ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/110923/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Demontis, Ditte, Walters, Raymond K., Martin, Joanna , Mattheisen, Manuel, Als, Thomas Damm, Agerbo, Esben, Baldursson, Gísli, Belliveau, Rich, Bybjerg-Grauholm, Jonas, Bækvad-Hansen, Marie, Cerrato, Felecia, Chambert, Kimberly, Churchhouse, Claire, Dumont, Ashley, Eriksson, Nicholas, Gandal, Michael, Goldstein, Jacqueline, Grasby, Katrina L., Grove, Jakob, Gudmundsson, Olafur O., Hansen, Christine S., Hauberg, Mads Engel, Hollegaard, Mads V., Howrigan, Daniel P., Huang, Hailiang, Maller, Julian, Martin, Alicia R. , Martin, Nicholas G. , Moran, Jennifer, Pallesen, Jonatan, Palmer, Duncan S., Bøcker Pedersen, Carsten, Giørtz Pedersen, Marianne, Poterba, Timothy, Buchhave Poulsen, Jesper, Ripke, Stephan, Robinson, Elise B., Satterstrom, F. Kyle, Stefansson, Hreinn, Stevens, Christine, Turley, Patrick, Walters, G. Bragi, Won, Hyejung, Wright, Margaret J., Andreassen, Ole A., Asherson, Philip, Burton, Christie, Boomsma, Dorret I., Cormand, Bru, Dalsgaard, Soren, Franke, Barbara, Gelernter, Joel, Geschwind, Daniel, Hakonarson, Hakon, Haavik, Jan, Kranzler, Henry, Kuntsi, Jonna, Langley, Kate , Lesch, Klaus-Peter, Middeldorp, Christel, Reif, Andreas, Rohde, Luis Augusto, Roussos, Panos, Schacha, Russell, Sklar, Pamela, Sonuga-Barke, Edmund J. S., Sullivan, Patrick F., Thapar, Anita , Tung, Joyce Y., Waldman, Irwin ., Medland, Sarah E., Stefansson, Kari, Nordentoft, Merete, Hougaard, David M., Werge, Thomas, Mors, Ole, Bo Mortensen, Preben, Daly, Mark J., Faraone, Stephen V., Børglum, Anders D., Neale, Benjamin M. and Anney, Richard 2019. Discovery of the first genome-wide significant risk loci for attention-deficit/hyperactivity disorder. Nature Genetics 51 , pp. 63-75. 10.1038/s41588-018-0269-7

Publishers page: https://doi.org/10.1038/s41588-018-0269-7

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

Discovery of the first genome-wide significant risk loci for attention-deficit/hyperactivity disorder

Ditte Demontis[†1,2,3], Raymond K. Walters[†4,5], Joanna Martin[5,6,7], Manuel Mattheisen[1,2,3,8,9], Thomas Damm Als[1,2,3], Esben Agerbo[1,10,11], Gísli Baldursson[12], Rich Belliveau[5], Jonas Bybjerg-Grauholm[1,13], Marie Bækvad-Hansen[1,13], Felecia Cerrato[5], Kimberly Chambert[5], Claire Churchhouse[4,5,14], Ashley Dumont[5], Nicholas Eriksson[15], Michael Gandal[16,17,18,19], Jacqueline Goldstein[4,5,14], Katrina L. Grasby[20], Jakob Grove[1,2,3,21], Olafur O. Gudmundsson[22,23,12], Christine S. Hansen[1,13,24], Mads Engel Hauberg[1,2,3], Mads V. Hollegaard[1,13], Daniel P. Howrigan[4,5], Hailiang Huang[4,5], Julian Maller[5,25], Alicia R. Martin[4,5,14], Nicholas G. Martin[20], Jennifer Moran[5], Jonatan Pallesen[1,2,3], Duncan S. Palmer[4,5], Carsten Bøcker Pedersen[1,10,11], Marianne Giørtz Pedersen[1,10,11], Timothy Poterba[4,5,14], Jesper Buchhave Poulsen[1,13], Stephan Ripke[4,5,14,26], Elise B. Robinson[4,27], F. Kyle Satterstrom[4,5,14], Hreinn Stefansson[22], Christine Stevens[5], Patrick Turley[4,5], G. Bragi Walters[22,23], Hyejung Won[16,17], Margaret J. Wright[28], ADHD Working Group of the Psychiatric Genomics Consortium (PGC), Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium, 23andMe Research Team, Ole A. Andreassen[29], Philip Asherson[30], Christie Burton[31], Dorret Boomsma[32,33], Bru Cormand[34,35,36,37], Søren Dalsgaard[10], Barbara Franke[38], Joel Gelernter[39,40], Daniel Geschwind[16,17,18], Hakon Hakonarson[41], Jan Haavik[42,43], Henry Kranzler[44,45], Jonna Kuntsi[30], Kate Langley[7], Klaus-Peter Lesch[46,47,48], Christel Middeldorp[32,49,50], Andreas Reif[51], Luis Augusto Rohde[52,53], Panos Roussos[54,55,56,57], Russell Schachar[31], Pamela Sklar[54,55,56], Edmund Sonuga-Barke[58,59,60], Patrick F. Sullivan[61,6], Anita Thapar[7], Joyce Tung[15], Irwin Waldman[62], Sarah E. Medland[20], Kari Stefansson[22,23], Merete Nordentoft[1,63], David M. Hougaard[1,13], Thomas Werge[1,24,64], Ole Mors[1,65], Preben Bo Mortensen[1,2,10,11], Mark J. Daly[4,5,14], Stephen V. Faraone[*66], Anders D. Børglum[*1,2,3], and Benjamin M. Neale[*4,5,14]

† Equal contributions. * Co-last authors.

Correspond with: Benjamin M. Neale (bneale@broadinstitute.org), Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. Anders D. Børglum (anders@biomed.au.dk) Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark . Stephen V. Faraone (sfaraone@childpsychresearch.org)Departments of Psychiatry and Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, New York, USA.

Author Affiliations:

1 The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark
2 Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark
3 Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark
4 Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA
5 Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA

6 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
7 MRC Centre for Neuropsychiatric Genetics & Genomics, School of Medicine, Cardiff University, Cardiff, United Kingdom
8 Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
9 Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden
10 National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark
11 Centre for Integrated Register-based Research, Aarhus University, Aarhus, Denmark
12 Department of Child and Adolescent Psychiatry, National University Hospital, Reykjavik, Iceland
13 Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark
14 Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA
15 23andMe, Mountain View, California, USA
16 Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA
17 Center for Autism Research and Treatment and Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, California, USA
18 Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA
19 Department of Psychiatry, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, California, USA
20 QIMR Berghofer Medical Research Institute, Royal Brisbane Hospital, Brisbane, Australia
21 Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark
22 deCODE genetics/Amgen, Reykjavík, Iceland
23 Faculty of Medicine, University of Iceland, Reykjavík, Iceland
24 Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, Roskilde, Denmark
25 Genomics plc, Oxford, United Kingdom
26 Department of Psychiatry, Charite Universitatsmedizin Berlin Campus Benjamin Franklin, Berlin, Germany
27 Department of Epidemiology, Harvard Chan School of Public Health, Boston, Massachusetts, USA
28 Queensland Brain Institute, the University of Queensland, Brisbane, Australia
29 KG Jebsen Centre for Psychosis Research, Norway Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway
30 Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
31 Psychiatry, Neurosciences and Mental Health, The Hospital for Sick Children, University of Toronto, Toronto, Canada

32 Department of Biological Psychology, Neuroscience Campus Amsterdam, VU University, Amsterdam, The Netherlands
33 EMGO Institute for Health and Care Research, Amsterdam, The Netherlands
34 Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain
35 Biomedical Network Research Centre on Rare Diseases (CIBERER), Barcelona, Spain
36 Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Catalonia, Spain
37 Institut de Recerca Sant Joan de Déu (IRSJD), Barcelona, Catalonia, Spain
38 Departments of Human Genetics (855) and Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, The Netherlands
39 Department of Psychiatry, Genetics, and Neuroscience, Yale University School of Medicine, New Haven, Connecticut, USA
40 Veterans Affairs Connecticut Healthcare Center, West Haven, Connecticut, USA
41 The Center for Applied Genomics, The Children´s Hospital of Philadelphia, Philadelphia, PA, USA
42 K.G. Jebsen Centre for Neuropsychiatric Disorders, University of Bergen, Norway.
43 Haukeland University Hospital, Bergen, Norway
44 Department of Psychiatry, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
45 Veterans Integrated Service Network (VISN4) Mental Illness Research, Education, and Clinical Center (MIRECC), Crescenz VA Medical Center, Philadephia, PA, USA
46 Division of Molecular Psychiatry, Clinical Research Unit on Disorders of Neurodevelopment and Cognition. Laboratory of Translational Neuroscience. Center of Mental Health, University of Wuerzburg, Germany
47 Department of Neuroscience, School for Mental Health and Neuroscience (MHENS), Maastricht University, The Netherlands
48 Laboratory of Psychiatric Neurobiology, Institute of Molecular Medicine, I.M. Sechenov First Moscow State Medical University, Moscow, Russia
49 Child Health Research Centre, University of Queensland, Brisbane Australia
50 Child and Youth Mental Health Service, Children's Health Queensland Hospital and Health Service, Brisbane, Australia
51 Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital, Frankfurt, Germany
52 Department of Psychiatry, Faculty of Medicine, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
53 ADHD Outpatient Clinic, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil
54 Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA
55 Institute for Genomics and Multiscale Biology, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
56 Friedman Brain Institute, Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA

57 Mental Illness Research Education and Clinical Center (MIRECC), James J. Peters VA Medical Center, Bronx, New York, USA
58 School of Psychology, University of Southampton, Southampton, United Kingdom
59 Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium
60 Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom
61 Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC, USA
62 Department of Psychology, Emory University, Atlanta, Georgia, USA
63 Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Copenhagen, Denmark
64 Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
65 Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark
66 Departments of Psychiatry and Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, New York, USA

Abstract

Attention-Deficit/Hyperactivity Disorder (ADHD) is a highly heritable childhood behavioral disorder affecting 5% of children and 2.5% of adults. Common genetic variants contribute substantially to ADHD susceptibility, but no variants have been robustly associated with ADHD. We report a genome-wide association meta-analysis of 20,183 ADHD diagnosed cases and 35,191 controls that identifies variants surpassing genome-wide significance in 12 independent loci, revealing new and important information on the underlying biology of ADHD. Associations are enriched in evolutionarily constrained genomic regions, loss-of-function intolerant genes and around brain-expressed regulatory marks. Analyses of three replication studies; a cohort of diagnosed ADHD, a self-reported ADHD sample and a meta-analysis of quantitative measures of ADHD symptoms in the population, support these findings while highlighting study-spcific differences on genetic overlap with educational attainment. Strong concordance with GWAS of quantitative population measures of ADHD symptoms supports that clinical diagnosis of ADHD is an extreme expression of continuous heritable traits.

Introduction

Attention-Deficit/Hyperactivity Disorder (ADHD) is a neurodevelopmental psychiatric disorder, that affects around 5% of children and adolescents and 2.5% of adults worldwide[1]. ADHD is often persistent and markedly impairing with increased risk of harmful outcomes such as injuries[2], traffic accidents[3], increased health care utilization[4,5], substance abuse[6], criminality[7], unemployment[8], divorce[4], suicide[9], AIDS risk behaviors[8], and premature mortality[10]. Epidemiologic and clinical studies implicate genetic and environmental risk factors that affect

the structure and functional capacity of brain networks involved in behavior and cognition[1], in the etiology of ADHD.

Consensus estimates from over 30 twin studies indicate that the heritability of ADHD is 70-80% throughout the lifespan[11,12] and that environmental risks are those not shared by siblings[13]. Twin studies also suggest that diagnosed ADHD represents the extreme tail of one or more heritable quantitative traits[14]. Additionally, family and twin studies report genetic overlap between ADHD and other conditions including antisocial personality disorder/behaviours[15], cognitive impairment[16], autism spectrum disorder[17,18], schizophrenia[19], bipolar disorder[20], and major depressive disorder[21].

Thus far genome-wide association studies (GWASs) to identify common DNA variants that increase the risk of ADHD have not been successful[22]. Nevertheless, genome-wide SNP heritability estimates range from 0.10 – 0.28[23,24] supporting the notion that common variants comprise a significant fraction of the risk underlying ADHD[25] and that with increasing sample size, and thus increasing statistical power, genome-wide significant loci will emerge.

Previous studies have demonstrated that the common variant risk, also referred to as the single nucleotide polymorphism (SNP) heritability, of ADHD is also associated with depression[25], conduct problems[26], schizophrenia[27], continuous measures of ADHD symptoms[28,29] and other neurodevelopmental traits[29] in the population. Genetic studies of quantitative ADHD symptom scores in children further support the hypothesis that ADHD is the extreme of a quantitative trait[30].

Here we present a genome-wide meta-analysis identifying the first genome-wide significant loci for ADHD using a combined sample of 55,374 individuals from an international collaboration. We also strengthen the case that the clinical diagnosis of ADHD is the extreme expression of one or more heritable quantitative traits, at least as it pertains to common variant genetic risk, by integrating our results with previous GWAS of ADHD-related behavior in the general population.

Results

Genome-wide significantly associated ADHD risk loci

Genotype array data for 20,183 ADHD cases and 35,191 controls were collected from 12 cohorts (Supplementary Table 1). These samples included a population-based cohort of 14,584 cases and 22,492 controls from Denmark collected by the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH; Supplementary Figure 1), and 11 European, North American and Chinese cohorts aggregated by the Psychiatric Genomics Consortium (PGC). ADHD cases in iPSYCH were identified from the national Psychiatric Central Research Register psychiatric and diagnosed by psychiatrists at a psychiatric hospital according to ICD10 (F90.0), and genotyped using Illumina PsychChip. Designs for the PGC cohorts have been described previously[24,25,31,32,22] (see Supplementary Information for detailed cohort descriptions).

Prior to analysis, stringent quality control procedures were performed on the genotyped markers and individuals in each cohort using a standardized pipeline[33] (Online Methods). Related individuals were removed, and genetic outliers within each cohort were excluded based on

principal component analysis. Non-genotyped markers were imputed using the 1000 Genomes Project Phase 3 reference panel[34] (Online Methods).

GWAS was conducted in each cohort using logistic regression with the imputed additive genotype dosages. Principal components were included as covariates to correct for population stratification[35] (Supplementary Information), and variants with imputation INFO score < 0.8 or minor allele frequency (MAF) < 0.01 were excluded. The GWAS were then meta-analyzed using an inverse-variance weighted fixed effects model[36]. The included single Chinese cohort had insufficient sample size for well-powered trans-ethnic modelling (Supplementary Figure 2). Association results were considered only for variants with an effective sample size greater than 70% of the full meta-analysis, leaving 8,047,421 variants in the final meta-analysis. A meta-analysis restricted to European-ancestry individuals (19,099 cases, 34,194 controls) was also performed to facilitate secondary analyses (Supplementary Information).

In total, 304 genetic variants in 12 loci surpassed the threshold for genome-wide significance ($P<5\times10^{-8}$; Figure 1, Table 1, Supplementary Figure 3.A2 – 3.N2). Results for the European ancestry meta-analysis were substantively similar (Supplementary Figure 4). No marker demonstrated significant heterogeneity between studies (Supplementary Figure 5 and 6) and no heterogeneity was observed between the Chinese and European ancestry cohorts (Supplementary Figure 2). Conditional analysis within each locus did not identify any independent secondary signals meeting genome-wide significance (Online Methods, Supplementary Table 2).

Homogeneity of effects between cohorts

No genome-wide significant heterogeneity was observed in the ADHD GWAS meta-analysis (Supplementary Information). Genetic correlation analysis (Online Methods) provided further evidence that effects were consistent across cohort study designs. The estimated genetic correlation between the European ancestry PGC samples and the iPSYCH sample from LD score regression[37] was not significantly less than one (r_g = 1.17, SE = 0.20). The correlation between European ancestry PGC case/control and trio cohorts estimated with bivariate GREML was similarly close to one (r_g = 1.02, SE = 0.32; Supplementary Table 3).

Polygenic risk scores (PRS)[38] were also consistent across target samples. PRS computed in each PGC study using iPSYCH as the training sample were consistently higher in ADHD cases as compared to controls or pseudo-controls (Supplementary Figure 7). Increasing deciles of PRS in the PGC were associated with higher odds ratio (OR) for ADHD (Figure 2). A similar pattern was seen in five-fold cross validation in the iPSYCH cohort, with PRS for each subset computed from the other four iPSYCH subsets and the PGC samples used as training samples (Online Methods; Figure 2). Across iPSYCH subsets, the mean of the maximum variance explained by the estimated PRS (Nagelkerke's R^2) was 5.5% (SE = 0.0012) (Supplementary Figure 8). The difference in standardized PRS between cases and controls was stable across iPSYCH subsets (OR = 1.56, 95% confidence interval [CI]: 1.53 – 1.60; Supplementary Figure 9) and across waves and PGC cohorts (Supplementary Figure 10). These results further support the highly polygenic architecture of ADHD and demonstrate that ADHD risk is significantly associated with PRS in a dose-dependent manner.

Polygenic Architecture of ADHD

To assess the proportion of phenotypic variance explained by common variants we applied LD score regression[37] to results from the European ancestry meta-analysis (Online Methods). Assuming a population prevalence of 5% for ADHD[39], we estimate that the liability-scale SNP heritability h^2_{snp} = 0.216 (SE = 0.014, P = 8.18×10^{-54}; Supplementary Table 4). These estimated polygenic effects account for 88% (SE = 0.0335) of observed genome-wide inflation of the test statistics in the meta-analysis (λ = 1.200; see Supplementary Figure 11 for quantile-quantile plots); the remaining inflation, which may reflect confounding factors such as cryptic relatedness and population stratification, is significant but modest (intercept=1.0362, SE = 0.0099, P=2.27×10^{-4}).

To further characterize the patterns of heritability from the genome-wide association data, we partitioned SNP heritability by functional annotations as described in Finucane et al.[40] using partitioned LD Score regression (Online Methods). The analysis revealed significant enrichment in the heritability from SNPs located in conserved regions (P = 8.49×10^{-10}; Supplementary Figure 12), supporting their biological importance. Enrichment of the SNP heritabilty in cell-type-specific regulatory elements was evaluated using the cell-type-specific group annotations described in Finucane et al[40]. We observed a significant enrichment of the average per SNP heritability for variants located in central nervous system specific regulatory elements (enrichment = 2.44, SE = 0.35, P = 5.81×10^{-5}; Supplementary Figures 13 and 14).

Genetic correlation with other traits

Pairwise genetic correlation with ADHD was estimated for 219 phenotypes using LD score regression[41,42] (Online Methods, Supplementary Data 1). Fourty-three phenotypes demonstrated

significant genetic overlap with ADHD ($P < 2.28 \times 10^{-4}$), including major depressive disorder[43], anorexia nervosa[44], educational outcomes[45-49], obesity-related phenotypes[50-55], smoking[56-58], reproductive success[59], insomnia[60], and mortality[61] (Figure 3; Supplementary Table 5). In most domains the genetic correlation is supported by GWAS of multiple related phenotypes. For the positive genetic correlation with major depressive disorder ($r_g = 0.42$, $P = 7.38 \times 10^{-38}$), we also observe a positive correlation with depressive symptoms ($r_g = 0.45$, $P = 7.00 \times 10^{-19}$), neuroticism ($r_g = 0.26$, $P = 1.02 \times 10^{-8}$) and a negative correlation with subjective well-being ($r_g = -0.28$, $P = 3.73 \times 10^{-9}$). The positive genetic correlations with ever smoked ($r_g = 0.48$, $P = 4.33 \times 10^{-16}$) and with number of cigarettes smoked ($r_g = 0.45$, $P = 1.07 \times 10^{-5}$) are reinforced by significant positive correlation with lung cancer ($r_g = 0.39$, $P = 6.35 \times 10^{-10}$). Similarly, genetic correlations related to obesity include significant relationships with body mass index (BMI; $r_g = 0.26$, $P = 1.68 \times 10^{-15}$), waist-to-hip ratio ($r_g = 0.30$, $P = 1.16 \times 10^{-17}$), childhood obesity ($r_g = 0.22$, $P = 3.29 \times 10^{-6}$), HDL cholesterol ($r_g = -0.22$, $P = 2.44 \times 10^{-7}$), and Type 2 Diabetes ($r_g = 0.18$, $P = 7.80 \times 10^{-5}$). Additionally the negative correlation with years of schooling ($r_g = -0.53$, $P = 6.02 \times 10^{-80}$) is supported by a negative genetic correlation with human intelligence ($r_g = -0.41$, $P = 7.03 \times 10^{-26}$). Finally the genetic correlation with reproduction include a negative correlation with age of first birth ($r_g = -0.612$, $P = 3.70 \times 10^{-61}$) and a positive correlation with number of children ever born ($r_g = 0.42$, $P = 8.51 \times 10^{-17}$).

Biological annotation of significant loci

For the 12 genome-wide significant loci, Bayesian credible sets were defined to identify the set of variants at each locus most likely to include a variant with causal effect (Online Methods, Supplementary Data 2; Supplementary Table 6). Biological annotations of the variants in the

credible set were then considered to identify functional or regulatory variants, common chromatin marks, and variants associated with gene expression (eQTLs) or in regions with gene interactions observed in Hi-C data (Online Methods, Supplementary Data 3). Broadly, the significant loci do not coincide with candidate genes proposed to play a role in ADHD[62].

Here we highlight genes that are identified in the regions of association (see also Supplementary Table 7). The loci on chromosomes 2, 7, and 10 each have credible sets localized to a single gene with limited additional annotations. In the chromosome 7 locus, *FOXP2* encodes a forkhead/winged-helix transcription factor and is known to play an important role in synapse formation and neural mechanisms mediating the development of speech and learning[63-65]. Comorbidity of ADHD with specific developmental disorders of language and learning is common (7 – 11%)[66,67], and poor language skills have been associated with higher inattention/hyperactivity symptoms in primary school[68]. On chromosome 10, the ADHD association is intronic, located in *SORCS3*, which encodes a brain-expressed transmembrane receptor that is important for neuronal development and plasticity[69] and has previously been associated with depression[43,70].

Genome-wide significant loci on chromosomes 12 and 15 have more biological annotations supporting the co-localized genes. The credible set on chromosome 12 spans *DUSP6*, and includes an annotated missense variant in the first exon and an insertion near the transcription start site, though neither is the lead variant in the locus (Supplementary Data 4). *DUSP6* encodes a dual specificity phosphatase[71], and may play a role in regulating neurotransmitter homeostasis by affecting dopamine levels in the synapses[72,73]. Regulation of dopamine levels is likely to be

relevant to ADHD since widely used ADHD medications have dopaminergic targets[74,75] that increase the availability of synaptic dopamine. The chromosome 15 locus is located in *SEMA6D*, and the majority of variants in the credible set are strongly associated with expression of *SEMA6D* in fibroblasts[76]. *SEMA6D* is active in the brain during embryonic development, and may play a role in neuronal wiring[77]. Furthermore, variants in *SEMA6D* have previously been associated with eduational attainment[78].

Credible set annotations at the remaining loci are more diverse (Supplementary Data 3). The most strongly associated locus on chromosome 1 (index variant rs112984125) covers a gene-rich 250kb region of strong LD. The index variant is intronic to *ST3GAL3*, and most SNPs in the credible set are strongly associated with expression of *ST3GAL3* in whole blood[79] (Supplementary Data 3). Missense mutations in *ST3GAL3* have been shown to cause autosomal recessive intellectual disability[80]. Hi-C and eQTL annotations suggest multiple alternative genes however, including *PTPRF* (Supplementary Data 4). The locus also includes an intergenic variant, rs11210892, that has previously been associated with schizophrenia[33].

On chromosome 5, the credible set includes links to *LINC00461* and *TMEM161B* (Supplementary Data 3). The function of *LINC00461* is unclear, but the RNA has highly localized expression in the brain[81] and the genome-wide significant locus overlaps with variants in *LINC00461* associated with educational attainment[78]. Alternatively, a genome-wide significant SNP in this locus (rs304132) is located in *MEF2C-AS1*, of strong interest given previous associations between *MEF2C* and severe intellectual disability,[82-84] cerebral malformation[83], depression[70], schizophrenia[33] and Alzheimer's disease[85], but the corresponding

variant is not supported by the credible set analysis. Credible set annotations for other significant loci are similarly cryptic.

Analysis of gene sets

Competitive gene based tests were performed for *FOXP2* target genes, highly constrained genes, and for all Gene Ontology terms[86] from MsigDB 6.0[87] using MAGMA[88] (Online Methods). Association results for individual genes are consistent with the genome-wide significant loci for the GWAS (Supplementary Table 8), however four new genes passed the threshold for exome-wide significant association (Supplementary Figure 15.A-D). Three independent sets of *FOXP2* downstream target genes[89,90] were tested (Online Methods), none of which demonstrated significant association to ADHD (Supplementary Table 9). The lack of association may be caused by unknown functions of *FOXP2* driving ADHD risk, insufficient power to detect relevant downstream genes, or because only a small subset of biological functions regulated by FOXP2 are relevant to ADHD pathogenesis.

Consistent with the partitioning of heritability, a set of 2,932 genes that are highly constrained and show high intolerance to loss of function[91] showed significant association with ADHD (β = 0.062, P = 2.6×10^{-4}; Supplementary Table 10). We also find little evidence for effects in previously proposed candidate genes for ADHD[62]; of the nine proposed genes only *SLC9A9* showed weak association with ADHD (P = 3.4×10^{-4}; Supplementary Table 11). None of the Gene Ontology gene sets were significant after correcting for multiple testing, although the most associated included interesting nominally significant pathways such as "dopamine receptor binding" (P = 0.0010) and "Excitatory Synapse" (P = 0.0088; Supplementary Data 5).

Replication of GWAS loci

For replication we evaluated the comparison of the GWAS meta-analysis of ADHD with three other independent ADHD-related GWASs: replication of top loci in an Icelandic cohort with ADHD status derived from medical records of ICD codes and medication history by deCODE (5,085 cases, 131,122 controls), a GWAS of self-reported ADHD status among 23andMe research participants (5,857 cases, 70,393 controls) and a meta-analysis of GWAS of childhood rating scales of ADHD symptoms performed by the EAGLE consortium (17,666 children < 13 years of age)[30] and QIMR[92] (2,798 adolescents), referred to as EAGLE/QIMR throughout the text. Although the phenotyping and cohort ascertainment of the 23andMe and EAGLE/QIMR studies differ from the PGC and iPSYCH ADHD meta-analysis (Supplemenatry Information), they have clear relevance to understanding how the ADHD GWAS results generalize to closely related phenotypes.

Top loci from the ADHD GWAS showed moderate concordance across the three replication studies. Sign concordance between each of the three replication cohorts and the ADHD GWAS was significantly greater than would be expected by chance (range 72–82% concordant; $P < 0.0167 = 0.05/3$ replication cohorts; Supplementary Table 12) for nominally associated loci from the ADHD GWAS ($P < 1 \times 10^{-6}$), with the highest concordance observed in EAGLE/QIMR. The deCODE and 23andMe results also permit direct comparisons of the magnitude of effect sizes for the top loci in the ADHD loci (Supplementary Table 13). Regressing effect size estimates from each replication cohort on estimates from the ADHD GWAS adjusted for winner's curse yields significantly positive slopes (deCODE slope = 0.664, $P = 1.2 \times 10^{-4}$; 23andMe slope =

0.417, P = 1.11×10^{-3}), although these slopes are less than one, suggesting imperfect replication. Among the genome-wide significant loci, rs9677504 (*SPAG16* locus) in deCODE and rs112984125 (*ST3GAL3/PTPRF* locus) and rs212178 (*LINC01572* locus) in 23andMe are noteable outlers with weak replication results (Online Methods, Supplementary Figure 16-17).

The genome-wide data available from 23andMe and EAGLE/QIMR showed similar trends for replication. The genetic correlation between EAGLE/QIMR and the ADHD GWAS was extremely strong (r_g = 0.970, SE = 0.207, P = 2.66×10^{-6}) and not significantly different from one (one-sided P = 0.442). Genetic correlation with 23andMe was weaker but still strongly positive (r_g = 0.653, SE = 0.114, P = 1.11×10^{-8}), although also significantly less than 1 (one-sided P= 1.17×10^{-3}). To explore this lower correlation we evaluated the genetic correlation between 23andMe and traits from LD Hub (see URLs)[42] to potentially identify differences in the profile of genetic correlation compared to the ADHD GWAS (Online Methods). This comparison identified striking differences (Supplementary Table 14), most notably that the 23andMe GWAS show little to no genetic correlation with college completion (r_g = 0.056, compared to r_g = -0.54 for the primary ADHD GWAS; approximate P = 1.1×10^{-9} for difference) and other education-related phenotypes. Genetic correlations with obesity-related phenotypes were similarly smaller for the 23andMe cohort. The one domain where 23andMe exhibited a trend toward stronger genetic correlations were schizophrenia (r_g = 0.27, vs. r_g = 0.12 in ADHD, P = 0.053) and bipolar disorder (r_g = 0.029, vs. r_g = 0.095 in ADHD, P = 0.09), though these trends are not significant with the approximated test of the difference in genetic correlation.

Finally, we meta-analyzed the ADHD GWAS with each replication cohort. For EAGLE/QIMR, we developed a novel model to meta-analyze the GWAS of the continuous measure of ADHD with the clinical diagnosis in the ADHD GWAS. In brief, we perform a Z-score based meta-analysis using a weighting scheme derived from the SNP heritability and effective sample size for each phenotype that fully accounts for the differences in measurement scale (detailed description in Supplementary Information, and Supplementary Figures 24-26). This calibration based on the genome-wide estimate of heritability prevents joint meta-analysis of all replication cohorts since genome-wide data is not available for the deCODE study.

Meta-analyses of the ADHD GWAS with each replication study identified 10 genome-wide significant loci ($P < 5 \times 10^{-8}$, without multiple testing correction) in meta-analysis with deCODE, 10 significant loci with 23andMe, and 15 significant loci with EAGLE/QIMR (Supplementary Data 6, Supplementary Figures 18 and 19). Of the 12 significant loci from the primary ADHD GWAS, four were significant in all three of these replication meta-analyses: index variants rs11420276 (*ST3GAL3/PTPRF*), rs5886709 (*FOXP2*), rs11591402 (*SORCS3*), and rs1427829 (intergenic). The remaining loci were all significant in at least one of the replication meta-analyses. In addition, ten novel loci reached genome-wide significance in the replication meta-analyses, of which three loci were significant in two of these analyses (Supplementary Data 6): index variants rs1592757 / rs30266 (Refseq *LOC105379109*), rs28452470 / rs1443749 (*CADPS2*), and rs2243638 / rs9574218 (*RNF219-AS1*). The *CADPS2* locus has recently been identified in autism spectrum disorder as a novel locus shared with educational attainment[93].

Meta-analysis with the 23andMe cohort also found genome-wide significant heterogeneity at the lead Chromosome 1 locus from the ADHD GWAS meta-analysis (rs12410155: $I^2 = 97.2$, $P = 2.29 \times 10^{-9}$; Supplementary Figures 20-21). This heterogeneity is consistent with the moderate sign concordance, effect size replication, and genetic correlation of the 23andMe cohort with the ADHD GWAS. Notably, the lead chromosome 1 locus in the ADHD GWAS overlaps a reported association with educational attainment[78], suggesting this heterogeneity is consistent with the much weaker genetic correlation between the 23andMe results and published GWAS of education-related outcomes. No genome-wide significant heterogeneity was observed in the replication meta-analyses with deCODE or EAGLE/QIMR (Supplementary Figures 22-23, Supplementary Data 6).

Discussion

GWAS meta-analysis of ADHD revealed the first genome-wide significant risk loci, and indicates an important role for common variants in the polygenic architecture of ADHD. Several of the loci are located in or near genes that implicate neurodevelopmental processes that are likely to be relevant to ADHD, including *FOXP2*, *SORCS3*, and *DUSP6*. Future work may focus on refining the source of the strong association in each locus, especially the lead locus on chromosome 1 which is complicated by broad LD and substantial heterogeneity between ADHD the main meta-analysis and analysis of self-reported ADHD status in 23andMe.

The 12 significant loci are compelling, but only capture a tiny fraction of common variant risk for ADHD. The odds ratios for the risk increasing allele at the index SNPs in the 12 significant loci are modest, ranging from 1.077 to 1.198 (Table 1). This is within the range of effect sizes

for common genetic variants that has been observed for other highly polygenic psychiatric disorders e.g. schizophrenia[33]. A considerably larger proportion of the heritability of ADHD can be explained by all common variants (h^2_{snp} = 0.22, SE = 0.01). This is consistent with previous estimates of h^2_{snp} for ADHD in smaller studies (h^2_{snp}: 0.1 - 0.28)[23,24], and also comparable to SNP heritability estimates for schizophrenia (h^2_{snp} 0.23 - 0.26)[23,24]. As would be hypothesized for a psychiatric disorder, these effects are enriched in conserved regions and regions containing enhancers and promoters of expression in central nervous system tissues, consistent with previous observations in schizophrenia and bipolar disorder[40]. On the other hand, we do not observe substantial effects in most previously reported candidate genes for ADHD[62].

Along with polygenicity, selection and evolutionary pressures may be an important feature of the architecture of ADHD genetics. We observe that ADHD risk variants are strongly enriched in genomic regions conserved in mammals[94], and constrained genes likely to be intolerant of loss-of-function mutations[91] are associated with ADHD. We also find that common variant risk for ADHD is genetically correlated with having children younger and having more children, in line with epidemiological findings of increased risky sexual behaviour[95-97] and increased risk of ADHD for children born to young parents[98-100]. Given the phenotypic[101,102] and genetic[103] correlation of ADHD with reduced educational attainment, positive selective pressure on the genetics of ADHD would be consistent with recent work suggesting that variants associated with educational attainment are under negative selection in Iceland[104]. Future studies of fecundity and the role of rare and *de novo* variants in ADHD may provide more insight on selective pressures in ADHD-associated loci.

The observed genetic correlations with educational outcomes and other phenotypes suggest a strong genetic component to the epidemiological correlates of ADHD. The significant positive genetic correlation of ADHD with major depressive disorder and depressive symptoms supports previous findings suggesting a positive genetic overlap between those phenotypes[24,42], as well as the broader genetic overlap of psychiatric disorders[23,24]. Positive genetic correlations between ADHD and health risk behaviors such as smoking and obesity are consistent with the observed increase in those behaviors among individuals with ADHD[105-108] and are indicative of a shared genetic basis for these traits. We also observe a positive genetic correlation of ADHD with insomnia, consistent with reports of sleep disturbances in ADHD[109], but this relationship does not appear to generalize to other sleep-related phenotypes.

These genetic correlations may not generalize to all settings. We observe much weaker genetic correlation of the 23andMe ADHD results with educational attainment, with only partial genetic correlation between 23andMe and the current ADHD GWAS, including significant heterogeneity in the lead chromosome 1 locus. The pattern of replication for the top loci in the deCODE study is stronger but still mixed. These differences may reflect dissimilarities in phenotyping (e.g. self-report vs. medical records), exclusion of individuals with comorbid psychiatric disorders (deCODE), study population (e.g. higher average education and socio-economic status among 23andMe research participants possibly under-representing the proportion of individuals with ADHD with poor educational outcomes in the general population), or other study factors that should be a focus of future work.

On the other hand, the replication results from EAGLE[30]/QIMR[92] are much stronger and support the hypothesis that ADHD is the extreme expression of one or more heritable quantitative traits[110]. We observe strong concordance between the GWAS of ADHD and the previous GWASs of ADHD-related traits in the population, both in terms of genome-wide genetic correlation and concordance at individual loci. Polygenic risk for ADHD has previously been associated with inattentive and hyperactive/impulsive trait variation below clinical thresholds in the population[29]. Shared genetic risk with health risk behaviors may similarly be hypothesized to reflect an impaired ability to self-regulate and inhibit impulsive behavior[111,112]. The observed negative correlation between ADHD and anorexia nervosa may also be related to these behavioral factors.

In summary, we report 12 independent genome-wide significant loci associated with ADHD in GWAS meta-analysis of 55,374 individuals from 12 study cohorts. The GWAS meta-analysis implicates *FOXP2* and other biologically informative genes as well as constrained regions of the genome as important contributors to the etiology of ADHD. The results also highlight strong overlap with the genetics of ADHD-related traits and health risk behaviors in the population, encouraging a dimensional view of ADHD as the extreme end of a continuum of symptoms.

URLs

LD-Hub: http://ldsc.broadinstitute.org/ldhub/

LD score regression: https://github.com/bulik/ldsc

Pre-computed European LD scores: https://data.broadinstitute.org/alkesgroup/LDSCORE/

PGC Ricopili GWA pipeline: https://github.com/Nealelab/ricopili

Credible set analysis: https://github.com/hailianghuang/FM-summary

FUMA: http://fuma.ctglab.nl

Supplementary Information is linked to the online version of the paper at www.nature.com/nature

Acknowledgements

The iPSYCH team acknowledges funding from the Lundbeck Foundation (grant no R102-A9118 and R155-2014-1724), the Stanley Medical Research Institute, the European Research Council (project no: 294838), the European Community's Horizon 2020 Programme (H2020/2014-2020) under Grant No. 667302 (CoCA), the Novo Nordisk Foundation for supporting the Danish National Biobank resource, and grants from Aarhus and Copenhagen Universities and University Hospitals, including support to the iSEQ Center, the GenomeDK HPC facility, and the CIRRAU Center.
The Broad Institute and Massachusetts General Hospital investigators would like to acknowledge support from the Stanley Medical Research Institute and NIH grants: 5U01MH094432-04(PI: Daly), 1R01MH094469 (PI: Neale), 1R01MH107649-01 (PI: Neale), 1R01MH109539-01 (PI: Daly). We thank T., Lehner, A. Addington and G. Senthil for their support in the Psychiatric Genomics Consortium.
Dr. J. Martin was supported by the Wellcome Trust (Grant No: 106047).
Dr. Faraone is supported by the K.G. Jebsen Centre for Research on Neuropsychiatric Disorders, University of Bergen, Bergen, Norway, the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602805, the European Union's Horizon 2020 research and innovation programme under grant agreement No 667302 and NIMH grants 5R01MH101519 and U01 MH109536-01.
Dr. Haavik is supported by grants from Stiftelsen K.G. Jebsen, University of Bergen and The Research Council of Norway.
Dr. Cormand received financial support for this research from the Spanish 'Ministerio de Economía y Competitividad' (SAF2015-68341-R) and 'Generalitat de Catalunya/AGAUR' (2014SGR932). Dr. Cormand, Dr. Reif and collaborators received funding from the European Community's Seventh Framework Programme (under grant agreement number 602805, Aggressotype), the European Community's H2020 Programme (under grant agreements number 667302, CoCA, and 402003, MiND), the ECNP network 'ADHD across the lifespan' and DFG CRC 1193, subproject Z03.
Dr. Andreassen is supported by the Research Council of Norway (grant nos: 223273, 248778, 213694, 249711), and KG Jebsen Stiftelsen.
Philip Asherson is supported by a National Institute of Health Research Senior Investigator Award (NF-SI-0616-10040), and the Biomedical Research Centre for Mental Health
Dr. Kuntsi's research on ADHD is supported by the European Commission (grant agreements no. 643051 MiND, 667302 CoCA and 602805 Aggressotype); Action Medical Research (GN2080 and GN2315); 4 Medical Research Council and SGDP Centre PhD studentships; and by the ECNP Network ADHD Across the Lifespan.
Dr. Langley was funded by Wellcome Trust (Grant No: 079711)
Dr. Thapar received ADHD funding from the Wellcome Trust, Medical Research Council (MRC UK), Action Medical Research.
Barbara Franke's research is supported by funding from a personal Vici grant of the Netherlands Organisation for Scientific Research (NWO; grant 016-130-669, to BF), from the European Community's

Seventh Framework Programme (FP7/2007 – 2013) under grant agreements n° 602805 (Aggressotype), n° 602450 (IMAGEMEND), and n° 278948 (TACTICS), and from the European Community's Horizon 2020 Programme (H2020/2014 – 2020) under grant agreements n° 643051 (MiND) and n° 667302 (CoCA). In addition, this work was supported by the European College of Neuropsychopharmacology (ECNP Network "ADHD across the Lifespan").
Dr. Schachar received support from Bank Chair in Child Psychiatry, Canadian Institutes of Health Research (MOP-106573 and MOP – 93696).
Dr. Roussos was supported by the National Institutes of Health (R01AG050986 Roussos and R01MH109677), Brain Behavior Research Foundation (20540), Alzheimer's Association (NIRG-340998) and the Veterans Affairs (Merit grant BX002395).
We thank the customers of 23andMe who answered surveys, as well as the employees of 23andMe, who together made this research possible.
The QIMR studies were supported by funding from the Australian National Health and Medical Research Council (grant numbers: 241944, 339462, 389927, 389875, 389891, 389892, 389938, 443036, 442915, 442981, 496739, 552485, and 552498, and most recently 1049894) and the Australian Research Council (grant numbers: A7960034, A79906588, A79801419, DP0212016, and DP0343921). SEM is supported by an NHMRC fellowship (1103623).
Additional acknowledegements can be found in the Supplementary Note

Author Contributions

Analysis:
DD, RKW, JMar, MM, TDA, CC, NE, MG, KLG, MEH, DPH, HHai, JMal, ARM, JP, DSP, TP, SR, EBR, FKS, HS, SS, PT, GBW, HW, DB, DG, CM, PR, PFS, JT, SEM, KS. ADB and BMN supervised and coordinated analyses.

Sample and/or data provider and processing:
DD, RKW, JMar, MM, EA, GB, RB, JB-G, MB-H, FC, KC, AD, NE, JGo, JGr, OOG, CSH, MVH, JMal, NGM, JMo, CBP, MGP, JBP, SR, CS, MJW, OA, PA, CB, DB, BC, SD, BF, JGe, HHak, JH, HK, JK, KL, KPL, CM, AR, LAR, RS, PS, ESB, AT, JT, IW, SEM, DMH, OM, PBM, ADB, ADHD Working Group of the Psychiatric Genomics Consortium, Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium, 23andMe Research Team.

Core PI group:
SEM, KS, MN, DMH, TW, OM, PBM, MJD, SVF, ADB, BMN.

Core writing group:
DD, RKW, JM, SVF, ADB, BMN.

Direction of study: ADB, SVF, BMN.

References

1 Faraone, S. V. *et al.* Attention-deficit/hyperactivity disorder. *Nature Reviews Disease Primers*, 15020, doi:10.1038/nrdp.2015.20 (2015).

2 Dalsgaard, S., Leckman, J. F., Mortensen, P. B., Nielsen, H. S. & Simonsen, M. Effect of drugs on the risk of injuries in children with attention deficit hyperactivity disorder: a prospective cohort study. *Lancet Psychiatry* **2**, 702-709, doi:10.1016/S2215-0366(15)00271-0 (2015).

3 Chang, Z., Lichtenstein, P., D'Onofrio, B. M., Sjolander, A. & Larsson, H. Serious transport accidents in adults with attention-deficit/hyperactivity disorder and the effect of medication: a population-based study. *JAMA Psychiatry* **71**, 319-325, doi:10.1001/jamapsychiatry.2013.4174 (2014).

4 Biederman, J. & Faraone, S. V. Attention-deficit hyperactivity disorder. *Lancet* **366**, 237-248, doi:10.1016/S0140-6736(05)66915-2 (2005).

5 Dalsgaard, S., Nielsen, H. S. & Simonsen, M. Consequences of ADHD medication use for children's outcomes. *J Health Econ* **37**, 137-151, doi:10.1016/j.jhealeco.2014.05.005 (2014).

6 Dalsgaard, S., Mortensen, P. B., Frydenberg, M. & Thomsen, P. H. ADHD, stimulant treatment in childhood and subsequent substance abuse in adulthood - a naturalistic long-term follow-up study. *Addict Behav* **39**, 325-328, doi:10.1016/j.addbeh.2013.09.002 (2014).

7 Lichtenstein, P. & Larsson, H. Medication for attention deficit-hyperactivity disorder and criminality. *N Engl J Med* **368**, 776, doi:10.1056/NEJMc1215531 (2013).

8 Barkley, R. A., Murphy, K. R. & Fischer, M. *ADHD in adults What the Science Says*. (The Guilford Press, 2007).

9 Furczyk, K. & Thome, J. Adult ADHD and suicide. *Atten Defic Hyperact Disord* **6**, 153-158, doi:10.1007/s12402-014-0150-1 (2014).

10 Dalsgaard, S., Ostergaard, S. D., Leckman, J. F., Mortensen, P. B. & Pedersen, M. G. Mortality in children, adolescents, and adults with attention deficit hyperactivity disorder: a nationwide cohort study. *Lancet* **385**, 2190-2196, doi:10.1016/S0140-6736(14)61684-6 (2015).

11 Franke, B. *et al.* The genetics of attention deficit/hyperactivity disorder in adults, a review. *Molecular psychiatry* **17**, 960-987, doi:10.1038/mp.2011.138 (2012).

12 Faraone, S. V. *et al.* Molecular genetics of attention-deficit/hyperactivity disorder. *Biological psychiatry* **57**, 1313-1323, doi:10.1016/j.biopsych.2004.11.024 (2005).

13 Burt, S. A. Rethinking environmental contributions to child and adolescent psychopathology: a meta-analysis of shared environmental influences. *Psychological bulletin* **135**, 608-637, doi:10.1037/a0015702 (2009).

14 Larsson, H., Anckarsater, H., Rastam, M., Chang, Z. & Lichtenstein, P. Childhood attention-deficit hyperactivity disorder as an extreme of a continuous trait: a quantitative genetic study of 8,500 twin pairs. *J Child Psychol Psychiatry* **53**, 73-80, doi:10.1111/j.1469-7610.2011.02467.x (2012).

15 Christiansen, H. *et al.* Co-transmission of conduct problems with attention-deficit/hyperactivity disorder: familial evidence for a distinct disorder. *J Neural Transm (Vienna)* **115**, 163-175, doi:10.1007/s00702-007-0837-y (2008).

16 Kuntsi, J. *et al.* The separation of ADHD inattention and hyperactivity-impulsivity symptoms: pathways from genetic effects to cognitive impairments and symptoms. *Journal of abnormal child psychology* **42**, 127-136, doi:10.1007/s10802-013-9771-7 (2014).

17 Rommelse, N. N., Franke, B., Geurts, H. M., Hartman, C. A. & Buitelaar, J. K. Shared heritability of attention-deficit/hyperactivity disorder and autism spectrum disorder. *European child & adolescent psychiatry* **19**, 281-295, doi:10.1007/s00787-010-0092-x (2010).

18 Ghirardi, L. *et al.* The familial co-aggregation of ASD and ADHD: a register-based cohort study. *Molecular psychiatry*, doi:10.1038/mp.2017.17 (2017).

19 Larsson, H. *et al.* Risk of bipolar disorder and schizophrenia in relatives of people with attention-deficit hyperactivity disorder. *The British journal of psychiatry : the journal of mental science* **203**, 103-106, doi:10.1192/bjp.bp.112.120808 (2013).
20 Faraone, S. V., Biederman, J. & Wozniak, J. Examining the comorbidity between attention deficit hyperactivity disorder and bipolar I disorder: a meta-analysis of family genetic studies. *The American journal of psychiatry* **169**, 1256-1266, doi:10.1176/appi.ajp.2012.12010087 (2012).
21 Faraone, S. V. & Biederman, J. Do attention deficit hyperactivity disorder and major depression share familial risk factors? *The Journal of nervous and mental disease* **185**, 533-541 (1997).
22 Neale, B. M. *et al.* Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry* **49**, 884-897, doi:10.1016/j.jaac.2010.06.008 (2010).
23 Anttila, V. *et al.* Analysis of shared heritability in common disorders of the brain. *BioRxiv*, 1-22 (2016).
24 Cross-Disorder Group of the Psychiatric Genomics, C. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature genetics* **45**, 984-994, doi:10.1038/ng.2711 (2013).
25 Cross-Disorder Group of the Psychiatric Genomics, C. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371-1379, doi:10.1016/S0140-6736(12)62129-1 (2013).
26 Hamshere, M. L. *et al.* High loading of polygenic risk for ADHD in children with comorbid aggression. *The American journal of psychiatry* **170**, 909-916, doi:10.1176/appi.ajp.2013.12081129 (2013).
27 Hamshere, M. L. *et al.* Shared polygenic contribution between childhood attention-deficit hyperactivity disorder and adult schizophrenia. *The British journal of psychiatry : the journal of mental science* **203**, 107-111, doi:10.1192/bjp.bp.112.117432 (2013).
28 Groen-Blokhuis, M. M. *et al.* Attention-deficit/hyperactivity disorder polygenic risk scores predict attention problems in a population-based sample of children. *Journal of the American Academy of Child and Adolescent Psychiatry* **53**, 1123-1129 e1126, doi:10.1016/j.jaac.2014.06.014 (2014).
29 Martin, J., Hamshere, M. L., Stergiakouli, E., O'Donovan, M. C. & Thapar, A. Genetic risk for attention-deficit/hyperactivity disorder contributes to neurodevelopmental traits in the general population. *Biological psychiatry* **76**, 664-671, doi:10.1016/j.biopsych.2014.02.013 (2014).
30 Middeldorp, C. M. *et al.* A Genome-Wide Association Meta-Analysis of Attention-Deficit/Hyperactivity Disorder Symptoms in Population-Based Pediatric Cohorts. *Journal of the American Academy of Child and Adolescent Psychiatry* **55**, 896-905 e896, doi:10.1016/j.jaac.2016.05.025 (2016).
31 Yang, L. *et al.* Polygenic transmission and complex neuro developmental network for attention deficit hyperactivity disorder: genome-wide association study of both common and rare variants. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **162B**, 419-430, doi:10.1002/ajmg.b.32169 (2013).
32 Zayats, T. *et al.* Genome-wide analysis of attention deficit hyperactivity disorder in Norway. *PloS one* **10**, e0122501, doi:10.1371/journal.pone.0122501 (2015).
33 Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427, doi:10.1038/nature13595 (2014).
34 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).

35 Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38**, 904-909, doi:10.1038/ng1847 (2006).
36 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).
37 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics* **47**, 291-295, doi:10.1038/ng.3211 (2015).
38 Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-752, doi:0.1038/nature08185 (2009).
39 Polanczyk, G., de Lima, M. S., Horta, B. L., Biederman, J. & Rohde, L. A. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *The American journal of psychiatry* **164**, 942-948, doi:10.1176/ajp.2007.164.6.942 (2007).
40 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics* **47**, 1228-1235, doi:10.1038/ng.3404 (2015).
41 Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279, doi:10.1093/bioinformatics/btw613 (2017).
42 Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nature genetics* **47**, 1236-1241, doi:10.1038/ng.3406 (2015).
43 Wray, N. R. & Sullivan, P. F. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *bioRxiv*, doi:10.1101/167577 (2017).
44 Duncan, L. *et al.* Significant Locus and Metabolic Genetic Correlations Revealed in Genome-Wide Association Study of Anorexia Nervosa. *The American journal of psychiatry* **174**, 850-858, doi:10.1176/appi.ajp.2017.16121402 (2017).
45 Benyamin, B. *et al.* Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Molecular psychiatry* **19**, 253-258, doi:10.1038/mp.2012.184 (2014).
46 Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature genetics* **48**, 624-633, doi:10.1038/ng.3552 (2016).
47 Rietveld, C. A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467-1471, doi:10.1126/science.1235488 (2013).
48 Rietveld, C. A. *et al.* Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 13790-13794, doi:10.1073/pnas.1404623111 (2014).
49 Davies, G. *et al.* Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). *Molecular psychiatry* **21**, 758-767, doi:10.1038/mp.2016.45 (2016).
50 Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707-713, doi:10.1038/nature09270 (2010).
51 Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature genetics* **44**, 981-990, doi:10.1038/ng.2383 (2012).
52 Bradfield, J. P. *et al.* A genome-wide association meta-analysis identifies new childhood obesity loci. *Nature genetics* **44**, 526-531, doi:10.1038/ng.2247 (2012).
53 Berndt, S. I. *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nature genetics* **45**, 501-512, doi:10.1038/ng.2606 (2013).
54 Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics* **42**, 937-948, doi:10.1038/ng.686 (2010).

55 Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187-196, doi:10.1038/nature14132 (2015).
56 Tobacco & Genetics, C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature genetics* **42**, 441-447, doi:10.1038/ng.571 (2010).
57 Patel, Y. M. *et al.* Novel Association of Genetic Markers Affecting CYP2A6 Activity and Lung Cancer Risk. *Cancer Res* **76**, 5768-5776, doi:10.1158/0008-5472.CAN-16-0446 (2016).
58 Wang, Y. *et al.* Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nature genetics* **46**, 736-741, doi:10.1038/ng.3002 (2014).
59 Barban, N. *et al.* Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nature genetics* **48**, 1462-1472, doi:10.1038/ng.3698 (2016).
60 Hammerschlag, A. R. *et al.* Genome-wide association analysis of insomnia complaints identifies risk genes and genetic overlap with psychiatric and metabolic traits. *Nature genetics* **49**, 1584-1592, doi:10.1038/ng.3888 (2017).
61 Pilling, L. C. *et al.* Human longevity is influenced by many genetic variants: evidence from 75,000 UK Biobank participants. *Aging (Albany NY)* **8**, 547-560, doi:10.18632/aging.100930 (2016).
62 Hawi, Z. *et al.* The molecular genetic architecture of attention deficit hyperactivity disorder. *Molecular psychiatry* **20**, 289-297, doi:10.1038/mp.2014.183 (2015).
63 Sia, G. M., Clem, R. L. & Huganir, R. L. The human language-associated gene SRPX2 regulates synapse formation and vocalization in mice. *Science* **342**, 987-991, doi:10.1126/science.1245079 (2013).
64 Tsui, D., Vessey, J. P., Tomita, H., Kaplan, D. R. & Miller, F. D. FoxP2 regulates neurogenesis during embryonic cortical development. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **33**, 244-258, doi:10.1523/JNEUROSCI.1665-12.2013 (2013).
65 Schreiweis, C. *et al.* Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 14253-14258, doi:10.1073/pnas.1414542111 (2014).
66 Jensen, C. M. & Steinhausen, H. C. Comorbid mental disorders in children and adolescents with attention-deficit/hyperactivity disorder in a large nationwide study. *Atten Defic Hyperact Disord* **7**, 27-38, doi:10.1007/s12402-014-0142-1 (2015).
67 Larson, K., Russ, S. A., Kahn, R. S. & Halfon, N. Patterns of comorbidity, functioning, and service use for US children with ADHD, 2007. *Pediatrics* **127**, 462-470, doi:10.1542/peds.2010-0165 (2011).
68 Peyre, H. *et al.* Relationship between early language skills and the development of inattention/hyperactivity symptoms during the preschool period: Results of the EDEN mother-child cohort. *BMC psychiatry* **16**, 380, doi:10.1186/s12888-016-1091-3 (2016).
69 Breiderhoff, T. *et al.* Sortilin-related receptor SORCS3 is a postsynaptic modulator of synaptic depression and fear extinction. *PloS one* **8**, e75006, doi:10.1371/journal.pone.0075006 (2013).
70 Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature genetics* **48**, 1031-1036, doi:10.1038/ng.3623 (2016).
71 Caunt, C. J. & Keyse, S. M. Dual-specificity MAP kinase phosphatases (MKPs): shaping the outcome of MAP kinase signalling. *FEBS J* **280**, 489-504, doi:10.1111/j.1742-4658.2012.08716.x (2013).
72 Mortensen, O. V. MKP3 eliminates depolarization-dependent neurotransmitter release through downregulation of L-type calcium channel Cav1.2 expression. *Cell Calcium* **53**, 224-230, doi:10.1016/j.ceca.2012.12.004 (2013).
73 Mortensen, O. V., Larsen, M. B., Prasad, B. M. & Amara, S. G. Genetic complementation screen identifies a mitogen-activated protein kinase phosphatase, MKP3, as a regulator of dopamine transporter trafficking. *Mol Biol Cell* **19**, 2818-2829, doi:10.1091/mbc.E07-09-0980 (2008).

74 Volkow, N. D., Fowler, J. S., Wang, G., Ding, Y. & Gatley, S. J. Mechanism of action of methylphenidate: insights from PET imaging studies. *J Atten Disord* **6 Suppl 1**, S31-43 (2002).
75 Volkow, N. D. *et al.* Therapeutic doses of oral methylphenidate significantly increase extracellular dopamine in the human brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **21**, RC121 (2001).
76 Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nature genetics* **45**, 580-585, doi:10.1038/ng.2653 (2013).
77 Qu, X. *et al.* Identification, characterization, and functional study of the two novel human members of the semaphorin gene family. *The Journal of biological chemistry* **277**, 35574-35585, doi:10.1074/jbc.M206451200 (2002).
78 Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539-542, doi:10.1038/nature17671 (2016).
79 Zhernakova, D. V. *et al.* Identification of context-dependent expression quantitative trait loci in whole blood. *Nature genetics* **49**, 139-145, doi:10.1038/ng.3737 (2017).
80 Hu, H. *et al.* ST3GAL3 mutations impair the development of higher cognitive functions. *American journal of human genetics* **89**, 407-414, doi:10.1016/j.ajhg.2011.08.008 (2011).
81 Oliver, P. L. *et al.* Disruption of Visc-2, a Brain-Expressed Conserved Long Noncoding RNA, Does Not Elicit an Overt Anatomical or Behavioral Phenotype. *Cereb Cortex* **25**, 3572-3585, doi:10.1093/cercor/bhu196 (2015).
82 Sobreira, N., Walsh, M. F., Batista, D. & Wang, T. Interstitial deletion 5q14.3-q21 associated with iris coloboma, hearing loss, dental anomaly, moderate intellectual disability, and attention deficit and hyperactivity disorder. *Am J Med Genet A* **149A**, 2581-2583, doi:10.1002/ajmg.a.33079 (2009).
83 Le Meur, N. *et al.* MEF2C haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. *J Med Genet* **47**, 22-29, doi:10.1136/jmg.2009.069732 (2010).
84 Novara, F. *et al.* Refining the phenotype associated with MEF2C haploinsufficiency. *Clin Genet* **78**, 471-477, doi:10.1111/j.1399-0004.2010.01413.x (2010).
85 Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics* **45**, 1452-1458, doi:10.1038/ng.2802 (2013).
86 Gene Ontology, C. Gene Ontology Consortium: going forward. *Nucleic acids research* **43**, D1049-1056, doi:10.1093/nar/gku1179 (2015).
87 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
88 de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219, doi:10.1371/journal.pcbi.1004219 (2015).
89 Vernes, S. C. *et al.* Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS genetics* **7**, e1002145, doi:10.1371/journal.pgen.1002145 (2011).
90 Spiteri, E. *et al.* Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. *American journal of human genetics* **81**, 1144-1157, doi:10.1086/522237 (2007).
91 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).
92 Ebejer, J. L. *et al.* Genome-wide association study of inattention and hyperactivity-impulsivity measured as quantitative traits. *Twin Res Hum Genet* **16**, 560-574, doi:10.1017/thg.2013.12 (2013).

93 Grove, J. *et al.* Common risk variants identified in autism spectrum disorder. *bioRxiv*, doi:10.1101/224774 (2017).
94 Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476-482, doi:10.1038/nature10530 (2011).
95 Flory, K., Molina, B. S., Pelham, W. E., Jr., Gnagy, E. & Smith, B. Childhood ADHD predicts risky sexual behavior in young adulthood. *J Clin Child Adolesc Psychol* **35**, 571-577, doi:10.1207/s15374424jccp3504_8 (2006).
96 Marsh, L. E., Norvilitis, J. M., Ingersoll, T. S. & Li, B. ADHD symptomatology, fear of intimacy, and sexual anxiety and behavior among college students in China and the United States. *J Atten Disord* **19**, 211-221, doi:10.1177/1087054712453483 (2015).
97 Hosain, G. M., Berenson, A. B., Tennen, H., Bauer, L. O. & Wu, Z. H. Attention deficit hyperactivity symptoms and risky sexual behavior in young adult women. *J Womens Health (Larchmt)* **21**, 463-468, doi:10.1089/jwh.2011.2825 (2012).
98 Chudal, R. *et al.* Parental age and the risk of attention-deficit/hyperactivity disorder: a nationwide, population-based cohort study. *Journal of the American Academy of Child and Adolescent Psychiatry* **54**, 487-494 e481, doi:10.1016/j.jaac.2015.03.013 (2015).
99 Chang, Z. *et al.* Maternal age at childbirth and risk for ADHD in offspring: a population-based cohort study. *Int J Epidemiol* **43**, 1815-1824, doi:10.1093/ije/dyu204 (2014).
100 Ostergaard, S. D., Dalsgaard, S., Faraone, S. V., Munk-Olsen, T. & Laursen, T. M. Teenage Parenthood and Birth Rates for Individuals With and Without Attention-Deficit/Hyperactivity Disorder: A Nationwide Cohort Study. *Journal of the American Academy of Child and Adolescent Psychiatry* **56**, 578-584 e573, doi:10.1016/j.jaac.2017.05.003 (2017).
101 Barbaresi, W. J., Katusic, S. K., Colligan, R. C., Weaver, A. L. & Jacobsen, S. J. Long-term school outcomes for children with attention-deficit/hyperactivity disorder: a population-based perspective. *J Dev Behav Pediatr* **28**, 265-273, doi:10.1097/DBP.0b013e31811ff87d (2007).
102 Faraone, S. V. *et al.* Intellectual performance and school failure in children with attention deficit hyperactivity disorder and in their siblings. *Journal of abnormal psychology* **102**, 616-623 (1993).
103 Sniekers, S. *et al.* Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature genetics* **49**, 1107-1112, doi:10.1038/ng.3869 (2017).
104 Kong, A. *et al.* Selection against variants in the genome associated with educational attainment. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E727-E732, doi:10.1073/pnas.1612113114 (2017).
105 Lee, S. S., Humphreys, K. L., Flory, K., Liu, R. & Glass, K. Prospective association of childhood attention-deficit/hyperactivity disorder (ADHD) and substance use and abuse/dependence: a meta-analytic review. *Clin Psychol Rev* **31**, 328-341, doi:10.1016/j.cpr.2011.01.006 (2011).
106 Halfon, N., Larson, K. & Slusser, W. Associations between obesity and comorbid mental health, developmental, and physical health conditions in a nationally representative sample of US children aged 10 to 17. *Acad Pediatr* **13**, 6-13, doi:10.1016/j.acap.2012.10.007 (2013).
107 Chen, A. Y., Kim, S. E., Houtrow, A. J. & Newacheck, P. W. Prevalence of obesity among children with chronic conditions. *Obesity (Silver Spring)* **18**, 210-213, doi:10.1038/oby.2009.185 (2010).
108 Cortese, S. *et al.* Association Between ADHD and Obesity: A Systematic Review and Meta-Analysis. *The American journal of psychiatry* **173**, 34-43, doi:10.1176/appi.ajp.2015.15020266 (2016).
109 Owens, J. A. A clinical overview of sleep and attention-deficit/hyperactivity disorder in children and adolescents. *J Can Acad Child Adolesc Psychiatry* **18**, 92-102 (2009).
110 Lubke, G. H., Hudziak, J. J., Derks, E. M., van Bijsterveldt, T. C. & Boomsma, D. I. Maternal ratings of attention problems in ADHD: evidence for the existence of a continuum. *Journal of the*

American Academy of Child and Adolescent Psychiatry **48**, 1085-1093, doi:10.1097/CHI.0b013e3181ba3dbb (2009).

111 Cortese, S., Comencini, E., Vincenzi, B., Speranza, M. & Angriman, M. Attention-deficit/hyperactivity disorder and impairment in executive functions: a barrier to weight loss in individuals with obesity? *BMC psychiatry* **13**, 286, doi:10.1186/1471-244X-13-286 (2013).

112 Ortal, S. *et al.* The Role of Different Aspects of Impulsivity as Independent Risk Factors for Substance Use Disorders in Patients with ADHD: A Review. *Curr Drug Abuse Rev* **8**, 119-133 (2015).

Figure legends

Figure 1. Manhattan plot of the results from the GWAS meta-analysis of ADHD.

The index variants in the 12 genome-wide significant loci are highlighted as a green diamond. Index variants located with a distance less than 400kb are considered as one locus. The y-axis represents –log(two-sided P-values) for association of variants with ADHD, from meta-analysis using an inverse-variance weighted fixed effects model, and a total sample size of 20,183 ADHD cases and 35,191 controls. The vertical red line represents the threshold for genome-wide significance.

Figure 2. Odds Ratio by PRS for ADHD

Odds Ratio (OR) by PRS within each decile estimated for n = 18,298 biological independent individuals in the PGC samples (red dots) and in n = 37,076 biological independent individuals in the iPSYCH sample (blue dots). PRSs in the iPSYCH sample were obtained by five leave-one-out analyses, using 4 of 5 groups as training datasets for estimation of SNP weights, while estimating Polygenic Risk Scores (PRS) for the remaining target group. Odds ratios and 95% confidence limits (error bars) were estimated using logistic regression on the continuous scores.

Figure 3. Genetic correlations of ADHD with other phenotypes

Significant genetic correlations between ADHD (results from Europena GWAS meta-analysis of 19,099 cases, 34,194 controls) and other traits reveal overlap of genetic risk factors for ADHD across several groups of traits (grouping indicated by a horizontal line): educational, psychiatric/personality, weight (and possible weight related traits), smoking behaviour/smoking-related cancer, reproductive traits and parental longevity (Sample size of the external GWASs are presented in Supplementary Table 5) . In total 219 traits were tested and only traits significant after Bonferroni correction are presented in the figure. Two significant educational phenotypes are omitted due to substantial overlap with years of schooling. Genetic correlation is presented as a dot and error bars indicate 95% confidence limits.

Table 1. Results for the genome-wide significant index variants in the 12 loci associated with ADHD identified in the GWAS meta-analysis of 20,183 cases and 35,191 controls. Index variants are LD independent ($r^2 < 0.1$), and are merged into one locus when located with a distance less than 400kb. The location (chromosome [Chr] and base position [BP]), alleles (A1 and A2), allele frequency (A1 Freq), odds ratio (OR) of the effect with respect to A1, and association P-values from inverse-variance weighted fixed effects model, of the index variant are given, along with genes within 50kb of the credible set for the locus.

Locus	Chr	BP	Index Variant	Genes	A1	A2	A1 Freq	OR	P-value
1	1	44184192	rs11420276	*ST3GAL3, KDM4A, KDM4A-AS1, PTPRF, SLC6A9, ARTN, DPH2, ATP6V0B, B4GALT2, CCDC24, IPO13*	G	GT	0.696	1.113	2.14×10^{-13}
2	1	96602440	rs1222063	Intergenic	A	G	0.328	1.101	3.07×10^{-8}
3	2	215181889	rs9677504	*SPAG16*	A	G	0.109	1.124	1.39×10^{-8}
4	3	20669071	rs4858241	Intergenic	T	G	0.622	1.082	1.74×10^{-8}
5	4	31151456	rs28411770	*PCDH7, LINC02497*	T	C	0.651	1.090	1.15×10^{-8}
6	5	87854395	rs4916723	*LINC00461, MIR9-2, LINC02060, TMEM161B-AS1*	A	C	0.573	0.926	1.58×10^{-8}
7	7	114086133	rs5886709	*FOXP2, MIR3666*	G	GTC	0.463	1.079	1.66×10^{-8}
8	8	34352610	rs74760947	*LINC01288*	A	G	0.957	0.835	1.35×10^{-8}
9	10	106747354	rs11591402	*SORCS3*	A	T	0.224	0.911	1.34×10^{-8}
10	12	89760744	rs1427829	*DUSP6, POC1B*	A	G	0.434	1.083	1.82×10^{-9}
11	15	47754018	rs281324	*SEMA6D*	T	C	0.531	0.928	2.68×10^{-8}
12	16	72578131	rs212178	*LINC01572*	A	G	0.883	0.891	7.68×10^{-9}

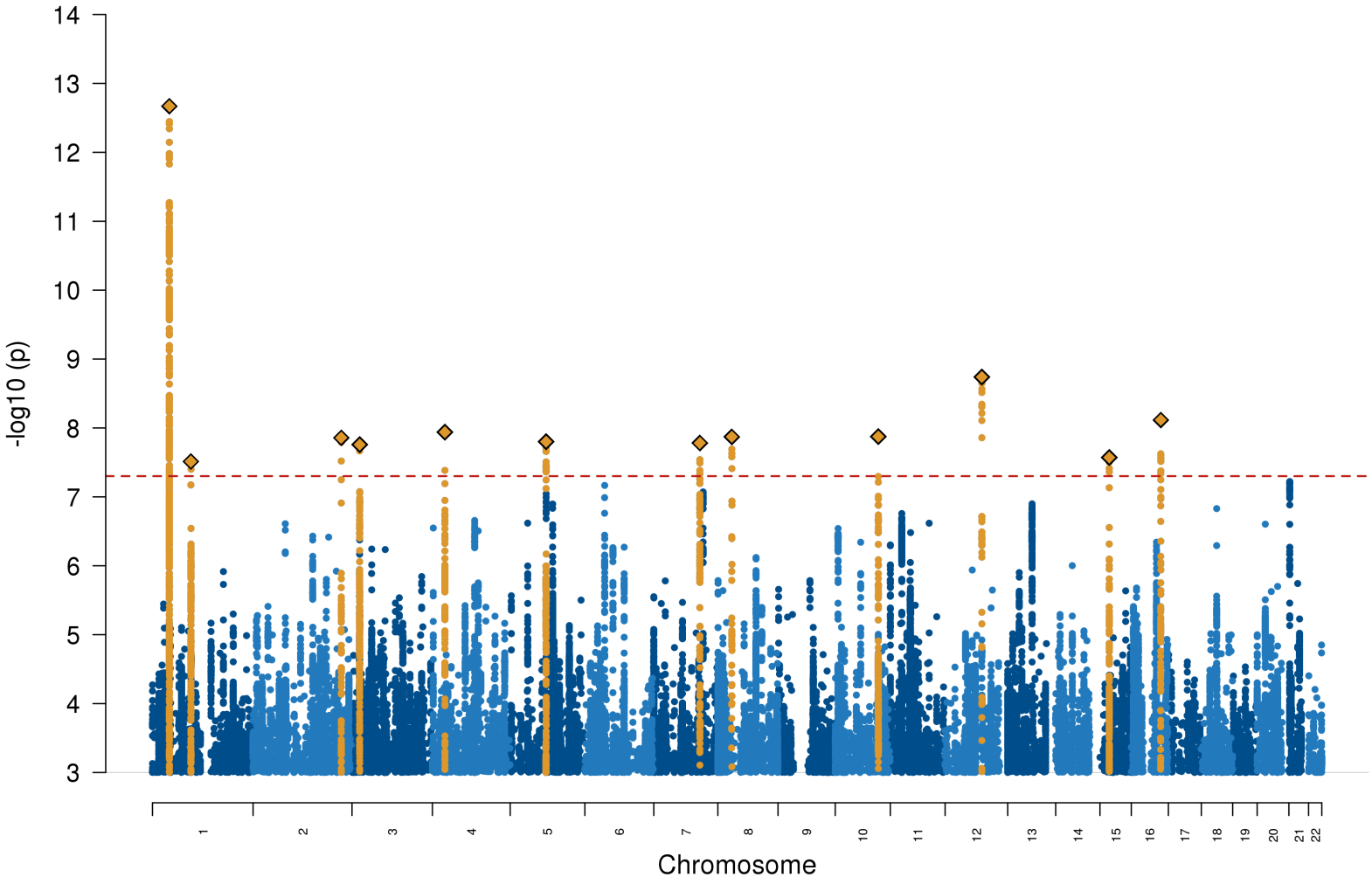
-log10 (p)
14
13
12
11
10
9
8
7
6
5
4
3
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

Figure 1

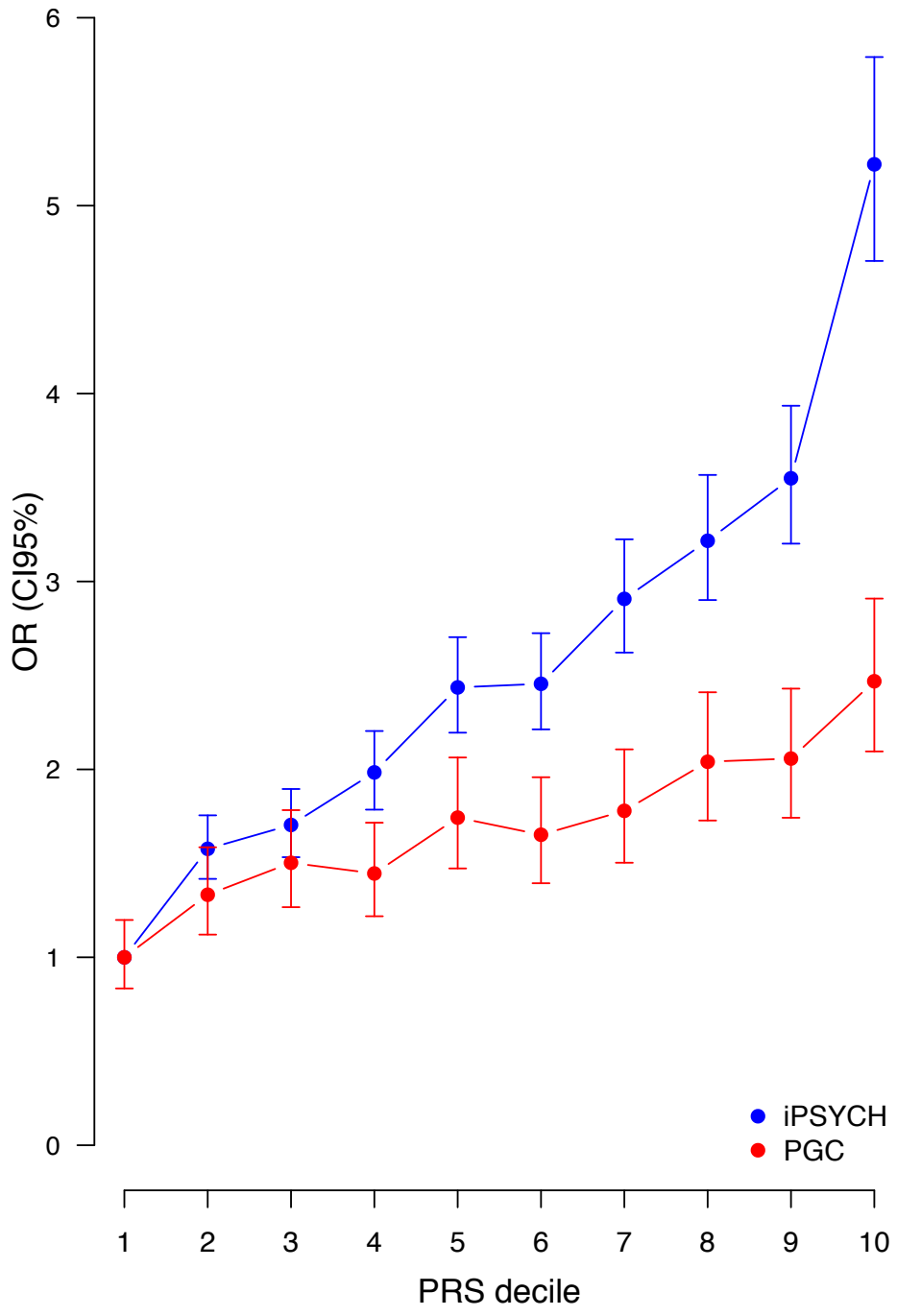

Figure 2

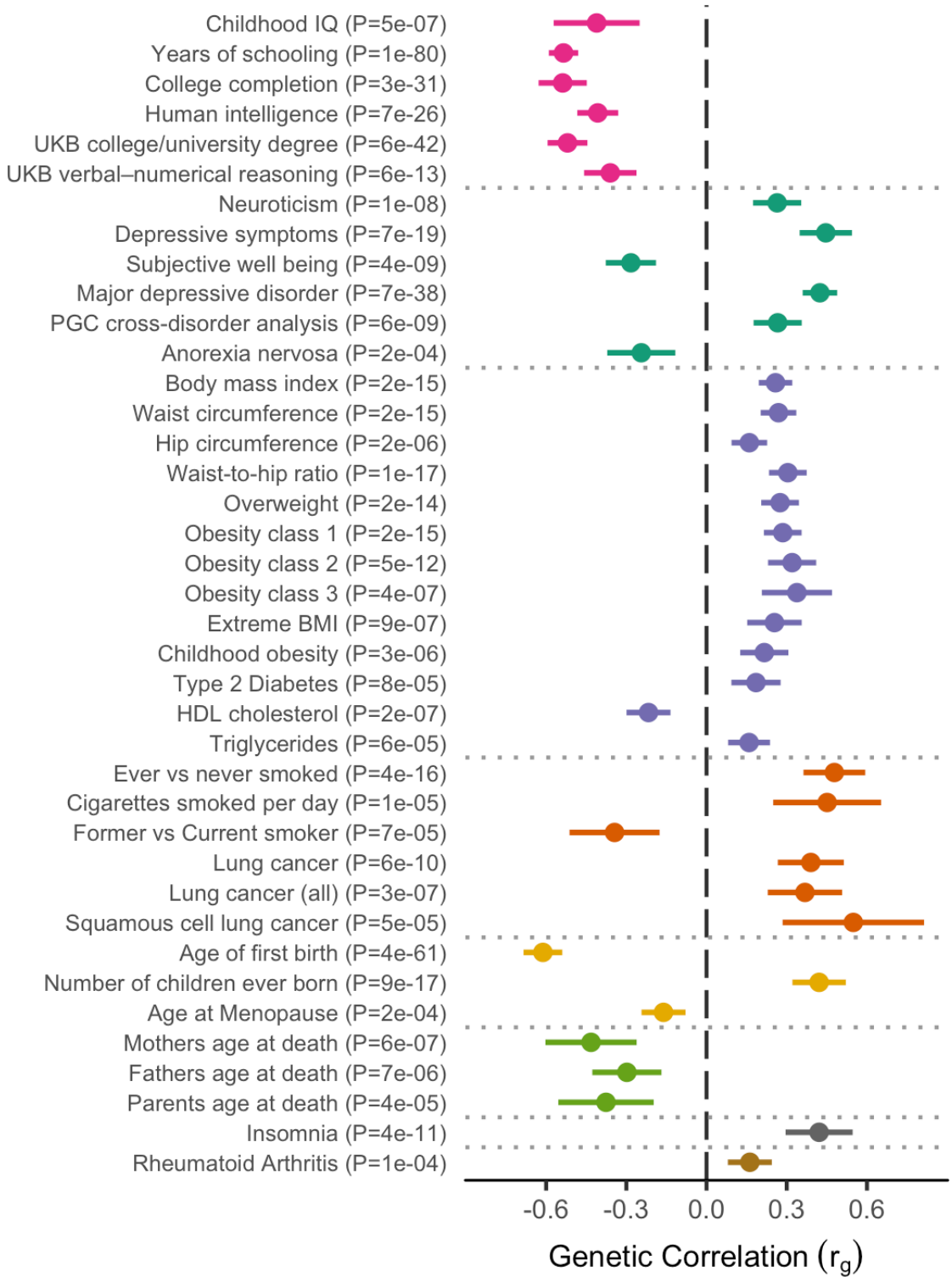
Childhood IQ (P=5e-07)
Years of schooling (P=1e-80)
College completion (P=3e-31)
Human intelligence (P=7e-26)
UKB college/university degree (P=6e-42)
UKB verbal–numerical reasoning (P=6e-13)
Neuroticism (P=1e-08)
Depressive symptoms (P=7e-19)
Subjective well being (P=4e-09)
Major depressive disorder (P=7e-38)
PGC cross-disorder analysis (P=6e-09)
Anorexia nervosa (P=2e-04)
Body mass index (P=2e-15)
Waist circumference (P=2e-15)
Hip circumference (P=2e-06)
Waist-to-hip ratio (P=1e-17)
Overweight (P=2e-14)
Obesity class 1 (P=2e-15)
Obesity class 2 (P=5e-12)
Obesity class 3 (P=4e-07)
Extreme BMI (P=9e-07)
Childhood obesity (P=3e-06)
Type 2 Diabetes (P=8e-05)
HDL cholesterol (P=2e-07)
Triglycerides (P=6e-05)
Ever vs never smoked (P=4e-16)
Cigarettes smoked per day (P=1e-05)
Former vs Current smoker (P=7e-05)
Lung cancer (P=6e-10)
Lung cancer (all) (P=3e-07)
Squamous cell lung cancer (P=5e-05)
Age of first birth (P=4e-61)
Number of children ever born (P=9e-17)
Age at Menopause (P=2e-04)
Mothers age at death (P=6e-07)
Fathers age at death (P=7e-06)
Parents age at death (P=4e-05)
Insomnia (P=4e-11)
Rheumatoid Arthritis (P=1e-04)
-0.6
-0.3
0.0
0.3
0.6
Genetic Correlation (r_g)

Figure 3

Online Methods

GWAS meta-analysis

Quality control, imputation and primary association analyses were done using the bioinformatics pipeline Ricopili (available at https://github.com/Nealelab/ricopili), developed by the Psychiatric Genomics Consortium (PGC)[33]. In order to avoid potential study effects the 11 PGC samples and the 23 genotyping batches within iPSYCH were each processed separately unless otherwise stated (Supplementary Information).

Stringent quality control was applied to each cohort following standard procedures for GWAS, including filters for call rate, Hardy-Weinberg equilibrium, and heterozygosity rates (Supplementary Information). Each cohort was then phased and imputed using the 1000 Genomes Project phase 3 (1KGP3)[34,113] imputation reference panel using SHAPEIT[114] and IMPUTE2[115], respectively. For trio cohorts, pseudocontrols were defined from phased haplotypes prior to imputation.

Cryptic relatedness and population structure were evaluated using a set of high quality markers pruned for linkage disequilibrium (LD). Genetic relatedness was estimated using PLINK v1.9[116,117] to identify first and second-degree relatives ($\hat{\pi} > 0.2$) and one individual was excluded from each related pair. Genetic outliers were identified for exclusion based on principal component analyses using EIGENSOFT[35,118]. This was done separately for each of the PGC cohorts and on a merged set of genotypes for the iPSYCH cohort (Supplementary Information). Across studies, a total of 20,183 cases and 35,191 controls remained for analysis after QC.

Genome-wide association analyses for the 11 PGC samples and the 23 waves in iPSYCH were performed using logistic regression model with the imputed marker dosages in PLINK v1.9[116,117]. Principal components were included as covariates to control for population stratification[35,118], along with relevant study-specific covariates where applicable (Supplementary Information, Supplementary Table 1). Subsequently the results were meta-analysed using an inverse-variance weighted fixed effects model, implemented in METAL (version 2011-03-25)[36]. Variants were filtered and included if imputation quality (INFO score) was > 0.8 and MAF > 0.01. Only markers supported by an effective sample size $N_{eff} = 4/(1/N_{cases} + 1/N_{controls})$[119] greater than 70% were included. After filtering, the meta-analysis included results for 8,047,421 markers.

Conditional analysis

Twelve independent genome-wide significant loci were identified by LD clumping and merging loci within 400 kb (Supplementary Information). In two of these loci a second index variant persisted after LD clumping. The two putative secondary signals were evaluated by considering analysis conditional on the lead index variant in each locus. In each cohort, logistic regression was performed with the imputed genotype dosage for the lead index variant included as a covariate. All covariates from the primary GWAS (e.g. principal components) were also included. The conditional association results were then combined in an inverse-variance weighted meta-analysis.

Genetic correlations between ADHD samples

Genetic correlation between the European-ancestry PGC and iPSYCH GWAS results was calculated using LD Score regression[37]. The regression was performed using pre-computed LD scores for

HapMap3 SNPs calculated based on 378 European-ancestry individuals from the 1000 Genomes Project (available on https://github.com/bulik/ldsc). Only results for markers with an imputation INFO score > 0.90 were included in the analysis. In addition, a bivariate GREML analysis was conducted using GCTA[120] in order to estimate the genetic correlation between PGC case/control and trio study designs.

Polygenic Risk Scores for ADHD

The iPSYCH sample were split into five groups, and subsequently five leave-one-out association analyses were conducted, using four out of five groups and the PGC samples as training datasets[38]. PRS were estimated for each target sample using variants passing a range of association P-value thresholds in the training samples. PRS were calculated by multiplying the natural log of the odds ratio of each variant by the allele-dosage (imputation probability) and whole-genome polygenic risk scores were obtained by summing values over variants for each individual.

For each of the five groups of target samples, PRS were normalized and the significance of the case-control score difference was tested by standard logistic regression including principal components. For each target group and for each P-value threshold the proportion of variance explained (i.e. Nagelkerke's R^2) was estimated by comparing the regression with PRS to a reduced model with covariates only. The OR for ADHD within each PRS decile group was estimated based on the normalized score across groups (using the P-value threshold with the highest Nagelkerke's R^2 within each target group) (Figure 3). OR was also estimated using logistic regression on the continuous scores for each target group separately and an OR based on all samples using the normalized PRS score across all groups (Supplementary Figure 9). Additionally PRS were evaluated in the PGC samples using the iPSYCH sample as training sample, following the approach described above (Supplementary Information).

SNP heritability and intercept evaluation

LD score regression[37] was used to evaluated the relative contribution of polygenic effects and confounding factors, such as cryptic relatedness and population stratification, to deviation from the null in the genome-wide distribution of GWAS χ^2 statistics. Analysis was performed using pre-computed LD scores from European-ancestry samples in the 1000 Genomes Project (available on https://github.com/bulik/ldsc) and summary statistics for the European-ancestry ADHD GWAS to ensure matching of population LD structure. The influence of confounding factors was tested by comparing the estimated intercept of the LD score regression to one, it's expected value under the null hypothesis of no confounding from e.g. population stratification. The ratio between this deviation and the deviation of the mean χ^2 from one (i.e. it's expected value under the null hypothesis of no association) was used to estimate the proportion of inflation in χ^2 attributable to confounding as opposed to true polygenic effects (ratio = (intercept-1)/(mean χ^2-1)). SNP heritability was estimated based on the slope of the LD score regression, with heritability on the liability scale calculated assuming a 5% population prevalence of ADHD[39].

Partitioning of the heritability

SNP heritability was partitioned by functional category and tissue association using LD score regression[40]. Partitioning was performed for 53 overlapping functional categories, as well as 220 cell-type-specific annotations grouped into 10 cell-type groups, as described in Finucane et al. [40]. For both sets of annotations we used previously computed LD scores and allele frequencies from European

ancestry samples in the 1000 Genomes Project (available on https://data.broadinstitute.org/alkesgroup/LDSCORE/).

Additionally we expanded the cell-type specific heritability analysis by including an annotation based on information about H3K4Me1 imputed gapped peaks excluding the broad MHC-region (chr6:25-35MB), generated by the Roadmap Epigenomics Mapping Consortium[121,122] (Supplementary Information). The analyses were restricted to the European GWAS meta-analysis results to ensure matching of population LD structure. Results for each functional category were evaluated based on marginal enrichment, defined as the proportion of SNP heritability explained by SNPs in the annotation divided by the proportion of genome-wide SNPs in the annotation[40]. For each cell-type group and each H3K4Me1 cell-type annotations, the contribution to SNP heritability was tested conditional on the baseline model containing the 53 functional categories.

Genetic correlations of ADHD with other traits

The genetic correlations of ADHD with other phenotypes were evaluated using LD Score regression[42]. For a given pair of traits, LD score regession estimates the expected population correlation between the best possible linear SNP-based predictor for each trait, restricting to common SNPs. Such correlation of genetic risk may reflect a combination of colocalization, pleiotropy, shared biological mechanisms, and causal relationships between traits. Correlations were tested for 211 phenotypes with publically available GWAS summary statistics using LD Hub[41] (Supplementary Information). Additonally, we analysed on our local computer cluster, the genetic correlation of ADHD with eight phenotypes: human intelligence[103], four phenotypes related to education and cognition analyzed in samples from the UK_Biobank[49] (college/university degree, verbal–numerical reasoning, memory and reaction time),

insomnia[60], anorexia nervosa[44], and major depressive disorder[43]. The genetic correlation with major depressive disorder was tested using GWAS results from an updated analysis of 130,664 cases with major depressive disorder and 330,470 controls from the Psychiatric Genomics Consortium. As in the previous LD score regression analyses, this estimation was based on summary statistics from the European GWAS meta-analysis, and significant correlations reported are for traits analysed using individuals with European ancestry.

Credible set analysis

We defined a credible set of variants in each locus using the method described by Maller et al.[123] (Supplementary Information), implemented by a freely available R script (https://github.com/hailianghuang/FM-summary). Under the assumption that (a) there is one causal variant in each locus, and (b) the causal variant is observed in the genotype data, the credible set can be considered to have a 99% probability of containing the causal variant. For each the 12 genome-wide significant loci, variants within 1MB and in LD with correlation $r^2 > 0.4$ to the index variant were considered for inclusion in the credible set analysis. The credible set analysis was done using the European GWAS meta-analysis to ensure consistent LD structure in the analyzed cohorts.

Biological annotation of variants in credible set

The variants in the credible set for each locus, were annotated based on external reference data in order to evaluate potential functional consequences. In particular, we identify: (a) Gene and regulatory consequences annotated by Variant Effect Predictor (VEP) using Ensembl with genome build GRCh37[124]. We exclude upstream and downstream consequences, and consequences for transcripts that

lack a HGNC gene symbol (e.g. vega genes). (b) Variants within 2kb upstream of the transcription start site (TSS) of at least one gene isoform based on Gencode v19[125]. (c) Variants annotated as interacting with a given gene in Hi-C data from samples of developing human cerebral cortex during neurogenesis and migration[126]. Annotations are considered for both the germinal zone (GZ), primarily consisting of actively dividing neural progenitors, and the cortical and subcortical plate (CP), primarily consisting of post-mitotic neurons. (d) Variants identified as eQTLs based on gene expression in GTEx[127] or BIOS[79]. Expression quantitative trait loci were annotated using FUMA (http://fuma.ctglab.nl/). We restricted to eQTL associations with false discovery fate (FDR) < 1e-3 within each dataset. (e) Chromatin states of each variant based on the 15-state chromHMM analysis of epigenomics data from Roadmap[128]. The 15 states summarize to annotations of active chromatin marks (i.e. Active TSS, Flanking Active TSS, Flanking Transcription, Strong Transcription, Weak Transcription, Genic Enhancer, Enhancer, or Zinc Finger [ZNF] gene), repressed chromatin marks (Heterochromatin, Bivalent TSS, Flanking Bivalent TSS, Bivalent Enhancer, Repressed Polycomb, or Weak Repressed Polycomb), or quiescent. The most common chromatin state across 127 tissue/cell types was annotated using FUMA (http://fuma.ctglab.nl/). We also evalauted the annotated chromatin state from fetal brain.

Gene-set analyses

Gene-based association with ADHD was estimated with MAGMA 1.05[88] using the summary statistics from the European GWAS meta-analysis (N_{cases} = 19,099, $N_{controls}$ = 34,194; Supplementary Information, Supplementary Information Table 1). Association was tested using the SNP-wise mean model, in which the sum of -log(SNP P-value) for SNPs located within the transcribed region (defined using NCBI 37.3 gene definitions) was used as the test statistic. MAGMA accounts for gene-size,

number of SNPs in a gene and LD between markers when estimating gene-based P-values. LD correction was based on estimates from the 1000 genome phase 3 European ancestry samples[34].

The generated gene-based P-values were used to analyze sets of genes in order to test for enrichment of association signals in genes belonging to specific biological pathways or processes. In the analysis only genes on autosomes, and genes located outside the broad MHC region (hg19:chr6:25-35M) were included. We used the gene names and locations and the European genotype reference panel provided with MAGMA. For gene sets we used sets with 10-1000 genes from the Gene Ontology sets[86] currated from MsigDB 6.0[87].

Targeted *FOXP2* downstream target gene sets were analysed for association with ADHD. Three sets were examined: 1) Putative target genes of *Foxp2* that were enriched in wild type compared to control *Foxp2* knockout mouse brains in ChIP-chip experiments (219 genes), 2) Genes showing differential expression in wild type compared to *Foxp2* knockout mouse brains (243 genes), and 3) *FOXP2* target genes that were enriched in either or both basal ganglia (BG) and inferior frontal cortex (IFC) from human fetal brain samples in ChIP-chip experiments (258 genes). Curated short lists of high-confidence genes were obtained from Vernes et al.[89] and Spiteri et al[90].

A set of evolutionarily highly constrained genes were also analysed. The set of highly constrained genes was defined using a posterior probability of being loss-of-function intolerant (pLI) based on the observed and expected counts of protein-truncating variants (PTV) within each gene in a large study of over 60,000 exomes from the Exome Aggregation Consortium (ExAC)[91]. Genes with pLI $\geq$0.9 were selected as the set of highly constrained genes (2932 genes).

Replication of GWAS loci

To replicate the results of the ADHD GWAS meta-analysis we compared the results to analyses of cohorts from deCODE and 23andMe, and a meta-analysis of two independent studies conducted by EAGLE and QIMR (referred to as EAGLE/QIMR). We evaluated evidence for replication based on: (a) sign tests of concordance between the ADHD GWAS meta-analysis and each replication cohort; (b) comparison of bias-corrected effect sizes between the ADHD GWAS and the deCODE and 23andMe replication cohorts; (c) genetic correlation between the ADHD GWAS and the 23andMe and EAGLE/QIMR replication cohorts; (d) meta-analysis of the ADHD GWAS meta-analysis results with the results from each replication cohort; and (e) tests of heterogeneity between the ADHD GWAS and each replication cohorts.

For the sign test, we first identified the overlapping SNPs present in the ADHD GWAS and each of the three replication analyses (i.e. deCODE, 23andMe, and EAGLE/QIMR). For each replication cohort intersecting SNPs were then clumped for LD ($r^2 > 0.05$ within 1 Mb) for all variants with $P < 1 \times 10^{-4}$ in the ADHD GWAS (or $P < 1 \times 10^{-5}$ for the deCODE replication) using 1000 Genomes Phase 3 data on European ancestry populations. After clumping, sign tests were performed to test the proportion of loci with a concordant direction of effect in the replication cohort (π) using a one sample test of the proportion with Yates' continuity correction[129] against a null hypothesis of $\pi = 0.50$ (i.e. the signs are concordant between the two analyses by chance) in R[130]. This test was evaluated separately for concordance in deCODE, 23andMe, and EAGLE/QIMR for loci passing P-value thresholds of $P < 5 \times 10^{-8}$ (i.e. genome-wide significant loci), $P < 1 \times 10^{-7}$, $P < 1 \times 10^{-6}$, $P < 1 \times 10^{-5}$, and $P < 1 \times 10^{-4}$ in the ADHD GWAS meta-analysis (Supplementary Information).

In addition to testing concordance for the direction of effect, we also evaluate replication for the magnitude of the effect sizes. Specifically, for each of deCODE and 23andMe we regressed the effect size in the replication cohort (i.e. the log odds ratio) on the estimated effect size from the ADHD GWAS after adjustment for winner's curse for loci with P < 1e-6. Winner's curse correction is perfomed by computing posterior mean estimates of marginal SNP effects β_j after fitting a spike-and-alab distribution

$$\beta_j \sim \begin{cases} 0 & \text{with probability } \pi \\ N(0, \tau^2) & \text{otherwise} \end{cases}$$

by maximum likelihood as described by Okbay et al.[78] (Supplementary Information). For the regression of effect sizes we oriented all variants in the direction of the risk increasing allele estimated from the ADHD GWAS, constrained the intercept to zero, and weighted the variants proportional to the inverse of their squared standard error from the ADHD GWAS. A regression slope of one indicates "ideal" replication of all loci in the regression, whereas a slope of zero indicates no replication.

Genetic correlation of the ADHD GWAS with the 23andMe and EAGLE/QIMR results was computed using LD score regression[37] with pre-computed European ancestry LD scores following the same procedure as described above for other genetic correlation analyses. Genetic correlation could not be computed for deCODE since results were only available for top loci from the ADHD GWAS. To further explore the moderate genetic correlation between the 23andMe results and the ADHD GWAS we also evaluated the genetic correlation between 23andMe and traits from LD Hub (http://ldsc.broadinstitute.org/ldhub/)[42]. To evaluate the magnitude of the observed differences in r_g we consider both the absolute difference (i.e. $|r_{g,ADHD} - r_{g,23andMe}|$) and the test of an approximate Z score for this difference (Supplementary Information):

$$Z = \frac{r_{g,ADHD} - r_{g,23andMe}}{\sqrt{SE^2_{ADHD} + SE^2_{23andMe}}}$$

We do not expect this to be an ideal formal test for the difference between two genetic correlations, and therefore emphasize caution in interpreting the precise results. Nevertheless, it does offer a useful benchmark for evaluating the magnitude of the difference between the r_g estimates in the context of the uncertainty in those values.

Finally, we meta-analyzed the ADHD GWAS with the results from each replication cohort. For deCODE and 23andMe inverse variance-weighted meta-analyses were performed. For meta-analysis with the EAGLE/QIMR GWAS of ADHD-related behaviors in childhood population samples we used a modified sample size-based weighting method. Modified sample size-based weights were derived to accounts for the respective heritabilities, genetic correlation, and measurement scale of the GWASs (Supplementary Information). To summarize, given *z*-scores Z_{1j} and Z_{2j} resulting from GWAS of SNP *j* in a dichotomous phenotype (e.g. ADHD) with sample size N_1 and a continuous phenotype (e.g. ADHD-related traits) with sample size N_2, respectively, we calculate

$$Z_{j,meta} = \frac{\sqrt{\tilde{N}_{1j}} Z_{1j} + \sqrt{\tilde{N}_{2j}} \tilde{Z}_{2j}}{\sqrt{\tilde{N}_{1j} + \tilde{N}_{2j}}}$$

where

$$\tilde{Z}_{2j} = sign(r_g) \times \frac{Z_{2j}}{\sqrt{1 + (1 - r_g^2) N_{2j} h_2^2 \, l_j / M}}$$

$$\tilde{N}_{1j} = N_{1j} \frac{P(1-P)\, \phi(\Phi^{-1}[K])^2}{[K(1-K)]^2}$$

$$\widetilde{N}_{2j} = N_{2j} \frac{r_g^2 h_2^2 / h_1^2}{1 + (1 - r_g^2)\, N_{2j} h_2^2 l_j / M}$$

The adjusted sample sizes $\widetilde{N_1}$ and $\widetilde{N_2}$ reflect differences in power between the studies due to measurement scale and relative heritability that is not captured by sample size. The calculation of $\widetilde{Z_2}$ reduces the contribution of the continuous phenotype's GWAS to the meta-analysis based on imperfect genetic correlation with the dichotomous phenotype of interest (i.e. ADHD). The adjustments are computed based on the sample prevalence (*P*) and population prevalence (*K*) of the dichotomous phenotype, the estimated liability scale SNP heritability of the two phenotypes (h_1^2 and h_2^2), and the genetic correlation (r_g) between the two phenotypes, as well as the average SNP LD score (l_j) and the number of SNPs (*M*). Heritability and genetic correlation values to compute these weights are computed using LD score regression. This meta-analysis weighting scheme is consistent with weights alternatively derived based on modelling the joint distribution of marginal GWAS beta across traits[131].

To test heterogeneity with each replication cohort, we considered Cochran's *Q* test of heterogeneity in the meta-analyses. Specifically, we evaluated the one degree of freedom test for heterogeneity between the ADHD GWAS meta-analysis and the replication cohort.

Data Avalibility Statement

The PGC's policy is to make genome-wide summary results public. Summary statistics with the results from the ADHD GWAs meta-analysis of iPSYCH and the PGC samples are available on the PGC website (https://www.med.unc.edu/pgc/results-and-downloads). GWA summary statistics with results from the GWAS of ADHD symptom scores analyzed in the EAGLE sample can be accessed at the PGC

website (see link above). Summary statistics for the 23andMe dataset can be obtained by qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants.

Availability of genotype data and summary statistics

For access to genotypes from the PGC cohorts and the iPSYCH sample interested researchers should contact the lead PIs (iPSYCH: lead PI Anders D. Børglum; PGC: Benjamin Neale and Stephen Faraone). Summary statitistics can be downloaded from: https://www.med.unc.edu/pgc/results-and-downloads

http://ipsych.au.dk/downloads/

http://www.wikigenes.org/e/art/e/348.html

Methods-only references

113 Sudmant, P. H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75-81, doi:10.1038/nature15394 (2015).

114 Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nature methods* **9**, 179-181, doi:10.1038/nmeth.1785 (2011).

115 Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3* **1**, 457-470, doi:10.1534/g3.111.001198 (2011).

116 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559-575, doi:10.1086/519795 (2007).

117 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).

118 Galinsky, K. J. *et al.* Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. *American journal of human genetics* **98**, 456-472, doi:10.1016/j.ajhg.2015.12.022 (2016).

119 Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protocols* **9**, 1192-1212, doi:10.1038/nprot.2014.071 (2014).

120 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics* **88**, 76-82, doi:10.1016/j.ajhg.2010.11.011 (2011).

121 Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).
122 Ernst, J. & Kellis, M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nature biotechnology* **33**, 364-376 (2015).
123 Wellcome Trust Case Control, C. *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nature genetics* **44**, 1294-1301, doi:10.1038/ng.2435 (2012).
124 McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122, doi:10.1186/s13059-016-0974-4 (2016).
125 Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760-1774, doi:10.1101/gr.135350.111 (2012).
126 Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523-527, doi:10.1038/nature19847 (2016).
127 Consortium, G. T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-660, doi:10.1126/science.1262110 (2015).
128 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).
129 Yates, F. Contingency Tables Involving Small Numbers and the χ2 Test. *Supplement to the Journal of the Royal Statistical Society* **1**, 217-235, doi:10.2307/2983604 (1934).
130 R Core Team. *R: A language and environment for statistical computing.* , <http://www.r-project.org/> (2014).
131 Turley, P. *et al.* MTAG: Multi-Trait Analysis of GWAS. *bioRxiv*, doi:10.1101/118810 (2017).

End notes

Reprints and permissions information is available at www.nature.com/reprints

Competing Interests

In the past year, Dr. Faraone received income, potential income, travel expenses continuing education support and/or research support from Lundbeck, Rhodes, Arbor, KenPharm, Ironshore, Shire, Akili Interactive Labs, CogCubed, Alcobra, VAYA, Sunovion, Genomind and Neurolifesciences. With his institution, he has US patent US20130217707 A1 for the use of sodium-hydrogen exchange inhibitors in the treatment of ADHD. In previous years, he received support from: Shire, Neurovance, Alcobra, Otsuka, McNeil, Janssen, Novartis, Pfizer and Eli Lilly. Dr. Faraone receives royalties from books published by Guilford Press: Straight Talk about Your Child's Mental Health, Oxford University Press: Schizophrenia: The Facts and Elsevier: ADHD: Non-Pharmacologic Interventions. He is principal investigator of www.adhdinadults.com.
Dr. Neale is a member of Deep Genomics Scientific Advisory Board and has received travel expenses from Illumina. He also serves as a consultant for Avanir and Trigeminal solutions.

Olafur O. Gudmundsson, G. Bragi Walters, Hreinn Stefansson and Kari Stefansson are employees of deCODE genetics/Amgen.
Members of the 23andMe research team are employees of 23andMe.
Dr. Rohde has received honoraria, has been on the speakers' bureau/advisory board and/or has acted as a consultant for Eli-Lilly, Janssen-Cilag, Novartis, Medice and Shire in the last three years. He receives authorship royalties from Oxford Press and ArtMed. He also received travel award for taking part of 2015 WFADHD meeting from Shire. The ADHD and Juvenile Bipolar Disorder Outpatient Programs chaired by him received unrestricted educational and research support from the following pharmaceutical companies in the last three years: Eli-Lilly, Janssen-Cilag, Novartis, and Shire.
Over the last three years Dr. Sonuga-Barke has received speaker fees, consultancy, research funding and conference support from Shire Pharma and speaker fees from Janssen Cilag. He has received consultancy fees from Neurotech solutions, Aarhus University, Copenhagen University and Berhanderling, Skolerne, Copenhagen, KU Leuven. Book royalties from OUP and Jessica Kingsley. He is the editor-in-chief of the Journal of Child Psychology and Psychiatry for which his University receives financial support.
Barbara Franke has received educational speaking fees from Merz and Shire.
Dr. Schachar's disclosures: ehave equity and advisory board, Ironshore Pharmaceuticals Advisory Board.
Dr. Reif has received a research grant from Medice, and speaker's honorarium from Medice and Servier.
Dr. Haavik has received speaker fees from Shire, Lilly and Novartis.

Correspondence and requests for materials should be addressed to:

Benjamin Neale, bneale@broadinstitute.org
Anders D. Børglum, anders@biomed.au.dk
Stephen Faraone, sfaraone@childpsychresearch.org