the 23 and Me Research Team, ADHD Working Group of the Psychiatric Genomics Consortium (PGC), Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium, Demontis, D., Walters, R. K., Martin, J., ... Medland, S. E. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nature Genetics, 51(1), 63-75. https://doi.org/10.1038/s41588-018-0269-7

Peer reviewed version

License (if available):
Other

Link to published version (if available):
10.1038/s41588-018-0269-7

Link to publication record in Explore Bristol Research
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via Springer Nature at https://doi.org/10.1038/s41588-018-0269-7. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/pure/about/ebr-terms
Published in final edited form as:


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

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.

1Equally contributed.

Co-last authors.

**Author Contributions**

Analysis:

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

Sample and/or data provider and processing:


SEM, DMH, OM, PBM, ADB, ADHD Working Group of the Psychiatric Genomics Consortium, Early Life course & 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.

All authors contributed with critical revision of the manuscript.

**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 NeurilifeSciences. 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](http://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 TriGenimal solutions.

Olafur O. Gudmundsson, G. Bragi Walters, Hreinn Stefansson and Kari Stefansson are employees of deCODE genetics/Amgen. Nicholas Eriksson, Joyce Y Tung, and the 23andMe Research Team are employees of 23andMe, Inc., and hold stock or stock options in 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 Berherdling, 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.

Dr. Franka 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.

Dr. Kranzler has been an advisory board member, consultant, or CME speaker for Alkermes, Indivior, and Lundbeck. He is also a member of American Society of Clinical Psychopharmacology’s Alcohol Clinical Trials Initiative, which was supported in the last three years by AbbVie, Alkermes, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, Pfizer, Arbor, and Amygdala Neurosciences.


Dr Asherson received honoraria paid to King’s College London by Shire, Flynn Pharma, Lilly, Janssen, Novartis and Lunebeck for research, speaker fees, education events, advisory board membership or consultancy.

Dr. Andreassen has received speaker fees from Lundbeck and Sunovion. Dr Kuntsi has received speaker’s honorarium from Medice; all funds are received by King’s College London and used for studies of ADHD.

Thomas Were has acted as lecturer and scientific advisor to H. Lundbeck A/S.

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 statistics can be downloaded from:
Discovery of the first genome-wide significant risk loci for

https://www.med.unc.edu/pgc/results-and-downloads
http://ipsych.au.dk/downloads/
http://www.wikigenes.org/e/art/e/348.html

Further information about study design is available in the Life Sciences Reporting Summary linked to this article.

Reporting summary

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

Consortium Members

Consortium authors are listed alphabetically by surname for each contributing group.

ADHD Working Group of the Psychiatric Genomics Consortium (PGC)


70 Department of Psychosomatic Medicine and Psychotherapy, Hannover Medical School (MHH), Hannover, Germany

71 MRC Centre for Neuropsychiatric Genetics and Genomics, Department of Psychological Medicine and Neurology, School of Medicine, Cardiff University, Cardiff, Wales, UK

72 University Hospital Mutua Terrassa, Barcelona, Spain

73 Department of Child and Adolescent Psychiatry, Central Institute of Mental Health and Mannheim Medical Faculty, University of Heidelberg, Heidelberg, Germany

74 Department of Genetics, Instituto de Biocències, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

75 Pediatric Psychopharmacology Unit, Massachusetts General Hospital, Boston, MA, USA

76 Department of Psychiatry, Harvard Medical School, Boston, MA, USA

77 Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Centre, Nijmegen, The Netherlands

78 Karakter Child and Adolescent Psychiatry University Center, Nijmegen, The Netherlands

79 Universitat Autònoma de Barcelona, Barcelona, Spain

80 Programa Corporatiu “Neurodevelopment Disorders along Life Span”, Institut Català de la Salut, Barcelona, Spain

81 Department of Psychiatry, Hospital Universitari Vall d’Hebron, Barcelona, Spain

82 Clínica Galatea y PAIMM, Mental Health Program for Impaired Physicians, Barcelona, Spain

83 The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada

84 Institute of Medical Informatics and Statistics, Kiel University, Kiel, Germany

85 Massachusetts General Hospital, Boston, MA, USA.

86 Harvard Medical School, Boston, MA, USA.

Nat Genet. Author manuscript; available in PMC 2019 July 01.
attention-deficit/hyperactivity disorder
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 diagnosed ADHD 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

134 Peking University Institute of Mental Health, Beijing Shi, China
135 Cardiff University, Medical Research Council Center for Neuropsychiatric Genetics & Genomics, Institute of Psychology, Medicine & Clinical Neuroscience, Cardiff, UK
136 Analytic Translational Genetics Unit, Massachusetts General Hospital Harvard Medical School, Boston, Massachusetts, USA
137 Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK.
138 Avera Institute for Human Genetics, Sioux Falls, SD, USA
139 University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia
140 Radboud University Medical Center, Donders Institute for Brain, Cognition and Behaviour, Department of Cognitive Neuroscience, Nijmegen, The Netherlands
141 Medical Research Council Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neurosciences, King’s College London, London, UK
142 GGZ inGeest, Amsterdam, the Netherlands
143 ISGlobal - Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain
144 Universitat Pompeu Fabra (UPF), Barcelona, Spain
145 Department of Child and Adolescent Psychiatry/Psychology, Erasmus University Medical Centre–Sophia Children’s Hospital, Rotterdam, The Netherlands
146 Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam, The Netherlands
147 Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
148 Institute of Epidemiology I, Helmholtz Zentrum München – German Research Centre for Environmental Health, Neuherberg, Germany
149 Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Inner City Clinic, University Hospital of Munich (LMU), Munich, Germany
150 Biological Psychology, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
151 Vermont Center for Children Youth and Families and University of Vermont Medical Center, University of Vermont, Burlington, VT, USA
152 Child Psychiatry, School of Medicine, Washington University, St. Louis, MO, USA
153 Erasmus University Medical Centre–Sophia Children’s Hospital, Rotterdam, The Netherlands
154 Geisel School of Medicine, Dartmouth, Hanover, NH, USA
155 Department of Genetic Research and Bioinformatics, Norwegian Institute of Public Health, Oslo, Norway
156 Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway
157 Epidemiology and Environmental Health Joint Research Unit, FISABIO-Universitat Jaume I-Universitat de València, Valencia, Spain.
158 Department of Genes and Environment, Norwegian Institute of Public Health, Oslo, Norway
159 Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
160 Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia
161 Generation R Study Group, Erasmus Medical Center, Rotterdam, The Netherlands
162 School of Women’s and Infants’ Health, The University of Western Australia, Crawley, Western Australia, Australia
163 School of Social and Community Medicine, University of Bristol, Bristol, UK
164 Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
165 Norwegian Institute of Public Health, Oslo, Norway
166 CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain
167 ISGlobal Barcelona Institute for Global Health, Barcelona, Spain
168 Division of Metabolic Diseases and Nutritional Medicine, Dr. von Hauner Children’s Hospital, Ludwig-Maximilians-University of Munich, Munich, Germany
169 Telethon Kids Institute, University of Western Australia, West Perth, Western Australia, Australia

Nat Genet. Author manuscript; available in PMC 2019 July 01.
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-specific 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\(^9\), 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 previously24,25,31,32,22 (see Supplementary Information for detailed cohort descriptions). All relevant ethical permissions and informed consent were obtained for the included cohorts (See Supplementary Information for details about approval authorities).

Prior to analysis, stringent quality control procedures were performed on the genotyped markers and individuals in each cohort using a standardized pipeline33 (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 panel34 (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 stratification35 (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 model36. 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<5×10^{-8}; Figure 1, Table 1, Supplementary Figure 3). 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\textsuperscript{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)\textsuperscript{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\textsuperscript{37} to results from the European ancestry meta-analysis (Online Methods). Assuming a population prevalence of 5% for ADHD\textsuperscript{39}, we estimate that the liability-scale SNP heritability $h^2_{\text{snp}} = 0.216$ (SE = 0.014, $P = 8.18\times10^{-54}$; 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 ($\lambda = 1.200$; 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, $P=2.27 \times 10^{-4}$).

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.\textsuperscript{40} using partitioned LD Score regression (Online Methods). The analysis revealed significant enrichment in the heritability from SNPs located in conserved regions ($P = 8.49 \times 10^{-10}$; Supplementary Figure 12), supporting their biological importance. Enrichment of the SNP heritability in cell-type-specific regulatory elements was evaluated using the cell-type-specific group annotations described in Finucane et al\textsuperscript{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, $P = 5.81 \times 10^{-5}$; 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). Forty-three phenotypes demonstrated significant genetic overlap with ADHD (P < 2.28 × 10\(^{-4}\)) including major depressive disorder\(^3\), 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 × 10\(^{-38}\)), we also observe a positive correlation with depressive symptoms (r\(_g\) = 0.45, P = 7.00 × 10\(^{-19}\)), neuroticism (r\(_g\) = 0.26, P = 1.02 × 10\(^{-8}\)) and a negative correlation with subjective well-being (r\(_g\) = −0.28, P = 3.73 × 10\(^{-9}\)). The positive genetic correlations with ever smoked (r\(_g\) = 0.48, P = 4.33 × 10\(^{-16}\)) and with number of cigarettes smoked (r\(_g\) = 0.45, P = 1.07 × 10\(^{-5}\)) are reinforced by significant positive correlation with lung cancer (r\(_g\) = 0.39, P = 6.35 × 10\(^{-10}\)). Similarly, genetic correlations related to obesity include significant relationships with body mass index (BMI; r\(_g\) = 0.26, P = 1.68 × 10\(^{-15}\)), waist-to-hip ratio (r\(_g\) = 0.30, P = 1.16 × 10\(^{-17}\)), childhood obesity (r\(_g\) = 0.22, P = 3.29 × 10\(^{-6}\)), HDL cholesterol (r\(_g\) = −0.22, P = 2.44 × 10\(^{-7}\)), and Type 2 Diabetes (r\(_g\) = 0.18, P = 7.80 × 10\(^{-5}\)). Additionally the negative correlation with years of schooling (r\(_g\) = −0.53, P = 6.02 × 10\(^{-80}\)) is supported by a negative genetic correlation with human intelligence (r\(_g\) = −0.41, P = 7.03 × 10\(^{-26}\)). Finally the genetic correlation with reproduction include a negative correlation with age of first birth (r\(_g\) = −0.612, P = 3.70 × 10\(^{-61}\)) and a positive correlation with number of children ever born (r\(_g\) = 0.42, P = 8.51 × 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, and may play a role in regulating neurotransmitter homeostasis by affecting dopamine levels in the synapses. Regulation of dopamine levels is likely to be relevant to ADHD since widely used ADHD medications have dopaminergic targets 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. SEMA6D is active in the brain during embryonic development, and may play a role in neuronal wiring. Furthermore, variants in SEMA6D have previously been associated with educational attainment.

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 (Supplementary Data 3). Missense mutations in ST3GAL3 have been shown to cause autosomal recessive intellectual disability. 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.

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 and the genome-wide significant locus overlaps with variants in LINC00461 associated with educational attainment. 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, cerebral malformation, depression, schizophrenia and Alzheimer’s disease, 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 from MsigDB 6.0 using MAGMA (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 15a–d). Three independent sets of FOXP2 downstream target genes 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 showed significant association with ADHD ($\beta = 0.062$, $P = 2.6 \times 10^{-4}$; Supplementary Table 10). We also find little evidence for effects in previously proposed candidate genes for ADHD; of the nine proposed genes only SLC9A9 showed weak association with ADHD ($P = 3.4 \times 10^{-4}$; 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) and QIMR (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 (Supplementary 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, $P = 1.11 \times 10^{-3}$), 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 notable outliers 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, $P = 2.66 \times 10^{-6}$) 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, $P = 1.11 \times 10^{-8}$), although also significantly less than 1 (one-sided $P= 1.17 \times 10^{-3}$). To explore this lower correlation we evaluated the genetic correlation between 23andMe and traits from LD Hub (see URLs) to potentially identify
differences in the profile of genetic correlation compared to the ADHD GW AS (Online Methods). This comparison identified striking differences (Supplementary Table 14), most notably that the 23andMe GW AS show little to no genetic correlation with college completion ($r_g = 0.056$, compared to $r_g = -0.54$ for the primary ADHD GW AS; approximate $P = 1.1 \times 10^{-9}$ 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 GW AS with each replication cohort. For EAGLE/QIMR, we developed a novel model to meta-analyze the GW AS of the continuous measure of ADHD with the clinical diagnosis in the ADHD GW AS. 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 18–20). 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 GW AS 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 21–22). Of the 12 significant loci from the primary ADHD GW AS, 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 attainment93.

Meta-analysis with the 23andMe cohort also found genome-wide significant heterogeneity at the lead Chromosome 1 locus from the ADHD GW AS meta-analysis (rs12410155: $I^2 = 97.2$, $P = 2.29 \times 10^{-9}$; Supplementary Figures 23–24). This heterogeneity is consistent with the moderate sign concordance, effect size replication, and genetic correlation of the 23andMe cohort with the ADHD GW AS. Notably, the lead chromosome 1 locus in the ADHD GWAS
overlaps a reported association with educational attainment\(^7\)\(^8\), 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 25–26, 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\(^3\)\(^3\). A considerably larger proportion of the heritability of ADHD can be explained by all common variants (\(h^2_{\text{snp}} = 0.22\), SE = 0.01). This is consistent with previous estimates of \(h^2_{\text{snp}}\) for ADHD in smaller studies (\(h^2_{\text{snp}}: 0.1 – 0.28\))\(^2\)\(^3\)\(^,\)\(^4\), and also comparable to SNP heritability estimates for schizophrenia (\(h^2_{\text{snp}} 0.23 – 0.26\))\(^2\)\(^3\)\(^,\)\(^4\). 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\(^4\)\(^0\). On the other hand, we do not observe substantial effects in most previously reported candidate genes for ADHD\(^6\)\(^2\).

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\(^9\)\(^4\), and constrained genes likely to be intolerant of loss-of-function mutations\(^9\)\(^1\) 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\(^9\)\(^5\)–\(^9\)\(^7\) and increased risk of ADHD for children born to young parents\(^9\)\(^8\)–\(^1\)\(^0\). Given the phenotypic\(^1\)\(^0\)\(^1\)\(^,\)\(^1\)\(^2\) and genetic\(^1\)\(^3\) 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\(^1\)\(^0\)\(^4\). 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\textsuperscript{24,42}, as well as the broader genetic overlap of psychiatric disorders\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{30}/QIMR\textsuperscript{92} are much stronger and support the hypothesis that ADHD is the extreme expression of one or more heritable quantitative traits\textsuperscript{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\textsuperscript{29}. Shared genetic risk with health risk behaviors may similarly be hypothesized to reflect an impaired ability to self-regulate and inhibit impulsive behavior\textsuperscript{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 \textit{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.
Online Methods

GWAS meta-analysis

Quality control, imputation and primary association analyses were done using the bioinformatics pipeline Ricopili (see URLs), developed by the Psychiatric Genomics Consortium (PGC)\(^3\). 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 (\(\bar{\rho} > 0.2\)) 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 \(N_{\text{eff}} = 4/(1/N_{\text{cases}} + 1/N_{\text{controls}})\)\(^{119}\) 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\textsuperscript{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 (see URLs). 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\textsuperscript{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\textsuperscript{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\textsuperscript{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 $\chi^2$ statistics. Analysis was performed using pre-computed LD scores from European-ancestry samples in the 1000 Genomes Project (see URLs) 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 $\chi^2$ from one (i.e. it’s expected value under the null hypothesis of no association) was used to estimate the proportion of inflation in $\chi^2$ attributable to confounding as opposed to true polygenic effects (ratio = (intercept-1)/(mean $\chi^2$-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%

Partitioning of the heritability

SNP heritability was partitioned by functional category and tissue association using LD score regression. 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. 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 see URLs).

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 Consortium121,122 (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. 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. For a given pair of traits, LD score regression 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 (Supplementary Information; see URLs). Additionally, we analysed on our local computer cluster, the genetic correlation of ADHD with eight phenotypes: human intelligence103, four phenotypes related to education and cognition analyzed in samples from the UK_Biobank (college/university degree, verbal–numerical reasoning, memory and reaction time), insomnia60, anorexia nervosa44, and major depressive disorder43. 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. (Supplementary Information), implemented by a freely available R script (see URLs). 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\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{127} or BIOS\textsuperscript{79}. Expression quantitative trait loci were annotated using FUMA (see URLs). We restricted to eQTL associations with false discovery rate (FDR) < 1e-3 within each dataset. (e) Chromatin states of each variant based on the 15-state chromHMM analysis of epigenomics data from Roadmap\textsuperscript{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 (see URLs). We also evaluated the annotated chromatin state from fetal brain.

**Gene-set analyses**

Gene-based association with ADHD was estimated with MAGMA 1.05\textsuperscript{88} using the summary statistics from the European GWAS meta-analysis ($N_{\text{cases}} = 19,099$, $N_{\text{controls}} = 34,194$; 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\textsuperscript{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\textsuperscript{86} curated from MsigDB 6.0\textsuperscript{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.\textsuperscript{89} and Spiteri et al\textsuperscript{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)\textsuperscript{91}. Genes with pLI ≥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 ($\pi$) using a one sample test of the proportion with Yates’ continuity correction\textsuperscript{129} against a null hypothesis of $\pi = 0.50$ (i.e. the signs are concordant between the two analyses by chance) in R\textsuperscript{130}. This test was evaluated separately for concordance in deCODE, 23andMe, and EAGLE/QIMR for loci passing P-value thresholds of $P < 5 \times 10^{-6}$ (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 performed by computing posterior mean estimates of marginal SNP effects $\beta_j$ after fitting a spike-and-slab 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.\textsuperscript{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\textsuperscript{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 (see URLs)\textsuperscript{42}. To evaluate the magnitude of the observed differences in $r_g$ we consider both the absolute difference (i.e. $|r_g, \text{ADHD} - r_g, \text{23andMe}|$) and the test of an approximate $Z$ score for this difference (Supplementary Information):

$$
Z = \frac{r_g, \text{ADHD} - r_g, \text{23andMe}}{\sqrt{SE_{r_g, \text{ADHD}}^2 + SE_{r_g, \text{23andMe}}^2}}
$$

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 account 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 $N_j$ and a continuous phenotype (e.g. ADHD-related traits) with sample size $N_2$, respectively, we calculate
where

\[
\tilde{Z}_{2j} = \text{sign}(r_g) \frac{Z_{2j}}{\sqrt{\left(1 - r_g^2 \right) N_2 h_2^2 j / M}}
\]

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

\[
\tilde{N}_{2j} = N_{2j} \frac{r_g^2 h_1^2 h_2^2 j}{1 + (1 - r_g^2) N_2 h_2^2 j / M}
\]

The adjusted sample sizes \(\tilde{N}_{1j}\) and \(\tilde{N}_{2j}\) reflect differences in power between the studies due to measurement scale and relative heritability that is not captured by sample size. The calculation of \(\tilde{Z}_{2j}\) 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 \((l)\) 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 Availability 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.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**Authors**

Ditte Demontis†,1,2,3, Raymond K. Walters†,4,5, Joanna Martin5,6,7, Manuel Mattheisen1,2,3,8,9,10, Thomas D. Als1,2,3, Esben Agerbo1,11,12, Gisli Baldursson13, Rich Belliveau5, Jonas Bybjerg-Grauholm1,14, Marie Bækvad-Hansen1,14, Felecia Cerrato5, Kimberly Chamberl5, Claire Churchhouse4,5,15, Ashley Dumont5, Nicholas Eriksson16, Michael Gandal17,18,19,20, Jacqueline I. Goldstein4,5,15, Katrina L. Grasby21, Jakob Grove1,2,3,22, Olafur O. Gudmundsson23,24,13, Christine S. Hansen1,14,25, Mads Engel Hauberg1,2,3, Mads V. Hollegaard1,14, Daniel P. Howrigan4,5, Hailiang Huang4,5, Julian B. Maller5,26, Alicia R. Martin4,5,15, Nicholas G. Martin21, Jennifer Moran5, Jonatan Pallesen1,2,3, Duncan S. Palmer4,5, Carsten Bøcker Pedersen1,11,12, Marianne Giørtz Pedersen11,12, Timothy Poterba4,5,15, Jesper Buchhave Poulsen1,14, Stephan Ripke4,5,27, Elise B. Robinson4,28, F. Kyle Satterstrom4,5,15, Hreinn Stefansson23, Christine Stevens5, Patrick Turley4,5, G. Bragi Walters23,24, Hyejung Won17,18, Margaret J. Wright29, ADHD Working Group of the Psychiatric Genomics Consortium (PGC)30, Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium30, 23 and Me Research Team30, Ole A. Andreasen31, Philip Asherson32, Christie L. Burton33, Dorret I. Boomsma34,35, Bru Cormand36,37,38,39, Søren Dalsgaard11, Barbara Franke40, Joel Gelernter41,42, Daniel Geschwind17,18,19, Hakon Hakonarson43, Jan Haavik44,45, Henry R. Kranzler46,47, Jonna Kuntsi32, Kate Langley7,48, Klaus-Peter Lesch49,50,51, Christel Middeldorp34,52,53, Andreas Reif54, Luis Augusto Rohde55,56, Panos Roussos57,58,59,60, Russell Schachar33, Pamela Sklar57,58,59, Edmund J. S. Sonuga-Barke61, Patrick F. Sullivan62,6, Anita Thapar7, Joyce Y. Tung16, Irwin D. Waldman63, Sarah E. Medland21, Kari Stefansson23,24, Merete Nordentoft1,64, David M. Hougaard1,14, Thomas Werge1,25,65, Ole Mors1,66, Preben Bo Mortensen1,2,11,12, Mark J. Daly4,5,15,67, Stephen V. Faraone*,68, Anders D. Børglum*,1,2,3, and Benjamin M. Neale*4,5,15

**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 Department of Psychiatry, Psychosomotics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany 11 National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark 12 Centre for Integrated Register-based Research, Aarhus University, Aarhus, Denmark 13 Department of Child and Adolescent Psychiatry, National University Hospital, Reykjavik, Iceland 14 Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark 15 Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA 16 23andMe, Inc, Mountain View, California, USA 17 Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA 18 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 19 Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA 20 Department of Psychiatry, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, California, USA 21 QIMR Berghofer Medical Research Institute, Brisbane, Australia 22 Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark 23 deCODE genetics/Amgen, Reykjavik, Iceland 24 Faculty of Medicine, University of Iceland, Reykjavik, Iceland 25 Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, Roskilde, Denmark 26 Genomics plc, Oxford, United Kingdom 27 Dept. of Psychiatry and Psychotherapy, Charité - Universitätsmedizin, Berlin 10117, Germany 28 Department of Epidemiology, Harvard Chan School of Public Health, Boston, Massachusetts, USA 29 Queensland Brain Institute, University of Queensland, Brisbane, Australia 30 A full list of consortium members can be found at the end of the article 31 NORMENT KG Jebsen Centre for Psychosis Research, Division of Mental Health and Addiction, University of Oslo and Oslo University Hospital, Oslo, Norway 32 Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, UK 33 Psychiatry, Neurosciences and Mental Health, The Hospital for Sick Children, University of Toronto, Toronto, Canada 34 Department of Biological Psychology, Neuroscience Campus Amsterdam, VU University, Amsterdam, The Netherlands 35 EMGO Institute for Health and Care Research, Amsterdam, The Netherlands 36 Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain 37 Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, Madrid, Spain 38 Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Catalonia, Spain 39 Institut de Recerca Sant Joan de Déu (IRJSD), Esplugues de Llobregat, Barcelona, Catalonia, Spain 40 Departments of Human Genetics (855) and Psychiatry, Donders Institute for Brain, Cognition and Behaviour,
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 294838), the European Community (EC) Horizon 2020 Programme (grant 667302 (CoCA)), from EC Seventh Framework Program.
Programme (grant 602805 (Aggressotype)), 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. Faraone is supported by the K.G. Jebsen Centre for Research on Neuropsychiatric Disorders, University of Bergen, Norway, the EC’s Seventh Framework Programme (grant 602805), the EC’s Horizon 2020 (grant 667302) and NIMH grants 5R01MH01519 and U01 MH109536-01.

Dr. J. Martin was supported by the Wellcome Trust (grant 106047).

Dr. B. 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 EC’s Seventh Framework Programme (grant 602805 (Aggressotype), 602450 (IMAGEMEND), and 278948 (TACTICS)), and from the EC’s Horizon 2020 Programme (grant 643051 (MiND) and 667302 (CoCA)). In addition, this work was supported by the European College of Neuropsychopharmacology (ECNP Network “ADHD across the Lifespan”.

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” (2017-SGR-738). Dr. Cormand, Dr. Reif and collaborators received funding from the EC’s Seventh Framework Programme (grant 602805, Agressotype), the EC’s H2020 Programme (grants 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 (grants: 223273, 248778, 213694, 249711), and KG Jebsen Stiftelsen.

Dr. Thapar received ADHD funding from the Wellcome Trust, Medical Research Council (MRC UK), Action Medical Research.

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 acknowledgements can be found in the Supplementary Note.

References


Nat Genet. Author manuscript; available in PMC 2019 July 01.


[PubMed: 27663502]


Nat Genet. Author manuscript; available in PMC 2019 July 01.


Methods-only references


Nat Genet. Author manuscript; available in PMC 2019 July 01.


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 European 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.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Chr</th>
<th>BP</th>
<th>Index Variant</th>
<th>Genes</th>
<th>A1</th>
<th>A2</th>
<th>A1 Freq</th>
<th>OR</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>44184192</td>
<td>rs11420276</td>
<td><strong>ST3GAL3, KDM4A, KDM4A-AS1, PTPRF, SLC6A9, ARTN, DPH2, ATP6V0B, B4GALT2, CCDC24, IPO13</strong></td>
<td>G</td>
<td>GT</td>
<td>0.696</td>
<td>1.113</td>
<td>2.14 $\times 10^{-13}$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>96602440</td>
<td>rs1222063</td>
<td>Intergenic</td>
<td>A</td>
<td>G</td>
<td>0.328</td>
<td>1.101</td>
<td>3.07 $\times 10^{-8}$</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>215181889</td>
<td>rs9677504</td>
<td><strong>SPAG16</strong></td>
<td>A</td>
<td>G</td>
<td>0.109</td>
<td>1.124</td>
<td>1.39 $\times 10^{-8}$</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>20669071</td>
<td>rs4858241</td>
<td>Intergenic</td>
<td>T</td>
<td>G</td>
<td>0.622</td>
<td>1.082</td>
<td>1.74 $\times 10^{-8}$</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>31151456</td>
<td>rs28411770</td>
<td><strong>PCDH7, LINCO2497</strong></td>
<td>T</td>
<td>G</td>
<td>0.651</td>
<td>1.090</td>
<td>1.15 $\times 10^{-8}$</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>87854395</td>
<td>rs4916723</td>
<td><strong>LINCO0461, LINCO0260, LINCO0260, TMEM161B-AS1</strong></td>
<td>A</td>
<td>C</td>
<td>0.573</td>
<td>0.926</td>
<td>1.58 $\times 10^{-8}$</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>114086133</td>
<td>rs5886709</td>
<td><strong>FOXP2, MIR3666</strong></td>
<td>G</td>
<td>GTC</td>
<td>0.463</td>
<td>1.079</td>
<td>1.66 $\times 10^{-8}$</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>34352610</td>
<td>rs74760947</td>
<td><strong>LINCO1288</strong></td>
<td>A</td>
<td>G</td>
<td>0.957</td>
<td>0.835</td>
<td>1.35 $\times 10^{-8}$</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>106747354</td>
<td>rs11591402</td>
<td><strong>SORCS3</strong></td>
<td>A</td>
<td>T</td>
<td>0.224</td>
<td>0.911</td>
<td>1.34 $\times 10^{-8}$</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>89760744</td>
<td>rs1427829</td>
<td><strong>DUSP6, POC1B</strong></td>
<td>A</td>
<td>G</td>
<td>0.434</td>
<td>1.083</td>
<td>1.82 $\times 10^{-9}$</td>
</tr>
<tr>
<td>11</td>
<td>15</td>
<td>47754018</td>
<td>rs281324</td>
<td><strong>SEMA6D</strong></td>
<td>T</td>
<td>C</td>
<td>0.531</td>
<td>0.928</td>
<td>2.68 $\times 10^{-8}$</td>
</tr>
<tr>
<td>12</td>
<td>16</td>
<td>72578131</td>
<td>rs212178</td>
<td><strong>LINCO1572</strong></td>
<td>A</td>
<td>G</td>
<td>0.883</td>
<td>0.891</td>
<td>7.68 $\times 10^{-9}$</td>
</tr>
</tbody>
</table>