<sup>1,2,3,85</sup>, Edmund Sonuga-Barke<sup>55</sup>, Elizabeth Corfield<sup>42,43</sup>, Eugenio Horacio Grevet<sup>56,57</sup>, Georgios Athanasiadis<sup>2,6,7,85</sup>, Henrik Larsson<sup>58,59</sup>, Ian R. Gizer<sup>60</sup>, Irwin Waldman<sup>61</sup>, Isabell Brikell<sup>1,59</sup>, Jan Haavik<sup>62,63</sup>, Jennifer Crosbie<sup>51</sup>, Joanna Martin<sup>24</sup>, James McGough<sup>64</sup>, Jonna Kuntsi<sup>65</sup>, Joseph Glessner<sup>66,67</sup>, Kate Langley<sup>68</sup>, Klaus-Peter Lesch<sup>69,70,71</sup>, Luis Augusto Rohde<sup>72,73</sup>, Mara H. Hutz<sup>74</sup>, Maria Soler Artigas<sup>25,26,27,28</sup>, Marieke Klein<sup>31,32</sup>, Mark Bellgrove<sup>75</sup>, Marta Ribasés<sup>25,26,27,28</sup>, Martin Tesli<sup>42,76</sup>, Michael C. O'Donovan<sup>24</sup>, Nina Roth Mota<sup>31,32</sup>, Ole Andreas Andreassen<sup>77,78</sup>, Ole Mors<sup>2,36</sup>, Patrick W. L. Leung<sup>79</sup>, Pedro M. Pan<sup>80</sup>, Preben Bo Mortensen<sup>2,20,21</sup>, Raymond Walters<sup>8,9,85</sup>, Ridha Joobier<sup>81</sup>, Russel Schachar<sup>51</sup>, Sandra Loo<sup>64</sup>, Sarah E. Medland<sup>33</sup>, Stephanie H. Witt<sup>82</sup>, Stephen V. Faraone<sup>39</sup>, Ted Reichborn-Kjennerud<sup>42,83</sup>, Tetyana Zayats<sup>8,9,34</sup>, Thomas Werge<sup>2,6,86</sup>, Tobias Banaschewski<sup>84</sup>, Veera M. Rajagopal<sup>1,2,3</sup> & Ziarah Hawi<sup>75</sup>

<sup>42</sup>Department of Mental Disorders, Norwegian Institute of Public Health, Oslo, Norway. <sup>43</sup>Nic Waals Institute, Lovisenberg Diaconal Hospital, Oslo, Norway. <sup>44</sup>Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. <sup>45</sup>Department for Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Frankfurt am Main, Germany. <sup>46</sup>Department of Genetics, Microbiology and Statistics, University of Barcelona, Barcelona, Spain. <sup>47</sup>Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Madrid, Spain. <sup>48</sup>Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Spain. <sup>49</sup>Institut de Recerca Sant Joan de Déu (IR-SJD), Esplugues de Llobregat, Barcelona, Spain. <sup>50</sup>Department of Human Genetics, McGill University, Montreal, Quebec, Canada. <sup>51</sup>Neurosciences and Mental Health, Hospital for Sick Children, Toronto, Ontario, Canada. <sup>52</sup>Department of Genetics, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, Brazil. <sup>53</sup>ADHD and Developmental Psychiatry Programs, Hospital de Clínicas de Porto Alegre, Rio Grande do Sul, Brazil. <sup>54</sup>Departamento de Fisiología e Biofísica, Instituto de Ciencias Biomedicas, Universidade de São Paulo, São Paulo, Brazil. <sup>55</sup>School of Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. <sup>56</sup>Department of Psychiatry, Faculty of Medicine, Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, Brazil. <sup>57</sup>Adult ADHD Outpatient Program (ProDAH), Clinical Research Center, Hospital de Clínicas de Porto Alegre, Rio Grande do Sul, Brazil. <sup>58</sup>School of Medical Sciences, Örebro University, Örebro, Sweden. <sup>59</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. <sup>60</sup>Department of Psychological and Brain Sciences, University of Missouri, Columbia, MO, USA. <sup>61</sup>Department of Psychology, Emory University, Atlanta, GA, USA. <sup>62</sup>Department of Biomedicine, University of Bergen, Bergen, Norway. <sup>63</sup>Division of Psychiatry, Haukeland University Hospital, Bergen, Norway. <sup>64</sup>Semel Institute for Neuroscience and Human Behavior, UCLA David Geffen School of Medicine, Los Angeles, CA, USA. <sup>65</sup>Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. <sup>66</sup>Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, USA. <sup>67</sup>Department of Genetics, Perelman School of Medicine, Philadelphia, PA, USA. <sup>68</sup>School of Psychology, Cardiff University, Cardiff, UK. <sup>69</sup>Division of Molecular Psychiatry, Center of Mental Health, University of Würzburg, Würzburg, Germany. <sup>70</sup>Laboratory of Psychiatric Neurobiology, Institute of Molecular Medicine, I.M. Sechenov First Moscow State Medical University, Moscow, Russia. <sup>71</sup>Department of

Neuropsychology and Psychiatry, School for Mental Health and Neuroscience, Maastricht University, Maastricht, the Netherlands. <sup>72</sup>ADHD Outpatient Program & Developmental Psychiatry Program, Hospital de Clinica de Porto Alegre, Federal University of Rio Grande do Sul, Rio Grande do Sul, Brazil. <sup>73</sup>National Institute of Developmental Psychiatry, São Paulo, Brazil. <sup>74</sup>Department of Genetics, Universidade Federal do Rio Grande do Sul Porto Alegre, Rio Grande do Sul, Brazil. <sup>75</sup>Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Melbourne, Victoria, Australia. <sup>76</sup>Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. <sup>77</sup>NORMENT Centre, Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway. <sup>78</sup>KG Jebsen Centre for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway. <sup>79</sup>Department of Psychology, The Chinese University of Hong Kong, Shatin, China. <sup>80</sup>LiNC, Department of Psychiatry, Federal University of São Paulo, São Paulo, Brazil. <sup>81</sup>Douglas Mental Health University Institute, Department of Psychiatry, McGill University, Montreal, Quebec, Canada. <sup>82</sup>Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim/Heidelberg University, Mannheim, Germany. <sup>83</sup>Institute of Clinical Medicine, University of Oslo, Oslo, Norway. <sup>84</sup>Department of Child & Adolescent Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim/Heidelberg University, Mannheim, Germany.

## iPSYCH-Broad Consortium

**Anders D. Børglum**<sup>1,2,3,86</sup>, **Benjamin M. Neale**<sup>8,9,86</sup>, **Daniel Howrigan**<sup>8,9</sup>, **David M. Hougaard**<sup>2,17</sup>, **Ditte Demontis**<sup>1,2,3,85</sup>,  
**Esben Agerbo**<sup>2,20,21</sup>, **F. Kyle Satterstrom**<sup>8,9</sup>, **Georgios Athanasiadis**<sup>2,6,7,85</sup>, **Jakob Grove**<sup>1,2,16</sup>, **Jinjie Duan**<sup>1,2,3</sup>,  
**Jonas Bybjerg-Grauholt**<sup>2,17</sup>, **Leila Farajzadeh**<sup>1,2,3</sup>, **Marie Bækved-Hansen**<sup>2,17</sup>, **Mark J. Daly**<sup>8,9,36,37</sup>, **Merete Nordentoft**<sup>2,35</sup>,  
**Ole Mors**<sup>2,35,36</sup>, **Preben Bo Mortensen**<sup>2,20,21</sup>, **Raymond Walters**<sup>8,9,85</sup>, **Tetyana Zayats**<sup>8,9,34</sup>, **Thomas D. Als**<sup>1,2,3</sup>, **Thomas Werge**<sup>2,6,86</sup>,  
**Trine Tollerup Nielsen**<sup>1,2,3</sup> & **Veera M. Rajagopal**<sup>1,2,3</sup>

## Methods

The study was approved by the local scientific ethics committees and institutional review boards. The iPSYCH study was approved by the Scientific Ethics Committee in the Central Denmark Region (case no. 1-10-72-287-12) and the Danish Data Protection Agency. In accordance with Danish legislation, the Danish Scientific Ethics Committee has, for this study, waived the need for specific informed consent in biomedical research based on existing biobanks. This deCODE study was approved by the National Bioethics Committee of Iceland (VSN 15-047), and all participants provided informed consent.

### Samples, quality control and imputation

**iPSYCH.** The iPSYCH<sup>15,70</sup> cohort consists of 129,950 genotyped individuals, among whom 85,891 (cases) have been diagnosed with at least one of six mental disorders (that is, ADHD, SCZ, BD, MDD, ASD or postpartum disorder) and the remainder are population-based controls. Samples were selected from a baseline birth cohort comprising all singletons born in Denmark between 1 May 1981 to 31 December 2008, who were resident in Denmark on their first birthday and who have a known mother ( $n = 1,657,449$ ). ADHD cases were diagnosed by psychiatrists according to the ICD10 criteria (F90.0, F90.1, F98.8 diagnosis codes) identified using the Danish Psychiatric Central Research Register<sup>71</sup> and the Danish National Patient register<sup>72</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 were individuals not diagnosed with ADHD.

Detailed information on genotyping, imputation and quality control (QC) can be found in the Supplementary Note. After QC, the iPSYCH1 ADHD sample included 38,899 individuals, and iPSYCH2 included 24,144 individuals.

**deCODE.** The deCODE cohort consisted of 8,281 individuals with ADHD. These were either individuals with a clinical diagnosis of ADHD ( $n = 5,583$ ) according to the ICD10 criteria (ICD10-F90, F90.1, F98.8) or individuals who had been prescribed medication specific for ADHD symptoms (ATC-NA06BA, mostly methylphenidate) ( $n = 2,698$ ). The control sample did not contain individuals with a diagnosis of SCZ, BD, ASD or self-reported ADHD symptoms or diagnosis. All participants who donated samples gave informed consent. Information about genotyping, QC and evaluation of potential genetic heterogeneity among individuals identified based on diagnosis codes and medication can be found in the Supplementary Note.

**PGC cohorts.** We used summary statistics from the ten 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 the report by Demontis et al.<sup>14</sup>.

### GWAS meta-analysis of ADHD

GWAS 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 ten principal components (from the final principal component analyses) as covariates using PLINK v.1.9.

GWAS of deCODE samples (8,281 ADHD cases and 137,993 controls) was done using dosage data and logistic regression with sex, year of birth and county of origin as covariates and subsequently, alleles were converted to match Haplotype Reference Consortium alleles. 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 described previously<sup>55</sup>. Findings from analyses of the genetic structure of the Icelandic population by Price et al.<sup>73</sup> suggest that lambda correction will ensure proper correction without false positives.

For the PGC cohorts, we used GWAS summary statistics for each of the ten European PGC cohorts generated as a part of our previous GWAS meta-analysis<sup>14</sup>.

See Supplementary Note for sensitivity analyses related to the impact of using sex and age as covariates in the analyses and see Supplementary Fig. 14 for the impact of including or excluding sex as covariate in GWAS of iPSYCH data.

Summary statistics from GWAS of the individual cohorts, containing variants with imputation quality (INFO score)  $>0.8$  and minor allele frequency  $>0.01$ , were meta-analyzed with a fixed effects standard error weighted meta-analysis using METAL (v.2011-03-25)<sup>74</sup> and only variants supported by an effective sample size greater than 60% were retained in the final summary statistics (6,774,224 variants).

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

### Conditional analysis

We identified potentially independent genome-wide significant lead variants for four loci located on chromosomes 1 (two secondary lead variants), 5, 11 and 20. To evaluate whether these variants were independent from the lead variants, we performed association analyses of the secondary variants while conditioning on the index variant in the locus using COJO as implemented in GCTA<sup>21</sup>.

### Identification of sets of credible variants

To identify sets of causal variants, we fine-mapped each of the 27 genome-wide loci using three fine-mapping tools, FINEMAP v.1.3.1 (ref. <sup>75</sup>), PAINTOR v.3.0 (ref. <sup>76</sup>) and CAVIARBF v.0.2.1 (ref. <sup>77</sup>), using CAUSALdb-finemapping-pip downloaded from <https://github.com/mulinlab/CAUSALdb-finemapping-pip> (ref. <sup>78</sup>). As no secondary lead variants remained genome-wide significant after conditional analyses, one causal variant was assumed per locus. Variants located in a region of 1 Mb around index variants were included in the analyses. We used a threshold of 95% for the total posterior probability of the variants included in the credible sets, and only variants claimed to be within the set by all three methods were included in the final credible set for each locus.

### Genetic correlations among cohorts and $h^2_{\text{SNP}}$

$h^2_{\text{SNP}}$  and pairwise genetic correlations among the cohorts were calculated using LD score regression<sup>79</sup> analysis of summary statistics from GWAS of deCODE samples, meta-analysis of iPSYCH1 + iPSYCH2 and meta-analysis of the ten 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%. Testing 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>23</sup>. Protein-coding genes were mapped if they were located with a distance of 10 kb upstream or downstream of the index variants or if a credible variant was annotated to the gene based on eQTL data or chromatin interaction data from human brain (the datasets used are listed in the Supplementary Note). The mapping linked credible variants to 76 ADHD-prioritized risk genes. These genes were used in gene-set enrichment analyses to evaluate whether the candidate genes were enriched among (1) genes differentially expressed in specific brain tissues, (2) genes differentially expressed at specific brain developmental stages, (3) genes encoding proteins involved in synapses, and (4) genes encoding proteins in specific biological pathways. We corrected for multiple testing separately for each of these hypotheses.

The first two aims were addressed by performing enrichment analyses in the GENE2FUNC module in FUMA. Enrichment of ADHD risk genes was assessed among predefined sets of differentially expressed genes in GTEx (54 tissue types) and BrainSpan (29 samples from individuals representing 29 different ages of brains and from 11 general developmental stages) data using a hypergeometric test, and protein-coding genes were chosen as background genes.

The third aim was addressed using SynGO<sup>24</sup> (dataset version: 20210225) to test for enrichment of the 76 risk genes among genes involved in synaptic processes and locations. We analyzed for enrichment in two subsets: biological process (201 gene sets) and cellular component (92 gene sets). We controlled using a background set of ‘brain-expressed’ genes provided by the SynGo platform (defined as ‘expressed in any GTEx v7 brain tissues’) containing 18,035 unique genes, of which 1,225 overlap with SynGO annotated genes. For each ontology term, a one-sided Fisher’s 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 was selected where each ‘gene cluster’ (unique set of genes) was found, and then multiple testing correction was applied using the false discovery rate on the subset of terms that contained these ‘gene clusters’. Only ontology terms with gene sets with a minimum of three genes were included in the enrichment analysis.

The fourth aim was addressed by testing whether the 76 genes were enriched in pathways or gene sets using Enrichr<sup>25,26</sup> and its implemented databases (26 databases). Only pathways enriched with more than two genes were considered. We took 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 databases, no significantly enriched pathways were identified.

We also tested for enrichment of genes reported in the GWAS catalog (2019) and UK Biobank GWAS (v.1) among the 76 genes and used [https://appyters.maayanlab.cloud/Enrichr\\_Manhattan\\_Plot/](https://appyters.maayanlab.cloud/Enrichr_Manhattan_Plot/) to visualize the results.

Finally, we conducted pathway enrichment analysis using results from the full GWAS meta-analysis (that is, with no preselection of genes) by performing MAGMA<sup>27</sup> gene-set analysis in FUMA (see details in the Supplementary Note).

### Transcriptomic imputation model construction and TWAS

Transcriptomic imputation models were constructed as previously described<sup>28</sup> for DLPFC transcript levels<sup>80</sup>. The genetic dataset of the PsychENCODE cohort was uniformly processed for QC steps before genotype imputation. The analysis was restricted to samples with European ancestry as previously described<sup>28</sup>. Genotypes were imputed using the University of Michigan server<sup>81</sup> with the Haplotype Reference Consortium reference panel<sup>82</sup>. Gene expression information (at both the gene and transcript levels) was derived from RNA-sequencing counts, which were adjusted for known and hidden confounds, followed by quantile normalization<sup>80</sup>. For the construction of the transcriptomic imputation models, we used EpiXcan<sup>28</sup>, an elastic net-based method, which weighs SNPs based on available epigenetic annotation information<sup>83</sup>. We performed transcript–trait association analysis for ADHD as previously described<sup>28</sup>. Briefly, we applied the S-PrediXcan method<sup>28</sup> to integrate the ADHD GWAS meta-analysis summary statistics and the transcriptomic imputation models constructed above to obtain association results at both gene and transcript levels.

### Gene-based association and tissue-specific gene expression

We used MAGMA v.1.08 implemented in FUMA v.1.3.6a<sup>23</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* = 0.05/18,381 =  $2.72 \times 10^{-6}$ ).

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

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 were 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  $h^2_{\text{SNP}}$  of variants located in cell-type-specific epigenetic peaks by examining the overlap of common genetic risk variants with open chromatin from a DNase I hypersensitive sites study profiling major human cell types<sup>31</sup> and a single-cell assay for transposase accessible chromatin<sup>32</sup> study using an LD-score-partitioned heritability approach<sup>34</sup>. All regions of open chromatin were extended by 500 bp in either direction. The broad major histocompatibility complex (MHC) region (hg19 chr6:25–35 Mb) was excluded owing to its extensive and complex LD structure, but otherwise default parameters were used for the algorithm. We applied Bonferroni correction (correcting for 23 cell types), and results below *P* = 0.0022 were considered significant.

In addition, we performed cell-type-specific analyses implemented in FUMA, using data from 13 single-cell RNA-sequencing datasets from human brain. The method was described in detail by Watanabe et al.<sup>33</sup>. Datasets used and a short summary of the method can be found in the Supplementary Note.

### Overlap of common ADHD risk variants with rPTVs

We analyzed the overlap of common variants with rPTVs in a subset of iPSYCH samples that had also been whole-exome sequenced. A major part of the data (pilot 1, wave 1, wave 2) was also included in the recent study by Satterstrom et al.<sup>17</sup>, and the same QC procedure was applied in this study. Descriptions of the whole-exome-sequencing procedure, QC and annotation can be found in the Supplementary Note. Variants were defined as PTVs if they were annotated as having large effects on gene function (nonsense variant, frameshift, splice site). We defined a variant as being rare if it had an allele count of five or less across the combination of the full iPSYCH exome-sequencing dataset (*n* = 28,448) and non-Finnish Europeans in the nonpsychiatric gnomAD exome database (*n* = 44,779).

We tested for increased burden of rPTVs in ADHD compared with 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, and (3) genes with at least five credible variants within the coding regions. The requirement for five credible variants was chosen 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). Two other genes with fewer than five credible variants, located in two other loci on chromosome 3, were excluded.

The burden of rPTVs and rSYNs in cases compared with 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 rSYNs, percentage of target with coverage >20×, mean read depth at sites within the exome target passing Variant Quality Score Recalibration (VQSR), total number of variants, and sequencing wave.

Significant enrichment was found only in the set of 18 genes identified based on credible variants. We therefore looked specifically into these genes to identify whether the signal was driven by specific genes.

rPTVs were found in 13 of the genes, and eight of these genes had more rPTVs in cases than in controls when looking at raw counts (Supplementary Table 16). We performed a gene-based burden test using EPACTS (<https://genome.sph.umich.edu/wiki/EPACTS>) and logistic Wald test (correcting using the covariates as described above). In addition, in order to increase power to detect an increased burden of rPTVs at the gene level in ADHD cases, we combined iPSYCH controls with information about rPTVs in gnomAD (non-Finnish European nonpsychiatric subset of individuals), done as described previously<sup>17</sup>. We performed these gene-based tests using Fisher's exact test, and only the following genes were considered: (1) genes with a higher number of rSYNs in gnomAD controls compared with iPSYCH cases, and (2) genes with a higher rate of rPTVs in cases compared with controls in the iPSYCH data.

### Genetic overlap with other phenotypes

We estimated genetic correlations of ADHD with other phenotypes in LDhub<sup>35</sup> (published GWAS: 255 phenotypes; UK Biobank GWAS: 514 phenotypes). In addition, genetic correlations with three phenotypes not available in LDhub (cannabis use disorder<sup>55</sup>, smoking initiation<sup>85</sup> and education attainment<sup>51</sup>) were estimated locally using LD score regression<sup>79</sup>.

We applied MiXeR<sup>36</sup> to our ADHD GWAS summary statistics, to GWAS from a selection of complex traits showing high genetic correlation with ADHD (ASD<sup>86</sup>, SCZ<sup>54</sup>, BMI<sup>87</sup>, educational attainment<sup>88</sup>, age at first birth<sup>89</sup>, smoking initiation<sup>85</sup> and insomnia<sup>90</sup>) and to a new GWAS meta-analysis of depression including 371,184 cases and 978,703 controls<sup>91</sup> (Supplementary Table 19) to quantify (1) the number of variants influencing each trait and (2) 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. First, we ran a univariate model for each trait to estimate the number of common variants that had a nonzero genetic additive impact on the phenotype. The univariate model generated estimates of 'polygenicity' (that is, the proportion of nonnull variants) and 'discoverability' (that is, the variance of effect sizes of nonnull SNPs). Second, the variance estimates from the univariate step were used to run a bivariate model in a pairwise fashion (that is, ADHD versus each of the other traits), which produced estimates of SNPs with a specific effect on the first or the second trait and SNPs with a nonzero effect on both traits (for details of the method, see also ref. <sup>18</sup> and the Supplementary Note). The models were evaluated using the AIC<sup>52</sup> and illustrated with modeled versus observed conditional quantile–quantile plots (Supplementary Fig. 11). The AIC values can be found in Supplementary Table 19.

### PGS analysis of cognitive measures in PNC

PGS analysis was performed on 4,973 individuals of European ancestry from the PNC, aged 8–21 years. Information about imputation and QC of the PNC data can be found in the Supplementary Note.

The software PRS-CS<sup>92</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 (<https://github.com/getian107/PRScs>) was used. The following default settings were used for PRS-CS: parameter a in the  $\gamma$ - $\gamma$  prior = 1, parameter b in the  $\gamma$ -prior = 0.5, MCMC iterations = 1000, number of burn-in iterations = 500, and thinning of the Markov chain factor = 5. In addition, the global shrinkage parameter phi was determined using a fully Bayesian method. PLINK v.2.0 (ref. <sup>93</sup>) was then used to calculate individual-level ADHD-PGS. Linear regression was used to test the associations between ADHD-PGS and neurocognitive phenotypes measured in the PNC. Age (at time of neurocognitive testing)<sup>2</sup>, genotyping batch, sex, and the first ten MDS dimensions were used as covariates. The neurocognitive measures were obtained using the CNB, which consists of 14 tests in five domains: executive control, episodic memory, complex cognitive processing, social cognition and sensorimotor speed. The battery has been described in detail elsewhere<sup>40</sup>. In addition,

associations of ADHD-PGS with results from the WRAT-4 (ref. <sup>41</sup>) were analyzed. See the Supplementary Note regarding transformation of the CNB measures.

The total variance explained by ADHD-PGS and model covariates for each neurocognitive phenotype was reported using adjusted  $R^2$ . In addition, the variance explained by ADHD-PGS 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.

### Reporting summary

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

### Data availability

Summary statistics from the ADHD GWAS meta-analysis are available for download at the PGC website (<https://www.med.unc.edu/pgc/download-results/>). All relevant iPSYCH data are available from the authors after approval by the iPSYCH Data Access Committee and can only be accessed on the secured Danish server (GenomeDK; <https://genome.au.dk>) as the data are protected by Danish legislation. For data access and correspondence, please contact D.D. (ditte@biomed.au.dk) or A.D.B. (anders@biomed.au.dk).

### Code availability

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

### References

70. Bybjerg-Grauholt, J. et al. The iPSYCH2015 case-cohort sample: updated directions for unravelling genetic and environmental architectures of severe mental disorders. Preprint at medRxiv <https://doi.org/10.1101/2020.11.30.20237768> (2020).
71. Mors, O., Perto, G. P. & Mortensen, P. B. The Danish psychiatric central research register. *Scand. J. Public Health* **39**, 54–57 (2011).
72. Lynge, E., Sandegaard, J. L. & Rebolj, M. The Danish national patient register. *Scand. J. Public Health* **39**, 30–33 (2011).
73. Price, A. L. et al. The impact of divergence time on the nature of population structure: an example from Iceland. *PLoS Genet.* **5**, e1000505 (2009).
74. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
75. Benner, C. et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
76. 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 (2017).
77. Chen, W. et al. Fine mapping causal variants with an approximate Bayesian method using marginal test statistics. *Genetics* **200**, 719–736 (2015).
78. Wang, J. et al. CAUSALdb: a database for disease/trait causal variants identified using summary statistics of genome-wide association studies. *Nucleic Acids Res.* **48**, D807–D816 (2020).
79. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
80. Gandal, M. J. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018).
81. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

82. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
83. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
84. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
85. 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. *Nat. Genet.* **51**, 237–244 (2019).
86. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
87. Yengo, L. et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
88. Okbay, A. et al. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat. Genet.* **54**, 437–449 (2022).
89. Mills, M. C. et al. Identification of 371 genetic variants for age at first sex and birth linked to externalising behaviour. *Nat. Hum. Behav.* **5**, 1717–1730 (2021).
90. Watanabe, K. et al. Genome-wide meta-analysis of insomnia prioritizes genes associated with metabolic and psychiatric pathways. *Nat. Genet.* **54**, 1125–1132 (2022).
91. Als, T. D. et al. Identification of 64 new risk loci for major depression, refinement of the genetic architecture and risk prediction of recurrence and comorbidities. Preprint at medRxiv <https://doi.org/10.1101/2022.08.24.22279149> (2022).
92. Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
93. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

## Acknowledgements

We thank additional members of the ADHD working group of the PGC who are not named as coauthors under the working group banner for their contributions. We would like to thank the employees and research participants of 23andMe for making this work possible. D.D. was supported by the Novo Nordisk Foundation (NNF20OC0065561), the Lundbeck Foundation (R344-2020-1060) and the European Union's Horizon 2020 research and innovation program under grant agreement no. 965381 (TIMESPAN). The iPSYCH team was supported by grants from the Lundbeck Foundation (R102-A9118, R155-2014-1724 and R248-2017-2003), National Institutes of Health (NIH)/National Institute of Mental Health (NIMH) (1U01MH109514-01 and 1R01MH124851-01 to A.D.B.) and the Universities and University Hospitals of Aarhus and Copenhagen. The Danish National Biobank resource was supported by the Novo Nordisk Foundation. High-performance computer capacity for handling and statistical analysis of iPSYCH data on the GenomeDK HPC facility was provided by the Center for Genomics and Personalized Medicine and the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark (grant to A.D.B.). Research reported in this publication was supported by the NIMH of the NIH under award number R01MH124851. The content is solely the responsibility of the authors and does not

necessarily represent the official views of the NIH. 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 for the Dutch National Science Agenda NeurolabNL project (grant 400-17-602). S.E.M. was funded by National Health and Medical Research Council grants APP1172917, APP1158125 and APP1103623. This work was supported by the Instituto de Salud Carlos III (PI19/01224, PI20/0004); by the Pla Estratégic de Recerca i Innovació en Salut, 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 and by 'la Marató de TV3' (092330/31) and the European College of Neuropsychopharmacology Network 'ADHD across the Lifespan' ([## Author contributions](https://www.ecnp.eu/researchinnovation/ECNP-networks>List-ECNP-Networks/</a>). T.Z. was funded by NIH, grant no. R37MH107649-07S1 and by Research Council of Norway, NRC, Grant No. 288083. This study was also supported by the NIH, Bethesda, MD, under award numbers T32MH087004 (to K.T.), K08MH122911 (to G.V.), R01MH125246 (to P.R.) and U01MH116442 (to P.R.).</p></div><div data-bbox=)

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. performed the analysis. 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 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. performed sample and/or data provision and processing. D.D., G.B.W., K.T., G.A., G.V., J.B., H.S. and A.D.B. wrote the manuscript. 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.R., S.V.F., H.S., P.R., B.F., B.M.N., K.S. and A.D.B. performed core revision of the manuscript. D.D. and A.D.B. provided study direction. D.D., G.B.W., H.S., P.R., B.F., T.W., B.M.N., K.S. and A.D.B. supervised the study. All authors contributed to critical revision of the manuscript.

## Competing interests

B.M.N. currently serves as a member of the scientific advisory board at Deep Genomics and Neumora (previously RBNC) and as a consultant for Camp4 Therapeutics, Takeda Pharmaceutical and Biogen. All deCODE-affiliated authors are employees of deCODE/Amgen.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41588-022-01285-8>.

**Correspondence and requests for materials** should be addressed to Ditte Demontis or Anders D. Børglum.

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

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

## Reporting Summary

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

### Statistics

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

n/a Confirmed

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

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

#### Data collection

This is fully described in the Online Methods a. In brief:

iPSYCH genotype data: genotyping was done using the PsychChip and Global Screening Array from IlluminiGenCall and Birdseed and were used for genotype calling of variants with minor allele frequency > 0.01 i iPSYCH1. iPSYCH2 genotypes were called using GenTrain V3. Imputation was done using the prephasing/imputation stepwise approach implemented in EAGLE v2.3.5 and Minimac3, using the Haplotype Reference Consortium panel v1.0. deCODE genotype data: Samples were assayed with several Illumina arrays at deCODE genetics and genotypes called using GraphTyper2. Variant imputation was performed based on the IMPUTE HMM model and long-range phasing. PGC genotype data: These data has been published previously and well-described in Demontis and Walters et al. Nature Genetics 2019. Rare variants: alignment of sequence reads to the reference genome and calling of genotypes were done using BWA and GATK v.3.4. QC was done in Hail v. 0.1 (<https://github.com/hail-is/hail>), annotation of variants was done using SnpEff v. 4.3t and SnpSift v. 4.3t.

#### Data analysis

Quality control and association analyses were done using Plink 1.9 and Eigensoft 6.1.3. GWAS meta-analysis was done using METAL (version 2011-03-25). Genetic correlations and heritability were calculated using LD score regression v.1.0.1 (<https://github.com/bulik/lsc>), partitioned heritability using LDSC-SEG. Conditional analysis was done using COJO as implemented in GCTA. Finemapping was done using CAUSALdb-finemapping-pip downloaded from <https://github.com/mulinlab/CAUSALdb-finemapping-pip> which finemap loci using fine-mapping tools, FINEMAP v. 1.3.1, PAINTOR v.3.0 and CAVIARBF v.0.2.1. Variants were mapped to genes and analyzed in FUMA (<https://fuma.ctglab.nl/>). Enrichment analyses were done in FUMA, SynGO (<https://syngoportal.org/>) and Enrichr (<https://maayanlab.cloud/Enrichr/>). [https://appyters.maayanlab.cloud/Enrichr\\_Manhattan\\_Plot/](https://appyters.maayanlab.cloud/Enrichr_Manhattan_Plot/) to visualize the results. Construction of transcriptomic imputation models were done using EpiXcan (<https://bitbucket.org/roussoslab/epixcan/src/master/>), TWAS was done using S-PrediXcan (<https://github.com/hakyimlab/MetaXcan>). Gene-based association was done using MAGMA v1.08. Genetic outliers were detected using Eigensoft v1.6.4. Gene-based burden test was done using EPACTS (<https://genome.sph.umich.edu/wiki/EPACTS>) MixR (https://github.com/precimed/mixer) was used to estimate number of variants. PRS-CS (<https://github.com/getian107/PRScs>) was used to calculate per-allele posterior SNP effect sizes and Plink v2.0 was then used to calculate individual-level ADHD PGS. The variance explained was calculated using <https://github.com/>

GabrielHoffman/misc\_vp/blob/master/calcVarPart.R. 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.

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

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

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

Summary statistics from the ADHD GWAS meta-analysis is available for download at the iPSYCH and PGC websites: <https://ipsycho.dk/en/research/downloads>, <https://www.med.unc.edu/pgc/download-results/>. All relevant iPSYCH data are available from the authors after approval by the iPSYCH Data Access Committee and can only be accessed on the secured Danish server (GenomeDK <https://genome.au.dk>) as the data are protected by Danish legislation.

For data access please contact: Ditte Demontis or Anders D. Børglum.

Haplotype Reference Consortium panel v1.0: <http://www.haplotype-reference-consortium.org/data-access>

## Field-specific reporting

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

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

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

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

### Sample size

The GWAS meta-analysis of ADHD comprises 38,691 individuals with ADHD and 186,843 controls. No sample size calculation was made prior to this study. The number of ADHD cases in the iPSYCH cohort was fixed as it included all individuals born in Denmark between 1981 and 2008, which have received an ADHD diagnosis in 2016 or before. Additionally, we have previously published a GWAS meta-analysis of ADHD with around half the number of cases compared to this study, which identified 12 genome-wide significant loci, indicating that by increasing the sample size we would be able to discover more risk loci.

### Data exclusions

We aimed at analyzing genetically homogeneous individuals. Genetic outliers were excluded based on principal component analyses and related individuals were removed.

### Replication

We have chosen to include all available samples in one large hypothesis free GWAS meta-analysis to maximize power instead of leaving out a cohort and use that for replication of top associated loci from the GWAS meta-analysis. We have evaluated the consistency of the results across the cohorts included in the GWAS meta-analysis visualized as forest plots (Extended data 2), which demonstrate consistency of the signal across included cohorts for to associated loci.

### Randomization

Allocation into groups was not random. Individuals were allocated to the ADHD groups if they were diagnosed with ADHD (or alternatively were prescribed medication specific for ADHD treatment). The controls do not have a diagnosis of ADHD.

### Blinding

In iPSYCH, diagnoses are drawn from registries based on ICD10 diagnosis codes, and in deCODE from registries based on ICD10 diagnosis codes and medication perscription. These are administrative data bases populated by data from the clinicians long before the current study. The blood samples are pulled from a biobank. Hence, the study participants and diagnosing clinicians are blinded with respect to this study. Genotyping is done on a massive scale overall for the full iPSYCH and deCODE cohorts , and the data is generated without a specif goal or effect in mind except for an overall goal of investigating the genetic and environmental effects on psychiatric disorders and diseases. So although it is in principle possible for analysts in the lab to look up crude diagnostic data for a sample, it will not change the genotyping.

## Reporting for specific materials, systems and methods

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

**Materials & experimental systems**

n/a	Involved in the study
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology and archaeology
<input checked="" type="checkbox"/>	Animals and other organisms
<input type="checkbox"/>	Human research participants
<input checked="" type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	Dual use research of concern

**Methods**

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging

**Human research participants**Policy information about [studies involving human research participants](#)

## Population characteristics

Common variants, covariates used: iPSYCH - 10 PCs derived from principal component analysis (on a set of pruned variants); deCODE - sex, age county.  
 Rare variants, covariates used: 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. Polygenic score analyses, covariates used: Age (at time of neurocognitive testing), age squared, genotyping batch, sex, and the first 10 MDS dimensions.

## Recruitment

In iPSYCH, diagnoses (ICD10-F90, F90.1, F98.8) were drawn from national registries and the blood samples were pulled from the Danish Neonatal Screening Biobank. Hence, it is a population sample and bias from self-selection is not possible. The deCODE cohort consisted of ADHD cases with a clinical diagnosis of ADHD (ICD10-F90, F90.1, F98.8) or individuals that have been prescribed medication specific for ADHD symptoms (ATC-NA06BA, mostly methylphenidate)

## Ethics oversight

The study was approved by the Danish Data Protection Agency and the Scientific Ethics Committee in Denmark, and the National Bioethics Committee of Iceland

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