

Discovery of the first genome-wide significant risk loci for ADHD

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

† Equal contributions. * Co-last authors.

Correspond with: Benjamin M. Neale (), 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. Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark
13. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA
14. 23andMe, Mountain View, California, United States of America
15. Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA
16. 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
17. Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA
18. Department of Psychiatry, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, California, USA
19. Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark
20. Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, Roskilde, Denmark
21. Genomics plc, Oxford, United Kingdom
22. Department of Psychiatry, Charite Universitätsmedizin Berlin Campus Benjamin Franklin, Berlin, Germany
23. Department of Epidemiology, Harvard Chan School of Public Health, Boston, Massachusetts, USA
24. KG Jebsen Centre for Psychosis Research, Norway Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway
25. Psychiatry, Neurosciences and Mental Health, The Hospital for Sick Children, University of Toronto, Toronto, Canada
26. MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
27. Department of Biological Psychology, Neuroscience Campus Amsterdam, VU University, Amsterdam, The Netherlands
28. EMGO Institute for Health and Care Research, Amsterdam, The Netherlands
29. Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain
30. Biomedical Network Research Centre on Rare Diseases (CIBERER), Barcelona, Spain
31. Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Spain
32. Institut de Recerca Pediàtrica Hospital Sant Joan de Déu, Barcelona, Catalonia, Spain
33. Departments of Human Genetics (855) and Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, The Netherlands

34. Department of Psychiatry, Genetics, and Neuroscience, Yale University School of Medicine, New Haven, Connecticut, USA
35. Veterans Affairs Connecticut Healthcare Center, West Haven, Connecticut, USA
36. The Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA
37. Department of Psychiatry, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
38. Veterans Integrated Service Network (VISN4) Mental Illness Research, Education, and Clinical Center (MIRECC), Crescenzo VA Medical Center, Philadelphia, PA, USA
39. Division of Molecular Psychiatry, ADHD Clinical Research Unit, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Germany
40. Department of Neuroscience, School for Mental Health and Neuroscience (MHENS), Maastricht University, The Netherlands
41. Child Health Research Centre, University of Queensland, Brisbane Australia
42. Child and Youth Mental Health Service, Children's Health Queensland Hospital and Health Service, Brisbane, Australia
43. Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital, Frankfurt, Germany
44. Department of Psychiatry, Faculty of Medicine, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
45. ADHD Outpatient Clinic, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil
46. Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA
47. Institute for Genomics and Multiscale Biology, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
48. Friedman Brain Institute, Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA
49. Mental Illness Research Education and Clinical Center (MIRECC), James J. Peters VA Medical Center, Bronx, New York, USA
50. School of Psychology, University of Southampton, Southampton, United Kingdom
51. Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium
52. Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom
53. Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC, USA
54. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
55. Department of Psychology, Emory University, Atlanta, Georgia, USA
56. Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Copenhagen, Denmark
57. Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
58. Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark
59. 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 school-age children and 2.5% of adults. Common genetic variants contribute substantially to ADHD susceptibility, but no individual variants have been robustly associated with ADHD. We report a genome-wide association meta-analysis of 20,183 ADHD cases and 35,191 controls ascertained from clinical interviews and/or medical records 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 and loss-of-function intolerant genes, as well as around brain-expressed regulatory marks. Additional analyses of three replication cohorts; a cohort of diagnosed ADHD, a self-reported ADHD sample and a study of quantitative measures of ADHD symptoms in the population, broadly support these findings while highlighting potential study-specific effects on genetic overlap with educational attainment. The strong concordance with GWAS of quantitative population measures of ADHD symptoms supports the hypothesis that clinical diagnosis of ADHD is an extreme expression of one or more continuous heritable traits.

Background

Attention-Deficit/Hyperactivity Disorder (ADHD) is a neurodevelopmental psychiatric disorder, that affects around 5% of children and adolescents and 2.5% of adults worldwide¹. ADHD is often persistent and markedly impairing with increased risk of harmful outcomes such as injuries², traffic accidents³, increased health care utilization^{4,5}, substance abuse⁶, criminality⁷, unemployment⁸, divorce⁴, suicide⁹, AIDS risk behaviors⁸, and premature mortality¹⁰.

Epidemiologic and clinical studies implicate genetic and environmental risk factors that affect the structure and functional capacity of brain networks involved in behavior and cognition¹, 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 siblings¹³. Twin studies also suggest that diagnosed ADHD represents the extreme tail of one or more heritable quantitative traits¹⁴. Additionally, family and twin studies report genetic overlap between ADHD and other conditions including antisocial personality disorder/behaviours¹⁵, cognitive impairment¹⁶, autism spectrum disorder^{17,18}, schizophrenia¹⁹, bipolar disorder²⁰, and major depressive disorder²¹.

Thus far genome-wide association studies (GWASs) to identify common DNA variants that increase the risk of ADHD have not been successful²². 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 ADHD²⁵ 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 depression²⁵, conduct problems²⁶, schizophrenia²⁷, continuous measures of ADHD symptoms^{28,29} and other neurodevelopmental traits²⁹ in the population. Genetic studies of quantitative ADHD symptom

scores in children further support the hypothesis that ADHD is the extreme of a quantitative trait³⁰.

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.

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), 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 has 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 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 single Chinese cohort had insufficient sample size for well-powered trans-ethnic modelling. (Supplementary Figure 7.B). 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.

In total, 304 genetic variants in 12 loci surpassed the threshold for genome-wide significance ($P < 5 \times 10^{-8}$; Figure 1, Table 1, Supplementary Figure 3.A2 – 3.N2). Results for the European ancestry meta-analysis were substantively similar (Supplementary Figure 2). No marker demonstrated significant heterogeneity between studies (Supplementary Figures 6 and 7.A) and no heterogeneity was observed between the Chinese and European ancestry cohorts (Supplementary Figure 7.B). 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 regression³⁷ 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 ($r_g = 1.02$, $SE = 0.32$).

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 11). 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% ($SE = 0.0012$). 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). 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 regression³⁷ and results from the European ancestry meta-analysis (Online Methods). Assuming a population prevalence of 5% for ADHD³⁹, we estimate that the liability-scale SNP heritability $h^2_{\text{snp}} = 0.216$ (SE = 0.014, $P = 8.18 \times 10^{-54}$). These estimated polygenic effects account for 88% (SE = 0.0335) of observed genome-wide inflation of the test statistics in the meta-analysis ($\lambda = 1.200$); 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.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.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 eTable 5). Forty-three phenotypes

demonstrated significant genetic overlap with ADHD ($P < 2.28 \times 10^{-4}$), including major depressive disorder⁴³, anorexia nervosa⁴⁴, educational outcomes⁴⁵⁻⁴⁹, obesity-related phenotypes⁵⁰⁻⁵⁵, smoking⁵⁶⁻⁵⁸, reproductive success⁵⁹, insomnia⁶⁰, and mortality⁶¹ (Figure 3; Supplementary Table 11). 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 eTable 1). 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 eTable 2). Broadly, the significant loci do not coincide with candidate genes proposed to play a role in ADHD⁶².

Here we highlight genes that are identified in the regions of association (see also Supplementary Table 4). 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 learning⁶³⁻⁶⁵. 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 school⁶⁸. On chromosome 10, the ADHD association is located intronic in *SORCS3*, which encodes a brain-expressed transmembrane receptor that is important for neuronal development and plasticity⁶⁹ and has previously been associated with depression^{43,70} and schizophrenia.

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 eTable 3). *DUSP6* encodes a dual specificity phosphatase⁷¹, 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 fibroblasts⁷⁶. *SEMA6D* is active in the brain during embryonic development, and may play a role in neuronal wiring⁷⁷. Furthermore, variants in *SEMA6D* have previously been associated with educational attainment⁷⁸.

Credible set annotations at the remaining loci are more diverse (Supplementary eTable 2). 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 blood⁷⁹ (Supplementary eTable 2). Missense mutations in *ST3GAL3* have been shown to cause autosomal recessive intellectual disability⁸⁰. Hi-C and eQTL annotations suggest multiple alternative genes however, including *PTPRF* (Supplementary eTable 3). The locus also includes an intergenic variant, rs11210892, that has previously been associated with schizophrenia³³.

On chromosome 5, the credible set includes links to *LINC00461* and *TMEM161B* (Supplementary eTable 2). The function of *LINC00461* is unclear, but the RNA has highly localized expression in the brain⁸¹ and the genome-wide significant locus overlaps with variants in *LINC00461* associated with educational attainment⁷⁸. Alternatively, a genome-wide significant SNP in this locus (rs304132) is located in *MEF2C-ASI*, of strong interest given previous associations between *MEF2C* and severe intellectual disability,⁸²⁻⁸⁴ cerebral malformation⁸³, depression⁷⁰, schizophrenia³³ and Alzheimer's disease⁸⁵, 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 terms⁸⁶ from MsigDB 6.0⁸⁷ using MAGMA⁸⁸ (Online Methods). Association results for individual genes are consistent with the genome-wide significant loci for the GWAS (Supplementary Table 5). Three independent sets of *FOXP2* downstream target genes^{89,90} were tested (Online Methods), none of which demonstrated significant association to ADHD (Supplementary Table 7). 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 function⁹¹ showed significant association with ADHD ($\beta = 0.062$, $P = 2.6 \times 10^{-4}$). We also find little evidence for effects in previously proposed candidate genes for ADHD⁶²; of the nine proposed genes only *SLC9A9* showed weak association with ADHD ($P = 3.4 \times 10^{-4}$; Supplementary Table 6). 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 eTable 4).

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 QIMR⁹² (2,798 adolescents). 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 Figures 15-16).

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 (<http://ldsc.broadinstitute.org/ldhub/>)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 from the 23andMe GWAS showing little to no genetic correlation with college completion ($r_g = 0.056$, compared to $r_g = -0.54$ for the primary ADHD GWAS; approximate $P = 1.1 \times 10^{-9}$ 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 was 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). 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 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 eTable 6, Supplementary Figures 17 and 20). Of the 12 significant loci from the primary ADHD GWAS, 4 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 reach genome-wide significance in the replication meta-analyses, of which three loci are significant in two of these analyses (Supplementary eTable 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 one locus from the ADHD GWAS meta-analysis (rs12410155: $I^2 = 97.2$, $P = 2.29 \times 10^{-9}$; Supplementary Figures 18-19). 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 attainment⁷⁸, 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 21-22, Supplementary eTable 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 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. schizophrenia³³. 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_{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 the central nervous system tissues, consistent with previous observations in schizophrenia and bipolar disorder⁴⁰. On the other hand, we do not observe substantial effects in most previously reported candidate genes for ADHD⁶².

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 mammals⁹⁴, and constrained genes likely to be intolerant of loss-of-function mutations⁹¹ 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 behaviour⁹⁵⁻⁹⁷ and increased risk of ADHD for children born to young parents⁹⁸⁻¹⁰⁰. Given the phenotypic^{101,102} and genetic¹⁰³ 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 Iceland¹⁰⁴. 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 ADHD¹⁰⁵⁻¹⁰⁸ 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 ADHD¹⁰⁹, 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 differences 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 the EAGLE³⁰/QIMR⁹² are much stronger and support the hypothesis that ADHD is the extreme expression of one or more heritable quantitative traits¹¹⁰. 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 population²⁹. 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.

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 SHAPEIT¹¹⁴ and IMPUTE2¹¹⁵, 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 $N_{\text{eff}} = 4/(1/N_{\text{cases}} + 1/N_{\text{controls}})$ 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 regression³⁷. 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 GCTA¹²⁰ 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

datasets³⁸. 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 regression³⁷ was used to evaluate 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 ADHD³⁹.

Partitioning of the heritability

SNP heritability was partitioned by functional category and tissue association using LD score regression⁴⁰. 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 annotation⁴⁰. 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 regression⁴². 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 Hub⁴¹ (Supplementary Information). 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_Biobank⁴⁹ (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.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 GRCh37¹²⁴. 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 v19¹²⁵. (c) Variants annotated as interacting with a given gene in Hi-C data from samples of developing human cerebral cortex during neurogenesis and migration¹²⁶. 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 GTEx¹²⁷ or BIOS⁷⁹.

Expression quantitative trait loci were annotated using FUMA (<http://fuma.ctglab.nl/>). 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 Roadmap¹²⁸. 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 evaluated the annotated chromatin state from fetal brain.

Gene-set analyses

Gene-based association with ADHD was estimated with MAGMA 1.05⁸⁸ 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 $-\log(\text{SNP P-value})$ 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 samples³⁴.

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 sets⁸⁶ curated from MsigDB 6.0⁸⁷.

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 al⁹⁰.

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 from deCODE, 23andMe, and 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 correction¹²⁹ against a null hypothesis of $\pi = 0.50$ (i.e. the signs are concordant between the two analyses by chance) in R¹³⁰. 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 performed 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 regression³⁷ 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):

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

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_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

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

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

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

The adjusted sample sizes \tilde{N}_1 and \tilde{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 \tilde{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 traits¹³¹.

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.

Availability of results

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

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).

URLs

LD-Hub, <http://ldsc.broadinstitute.org>

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>

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:10.1038/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 FBNP1L. *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. & Hagan, 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., Dalgaard, 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).
- 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 χ^2 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).

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

End notes

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. Statistical analyses were carried out on the Genetic Cluster Computer (<http://www.geneticcluster.org>) hosted by SURFsara and financially supported by the Netherlands Scientific Organization (NWO 480-05-003) along with a supplement from the Dutch Brain Foundation and the VU University Amsterdam. The GRAS data collection was supported by the Max Planck Society, the Max-Planck-Förderstiftung, and the DFG Center for Nanoscale Microscopy & Molecular Physiology of the Brain (CNMPB), Göttingen, Germany.

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.

The Yale-Penn site study was supported by National Institutes of Health Grants RC2 DA028909, R01 DA12690, R01 DA12849, R01 DA18432, R01 AA11330, and R01 AA017535 and the Veterans Affairs Connecticut and Philadelphia Veterans Affairs Mental Illness Research, Educational, and Clinical Centers. Genotyping services for a part of our genome-wide association study were provided by the Center for Inherited Disease Research and Yale University (Center for Genome Analysis). Center for Inherited Disease Research is fully funded through a Federal contract from the National Institutes of Health to The Johns Hopkins University (contract number N01-HG-65403).

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.

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 SGBP 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.

We gratefully acknowledge all the studies and databases that made GWAS summary data available: ADIPOGen (Adiponectin genetics consortium), C4D (Coronary Artery Disease Genetics Consortium), CARDIOGRAM (Coronary ARtery DIsease Genome wide Replication and Meta-analysis), CKDGen (Chronic Kidney Disease Genetics consortium), dbGAP (database of Genotypes and Phenotypes), DIAGRAM (DIAbetes Genetics Replication And Meta-analysis), ENIGMA (Enhancing Neuro Imaging Genetics through Meta Analysis), EAGLE (EARly Genetics & Lifecourse Epidemiology Consortium, excluding 23andMe), EGG (Early Growth Genetics Consortium), GABRIEL (A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community), GCAN (Genetic Consortium for Anorexia Nervosa), GEFOS (GENetic Factors for OSteoporosis Consortium), GIANT (Genetic Investigation of ANthropometric Traits), GIS (Genetics of Iron Status consortium), GLGC (Global Lipids Genetics Consortium), GPC (Genetics of Personality Consortium), GUGC (Global Urate and Gout consortium), HaemGen (haematological and platelet traits genetics consortium), HRgene (Heart Rate consortium), IIBDGC (International Inflammatory Bowel Disease Genetics Consortium), ILCCO (International Lung Cancer Consortium), MSGC (International Multiple Sclerosis Genetic Consortium), MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium), MESA (Multi-Ethnic Study of Atherosclerosis), PGC (Psychiatric Genomics Consortium), Project MinE consortium, ReproGen (Reproductive Genetics Consortium), SSGAC (Social Science Genetics Association Consortium) and TAG (Tobacco and Genetics Consortium), TRICL (Transdisciplinary Research in Cancer of the Lung consortium), UK Biobank.

We gratefully acknowledge the contributions of Alkes Price (the systemic lupus erythematosus GWAS and primary biliary cirrhosis GWAS) and Johannes Kettunen (lipids metabolites GWAS).

Author Contributions

Analysis:

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

Sample and/or data provider and processing:

DD, RKW, JM, MM, EA, GB, RB, JB-G, MB-H, FC, KC, AD, NE, JGo, JGr, OOG, CSH, MVH, JM, JM, CBP, MGP, JBP, SR, CS, OA, CB, DB, BC, SD, BF, JG, HH, JH, HK, JK, KL, KPL, CM, AR, LAR, RS, PS, ESB, AT, JT, IW, DMH, OM, PBM, ADB, ADHD Working Group of the Psychiatric Genomics Consortium, Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium, 23andMe Research Team.

Core PI group:

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.

Author Information

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

Disclosures

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.

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 Berhandlerling, 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.

Members of the 23andMe research team are employees of 23andMe.

Olafur O. Gudmundsson, G. Bragi Walters, Stacy Steinberg, Hreinn Stefansson and Kari Stefansson are employees of deCODE genetics/Amgen.

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

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.

Figure 2. Odds Ratio (OR) by PRS within each decile estimated for individuals in the PGC samples (red dots) and in the iPSYCH sample (blue dots). Error bars indicate 95% confidence limits.

Figure 3. Significant genetic correlations between ADHD 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. In total 220 traits were tested. Two significant educational phenotypes are omitted due to substantial overlap with years of schooling. Error bars indicate 95% confidence limits.

Figure 1.

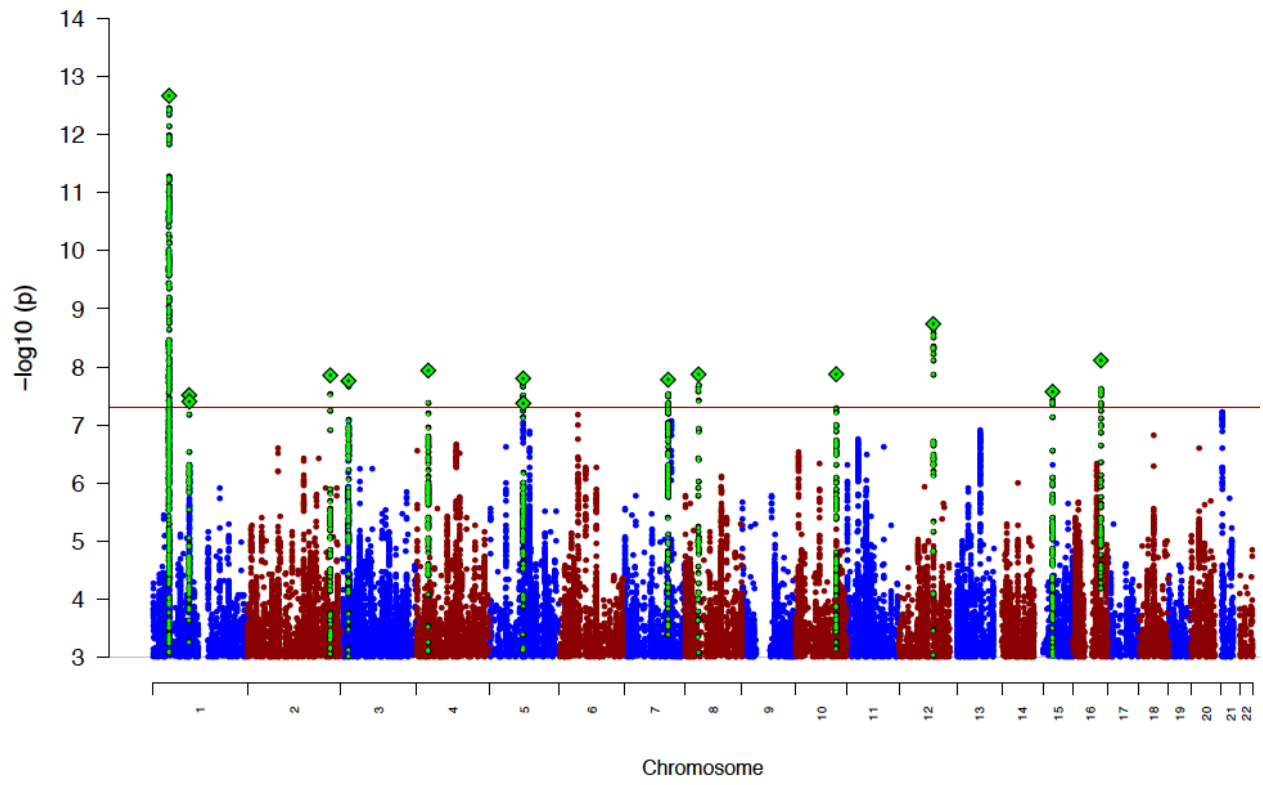


Figure 2.

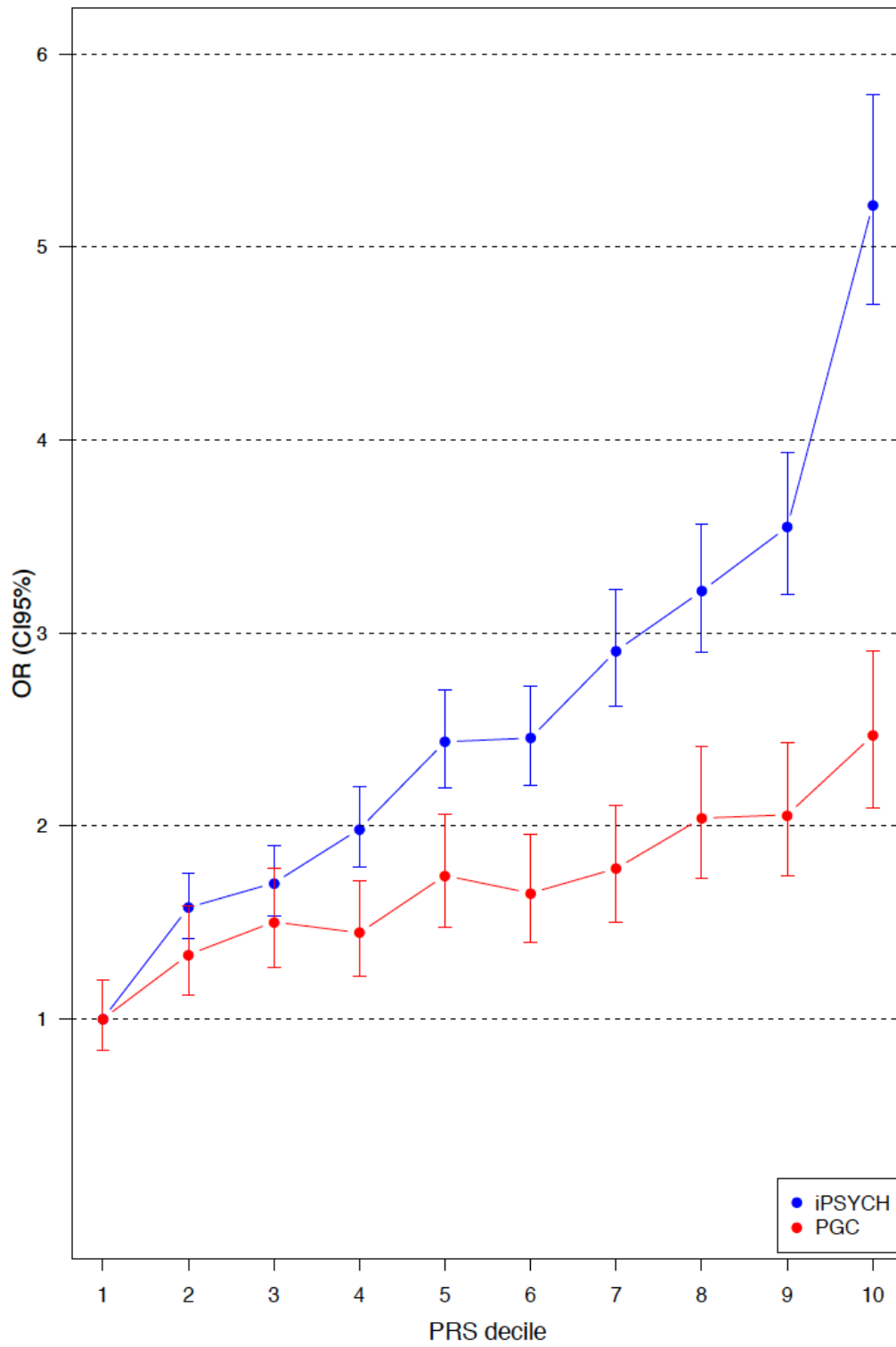


Figure 3.

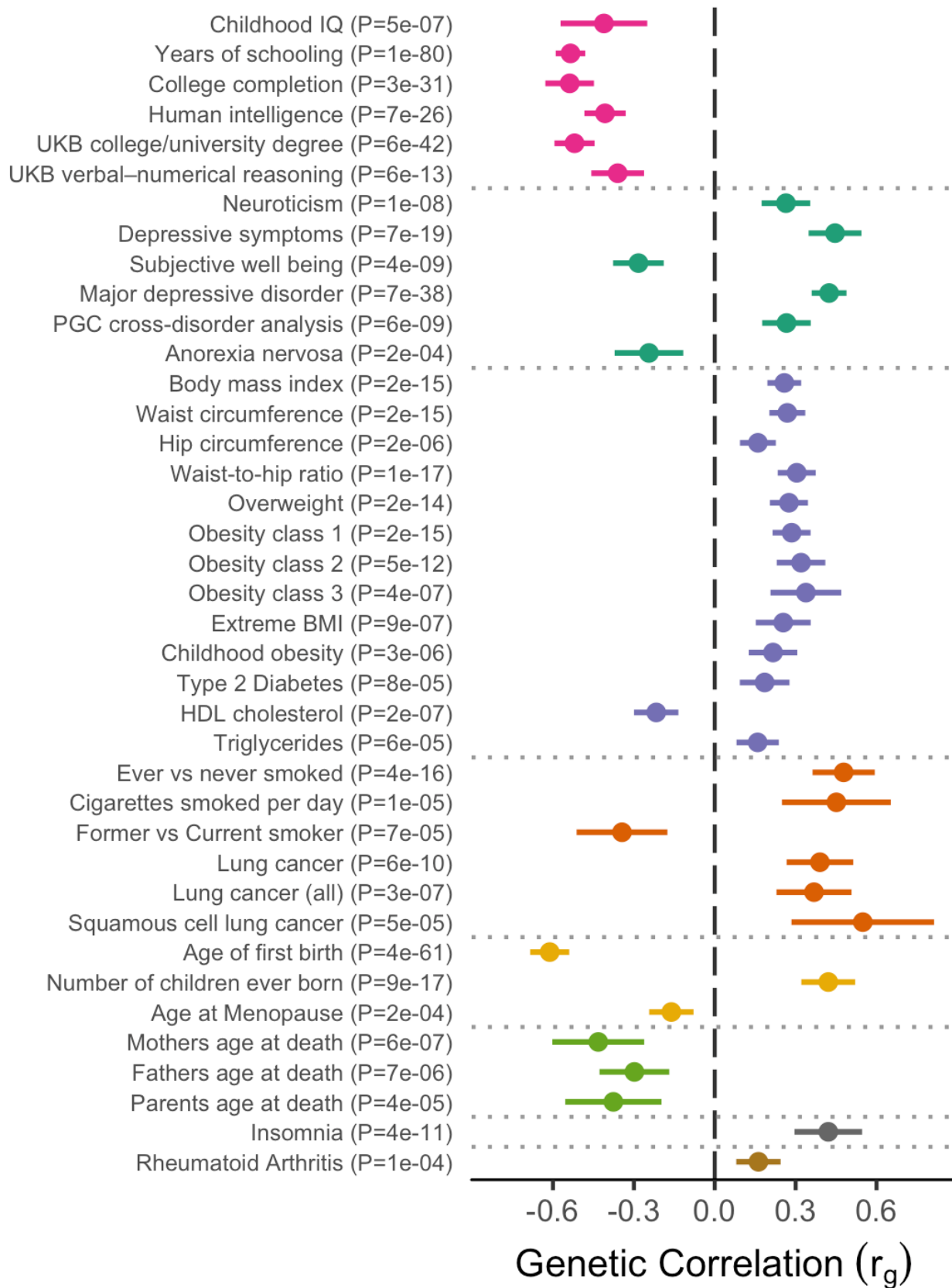


Table 1. Results for the genome-wide significant index variants in the 12 loci associated with ADHD identified in the GWAS meta-analysis. Index variants are LD independent ($r^2 < 0.1$), and are merged into one locus when located with a distance less than 400kb. The location (chromosome [Chr] and base position [BP]), alleles (A1 and A2), allele frequency (A1 Freq), odds ratio (OR) of the effect with respect to A1, and association P-value 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	<i>ST3GAL3, KDM4A, KDM4A-AS1, PTPRF, SLC6A9, ARTN, DPH2, ATP6V0B, B4GALT2, CCDC24, IPO13</i>	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	<i>SPAG16</i>	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	<i>PCDH7, LINC02497</i>	T	C	0.651	1.09	1.15×10^{-8}
6	5	87854395	rs4916723	<i>LINC00461, MIR9-2, LINC02060, TMEM161B-AS1</i>	A	C	0.573	0.926	1.58×10^{-8}
7	7	114086133	rs5886709	<i>FOXP2, MIR3666</i>	G	GTC	0.463	1.079	1.66×10^{-8}
8	8	34352610	rs74760947	<i>LINC01288</i>	A	G	0.957	0.835	1.35×10^{-8}
9	10	106747354	rs11591402	<i>SORCS3</i>	A	T	0.224	0.911	1.34×10^{-8}
10	12	89760744	rs1427829	<i>DUSP6, POC1B</i>	A	G	0.434	1.083	1.82×10^{-9}
11	15	47754018	rs281324	<i>SEMA6D</i>	T	C	0.531	0.928	2.68×10^{-8}
12	16	72578131	rs212178	<i>LINC01572</i>	A	G	0.883	0.891	7.68×10^{-9}

Extended Data

eTable 1. Bayesian credible sets of variants for each of the 12 genome-wide significant loci

eTable 2. Summary of the observed annotations for the credible set at each genome-wide significant locus

eTable 3. Variant-level annotations for the credible set at each genome-wide significant locus

eTable 4. Results of gene set analyses using sets from Gene Ontology

eTable 5. Extended results from genetic correlation analyses of ADHD and 219 phenotypes

eTable 6. Genome-wide significant index variants in meta-analyses of iPSYCH, PGC, deCODE, 23andMe and EAGLE

Supplementary Information

Discovery of the first genome-wide significant risk loci for ADHD

Table of Contents

Supplementary Methods and Results	4
Detailed description of individual samples	4
iPSYCH, Denmark	4
Samples from the Psychiatric Genomics Consortium (PGC).....	6
Parent-offspring trio samples	6
CHOP, USA	6
IMAGE-I, Europe	6
PUWMA, USA	7
Toronto, Canada	8
Case-control samples	9
Barcelona, Spain	9
Beijing, China	9
Bergen, Norway	10
Cardiff, UK.....	10
Germany	11
IMAGE-II, Europe & USA	12
Yale-Penn, USA	13
Replication samples.....	14
DeCODE, ADHD diagnoses from medical records.....	14
23andMe, self-reported ADHD diagnoses	15
EAGLE, ADHD symptom scores	16
QIMR, ADHD rating scale	17
Bioinformatics pipeline for quality control and association analyses.....	18
Pre-imputation quality control	19
Genotype imputation	19
Relatedness and population stratification	19
GWAS and meta-analysis.....	21
Defining independent genome-wide significant loci	22
Evaluating putative secondary signals.....	23
Correlation of secondary signals with their respective lead index variants	23
Conditional association analysis	24
Bayesian credible set analysis.....	24
Credible set estimation method.....	24
Variants considered for credible set analysis	26
Credible set results in PGC and iPSYCH data	26
Functional annotation of variants in credible set.....	27
Gene-based association analysis.....	28
Exome-wide association of single genes with ADHD.....	29

Gene-wise association of candidate genes for ADHD	30
Gene-set analyses	30
Hypothesis free gene set analyses	30
<i>FOXP2</i> downstream target gene set analysis	31
Highly constrained gene set analysis.....	31
LD Score intercept evaluation	32
Genetic correlations between PGC and iPSYCH ADHD samples	33
Genetic correlation between PGC case-control and trio samples	33
Polygenic risk scores for ADHD	34
Polygenic risk score prediction in iPSYCH samples.....	34
Polygenic risk score prediction in PGC samples	36
Leave-one-out analysis across cohorts.....	37
SNP heritability	37
Partitioning heritability by functional annotation and cell type	38
Genetic correlations of ADHD with other traits	40
Replication analysis in external cohorts	40
Sign test.....	41
Replication of effect sizes for top loci	43
Genetic correlation analysis of replication cohorts.....	44
Meta-analysis and heterogeneity test with each replication cohort	46
Meta-analysis across replication cohorts.....	49
Winner's curse correction for effect sizes.....	49
Method for meta-analysis of continuous and dichotomous ADHD measures.....	50
Basic genetic model for latent-scale phenotypes.....	51
Defining independent genetic factors	52
Effects of individual variants	53
GWAS test statistics for β_j	55
Test of β_{ij} in GWAS of dichotomous Y_1	57
Test of β_{ij} in GWAS of continuous Y_2	59
Meta-analysis for β_{ij} from GWAS of dichotomous Y_1 and continuous Y_2	62
Notes on Implementation	63
Supplementary Tables.....	66
Supplementary Table 1. Samples included in the GWAS meta-analyses of diagnosed ADHD	66
Supplementary Table 2. Conditional Analysis of Secondary GWAS Signals	68
Supplementary Table 3. Summary of Bayesian Credible Set Results	69
Supplementary Table 4. Biological function of potential ADHD risk genes located in genome-wide significantly associated loci	70
Supplementary Table 5. Results from MAGMA gene-based association with ADHD	75
Supplementary Table 6. Results from MAGMA gene-based association of ADHD candidate genes	76
Supplementary Table 7. Enrichment analysis of gene sets related to <i>FOXP2</i> downstream target genes.....	77
Supplementary Table 8. Enrichment analysis for a set of highly constrained genes	78
Supplementary Table 9. SNP heritability of ADHD	79
Supplementary Table 10. Heritability and genetic correlations for PGC ADHD samples	80
Supplementary Table 11. Genetic correlations of ADHD with other selected traits.....	81
Supplementary Table 12. Sign test results for each replication cohort	83
Supplementary Table 13. Results for significant ADHD loci in each replication cohort	84
Supplementary Table 14. Comparison of profile of genetic correlations for the ADHD GWAS and 23andMe..	85
Supplementary Figures.....	87
Supplementary Figure 1. Genotyping iPSYCH-ADHD sample, main steps and sample loss	87
Supplementary Figure 2. Manhattan plot from ADHD European GWAS meta-analysis	88
Supplementary Figure 3.A1 - M1. Forest plots for index SNPs in gws loci	89
Supplementary Figure 3.A2 - M2. Regional association plots for index SNPs in gws loci	89

Supplementary Figure 4.A – 4.D. Gene-based association, regional association plots.....	104
Supplementary Figure 5.A. – 5.B. Q-Q plot from GWAS meta-analyses	106
Supplementary Figure 6. Manhattan plot from test for heterogeneity between studies/waves in the ADHD GWAS meta-analysis.....	107
Supplementary Figure 7.A. Q-Q plot from test for heterogeneity between all samples/waves in the ADHD GWAS meta-analysis.....	108
Supplementary Figure 7.B. Q-Q plot from test for heterogeneity between Chinese and European ancestry cohorts in the ADHD GWAS meta-analysis.....	109
Supplementary Figure 8. Odds ratios by PRS within deciles in target groups.....	110
Supplementary Figure 9. Odds ratios within target groups in iPSYCH	111
Supplementary Figure 10. PRS-based odds ratios within each study/wave	112
Supplementary Figure 11. ADHD PRS stratified by case-control status and PGC study.....	113
Supplementary Figure 12. Partitioning of h^2 by functional annotations.....	114
Supplementary Figure 13. Partitioning of h^2 by tissue-group annotations.....	115
Supplementary Figure 14. Partitioning of h^2 by tissue-specific H3K4Me1 annotations.....	117
Supplementary Figure 15. Comparison of estimated effect sizes from deCODE	118
Supplementary Figure 16. Comparison of estimated effect sizes from 23andMe	119
Supplementary Figure 17. Manhattan plot of results from meta-analysis of ADHD+23andMe	120
Supplementary Figure 18. Q-Q plot from test for heterogeneity between ADHD GWAS meta-analysis and 23andMe	121
Supplementary Figure 19. Manhattan plot from test for heterogeneity between ADHD GWAS meta-analysis and 23andMe	122
Supplementary Figure 20. Manhattan plot of results from meta-analysis of ADHD+EAGLE/QIMR.....	123
Supplementary Figure 21. Q-Q plot from test for heterogeneity between ADHD GWAS meta-analysis and EAGLE/QIMR.....	124
Supplementary Figure 22. Manhattan plot from test for heterogeneity between ADHD GWAS meta-analysis and EAGLE/QIMR.....	125
Supplementary Figure 23. Distribution of 1000 Genomes Phase 3 European LD Scores.....	126
Supplementary Figure 24. Shrinkage factor for \tilde{Z}_j with varying l_j	127
Supplementary Figure 25. Relative effective sample size \tilde{N}_{ej} with varying l_j	129
References	131
Supplementary Notes.....	142
Consortium Members	142
ADHD Working Group of the Psychiatric Genomics Consortium (PGC)	142
Affiliations	142
Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium	145
Affiliations	145
23andMe Research Team	147

Supplementary Methods and Results

Detailed description of individual samples

iPSYCH, Denmark

A brief overview of the included samples can be found in Supplementary Table 1.

Since 1981 dried blood spot samples (Guthrie cards) from all newborn babies in Denmark have been stored in the Danish Newborn Screening Biobank (DNSB) at Statens Serum Institute (SSI). Samples from this nationwide biobank can be linked with the comprehensive Danish register system through the unique personal identification number (CPR-number), which is assigned to all live-born babies in Denmark. The CPR-number is stored in the Danish Civil Registration System (DCRS)¹ and is used in all contacts with the public sector, including all hospital contacts.

The iPSYCH-ADHD sample is a nationwide population based case-cohort sample selected from a baseline birth cohort comprising all singletons born in Denmark between May 1, 1981, and December 31, 2005, who were residents in Denmark on their first birthday and who have a known mother (N = 1,472,762). Cases were diagnosed by psychiatrists at in- or out-patient clinics, predominantly the latter according to ICD10 (F90.0 diagnosis code), identified using the Danish Psychiatric Central Research Register² (DPCRR). The DPCRR includes data on all people admitted to a psychiatric hospital for assessment, treatment, or both in Denmark since 1969 as well as people who attended psychiatric outpatient services since 1995. Diagnoses were given in 2013 or earlier for individuals at least 1 year old. Individuals with a diagnosis of moderate to severe mental retardation (ICD10 code F71-F79) were excluded. Controls were randomly selected from the same nationwide birth cohort and not diagnosed with ADHD (F90.0) or moderate-severe mental retardation (F71-F79). For both cases and controls all other comorbid diagnoses, except moderate-severe mental retardation, were allowed.

DNA was extracted from dried blood spot samples and whole genome amplified in triplicates as described previously^{3,4}. Genotyping was performed at the Broad Institute of Harvard and MIT (Cambridge, MA, USA) using Illumina's Beadarrays (PsychChip; Illumina, CA, San Diego, USA) according to the manufacturer's protocols. Genotypes were a result of merging callsets from three different calling algorithms (GenCall, Birdseed and Zcall). GenCall⁵ and Birdseed⁶ was used to call genotypes with minor allele frequency (MAF) > 0.01 and zCall⁷ was used to call genotypes with MAF < 0.01. The merging was done after pre-QC on individual call sets.

Processing of DNA, genotyping and genotype calling as well as imputing of genotypes of the iPSYCH-ADHD sample were carried out as a part of the genotyping of the full iPSYCH sample, which in total consists of around 79,492 individuals, including around 54,249 cases diagnosed with at least one of six mental disorders (schizophrenia, bipolar disorder, depression, ADHD, anorexia or autism spectrum disorder) and 26,248 randomly selected population controls (25,243 did not have any of the six psychiatric disorders investigated in iPSYCH). For the study of ADHD individuals with an ADHD diagnosis were exclude among the controls (N = 413). The data processing was done in 23 waves of approximately 3,500 individuals each. In order to control for potential batch effects, we included "wave" as a covariate in the regression models of all downstream analyses when relevant. Following genotyping all data processing, quality control, and downstream analyses were performed at secured servers in Denmark at the GenomeDK high performance-computing cluster (<http://genome.au.dk>). Overview of number samples in the iPSYCH study in the various steps, from identification in the registers to high quality genotypes included in the meta-analysis, can be found in Supplementary Figure 1.

The study was approved by the Danish Data Protection Agency and the Scientific Ethics Committee in Denmark.

Samples from the Psychiatric Genomics Consortium (PGC)

Parent-offspring trio samples

CHOP, USA

The CHOP (Children's Hospital of Philadelphia) ADHD trio sample (2,064 trios) were recruited from pediatric and behavioural health clinics in the Philadelphia area⁸ and included children aged 6–18 years from families of European with an ADHD diagnosis following the K-SADS (Schedule for Affective Disorders and Schizophrenia for School-Age Children; Epidemiologic Version) interview. Exclusion criteria were prematurity (<36 weeks), intellectual disability, major medical and neurological disorders, pervasive developmental disorder, psychoses and major mood disorders. Participants were assayed on the Illumina Infinium II HumanHap550 BeadChip (Illumina, San Diego, CA, USA) as previously described. The study was approved by The Children's Hospital of Philadelphia Institutional Review Board.

IMAGE-I, Europe

The IMAGE-I (International Multisite ADHD Genetics Project) trio samples^{9,10} were collected using a common protocol with centralized training and reliability testing of raters and centralized data management. Family members were Caucasians of European origin from countries in and around Europe including Belgium, Germany, Ireland, the Netherlands, Spain, Switzerland, and the United Kingdom, and Israel. At the IMAGE sites, parents of children were interviewed with the Parental Account of Childhood Symptom (PACS), a semi-structured, standardized, investigator-based interview developed as an instrument to provide an objective measure of child behaviour. Both parents and teachers completed the respective versions of the Conners ADHD rating scales and the Strengths and Difficulties Questionnaire (SDQ). Exclusion criteria were autism, epilepsy, IQ<70, brain disorders and any genetic or medical disorder associated with externalizing behaviours that

might mimic ADHD. Genotyping was conducted at Perlegen Sciences using their 600K genotyping platform, comprising approximately 600,000 tagging SNPs designed to be in high linkage disequilibrium with untyped SNPs for the three HapMap populations. The study was approved by the Institutional Review Board (IRB) or Ethical Committee at each site.

PUWMa, USA

The PUWMa (Pfizer-funded study from the University of California, Los Angeles (UCLA), Washington University, and Massachusetts General Hospital (MGH)) trio samples¹¹ were collected independently at those three sites using similar but slightly different methods.

309 families were recruited from clinics at MGH with children aged 6-17 years. Psychiatric assessments were made with the K-SADS-E. Exclusion criteria were major sensorimotor handicaps (deafness, blindness), psychosis/schizophrenia, autism, inadequate command of the English language, or a Full Scale IQ<80.

At Washington University, 272 families were selected from a population-representative sample identified through birth records of the state of Missouri, for a genetic epidemiologic study of the prevalence and heritability of ADHD. The original sample included 812 complete male and female twin pairs and six individual twins aged 7 to 19 years at the time of interview, identified from the Missouri Family Registry from 1996 to 2002. Families were invited into the study if at least one child exhibited three or more inattentive symptoms on a brief screening interview. Parents reported on their children and themselves, and the youths on themselves, using the Missouri Assessment of Genetics Interview for Children (MAGIC), a semi-structured psychiatric interview. DSM-IV diagnoses of ADHD were based upon parental reports (most of the time, maternal). Exclusion criteria were parent/guardian reported intellectual disability or if the parent/guardian and twins could not speak English.

At UCLA, 156 subjects were drawn from 540 children and adolescents aged 5 to 18 years and 519 of their parents ascertained from 370 families with ADHD-affected sibling pairs. Children and adolescents were assessed using the K-SADS-PL (Present and Lifetime version). Adult parents were assessed using the SADS-LA-IV (Lifetime version), supplemented with the K-SADS Behavioral Disorders module for diagnosis of ADHD and disruptive behavioural disorders. Direct interviews were supplemented with parent and teacher versions of the Swanson, Nolan, and Pelham, version IV (SNAP-IV) rating scale, as well as a parent-completed Childhood Behavior Checklist (CBCL) and Teacher Report Form (TRF). Exclusion criteria were neurological disorder, head injury resulting in concussion, lifetime diagnoses of schizophrenia or autism, or estimated Full Scale IQ<70. For all sites DNA was extracted from blood at each participating institution and Genizon BioSciences Inc. conducted genotyping with funding from Pfizer Inc. Genomic DNA samples from MGH and WASH-U were genotyped using the Illumina Human1M BeadChip (N = 1,057,265 SNPs), whereas the UCLA samples were genotyped using the Illumina Human 1M-Duo array (N = 1,151,846 SNPs). The study was approved by the subcommittee for human subjects of each site.

Toronto, Canada

The Canadian ADHD trio sample¹² was drawn from an outpatient clinic in an urban pediatric hospital and included children aged 6-16 years who were referred for attention, learning and/or behavioural problems. ADHD diagnostic data was obtained from parents and teachers in semi-structured clinical interviews including the Parent Interview for Child Symptoms (PICS) and the Teacher Telephone Interview (TTI). Exclusion criteria were an IQ<80 on both the verbal and the performance subscales of the Wechsler Intelligence Scale for Children (WISC). Samples were genotyped on the Affymetrix Genome-Wide Human SNP Array 6.0 with standard protocols as provided by the manufacturer. The study was approved by the Research Ethics Board of the Hospital for Sick Children, Toronto.

Case-control samples

Barcelona, Spain

The Barcelona sample¹³ comprised 607 ADHD cases. All patients were adults of Caucasian origin, recruited and evaluated at the Hospital Universitari Vall d'Hebron located in Barcelona (Spain). ADHD diagnostic criteria was assessed using the Structured Clinical Interview for DSM-IV and the Conner's Adult ADHD Diagnostic Interview for DSM-IV (CAADID). Impairment was measured with the Clinical Global Impression (CGI), included in the CAADID Part II, and the Sheehan Disability Inventory (SDI). Exclusion criteria were IQ<70, schizophrenia or other psychotic disorders, ADHD symptoms due to mood, anxiety, dissociative or personality disorders, adoption, sexual or physical abuse, birth weight <1.5 kg, and other neurological or systemic disorders that might explain ADHD symptoms. The control sample consisted of 584 unrelated blood donors frequency-matched for gender with the ADHD cases and screened to exclude those with lifetime ADHD symptoms or diagnosis.

Both cases and controls were genotyped on the Illumina HumanOmni1-Quad BeadChip platform. The study was approved by the relevant ethics committee.

Beijing, China

The Beijing, China sample¹⁴ comprised 1,040 ADHD cases aged between 6-16 years of Han Chinese decent. Cases were recruited from the Child and Adolescent Psychiatric Outpatient Department of the Sixth Hospital, Peking University. Clinical diagnoses from a senior child and adolescent psychiatrist were confirmed using the Chinese version of the Clinical Diagnostic Interview Scale. Exclusion criteria were those with major neurological disorders (e.g. epilepsy), schizophrenia, pervasive development disorder, and IQ<70. The 963 control individuals were students from local elementary schools, healthy blood donors from the Blood Center of the First Hospital, Peking University, and healthy volunteers from the institute of Han Chinese decent, screened using the

ADHD Rating Scale-IV (ADHD RS-IV) to exclude ADHD. Additional exclusion criteria were major psychiatric disorders, family history of psychosis, severe physical diseases, and substance abuse.

Both cases and controls were genotyped using the Affymetrix6.0 array at CapitalBio Ltd., Beijing, using the standard Affymetrix protocol.

The study was approved by the Institutional Review Board of the Peking University Health Science Center.

Bergen, Norway

The Bergen, Norway sample¹⁵ consisted of 300 adults with ADHD. Patients recruited through a Norwegian national medical registry, as well as by psychologists and psychiatrists working at outpatient clinics. Information regarding ADHD was obtained following systematic assessment of ADHD diagnostic criteria, developmental history, physical examination, evaluation of comorbidity, and, where possible, information from collateral informants. All gathered information was then sent to one of the expert committees for a definitive diagnostic assessment. There were no formal exclusion criteria. The 205 controls were recruited through the Medical Birth Registry of Norway above the age of 18 years with no known intellectual disability. Cases and controls were genotyped using the Human OmniExpress-12v1-1_B (Illumina, San Diego, CA, USA) platform. Genotyping was performed according to the standard Illumina protocol at Decode facility (Reykjavik, Iceland). The study was approved by the Norwegian Regional Medical Research Ethics Committee West (IRB #3 FWA00009490, IRB00001872).

Cardiff, UK

The Cardiff sample¹⁶ consisted of 727 Caucasian children aged 5-18 years old from Cardiff, Wales (N=510); St. Andrews, Scotland (N=35); and Dublin, Ireland (N=182). All children were recruited from community clinics and were assessed for ADHD using the Child and Adolescent Psychiatric Assessment (CAPA) Parent Version, a semi-structured research diagnostic interview, to assess

psychiatric diagnoses. Pervasiveness of ADHD symptoms (in school) was assessed using the Child ADHD Teacher Telephone Interview or the Conners Teacher Questionnaire. Exclusion criteria were intellectual disability (IQ <70), a major medical or neurological condition (e.g. epilepsy), autistic spectrum disorder, Tourette's syndrome, bipolar disorder, or known chromosomal abnormality. Control participants were obtained from the Wellcome Trust Case Control Consortium–Phase 2. They comprised 3,000 individuals born in the United Kingdom during 1 week in 1958 (the 1958 British Birth Cohort) and 3,000 individuals from the U.K. Blood Services collection (N=5,081 passed QC). The comparison subjects were not screened for psychiatric disorders. ADHD case subjects were genotyped on the Illumina (San Diego) Human660W-Quad BeadChip according to the manufacturer's instructions. Comparison subjects were genotyped by Wellcome Trust Case Control Consortium–Phase2 using the Illumina Human 1.2M BeadChip. The study was approved by the local research ethics committees at each site.

Germany

The German sample¹⁷ comprised 495 patients with ADHD (aged 6–18 years) recruited and phenotypically characterized in six psychiatric outpatient units for children and adolescents (Aachen, Cologne, Essen, Marburg, Regensburg, and Wurzburg). ADHD was assessed using the K-SADS-PL and a German teacher rating scale for ADHD (FBB-HKS). Exclusion criteria were IQ ≤ 75, potentially confounding psychiatric diagnoses such as schizophrenia, any pervasive developmental disorder, Tourette's disorder, and primary mood or anxiety disorder, neurological disorders such as epilepsy, a history of any acquired brain damage or evidence of the fetal alcohol syndrome, very preterm birth and/or (f) maternal reports of severe prenatal, perinatal or postnatal complications. The 1,300 adult controls were drawn from three population based epidemiological studies: (a) the Heinz Nixdorf Recall (Risk Factors, Evaluation of Coronary Calcification, and Lifestyle) study 3, (b) PopGen, (c) KORA (Cooperative Health Research in the Region of Augsburg). Ethnicity was assigned to patients

and controls according to self-reported ancestry (all German). The genome-wide genotyping was performed on HumanHap550v3 (Illumina; controls) and Human660W-Quadv1 BeadArrays (Illumina; cases). The study was approved by the ethics committees of all participating hospitals.

IMAGE-II, Europe & USA

The IMAGE-II ADHD case samples¹⁸ included some samples from the original IMAGE project (see IMAGE-I details above) along with samples provided by colleagues at other sites (Cardiff; St. Andrews, Dublin; MGH; Germany; and the Netherlands), using similar but not identical methods. Samples from Dublin and MGH followed the procedures described above for Cardiff and PUWMA, respectively. Case collection for the German and Dutch sites are described below.

In Germany, 351 participants were recruited in order of clinical referral in the outpatient clinics in Würzburg, Hamburg and Trier. Families were of German, Caucasian ancestry. All children were assessed by full semi-structured interview (Kiddie-Sads-PL-German Version or Kinder-DIPS) and parent and teacher ADHD DSM-IV based rating scales to ensure pervasiveness of symptoms. Exclusion criteria were IQ<80, comorbid autistic disorders or somatic disorders (hyperthyroidism, epilepsy, neurological diseases, severe head trauma etc.), primary affective disorders, Tourette's syndrome, psychotic disorders or other severe primary psychiatric disorders, and birth weight <2000 grams.

At the Dutch site, assessment data are available for 112 subjects aged 3-18 years with DSM-IV ADHD. Most of the sample was collected as part of a sib pair genome-wide linkage study in ADHD¹⁹. Subjects were assessed using the DSM-IV version of the Diagnostic Interview Schedule for Children (DISC-P) with both parents, supplemented by Conners Questionnaires (old versions), the CBCL and TRF. Exclusion criteria were autism, epilepsy, IQ <70, brain disorder, and any genetic or medical disorder associated with externalizing behaviours that might mimic ADHD.

Control samples (1,755 population controls of European ancestry) were assembled from an IRB approved genome-wide association study (GWAS) of myocardial infarction²⁰. Controls were collected from multiple sites in the US and Europe, including Seattle, Washington; Boston, Massachusetts; Gerona, Spain; Malmö, Sweden; and the United Kingdom. Sampling procedures for each cohort have been described previously²⁰. Control participants from the Wellcome Trust Case Control Consortium overlapping with the Cardiff, UK sample (described above) were removed. Cases were genotyped using the Affymetrix 5.0 array at the State University of New York Upstate Medical University, Syracuse using the standard protocol issued by Affymetrix. Controls were genotyped using the Affymetrix 6.0 array. The study was approved by the Institutional Review Board (IRB) or Ethical Committee at each site.

Yale-Penn, USA

The Yale-Penn sample consists of small nuclear families and unrelated individuals (2,020 individuals in 850 families and 6,951 unrelated individuals), collected to study the genetics of substance dependence²¹⁻²³. The case-control subjects were recruited from 2000 to 2013 from substance abuse treatment centers and through advertisements at the University of Connecticut Health Center, Yale University School of Medicine, the Medical University of South Carolina, the University of Pennsylvania, and McLean Hospital. The participants were identified through a family-based and a case-control protocol. Families were ascertained from treatment centers and advertisements that recruited affected sibling pairs (ASPs) meeting Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) criteria for cocaine or opioid dependence. Other family members of the ASPs were recruited when available, regardless of affection status and unaffected family members were included within the control subjects.

For this study, 182 individuals with ADHD and 1,315 unrelated controls of European ancestry were included. Unrelated individuals with ADHD and controls were selected from the family-based

protocol, with a focus on maximizing the number of ADHD cases retained for the analysis. *DSM-IV* diagnoses of ADHD case status, as well as other major psychiatric traits, were derived from the Semi-Structured Assessment for Drug Dependence and Alcoholism for all participants. Exclusion criteria were a clinical diagnosis of a major psychotic illness (for example, schizophrenia or schizoaffective disorder).

The sample was genotyped using one of two genotyping arrays: (1) the Illumina HumanOmni1-Quad v1.0 microarray containing 988,306 autosomal SNPs (Yale- Penn.1: performed at the Center for Inherited Disease Research (CIDR) and the Yale Center for Genome Analysis), (2) the Illumina Infinium Human Core Exome microarray (Yale-Penn.2 and Yale-Penn.3: performed at the Yale Center for Genome Analysis). The study was approved by the relevant institutional review boards.

Replication samples

DeCODE, ADHD diagnoses from medical records

The Icelandic ADHD cohort ($N = 5,085$) consists of individuals who either have a clinical ADHD diagnosis (mostly ICD10-F90) or who have been prescribed medication specific for ADHD symptoms (ATC-NA06BA, mostly methylphenidate). The Icelandic control individuals ($N = 131,122$) does not contain individuals with a diagnosis of schizophrenia, bipolar disorder, autism spectrum disorder or self-reported ADHD symptoms or diagnosis. All individuals used in the analysis have been chip genotyped, long range phased and genotypes imputed based on the Icelandic dataset as described previously²⁴.

GWAS in the Icelandic sample was performed for top loci with $P < 1 \times 10^{-5}$ (clumped variants with $r^2 < 0.25$) in the primary ADHD GWAS. Results from the Icelandic cohort were matched to the PGC and iPSYCH results based on rs-id and alleles. After filtering, results were available for 123 variants in the Icelandic sample from deCODE Genetics.

23andMe, self-reported ADHD diagnoses

The 23andMe sample consists of individuals who sent saliva samples (using the Oragene kit) to the genetic testing company 23andMe, Inc. and agreed to take part in research and answered questions about their ADHD history as part of a longer survey. All participants provided informed consent and answered surveys online according to 23andMe's human subjects protocol, which was reviewed and approved by Ethical & Independent Review Services, an AAHRPP-accredited institutional review board. As part of the "Your Medical History" survey, they were asked: "Have you ever been diagnosed by a doctor with any of the following psychiatric conditions: Attention deficit disorder (ADD) or Attention deficit hyperactivity disorder (ADHD)?" The response options were: "Yes", "No", "I don't know". A second question asked independently as a "Research Snippet" was: "Have you ever been diagnosed with attention deficit disorder (ADD) or attention deficit hyperactive disorder (ADHD)?" with the response options: "Yes", "No", "I'm not sure". Individuals who gave a positive response to these questions were classed as ADHD cases and controls were those who gave a negative response to these questions. Individuals with discordant responses were excluded.

Research participants were genotyped either on the Illumina HumanHap550k (13,030 controls, 840 cases) or HumanOmniExpress (57,363 controls, 5,017 cases) genotyping platforms by 23andMe. GWAS and imputation were performed separately for data generated by the two platforms. Within each platform, batches of 8,000-9,000 participants were imputed. Results were filtered for average and minimum imputation r^2 to exclude SNPs that showed batch effects. Covariates included in the GWAS by 23andMe were age, sex and the first four principal components to account for population stratification.

For the current study, the GWAS summary statistics were then aligned to the genotyped samples. The 23andMe summary statistics were verified to be consistent with genome build hg19. They were then

matched to the genotype data based on rsid, chromosome and base pair position. For SNPs, reported alleles were required to match the genotype data without a strand flip. For indels and multiallelic variants, alleles reported by 23andMe were evaluated heuristically for consistency with the genotype data and the alleles were matched accordingly (e.g. “I” or “D” alleles reported by 23andMe for indels were matched to the corresponding sequence of alleles for the insertion or deletion included in the genotype data). After alignment, the 23andMe GWAS results for the two platforms were combined in an inverse standard error-weighted meta-analysis to create a single 23andMe results set for use in the replication analyses. In total, 11,198,253 variants were matched from the 23andMe meta-analysis for inclusion in replication analyses.

EAGLE, ADHD symptom scores

The EARly Genetics and Lifecourse Epidemiology (EAGLE) consortium includes population-based birth cohorts from Europe, Australia, and the United States (<http://www.wikigenes.org/e/art/e/348.html>). The consortium focuses on a wide range of phenotypes in childhood including traits related to cognition and behaviour e.g. aggression²⁵, asthma allergy and atopy²⁶ and postnatal growth²⁷. In the study of ADHD symptoms, nine EAGLE cohorts were included with available ADHD symptom scores in childhood (age at measurement <13 years). An overview of the nine cohorts included in the EAGLE meta-analysis is provided in Middeldorp et al.²⁸. In order to assess ADHD symptoms different instruments were used across cohorts, including the Attention Problems scale of the Child Behavior Checklist (CBCL) and the Teacher Report Form (TRF), the Hyperactivity scale of the Strengths and Difficulties Questionnaire (SDQ), and the *DSM-IV* ADHD items as, for example, included in the Conners Rating Scale. For the meta-analysis, one phenotype was selected from each cohort. Based on the phenotype that was most available, school-age ratings were chosen over preschool-age ratings, parent ratings over teacher ratings, and the

measurement instrument with the largest information density was preferred over the other instruments²⁸.

Each of the included cohorts went through their own optimal pre-imputation QC and imputation was done using the March 2012 release of the Genomes Project (phase 1)²⁹. Detailed description of QC, imputation and the analysis procedures for the different cohorts can be found in Middeldorp et al.²⁸. Association analyses were done using linear regression and relevant principal components and subsequently meta-analysed using METAL³⁰. Summary statistics from the meta-analysis of N=17,666 individuals were provided for inclusion in the current study. These results were meta-analysed with results from QIMR as described below for inclusion in the replication analysis.

QIMR, ADHD rating scale

The QIMR (Queensland Institute of Medical Research) sample is drawn from the Brisbane Longitudinal Adolescent Twin Study which between 1992 and 2013 continuously recruited 12-year-old twins and their non-twin siblings from the greater Brisbane region³¹. Participants were recruited by contacting the principals of primary and secondary schools in the greater Brisbane area, media appeals and by word of mouth. It is estimated that approximately 50% of the eligible birth cohort were recruited into the study. The study used a longitudinal design in which participants were followed up at 14 and 16 years of age. In 2009 and 2015 additional cross-sectional follow-ups were conducted with participants over the ages of 19 and 25 respectively.

The phenotypic data used in the current analyses are maternal reports of ADHD symptoms for 2,798 individuals collected using the SWAN questionnaire³² which was introduced to the study protocol in 2010 and was completed during the first clinic visit after this date. The parents of participants who were over 14 years of age when the scale was added to the study were asked to complete the SWAN scale by online questionnaire in 2010.

Participants were genotyped using the Illumina Human610Quad BeadChip (Illumina, San Diego, CA, USA) as previously described. Genotypes were imputed to the 1000 genome references (Phase 3 Release 5) using the University of Michigan Imputation Server³³. Genome-wide association analysis was conducted in RAREMETALWORKER (<http://genome.sph.umich.edu/wiki/RAREMETALWORKER>) to correct for relatedness and included age, sex and ancestry PCs as covariates. The results of this GWAS have been previously described by Ebejer et al.³⁴. Ethical approval for the study was obtained from the QIMR Human Research Ethics Committee.

For the current study, summary statistics from the EAGLE and QIMR meta-analyses were aligned to the genotyped ADHD samples based on rsid and chromosome and base pair location. SNP alleles were required to be concordant without a strand flip. Indels and multiallelic variants with inconclusive matching to the ADHD GWAS were also excluded. After alignment and filtering for allele frequency and imputation quality, the EAGLE and QIMR results for 6,312,392 variants were meta-analysed to get a single genome-wide meta-analysis of ADHD-related behavioural traits. We denote this meta-analysis as EAGLE/QIMR, and focus on it for replication analyses.

Bioinformatics pipeline for quality control and association analyses

Quality control, imputation and primary association analyses were done using the bioinformatics pipeline “RicoPili”, which has been developed by the Psychiatric Genomics Consortium (PGC) Statistical Analysis Group³⁵. The pipeline generates high quality imputed data and performs GWAS and meta-analysis of large genetic data sets. In order to avoid potential study and “wave” effects the eleven PGC samples were processed separately and the iPSYCH sample was processed in 23 separate batches referred to as waves (see sample description above) unless otherwise is stated.

Pre-imputation quality control

Subjects and SNPs were included in the analyses based on the following quality control parameters: SNP call rate > 0.95 (before sample removal), subject call rate > 0.98 (> 0.95 for the iPSYCH samples), autosomal heterozygosity deviation ($|F_{het}| < 0.2$), SNP call rate > 0.98 (after sample removal), difference in SNP missingness between cases and controls < 0.02, and SNP Hardy-Weinberg equilibrium (HWE) ($P > 10^{-6}$ in controls or $P > 10^{-10}$ in cases).

Genotype imputation

In order to obtain information about non-genotyped markers, we used the pre-phasing software SHAPEIT³⁶ to estimate haplotypes and subsequently IMPUTE2³⁷ for imputing genotypes. Imputing was done in chunks of 3 Mb using default parameters. The imputation reference data consisted of 2,504 phased haplotypes from the 1000 Genomes Project, phase 3 (1KGP3)^{38,39} data (October 2014, 81,706,022 variants, release 20130502). Trio imputation was done with a case-pseudocontrol setup, where a pseudocontrol was defined to each affected offspring using the non-transmitted alleles from the two parents (estimated based on the haplotypes of the parents).

Relatedness and population stratification

Relatedness and population stratification were evaluated using a set of high quality markers (genotyped autosomal markers with minor allele frequency (MAF) > 0.05, HWE $p > 1 \times 10^{-4}$ and SNP call rate > 0.98), which were pruned for linkage disequilibrium (LD) resulting in a set of ~30,000 pruned markers (markers located in long range LD regions defined by Price et al.⁴⁰ were excluded). This was done separately for each of the PGC samples and on a merged set of genotypes from the 23 iPSYCH waves. In order to identify related individuals an identity by state analysis were performed

using PLINK v1.9^{41,42}, and one individual was excluded in pairs of subjects with $\hat{\pi} > 0.2$ (cases preferred kept over controls).

In order to identify genetic outliers, a principal component analysis (PCA) was performed using smartPCA incorporated in the software Eigensoft⁴³, and the same set of pruned autosomal markers as described above. For the iPSYCH sample a genetic homogenous sample was defined based on a subsample of individuals being Danes for three generations. This subsample was defined using register information about birth country of the individuals, their parents and grandparents, which was required to be Denmark in order to be included in the subsample. The subsample of Danes was used in order to define the center of an ellipsoid based on the mean values of principal component (PC) 1 and PC2. Subsequently PC1 and PC2 for all individuals in the iPSYCH sample were used to define a genetic homogenous population by excluding individuals with PC values greater than six standard deviations from the mean. For the PGC samples genetic outliers were removed based on visual inspection of the first six PCs. PCA including samples from the 1000 Genomes Project was also performed to confirm that the selected individuals matched the ancestry of European reference populations.

PCA was redone after exclusion of genetic outliers. The first 20 principal components were tested for association with the phenotype using logistic regression and their impact on the genome-wide test statistics were evaluated using λ . In the iPSYCH GWAS PC1-4 and significant PCs were included as covariates. For PGC samples, the number of PCs was adjusted based on the cohort's sample size in order to avoid overfitting and to reflect the differential power to capture true population structure by PCA. Specifically, the first five principal components were included as covariates for samples with fewer than 1000 individuals, and the first ten PCs were included for larger samples. Trio samples did not include PCs in the analysis unless strong population structure was evident (i.e. PUWMA). Where

necessary, study specific design covariates were also included (e.g. indicators variables for IMAGE-I sampling centres, alcohol dependence diagnosis for ascertainment in Yale-Penn).

GWAS and meta-analysis

Association analyses using the imputed marker dosages were performed separately for the 11 PGC samples and the 23 waves in iPSYCH by an additive logistic regression model using PLINK v1.9^{41,42}, with the derived principal components and necessary design factors included as covariates as described above (Supplementary Table 1). Adding sex as a covariate to the PGC and iPSYCH samples does not meaningfully alter the meta-analysis results (data not shown).

The meta-analysis included summary statistics from GWASs of the 23 waves in iPSYCH and 11 PGC samples, in total containing 20,183 cases and 35,191 controls. Only SNPs with imputation quality (INFO score) > 0.8 and MAF > 0.01 were included in the meta-analysis. Meta-analysis was performed using an inverse-weighted fixed effects model implemented in the software METAL (<http://csg.sph.umich.edu/abecasis/Metal/>)³⁰. Finally we filtered the GWAS meta-analysis, so only markers which were supported by an effective sample size ($N_{\text{eff}} = 4/(1/N_{\text{cases}} + 1/N_{\text{controls}})$)⁴⁴ greater than 70% were included (8,047,421 markers).

In some of the secondary analyses (e.g. when using LD score regression⁴⁵ and MAGMA⁴⁶ (see below)), information about LD structure in a reference genome reflecting the ancestry of the analysed population is used. Such analyses therefore require results from a GWAS meta-analysis based on a genetic homogenous group reflecting the same ancestry. We therefore performed a GWAS of the iPSYCH samples and the PGC samples with European ancestry (subsequently referred to as European GWAS meta-analysis). In this GWAS meta-analysis the Chinese PGC sample was excluded and the PUWMa sample was replaced with the PUWMa (strict) sample, in which individuals with non-

European genetic ancestry were excluded, resulting in 19,099 cases and 34,194 controls with European ancestry.

The GWAS meta-analysis identified 12 independent genome-wide significant (gws) loci (see Manhattan plot [Figure 1], forest plots [Supplementary Figure 3.A1 – 3.M1] and regional association plots [Supplementary Figure 3.A2 – 3.M2]). Independent loci were defined as described below. A more detailed description of potential risk genes located in the identified gws loci can be found in Supplementary Table 4.

In the European GWAS meta-analysis the number of independent gws loci decreased to 11. The gws locus on chromosome 2 (located in *SPAG16*) in the GWAS meta-analysis did not pass the significance threshold when only including individuals with European ancestry (see Manhattan plot, Supplementary Figure 2).

In addition, heterogeneity across studies/waves were tested with the Cochran's Q test and quantified with the I^2 heterogeneity index. No markers demonstrated significant heterogeneity between all cohorts (Supplementary Figures 6 and 7A) or between the Chinese and European ancestry cohorts (Supplementary Figure 7.B).

Defining independent genome-wide significant loci

303 variants reached genome-wide significance ($P < 5 \times 10^{-8}$) in the meta-analysis. We then identified independent loci from these markers based on LD clumping (--clump in PLINK 1.9^{41,42}). Beginning with the most significantly associated variant as the first index variant, we labelled variants as being part of the same locus if they were within 500 kb and correlated with an index variant ($r^2 > 0.2$). Variants not within 500 kb and not correlated with an existing index variant were labelled as a new index variant. Correlations were estimated from European-ancestry populations in the 1000 Genomes

Phase 3 reference panel³⁹. Clumping continued until all variants with $P < 5 \times 10^{-8}$ were either labelled as an index variant or assigned to a locus.

A gws locus was then defined as the physical region containing the identified LD independent index variants and their correlated variants ($r^2 > 0.6$) with $P < 0.001$. Associated loci located less than 400 kb apart were merged. The same process was applied to define independent genome-wide significant loci in subsequent meta-analyses.

Evaluating putative secondary signals

Correlation of secondary signals with their respective lead index variants

Two of the genome-wide significant loci defined by this process in the ADHD meta-analysis include more than one index variant (Supplementary Table 2). In other words, they contain two genome-wide significant variants that are within 500 kb but are not correlated ($r^2 < 0.1$). In this case, we label the less significantly associated index variant as a putative secondary signal and perform additional analyses to evaluate whether the second index variant can be confirmed as independent.

First, we confirmed that the putative secondary signals are not strongly correlated with their respective lead index variants in the current genotype data. The correlation with the index variant was evaluated in (1) imputed best-guess genotype data (hard-called genotypes derived from imputed genotype probabilities, for all variants with an imputation info score > 0.8) from the 11 PGC cohorts, and (2) imputed best-guess genotype data from iPSYCH. For both putative secondary effects, the correlation between the index variant and secondary effect is ($r^2 < 0.1$) in both the PGC and iPSYCH imputed genotype data (Supplementary Table 2). This confirms that the putative secondary signal does not reflect LD structure in the ADHD cohorts that is not well captured by the 1000 Genomes Phase 3 European reference panel³⁹.

Conditional association analysis

The two putative secondary signals were then 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, site indicators) were also included. The conditional association results were then combined in an inverse-variance weighted meta-analysis.

Neither of the putative secondary signals achieve genome-wide significance in the conditional association analysis (Supplementary Table 2). The decreased significance observed in the conditional analysis reflects modestly attenuated estimates of the odds ratio and increased standard errors compared to the marginal association analysis in the primary GWAS.

Based on the non-significant results for the putative secondary variants in the conditional analyses, we conclude that there is not yet sufficient evidence to confidently label these as independent effects in their respective loci.

Bayesian credible set analysis

In order to refine the genome-wide significant loci, we defined a credible set of variants in each locus using the method described by Maller et al.⁴⁷. 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.

Credible set estimation method

We summarize the method here following the description of Gormley et al.⁴⁸. Briefly, let D be the data including the genotype matrix X with P variants and the vector Y of phenotypes, and let β be the regression model parameters. Define P models A_j where variant j is causal and the remaining

variants are not causal, and define the null model A_0 where no variants are causal. Then by Bayes' rule the probability of model A_j is:

$$\Pr(A_j|D) = \int \Pr(D, \beta|A_j) \cdot \frac{\Pr(A_j)}{\Pr(D)} \cdot d\beta$$

Assuming a flat prior for the model parameters β , the integral can be approximated using the maximum likelihood estimates $\widehat{\beta}_j$, such that

$$\Pr(A_j|D) \approx \Pr(D|A_j, \widehat{\beta}_j) \cdot N^{-\frac{|\beta_j|}{2}} \cdot \frac{\Pr(A_j)}{\Pr(D)}$$

where N is the sample size and $|\beta_j|$ denotes the number of fitted parameters for model A_j . Given the assumption of one causal variant per locus, $|\beta_j|$ is a constant for all A_j . Next, note that the conventional likelihood ratio test of model A_j compared to the null model A_0 is defined as

$$\chi_j^2 \equiv -2 \log \frac{\Pr(D|A_0, \widehat{\beta}_0)}{\Pr(D|A_j, \widehat{\beta}_j)}$$

Thus by substitution,

$$\Pr(A_j|D) \approx \exp\left(\frac{\chi_j^2}{2}\right) \cdot l_0 \cdot N^{-\frac{|\beta_j|}{2}} \cdot \frac{\Pr(A_j)}{\Pr(D)}$$

with $l_0 = \Pr(D|A_0, \widehat{\beta}_0)$. Given a flat prior for models A_j the latter terms are constant, leaving

$$\Pr(A_j|D) \propto \exp\left(\frac{\chi_j^2}{2}\right).$$

Normalizing across all possible models A_j thus yields

$$\Pr(A_j) \equiv \Pr(A_j|D) / \sum_k \Pr(A_k|D).$$

Finally, the 99% credible set of variants is defined as the smallest set S of models such that

$$\sum_{A_j \in S} \Pr(A_j) \geq .99.$$

If the model assumptions are correctly specified, then this credible set S has a 99% probability of containing the true causal variant.

We implemented this approach using the published R script freely available online

(<https://github.com/haiianghuang/FM-summary>).

Variants considered for credible set analysis

We applied the Bayesian credible set analysis to each of the 12 genome-wide significant loci identified in the primary meta-analysis of ADHD as described above. For each locus, variants within 1MB and in linkage disequilibrium (LD) with correlation $r^2 > 0.4$ to the index variant were considered for inclusion in the credible set.

Because the credible set estimation is conditioned on LD structure, we performed the credible set analysis using the European GWAS meta-analysis to ensure consistent LD structure in the analysed cohorts. Credible sets were also estimated based on both (a) the observed LD in European ancestry PGC datasets, and (b) the observed LD in the iPSYCH dataset.

Observed LD with the index variant in each locus was computed using imputed best-guess genotype data (generated as described previously) with PLINK 1.9 (<https://www.cog-genomics.org/plink2>)⁴².

For the European ancestry PGC datasets, imputed genotype data was merged across cohorts prior to computed LD. For the iPSYCH dataset, imputed genotyped data for the 23 genotyping waves were similarly merged before computing LD.

Credible set results in PGC and iPSYCH data

Bayesian credible sets for each of the 12 genome-wide significant loci are reported in Supplementary eTable 1 (A-L). For the majority of the loci (7 of 12), there is no difference between the credible set results based on LD structure in the PGC datasets versus LD from the iPSYCH dataset (Supplementary Table 3). Differences between the credible sets for the remaining loci are modest,

with no more than six non-overlapping variants between the two sets for each locus. The non-overlapping variants also tend to have weak evidence for inclusion in the 99% credible set; of the 19 non-overlapping variants, only six would be included in an 90% credible set (i.e. the smallest set of variants with an 90% probability of containing the true causal variant under the Bayesian model), and only one would be included in an 80% credible set. To be conservative, we define the final credible set as the union of the credible sets estimated from the PGC and iPSYCH LD structure.

Functional annotation of variants in credible set

To evaluate the potential impact of the variants in the credible set for each locus, we consider annotations of predicted functional consequences for those variants based on external reference data.

In particular we evaluate:

- **Functional consequences:** Coding and regulatory consequences of each variant were annotated using the Ensembl Variant Effect Predictor (VEP⁴⁹) for genome build GRCh37 (hg19). Annotated consequences for transcripts without a HGNC gene symbol (e.g. clone-based vega genes) were excluded. Gene names were updated to the current HGNC gene symbol where applicable. For each variant, we summarize (a) annotated genes, excluding “upstream” and “downstream” annotations; (b) genes with an annotated consequence (i.e. excluding intronic annotations); and (c) annotated regulatory regions.
- **Transcription start site (TSS):** We annotate variants within 2kb upstream of the TSS of at least one gene isoform based on Gencode v19⁵⁰.
- **Hi-C interactions:** Variants annotated as physically interacting with a given gene were identified based on Hi-C data from samples of developing human cerebral cortex during neurogenesis and migration⁵¹. 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.

- Expression quantitative trait loci (eQTLs): SNPs associated with gene expression were annotated using FUMA (<http://fuma.ctglab.nl/>). Annotated eQTLs were identified from GTEx v6⁵² and BIOS⁵³, and filtered for false discovery rate (FDR) $< 1 \times 10^{-3}$ within each dataset. Annotations were updated to current HGNC gene symbols where applicable. For variants with multiple eQTL associations, we summarize the strongest eQTL association (i.e. the association with the lowest P-value) from each dataset.
- Chromatin state: Chromatin states for each variant were annotated based on the 15-state chromHMM analysis of epigenomics data from Roadmap⁵⁴. For each SNP, the most common chromatin state across 127 cell types was annotated using FUMA (<http://fuma.ctglab.nl/>). For all variants, we also annotate the predicted chromatin state in fetal brain. The 15 states are summarized 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), repressive chromatin marks (Heterochromatin, Bivalent TSS, Flanking Bivalent TSS, Bivalent Enhancer, Repressed Polycomb, or Weak Repressed Polycomb), or quiescent.
- CADD: Combined Annotation Dependent Depletion (CADD v1.3⁵⁵) scores were annotated for each SNP using FUMA (<http://fuma.ctglab.nl/>).

Supplementary eTable 2 summarizes the observed annotations for the credible set at each locus (see also Supplementary eTable 3 for variant-level annotations).

Gene-based association analysis

Exome-wide association of single genes with ADHD

Gene-based association with ADHD was estimated by MAGMA 1.05⁴⁶ using the summary statistics from the European GWAS meta-analysis ($N_{\text{cases}} = 19,099$; $N_{\text{controls}} = 34,194$; Supplementary Table 1) and summary statistics from the GWAS meta-analysis including the EAGLE sample ($N_{\text{total}} = 70,959$). We annotated SNPs to genes within their transcribed regions using the NCBI 37.3 gene definitions provided with MAGMA. We then calculated gene P-values using the SNP-wise mean model in which the sum of $-\log(\text{SNP P-value})$ is used as a test statistic. The gene P-value was calculated using a known approximation of the sampling distribution⁵⁶. MAGMA accounts for gene-size, number of SNPs in a gene and LD between markers. When using summary statistics in estimating gene-based P-values, MAGMA corrects for LD based on estimates from reference data with similar ancestry; for this we used the UKGP3, European ancestry samples, as the reference³⁹.

In total 20 genes demonstrated significant gene-wise association with ADHD after Bonferroni correction (correction for 17,877 genes; Supplementary Table 5). 11 genes were located in the complex region on chromosome 1 demonstrating strong gws association with ADHD in the single marker GWAS meta-analysis (see regional association plot for this region, Supplementary Figure 3.A2). Additional five genes overlapped with loci with gws single markers (*MEF2C*, *FOXP2*, *SORCS3*, *DUSP6* and *SEMA6D*). Four genes (*MANBA*, *CUBN*, *PIDDI*, *CDH8*) not located in single marker gws loci showed significant association (Supplementary Table 5). The LD region around three of the genes (*MANBA*, *CUBN*, *CDH8*) contains only the respective genes, indicating that the gene-based association signals were driven by markers in the genes and were unlikely to be caused by extended LD with markers in neighbouring gene loci (see regional association plots for the four new genes; Supplementary Figure 4.A. – 4.D.).

Gene-wise association of candidate genes for ADHD

Prior to the availability of large-scale whole-genome methods and technologies, many candidate genes have been examined in relation to ADHD. Such candidate gene studies frequently fail to replicate⁵⁷ and are likely to have been affected by publication bias, so it is unclear how many of the reported candidate genes are actually robustly associated with ADHD. As such, we set out to examine what the evidence for association is for the most highly studied candidate genes for ADHD, obtained from a recent review⁵⁸, in the current GWAS dataset. Annotated ADHD SNP results were tested for enrichment in each of these candidate genes using MAGMA to obtain overall gene P-values. The results do not show any support for the majority of the candidate genes that have been historically studied in relation to ADHD (Supplementary Table 6). The only exception is *SLC9A9* which shows nominal enrichment.

Gene-set analyses

Hypothesis free gene set analyses

For gene set analyses, we applied MAGMA⁴⁶. The analyses were based on the gene-based *P*-values generated as described above under “Gene-based association analysis”, based on summary statistics from the European GWAS meta-analysis. Those *P*-values were used to analyse sets of genes in order to test for enrichment in association signals in genes belong to specific to biological pathways or processes. MAGMA applies a competitive test to analyse if the genes of a gene set are more strongly associated with the trait than other genes, while correcting for a series of confounding effects such as gene length and size of the gene-set. In the analysis only genes on autosomes, and genes located outside the broad MHC region (hg19:chr6:25-35M) were included in the analysis. We applied no padding around genes. We used the gene names/locations and the European genotype reference panel

provided with the program. For gene sets we used the Gene Ontology⁵⁹ sets curated in MsigDB 6.0⁶⁰ keeping only gene sets with 10-1000 genes. No gene-sets remained significant after correction for multiple testing (Supplementary eTable 4).

***FOXP2* downstream target gene set analysis**

Targeted gene set analyses were run in MAGMA to determine whether *FOXP2* downstream target gene sets are enriched in ADHD. Three sets of genes 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, 2) Genes showing differential expression in wild type compared to *FOXP2* knockout mouse brains, 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. Curated lists of high-confidence genes were obtained from Vernes et al.⁶¹ and Spiteri et al.⁶². Mouse genes were mapped to human orthologues using MGI and NCBI. *FOXP2* was excluded, only 8 genes were present on more than one list and only one gene was present on all three lists (*NRN1*). ADHD SNP results were annotated using MAGMA and used for gene set analyses. Competitive P-values, using a conditional model to correct for confounding due to gene size and gene density were obtained for each gene set. The results showed no evidence of enrichment for any of these gene sets (Supplementary Table 7).

Highly constrained gene set analysis

We assessed whether genes that are intolerant to loss of function and thereby highly evolutionarily constrained are enriched in ADHD. The set of highly constrained genes was defined using a metric for probability of being loss-of-function (LoF) intolerant (pLI) based on the observed and expected protein-truncating variant (PTV) counts within each gene in a very large study of exome data (the

Exome Aggregation Consortium; ExAC)⁶³. Genes with observed <10% of expected PTVs were deemed haploinsufficient or highly constrained. Publically available results based on the full ExAC dataset were downloaded from:

ftp://ftp.broadinstitute.org/pub/ExAC_release/release0.3.1/functional_gene_constraint

Genes with $pLI \geq 0.9$ were selected as the set of highly constrained genes. Annotated ADHD SNP results were tested for enrichment in this gene set in MAGMA using a competitive gene set analysis. Results showed a significant enrichment of this set of genes (see Supplementary Table 8).

LD Score intercept evaluation

A strong deviation from null was observed in the distribution of the test statistics in the quantile-quantile plot (Q-Q plot) of the results from the GWAS meta-analyses (Supplementary Figure 5.A - 5.B.). When using LD score regression it is possible to distinguish the contribution of polygenicity from other confounding factors such as cryptic relatedness and population stratification to the deviation in the distribution of the test statistics⁴⁵. Under this model when regressing the chi-square statistics from GWAS against LD scores (pre-computed LD-scores downloaded from <https://github.com/bulik/ldsc>) for each SNP, the intercept minus one is an estimator for the mean contribution of confounding bias to the inflation in the test statistics. LD score regression analysis of the European GWAS meta-analysis estimated that the intercept was close to one (intercept = 1.04 (SE = 0.01)). Additionally, the ratio (ratio = (intercept-1)/(mean(chi²)-1)), which estimate of the proportion of the inflation in the mean chi-square that the LD Score regression intercept ascribes to causes other than polygenic heritability was estimated to ratio = 0.12 (SE = 0.03), indicating that the strong inflation in the distribution of the test statistics is caused primarily by polygenicity rather than confounding.

Genetic correlations between PGC and iPSYCH ADHD samples

In order to estimate the overlap in shared genetic risk factors between samples, genetic correlations r_g were calculated using LD Score regression⁴⁵. Pre-computed LD scores for HapMap3 SNPs calculated based on 378 phased European-ancestry individuals from the 1000 Genomes Project were used in the analysis (LD scores available on <https://github.com/bulik/ldsc>) and the summary statistics from European GWAS meta-analysis (iPSYCH + PGC European samples) and the PGC European samples²⁸. The r_g estimate was left unbounded in order to obtain unbiased estimates of SE. Only results for markers with an imputation INFO score > 0.90 were included in the analysis. The estimated genetic correlation between iPSYCH and PGC European GWAS was highly significant ($r_g = 1.17$; $SE = 0.2$; $P = 7.98 \times 10^{-9}$) and did not suggest imperfect correlation (i.e. $r_g < 1$) of common genetic risk factors between the studies.

Genetic correlation between PGC case-control and trio samples

The PGC European ancestry dataset consisted of two kinds of association study designs: case-control (Bergen, Cardiff, Germany, IMAGE-II, Spain, Yale-Penn) and trios (CHOP, Canada, IMAGE-I, PUWMA). A previous analysis of the PGC samples showed a genetic correlation of 0.71 ($SE=0.17$) across case-control and trio studies⁶⁴. We repeated this analysis within this newer set of PGC ADHD data.

For each of the PGC studies, best guess genotype data were generated using Ricopili and strictly filtered ($MAF > 0.05$, in addition to previous frequency, imputation quality and other filters). Genotypes were merged together across studies using PLINK. Asymmetric/ambiguous (AT, TA, CG, GC) and duplicate position SNPs were excluded. GCTA⁶⁵ was used to calculate a genomic relationship matrix for all individuals in this merged PGC sample for HapMap-3 SNPs. Analyses were based on 191,466 SNPs. One of each pair of individuals related at the level of 2nd cousins (pi-

hat>0.05) was excluded, preferentially keeping cases; this excluded: N=16 cases and N=91 controls. PCA was performed on the merged, unrelated samples using PLINK. The first 10 principal components as well as binary study/wave indicators were used as covariates for subsequent analyses. Univariate GREML analyses in GCTA were used to estimate SNP- h^2 on the liability scale (assuming a population prevalence of 5%) in the case-control samples and the trio samples separately. The GREML method was used for consistency with the previously published comparison of trio and case-control ADHD cohorts, and to accommodate the smaller sample sizes of the PGC subsets. As in the primary GWAS, trio studies were analysed using a case/pseudo-control design, where the pseudo-control is composed of the un-transmitted chromosomes from the parents of the proband. Bivariate GREML was then used to estimate the genetic correlation across these sub-cohorts.

The genetic correlation between the trio and case-control cohorts was strong and indistinguishable from 1 ($r_g=1.02$, $SE=0.32$), though the standard error remains quite large (Supplementary Table 10). The observed SNP- h^2 estimates were somewhat lower than the overall SNP- h^2 estimated in the primary analyses for the full meta-analysed results, consistent with the somewhat lower SNP- h^2 estimated from the PGC samples compared to iPSYCH (see SNP heritability analysis below).

Polygenic risk scores for ADHD

In addition to the genetic correlation analyses, we performed analyses of polygenic risk scores (PRS) to evaluate the consistency of common genetic effects and their predictive power across cohorts. We specifically considered PRS prediction within the iPSYCH samples, within the PGC cohorts, and in leave-one-out analysis across all cohorts.

Polygenic risk score prediction in iPSYCH samples

For analysis with the iPSYCH sample as the target cohort, the 23 genotyping-waves within the iPSYCH sample were split into five groups, aiming for approximately equal numbers of ADHD cases

within each group. We then conducted two sets of five leave-one-out analyses, with each leave-one-out analysis using four out of five iPSYCH groups as training datasets for estimation of SNP weights and then applying those weights to estimate PRS for the remaining target group⁶⁶. One set of leave-one-out analyses was performed with PGC European samples among the training datasets, while the other was without (only iPSYCH). The meta-analysis of the training samples was conducted using a SNP list filtered for minor allele frequency > 0.01 and an imputation threshold score above 0.8 intersecting across waves. Indels and variants in the extended MHC region (chromosome 6: 25-34 Mb) were also removed. Meta-analysis and “clumping” of significant SNPs was conducted using the Ricopili pipeline³⁵. PRS were then estimated for each target sample using a range of meta-analysis P-value thresholds (5×10^{-8} , 1×10^{-6} , 1×10^{-4} , 1×10^{-3} , 0.01, 0.05, 0.1, 0.2, 0.5, 1.0), multiplying the natural log of the odds ratio of each variant by the allele-dosage (imputation probability) of each variant. Whole genome PRS were obtained by summing values over variants for each individual. For each of the five groups of target samples PRS were normalized (subtracting the mean and dividing by the standard deviation), and the significance of the case-control score difference was tested by standard logistic regression including the first six principal components and a dummy variable indicating genotyping wave as covariates (using the `glm()` function of *R* 3.2.2). For each target group and for each P-value threshold the proportion of variance explained (i.e. Nagelkerke’s R^2) was estimated, comparing the full model to a reduced model without PRS and covariates only. The mean of the maximum Nagelkerke’s R^2 across P-value thresholds for each group was $R^2 = 0.055$ (SE = 0.055, range 0.047 – 0.06). For the P-value threshold with the highest Nagelkerke’s R^2 , odds ratios for PRS decile groups compared to the lowest decile were estimated for each target group (Supplementary Figure 8) and for the normalized score pooled across groups (Figure 2). Odds ratios were also estimated using logistic regression on the continuous scores for each target group separately

and finally an OR based on all samples was estimated by using the normalized PRS across all groups (Supplementary Figure 9).

Polygenic risk score prediction in PGC samples

Next the predictive utility of PRS was evaluated in the PGC samples. All European ancestry PGC best guess genotype data were merged together and iPSYCH-only summary statistics were used to calculate PRS in the PGC samples, using the approach described above. PRS in the PGC dataset were based on 9,323 clumped SNPs with $P < 0.1$ in the iPSYCH sample. The association between ADHD PRS and case status was significant in the merged PGC sample (OR=1.26 (1.22-1.31), variance explained on the liability scale (R^2) = 0.0103, $P = 2.4E-35$). Figure 2 displays odds ratios for ADHD case status by ADHD PRS decile for the PGC datasets. In the merged dataset, PRS were converted to deciles (where 1 was the lowest decile and 10 was the highest). Deciles 2-10 were then compared to the lowest decile using logistic regression including PCs as covariates. There is a clear pattern of increasing ORs with increasing decile.

To examine variation that could be related to differences in ascertainment of cases and controls within the PGC sample, mean PRS (residualised for PC covariates) were plotted stratified by case status and study (see Supplementary Figure 11). PGC cases had consistently higher PRS than PGC controls in the same study. There is some variation in PRS z-score across cases in different studies, for example with ADHD cases from the Cardiff (UK) sample having particularly high scores. Within controls, individuals within the Yale-Penn study have particularly high PRS; this may be due to this sample's ascertainment for the primary phenotype of substance abuse, with high levels of these problems in both the cases and controls^{21-23,67}. Variation in PRS in controls from different studies may be due to differences in ascertainment (e.g. pseudo-controls, screened or unscreened controls).

Leave-one-out analysis across cohorts

The odds ratio based on PRS over all PGC and iPSYCH waves/studies was also evaluated using a leave-one-study/wave-out approach. First, GWAS analyses of imputed dosage data were run for all samples in each PGC study and iPSYCH wave separately, as described previously, co-varying for relevant PCs. Meta-analyses using METAL³⁰ (with the STDERR scheme) were run excluding one set of summary results at a time, for each combination of studies. For each set of discovery results, LD-clumping was run to obtain a relatively independent set of SNPs, while retaining the most significant SNP in each LD block. The following parameters were applied in PLINK: --clump-kb 500 --clump-r2 0.3 --clump-p1 0.5 --clump-p2 0.5. Asymmetric/ambiguous (AT, TA, CG, GC) SNPs, indels and duplicate position SNPs were excluded. The SNP selection P-value threshold used was $P < 0.1$. The number of clumped SNPs for each study/wave varied from 20596-43427. Polygenic risk scores were calculated for each individual as described above. Scores were derived in best guess genotype data after filtering out SNPs with $MAF < 0.05$ and $INFO < 0.8$. The polygenic risk scores were standardized using z-score transformations. Logistic regression analyses including PCs tested for association of polygenic risk scores with case status. Finally, overall meta-analyses of the leave-one-out analyses were performed (Supplementary Figure 10).

SNP heritability

SNP heritability was estimated using LD score regression⁴⁵ in order to evaluate how much of the variation in the phenotypic trait could be ascribed to common additive genetic variation. Summary statistics from GWAS meta-analyses and pre-computed LD scores (available from <https://github.com/bulik/ldsc>) were used in the analyses. The SNP heritability for ADHD was calculated on the liability scale when using summary statistics from analyses of diagnosed ADHD and assuming a 5% prevalence of ADHD in the population⁶⁸. The SNP heritability (h^2_{SNP}) was

estimated to be 0.216 (SE = 0.014) based on the summary statistics from the European GWAS meta-analysis. LD score regression SNP heritability estimates for the iPSYCH and PGC samples respectively can be found Supplementary Table 9.

In order to evaluate the stability of our heritability estimates we also performed univariate GREML analyses in GCTA. However due to strict restrictions on access to individual genotypes, GREML analyses could only be performed separately in the PGC and iPSYCH cohorts. For each of these datasets, best guess genotype data were generated (hard called genotypes with imputation INFO-score > 0.8) and filtered to include genotypes with MAF>0.05. Asymmetric/ambiguous (AT, TA, CG, GC), multi-allelic and duplicate position SNPs were excluded. For each dataset, a genomic-relationship matrix was calculated, restricted to HapMap-3 SNPs. Analyses were based on the following numbers of SNPs: PGC-only: 191,466 SNPs; iPSYCH-only: 435,086 SNPs. One of each pair of individuals related (π -hat>0.05) was excluded, preferentially keeping cases. In the PGC dataset N=16 cases and N=91 controls were excluded and N=1,439 cases and N=3,170 controls in the iPSYCH dataset. PCA (after LD-pruning and removing SNPs located in long-range LD regions) was performed on the merged, unrelated samples using PLINK, to derive population covariates. The first 10 PCs as well as study/wave indicators were used as covariates in the univariate GREML analyses in GCTA.

Partitioning heritability by functional annotation and cell type

Partitioning of the heritability by functional categories was done using LD score regression and 53 functional overlapping annotations described in Finucane et al.⁶⁹ and the baseline model LD scores, regression weights and allele frequencies based on the 1KGP3 European ancestry samples were downloaded from <https://data.broadinstitute.org/alkesgroup/LDSCORE/>. The summary statistics from the European GWAS meta-analysis were used in the analysis. Enrichment in the heritability of

a functional category was defined as the proportion of SNP heritability explained divided by the proportion of SNPs⁶⁹. Results from analysis of the 24 main annotations (no window around the functional categories) are displayed in Supplementary Figure 12. The analysis revealed significant enrichment in the heritability by SNPs located in conserved regions ($P = 8.49 \times 10^{-10}$; Supplementary Figure 12).

Test for enrichment in the heritability of SNPs located in cell-type-specific regulatory elements was evaluated in two ways. One by using the 220 cell-type-specific annotations that have been grouped into 10 cell-type groups as described in Finucane et al.⁷⁰. These annotations are based on cell-specific histone markers, related to H3K4me1⁷¹, H3K4me3⁷¹, H3K9ac⁷¹ and H3K27ac⁷². The test was done using the summary statistics from the European GWAS meta-analysis and cell-type specific LD scores, baseline model LD scores, regression weights and allele frequencies based on 1KGP3 European ancestry samples available for download at: <https://data.broadinstitute.org/alkesgroup/LDSCORE/>. In the analyses, it was tested if the cell-group specific annotations contributed significantly to the SNP heritability when controlling for the annotations in the full baseline model (the coefficient P-value). The analysis revealed a significant enrichment in the heritability by SNPs located in central nervous system specific enhancers and promoters (enrichment = 2.44, SE=0.35, $P = 5.81 \times 10^{-5}$; Supplementary Figure 13).

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^{72,73}. This mark has previously been used with success in identifying significant enrichments in tissues/cells and often in a biologically plausible manner^{71,72}. This analysis identified enrichment in the heritability of SNPs located in specific regulatory elements of nine brain tissues as well as three stem-cell lines (Supplementary Figure 14).

Genetic correlations of ADHD with other traits

The genetic correlation of ADHD with other traits were evaluated using LD Score regression⁴⁵. Correlations with 211 phenotypes were tested using LD Hub (<http://ldsc.broadinstitute.org/ldhub/>)⁷⁴. This estimation was based on summary statistics from the European GWAS meta-analysis and summary statistics from published GWASs. In addition, we tested for genetic correlation with eight phenotypes on our local server: human intelligence⁷⁵, four phenotypes related to education and cognition analysed in samples from the UK-Biobank (college/university degree, verbal–numerical reasoning, memory and reaction time)⁷⁶, insomnia⁷⁷, anorexia nervosa⁷⁸, and correlation with Major Depressive Disorder was tested using GWAS results from an updated analysis of 130,664 cases and 330,470 controls from the Psychiatric Genomics Consortium⁷⁹. In total 219 phenotypes were tested for genetic overlap with ADHD and 43 demonstrated significant correlation after Bonferroni correction ($P < 2.28 \times 10^{-4}$). Detailed information about significant genetic correlations can be found in Supplementary Table 11 and extended results for all phenotypes tested can be found in Supplementary eTable 5.

Replication analysis in external cohorts

To replicate the results of the ADHD GWAS meta-analysis we compared the results to analyses of the deCODE, 23andMe, and EAGLE/QIMR cohorts. The sample design of these cohorts and the process for matching their GWAS results to the ADHD GWAS are described under the sample description section. Briefly, the deCODE analysis is based on ICD-10 diagnoses and prescription data from medical records, the 23andMe analysis is based on self-reported history of ADHD diagnosis, and EAGLE/QIMR involve analyses of continuous measures of ADHD-related behavioural traits in the general population. Given the phenotypic differences between these cohorts we evaluated

replication separately for each of these cohorts. The included replication analyses also depended on whether data is available for genome-wide variants (23andMe and EAGLE/QIMR) or only top hits (deCODE) and whether effect sizes are directly comparable to the ADHD GWAS (23andMe and deCODE only).

Taken together, we evaluated 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.

Sign test

To evaluate concordance of the direction of effect between the ADHD GWAS and the replication cohorts, we first identified the overlapping SNPs present in the ADHD GWAS and analysis of each of the three replication cohorts (i.e. deCODE, 23andMe, and EAGLE/QIMR). For each replication cohort, the ADHD GWAS results for the intersecting SNPs were then clumped to define independent loci using PLINK 1.9⁴². Given that the previous conditional analysis was unable to conclusively confirm putative independent signals in the loci defined with LD $r^2 > 0.1$ within 500 kb of the index variant, we apply more conservative clumping parameters to ensure independence for the sign test ($r^2 > 0.05$ within 1 Mb) and merge index variants within 1 Mb. This clumping was performed 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 as reference.

After clumping, sign tests were performed to compare the loci from the ADHD GWAS meta-analysis to each replication cohort. Specifically, for loci passing a given P-value threshold in the ADHD

GWAS meta-analysis, we tested the proportion with a concordant direction of effect in the replication cohort (π) using a one sample test of the proportion with Yates' continuity correction⁸⁰ against a null hypothesis of $\pi = 0.50$ (i.e. the signs are concordant between the two analyses by chance) in R⁸¹. 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. We note that the number of clumped loci at each P-value threshold varies depending on the availability of results for variants in each replication cohort. For example, the genome-wide significant variants ($P < 5 \times 10^{-8}$) from the chromosome 4 locus in the ADHD GWAS (index variant: rs28411770) are all absent in the 23andMe results, thus the sign test is limited to the other 11 loci.

Sign test results for each replication cohort are reported in Supplementary Table 12. Sign concordance between each of the three replication cohorts and the ADHD GWAS was significantly greater than would be expected by chance ($P < 0.0167 = 0.05/3$ replication cohorts) for nominally associated loci from the ADHD GWAS ($P < 1 \times 10^{-6}$). Power of the sign test at stricter P-value thresholds is limited by the number of loci at that level in the ADHD GWAS, but the sign test remained significant for EAGLE/QIMR which was sign concordant for all loci with $P < 1 \times 10^{-7}$ in the ADHD GWAS. Sign concordance was generally strongest in EAGLE/QIMR, followed by deCODE and 23andMe, though 23andMe showed stronger sign concordance for loci with $P < 1 \times 10^{-5}$ in the ADHD GWAS.

Focusing on the genome-wide significant loci ($P < 5 \times 10^{-8}$) for the ADHD GWAS, 11 of the 12 loci show sign concordance in at least 2 of the 3 replication cohorts (Supplementary Table 13). Eight of the 12 loci were sign concordant in all available replication results for either the index variant or a proxy variant in the locus. The weakest sign concordance was observed for the chromosome 16 locus (rs212178, *LINC01572*), which was only sign concordant in EAGLE/QIMR with discordant estimates observed from deCODE and 23andMe. Sign discordant estimates were also observed for

the chromosome 1 (index variant rs11420276, *PTPRF/ST3GAL3*) and chromosome 15 (rs281324, *SEMA6D*) loci in 23andMe, and for the chromosome 2 locus (rs9677504, *SPAG16*) in deCODE.

Replication of effect sizes for top loci

Although the sign test provides a useful check of concordance for direction of effect in the replication samples, it does not consider replication of the magnitude of the effect. For the deCODE and 23andMe replication cohorts that included GWAS of ADHD status from various sources, we anticipate that the true effect sizes should be similar to, though not necessarily identical to, the effect size for ADHD as defined in the current meta-analysis. Therefore, we looked at the replication of the effect sizes of the top loci in the ADHD GWAS in the replication cohorts.

Specifically, for each of deCODE and 23andMe we identified clumped loci that are nominally associated ($P < 1 \times 10^{-6}$) with ADHD in the primary ADHD GWAS. We then 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 (described below). For this regression, we oriented all variants in the direction of the risk increasing allele estimated from the ADHD GWAS, constrained the intercept term to zero, and weighed the variants proportional to the inverse of their squared standard error from the ADHD GWAS. If the included variants are all truly associated with ADHD with the same true population effect size in the ADHD GWAS and replication cohorts (i.e. regardless of the phenotyping measure for ADHD status) then the expected slope of this regression is one. On the other hand, if all SNPs are truly null in the ADHD GWAS and the replication cohorts then the expected slope of this regression is zero.

Comparison of effects sizes from the ADHD GWAS to the deCODE cohort suggests good but imperfect replication of top loci (Supplementary Figure 15). The deCODE effect size estimates are significantly correlated with the adjusted ADHD GWAS betas (slope = 0.664, SE=.154, $P=1.2 \times 10^{-4}$)

but this slope is also significantly less than 1 ($P=0.0178$). The estimated effect of rs9677504 (*SPAG16* locus) is a visible outlier in this comparison, following the discordant direction of association as noted in the sign test results. Omitting this effect size the regression increases the slope but is not sufficient to eliminate the significant difference from one (slope = 0.710, SE=0.149, $P = 0.030$ for comparison to slope=1).

Comparison of effect sizes with 23andMe follows a similar trend, but with somewhat weaker overall strength of replication (Supplementary Figure 16). Regressing 23andMe estimates of effect size on the adjusted ADHD GWAS effect sizes yields a slope of 0.417 (SE=0.117), which is significantly greater than zero ($P=1.11 \times 10^{-3}$) and significantly less than one ($P=9.52 \times 10^{-6}$). This lower slope compared to the deCODE replication is consistent with the weaker sign test results. The comparison of effect sizes also emphasizes the strong discordance at rs112984125 (*ST3GAL3/PTPRF*) and rs212178 (*LINC01572*). Removing those two loci from the regression increases the slope to 0.538 (SE=0.102, $P=3.86 \times 10^{-5}$ for slope less than 1).

Genetic correlation analysis of replication cohorts

Genetic correlation of the ADHD GWAS with the 23andMe and EAGLE/QIMR results was computed using LD score regression⁴⁵ 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.

The estimated genetic correlation between the 23andMe and the ADHD GWAS was large and significant ($r_g = 0.653$, SE = 0.114, $P = 1.11 \times 10^{-8}$), but also significantly less than 1 (one-sided $P=1.17 \times 10^{-3}$). Genetic correlation analysis with EAGLE/QIMR suggest a stronger overlap ($r_g = 0.970$, SE = 0.207, $P = 2.66 \times 10^{-6}$) with a genetic correlation not significantly different from one (one-sided

$P = 0.442$). These results are both consistent with sign test results, where both 23andMe and EAGLE/QIMR show significant concordance with the ADHD GWAS, with EAGLE/QIMR showing the most consistent results. The moderate but highly significant genetic correlation observed with 23andMe is also consistent with the moderate slope in the replication of effect sizes for top loci from the ADHD GWAS.

To further explore the moderate genetic correlation between the 23andMe results and the ADHD GWAS we evaluated the genetic correlation between 23andMe and traits from LD Hub (<http://ldsc.broadinstitute.org/ldhub/>)⁷⁴. The goal of this analysis is to potentially identify differences in the profile of genetic correlation of these studies to other traits, with the expectation that such differences may highlight domains where the 23andMe and ADHD GWAS results are particularly different. We focus this comparison on a curated set of 28 phenotypes with at least nominal ($P < 0.01$) evidence of genetic correlation with either the ADHD GWAS or 23andMe, with the set of phenotypes selected to emphasize domains of interest for ADHD while reducing the number of closely related traits.

Genetic correlation results for the 23andMe and ADHD GWAS with the 28 selected phenotypes are reported in Supplementary Table 14. 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:

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

The denominator is motivated by the expected standard error of the difference between the two means under the assumption that the samples are not correlated. In the context of mean differences this approximation may be expected to be conservative when the r_g estimates are positively correlated. We do not expect this to be an ideal formal test for the difference between two genetic correlations as estimated by LD score regression, and therefore emphasize the importance of 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. Comparison of the r_g results for the ADHD GWAS and 23andMe identifies some striking differences. Most notably, the 23andMe GWAS shows little to no genetic correlation with college completion ($r_g=0.056$, compared to $r_g=-0.54$ for the primary ADHD GWAS; $P=1.1 \times 10^{-9}$ for difference) and only limited correlation to childhood IQ ($r_g=-0.22$ vs. $r_g=-0.41$ in ADHD, $P=2.46 \times 10^{-1}$) and educational attainment ($r_g=-0.20$ vs. $r_g=-0.53$ in ADHD, $P=1.8 \times 10^{-7}$). The genetic correlation with age of first birth is also strongly attenuated in 23andMe ($r_g=-0.33$ vs. $r_g=-0.61$ in ADHD, $P=6.3 \times 10^{-4}$), while results for the other reproductive phenotypes are more similar. The 23andMe GWAS also shows evidence of weaker genetic correlations with the weight-related phenotypes, most notably class 1 obesity ($r_g=0.11$, vs. $r_g=0.29$ in ADHD, $P=0.024$) and childhood obesity ($r_g=-0.025$, vs. $r_g=0.22$ in ADHD, $P=0.036$). The one domain where 23andMe exhibits a trend toward stronger genetic correlations is schizophrenia ($r_g=0.27$, vs. $r_g=0.12$ in ADHD, $P=0.053$) and bipolar disorder ($r_g=0.29$, vs. $r_g=0.095$ in ADHD, $P=0.09$), these differences are not significant with the approximated Z score. There is a corresponding trend towards weaker genetic correlation with depressive symptoms compared to the ADHD GWAS, but it is also not significant with the current test ($r_g=0.30$, vs. $r_g=0.45$ in ADHD, $P=0.24$).

Meta-analysis and heterogeneity test with each replication cohort

For the replication analysis, we considered meta-analyses of the ADHD GWAS results with each of the three replication cohorts. The 23andMe and deCODE cohorts were each meta-analysed with the ADHD GWAS using conventional inverse variance-weighted meta-analyses. For EAGLE/QIMR, meta-analysis was performed using a modified sample size-based weighting method (see detailed description of methods below). For each of these three meta-analyses, we evaluated results for the

genome-wide significant loci from the ADHD GWAS and identified new loci reaching nominal genome-wide significance ($P < 5 \times 10^{-8}$, unadjusted for number of replication cohorts). In addition, we considered Cochran's Q test of heterogeneity, computing the 1 degree of freedom test for heterogeneity between the ADHD GWAS and each replication cohort.

Meta-analysis of the discovery ADHD GWAS with deCODE (hereafter ADHD+deCODE) yields 10 genome-wide significant loci (Supplementary eTable 6). Of these loci 5 were significant in the ADHD GWAS, leaving 7 loci that were significant in the ADHD GWAS but are not significant after meta-analysis with the deCODE replication cohort: rs1222063 (chr. 1, intergenic), rs9677504 (chr. 2, *SPAG16*), rs4858241 (chr. 3, *SGO1-ASI*), rs4916723 (chr. 5, *LINC00461*), rs74760947 (chr. 8, *LINC01288*), rs281324 (chr. 15, *SEMA6D*), and rs212178 (chr. 16, *LINC10572*). None of these loci show genome-wide significant heterogeneity between the ADHD GWAS and deCODE, though in many cases there is nominal evidence ($P < .05$) of heterogeneity. The remaining 5 loci significant in the ADHD+deCODE replication meta-analysis are novel: rs1592757 (chr. 5, Refseq gene *LOC105379109*), rs28452470 (chr. 7, *CADPS2*), rs10956838 (chr. 8, intergenic), rs4275621 (chr. 11, intergenic), and rs1848160 (chr. 11, intergenic).

The replication meta-analysis with 23andMe (hereafter ADHD+23andMe) identified 10 genome-wide significant loci (Supplementary eTable 6, Supplementary Figure 17). Three of these loci were novel: rs30266 (chr. 5, Refseq gene *LOC105379109*), rs62250537 (chr. 3, *CADM2*), and rs2243638 (chr. 13, *RNF219-ASI*). The chromosome 5 locus is the same region identified by the ADHD+deCODE meta-analysis. The other 7 genome-wide significant loci in ADHD+23andMe match loci from the discovery GWAS. The remaining 4 of the 12 genome-wide significant loci from the ADHD GWAS meta-analysis were no longer significant after meta-analysing with 23andMe: rs281324 (chr15, *SEMA6D*), rs212178 (chr16, *LINC01572*), rs4916723 (chr5, *LINC00461*), rs74760947 (chr8, *LINC01288*). The final significant locus from the ADHD GWAS, rs28411770

(chr4, *PCDH7*), is not present in the 23andMe results. Cochran's Q test of heterogeneity identified genome-wide significant heterogeneity between the ADHD GWAS and the 23andMe GWAS in the top locus on chromosome 1 from the ADHD GWAS (Supplementary Figure 18-19). The strongest evidence for discordance was at rs12410155 (*ST3GAL3*), with an estimated odds ratio of 1.111 ($P = 3.63 \times 10^{-13}$) in the ADHD GWAS, compared to an odds ratio of 0.954 ($P = 0.0244$) in the 23andMe analysis (heterogeneity $P = 2.28 \times 10^{-9}$, $I^2 = 97.2$).

Lastly, meta-analysis of the ADHD GWAS with EAGLE/QIMR (ADHD+EAGLE/QIMR) produced 15 genome-wide significant loci (Supplementary eTable 6, Supplementary Figure 20). All 12 of the genome-wide significant loci from the ADHD GWAS remain significant in the ADHD+EAGLE/QIMR meta-analysis, consistent with the sign concordance at all 12 loci. The three additional significant loci are located on chromosome 7 (rs1443749, *CADPS2*), chromosome 10 (rs9665567, intergenic) and chromosome 13 (rs7997529, *RNF219-AS1*). The chromosome 7 locus is concordant with the locus reaching genome-wide significance in ADHD+deCODE, and the chromosome 13 locus similarly matches the significant locus from ADHD+23andMe. No significant heterogeneity was observed between the EAGLE/QIMR and ADHD GWAS meta-analyses (Supplementary Figure 21-22), consistent with the strong genetic correlation between the two studies. Overall, 4 of the 12 significant loci from the ADHD GWAS are significant in all three of these replication meta-analyses: index variants rs11420276 (chr. 1, *ST3GAL3/PTPRF*), rs5886709 (chr. 7, *FOXP2*), rs11591402 (chr. 10, *SORCS3*), and rs1427829 (chr. 12, intergenic). The remaining loci are all significant in at least one of the replication meta-analyses. In addition, 3 novel loci reach genome-wide significance in two of these replication meta-analyses: index variants rs1592757/rs30266 (chr. 5, Refseq *LOC105379109*), rs28452470/rs1443749 (chr. 7, *CADPS2*), and rs2243638/rs9574218 (chr. 13, *RNF219-AS1*).

Meta-analysis across replication cohorts

We finally consider meta-analysis with all available data. Because replication results from deCODE are only available for top loci from the primary ADHD GWAS we are unable to estimate heritability and genetic correlation with EAGLE/QIMR to calibrate the meta-analysis weights as used for the other EAGLE/QIMR analyses (described below). Since we consequently cannot perform a meta-analysis simultaneously including both deCODE and EAGLE/QIMR, we instead separately evaluate the replication meta-analysis of all case/control cohorts (i.e. ADHD+deCODE+23andMe, omitting EAGLE/QIMR) and the replication meta-analysis of all genome-wide studies (i.e. ADHD+23andMe+EAGLE/QIMR, omitting deCODE).

Results for the ADHD+deCODE+23andMe and ADHD+23andMe+EAGLE/QIMR replication meta-analyses are generally consistent with the meta-analyses for each replication cohort (Supplementary eTable 6). Two new loci – located near *MAD1L1* (chr. 7) and *TM6SF2* (chr. 19) – nominally reach genome-wide significance in ADHD+23andMe+EAGLE/QIMR before consideration of multiple testing for the replication meta-analyses. No additional loci are nominated by the ADHD+deCODE+23andMe meta-analysis. Of the 12 significant loci in the ADHD GWAS, three are not significant in either of these pooled replication meta-analyses, primarily due to poor support for the locus in 23andMe: rs281324 (chr15, *SEMA6D*), rs212178 (chr16, *LINC01572*), and rs74760947 (chr8, *LINC01288*).

Winner's curse correction for effect sizes

To evaluate replication of the ADHD GWAS effect sizes it is necessary to account for the bias from the “winner's curse” in looking at only top loci nominally associated with ADHD. To correct for this bias, we compute posterior estimates for the expected value of β_j for each SNP given the observed

GWAS estimate and a spike-and-slab prior as described by Okbay et al.⁸² Briefly, we assume that marginal SNP effects follow the spike-and-slab distribution

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

where β_j is standardized for MAF such that τ^2 corresponds to average variance explained per SNP.

Following this model, we estimate $\hat{\pi}$ and $\hat{\tau}^2$ via maximum likelihood with the observed $\hat{\beta}_j$ from the ADHD GWAS. We then compute the posterior probability that each SNP is null $\hat{\pi}_j$. The posterior estimate of β_j corrected for winner's curse is then given by

$$\hat{\beta}_{adj,j} = (1 - \hat{\pi}_j) \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{s}_j^2} \hat{\beta}_j$$

where \hat{s}_j^2 is the squared standard error of $\hat{\beta}_j$ from the ADHD GWAS. As implied by this formula, the resulting $\hat{\beta}_{adj,j}$ corrected for winner's curse will be shrunken towards zero proportional to the probability that SNP j is null and the degree of uncertainty in its effect size as indicated by the standard error.

Method for meta-analysis of continuous and dichotomous ADHD measures

In order to integrate the EAGLE/QIMR data with the current analysis, we need to define a framework for comparing the GWAS of (continuous) measures of ADHD-related behaviour to the ADHD GWAS meta-analysis of (dichotomous) clinical diagnosis of ADHD.

As a starting point, motivated by the strong genetic correlation between the EAGLE/QIMR results and the ADHD GWAS meta-analysis, we could consider a conventional sample size-weighted meta-analysis of Z scores. Such an analysis however would not account for: (a) differences in power for continuous vs. dichotomous phenotypes, (b) differences in power from ascertainment on the dichotomous phenotype, (c) differences in the relative strength of overall genetic association (e.g.

SNP heritability) for the phenotype measures, or (d) imperfect correlation between the continuous ADHD-related behaviours measured in EAGLE/QIMR and clinical diagnosis of ADHD, with the matter being the phenotype of interest for the current study.

Therefore, we instead define a basic model for the genetic relationship between clinical diagnosis of ADHD and continuous ADHD-related behaviours that allows us to derive modified sample size-based weights that account for these factors. These weights should be better calibrated to provide a statistically efficient meta-analysis of the EAGLE/QIMR results with the ADHD GWAS.

Basic genetic model for latent-scale phenotypes

We begin by defining a joint model for the genetics of the two phenotypes. Let Y_1 be the observed dichotomous phenotype and Y_2 be the observed continuous phenotype. For dichotomous phenotype Y_1 , we assume there exists some latent continuous liability θ_1 such that

$$Y_1 = \begin{cases} 0, & \theta_1 < \tau_1 \\ 1, & \theta_1 \geq \tau_1 \end{cases}$$

where τ_1 is a threshold corresponding to the population prevalence K of Y_1 consistent with the standard liability threshold model⁸³. For convenience, assume that θ_1 is standardized with mean zero and unit variance. Similarly, let θ_2 denote the continuous phenotype Y_2 normalized to have mean 0 and unit variance in the population.

We describe a model for the genetics of the latent continuous phenotypes θ_1 and θ_2 before returning to the impact of the observed scale for each phenotype. Let g_1 and e_1 be genetic and environment components of θ_1 , respectively, and let g_2 and e_2 be corresponding components of θ_2 such that

$$\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

$$E \left(\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \right) = E \left(\begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \right) = E \left(\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\text{Cov}\left(\begin{bmatrix} g_1 \\ g_2 \\ e_1 \\ e_2 \end{bmatrix}\right) = \begin{bmatrix} rh_1^2 & \rho_g & 0 & 0 \\ \rho_g & rh_2^2 & 0 & 0 \\ 0 & 0 & \epsilon_1^2 & \rho_e \\ 0 & 0 & \rho_e & \epsilon_2^2 \end{bmatrix}$$

Note that we assume that not only are genetics and environment uncorrelated within phenotype (i.e. $\text{Cov}(g_1, e_1) = \text{Cov}(g_2, e_2) = 0$), but also between phenotypes ($\text{Cov}(g_1, e_2) = \text{Cov}(g_2, e_1) = 0$). Since θ_1 and Y_2^* are each defined to have unit variance, it follows that $\epsilon_1^2 = 1 - h_1^2$, $\epsilon_2^2 = 1 - h_2^2$, and ρ_g and ρ_e are the genetic and environmental covariances, respectively. The genetic correlation between the latent phenotypes θ_1 and θ_2 can then be defined as $r_g \equiv \text{cor}(g_1, g_2) = \rho_g / \sqrt{h_1^2 h_2^2}$.

Defining independent genetic factors

We next seek to specify the covariance of g_1 and g_2 in terms of two independent factors f_1 and f_2 . This transformation serves two purposes. First, for modeling the effects of individual SNPs it will allow us to define independent effects on f_1 and f_2 . Second, if we specify that g_1 depends only on f_1 while g_2 may depend on both f_1 and f_2 , then it will be possible to focus on the factor f_1 that determines the genetic component g_1 for Y_1 separate from any independent factors f_2 that contribute to Y_2 only. Separating these factors therefore allows us to move towards the goal of defining a scheme for meta-analysis that focuses only on genetic effects for the dichotomous phenotype (i.e. clinical ADHD diagnosis) while discounting any independent genetic effects that are only relevant to the continuous phenotype (i.e. ADHD-related behaviours).

The desired transformation is given by the inverse of the Cholesky decomposition of the covariance matrix for g_1 and g_2 , which yields

$$\begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \equiv \begin{bmatrix} 1 & 0 \\ \frac{\rho_g}{\sqrt{h_1^2}} & 1 \\ \frac{-\rho_g}{\sqrt{1 - r_g^2} \sqrt{h_1^2}} & \frac{1}{\sqrt{1 - r_g^2} \sqrt{h_2^2}} \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$$

It can then be shown that

$$\text{Cov} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

and $E(f_1) = E(f_2) = 0$. By substitution for g_1 and g_2 , the latent phenotypes are related to these factors by

$$\begin{aligned} \theta_1 &= \left(\sqrt{\overline{h_1^2}} \right) f_1 + e_1 \\ \theta_2 &= \left(r_g \sqrt{\overline{h_2^2}} \right) f_1 + \left(\sqrt{1 - r_g^2} \sqrt{\overline{h_2^2}} \right) f_2 + e_2. \end{aligned}$$

Effects of individual variants

The above model fully specifies the overall genetic components of the two phenotypes, but does not model the contribution of any specific variant. We now define the effects of individual variants so that we can work towards describing the GWAS results for each phenotype and the desired meta-analysis in terms of parameters for a given variant.

To consider effects for a given variant, let γ_{1j}^* be the causal effect of variant j on f_1 , and let γ_{2j}^* be the corresponding effect on f_2 . Then

$$\begin{aligned} f_1 &= \sum_j \gamma_{1j}^* x_j \\ f_2 &= \sum_j \gamma_{2j}^* x_j, \end{aligned}$$

where x_j is the standardized genotype of variant j . We denote these causal effects as γ to distinguish them from marginal effects β , and use the * superscript here to indicate that these are effect sizes on the latent genetic components f_1 and f_2 as opposed to the standardized phenotypes. The corresponding causal effects on θ_1 and θ_2 are

$$\gamma_{1j} = \sqrt{\overline{h_1^2}} \gamma_{1j}^*$$

$$\gamma_{2j} = \left(r_g \sqrt{h_2^2} \right) \gamma_{1j}^* + \left(\sqrt{1 - r_g^2} \sqrt{h_2^2} \right) \gamma_{2j}^*$$

respectively.

The marginal effect on f_1 and f_2 will depend not only on the variant's causal effect, but on the causal effect of other variants in LD with x_j . We define the corresponding marginal effects β_j^* as

$$\beta_{1j}^* = (X_j' X_j)^{-1} X_j' \mathbf{X} \boldsymbol{\gamma}_1^* = \sum_k \gamma_{1k}^* r_{jk}$$

$$\beta_{2j}^* = (X_j' X_j)^{-1} X_j' \mathbf{X} \boldsymbol{\gamma}_2^* = \sum_k \gamma_{2k}^* r_{jk}$$

where $r_{jk} = \text{cov}(x_j, x_k)$ and $\boldsymbol{\gamma}_1^*$ and $\boldsymbol{\gamma}_2^*$ are column vectors with elements γ_k^* for all k . As with γ^* , the superscript denotes these β_j^* as effect sizes on the latent genetic components f_1 and f_2 .

To get marginal effect sizes on the phenotypes, we can denote

$$f_{1,-j} = f_1 - \beta_{1j}^* x_j$$

$$f_{2,-j} = f_2 - \beta_{2j}^* x_j$$

to indicate f_1 and f_2 with the full marginal effect of x_j removed, so that we can then express the phenotypes in terms of the marginal effect of variant j as

$$\begin{aligned} \theta_1 &= (f_{1,-j} + \beta_{1j}^* x_j) \sqrt{h_1^2} + e_1 \\ &= (\beta_{1j}^* \sqrt{h_1^2}) x_j + (\sqrt{h_1^2}) f_{1,-j} + e_1 \end{aligned}$$

$$\begin{aligned} \theta_2 &= (r_g \sqrt{h_2^2}) (f_{1,-j} + \beta_{1j}^* x_j) + \left(\sqrt{h_2^2 (1 - r_g^2)} \right) (f_{2,-j} + \beta_{2j}^* x_j) + e_2 \\ &= \left[\beta_{1j}^* (r_g \sqrt{h_2^2}) + \beta_{2j}^* \left(\sqrt{h_2^2 (1 - r_g^2)} \right) \right] x_j + (r_g \sqrt{h_2^2}) f_{1,-j} + \left(\sqrt{h_2^2 (1 - r_g^2)} \right) f_{2,-j} + e_2 \end{aligned}$$

Therefore, given θ_1, θ_2 and x_j are all standardized and $x_j, f_{1,-j}, f_{2,-j}, e_1,$ and e_2 are all independent except for $cov(e_1, e_2) = \rho_e$.

$$\beta_{1j} = cor(x_j, \theta_1) = \beta_{1j}^* \sqrt{h_1^2}$$

$$\beta_{2j} = cor(x_j, \theta_2) = \beta_{1j}^* \left(r_g \sqrt{h_2^2} \right) + \beta_{2j}^* \left(\sqrt{h_2^2 (1 - r_g^2)} \right)$$

These are the standardized marginal effect of variant j on each phenotype (defined on the latent liability scale in the case of the dichotomous phenotype Y_1). We denote them as β_j because they are the effects of interest for the GWASs of θ_1 and θ_2 .

GWAS test statistics for β_j

Given the above parameterization, we can now focus on the behaviour of the test statistics from the GWAS of each phenotype. These test statistics are of primary interest since they are the intended input for the desired meta-analysis across the two phenotypes. In particular, we are focused on defining the relationship between these test statistics and β_{1j} , the effect of variant j on the dichotomous phenotype, so that we can calibrate meta-analysis of the observed Z scores to test the null hypothesis $\beta_{1j} = 0$.

Before discussing the test statistics for the observed phenotypes however, it is instructive to describe test statistics for a hypothetical GWAS of the latent phenotypes (θ_1 and θ_2), as a foundation evaluating the impact of e.g. dichotomizing θ_1 to a case/control phenotype. If both θ_1 and θ_2 were observed, we could define Z scores for the standardized effects β_j as

$$\begin{aligned} Z_{\theta,1j} &= \sqrt{N_1} \hat{\beta}_{1j} \\ &= \sqrt{N_{1j}} \beta_{1j} + \sqrt{N_{1j}} \delta_{1j} \\ Z_{\theta,2j} &= \sqrt{N_2} \hat{\beta}_{2j} \\ &= \sqrt{N_{2j}} \beta_{2j} + \sqrt{N_{2j}} \delta_{2j} \end{aligned}$$

where $\hat{\beta}_j$ is the observed GWAS estimate of β_j , N_j is the sample size of the data used to estimate $\hat{\beta}_j$, and δ_j is the corresponding sampling error of the estimate.

$$\delta_{1j} \sim N\left(0, \frac{1}{N_{1j}} \text{var}\left[\left(\sqrt{h_1^2}\right) f_{1,-j} + e_1\right]\right)$$

$$\delta_{2j} \sim N\left(0, \frac{1}{N_{2j}} \text{var}\left[\left(r_g \sqrt{h_2^2}\right) f_{1,-j} + \left(\sqrt{h_2^2(1-r_g^2)}\right) f_{2,-j} + e_2\right]\right)$$

Assuming the marginal effect of any given variant is small, we can approximate $f_{1,-j} \approx f_1$ and $f_{2,-j} \approx f_2$, giving

$$\delta_{1j} \sim N\left(0, \frac{1}{N_{1j}} [h_1^2 + \epsilon_1^2]\right)$$

$$\sim N\left(0, \frac{1}{N_{1j}}\right)$$

$$\delta_{2j} \sim N\left(0, \frac{1}{N_{2j}} [r_g^2 h_2^2 + (1-r_g^2) h_2^2 + \epsilon_2^2]\right)$$

$$\sim N\left(0, \frac{1}{N_{2j}} [h_2^2 + \epsilon_2^2]\right)$$

$$\sim N\left(0, \frac{1}{N_{2j}}\right)$$

based on f and e being independent, $\text{var}(f) = 1$, $\text{var}(e) = \epsilon^2$, and $h^2 + \epsilon^2 = 1$. To the extent that $f \neq f_{-j}$, then this is an overestimate of the variance of δ_j since $\text{var}(f_{-j}) \leq 1$. Returning to the Z scores defined previously, this gives

$$Z_{\theta,1j} \sim N(\sqrt{N_{1j}} \beta_{1j}, 1)$$

$$Z_{\theta,2j} \sim N(\sqrt{N_{2j}} \beta_{2j}, 1)$$

which have the desired standard normal distribution when $\beta_j = 0$. We can also express these in terms of the latent effects β_j^*

$$Z_{\theta,1j} \sim N(\sqrt{N_{1j} h_1^2} \beta_{1j}^*, 1)$$

$$\begin{aligned}
Z_{\theta,2j} &\sim N\left(r_g \sqrt{N_{2j} h_2^2 \beta_{1j}^*} + \sqrt{N_{2j} h_2^2 (1 - r_g^2) \beta_{2j}^*}, 1\right) \\
&\sim N\left(\sqrt{N_{2j} h_2^2 [r_g \beta_{1j}^* + \sqrt{1 - r_g^2} \beta_{2j}^*]}, 1\right)
\end{aligned}$$

We now consider how the actual observed Z statistics for each phenotype will differ from these idealized tests of association with the latent phenotypes.

Test of β_{ij} in GWAS of dichotomous Y_1

For the dichotomous phenotype (i.e. ADHD diagnosis) we do not observe the latent liability θ_1 , and thus we cannot compute $Z_{\theta,1j}$. Instead the GWAS results come from logistic regression of the observed phenotype Y_1 . The statistical properties of Z_{2j} will be affected by two key features of GWAS with case/control phenotypes: dichotomization and ascertainment. For dichotomization, recall that

$$\beta_{1j} = \text{cor}(x_j, \theta_1)$$

It has been shown that dichotomizing one variable attenuates the correlation such that if x and Y are approximated as bivariate normal

$$\text{cor}(x_j, Y_1) \approx \frac{\phi(\Phi^{-1}[K])}{\sqrt{K(1-K)}} \text{cor}(x_j, \theta_1)$$

where K is the prevalence of $Y_1 = 1$ after dichotomization, and $\phi(\cdot)$ and $\Phi^{-1}(\cdot)$ are the density function and inverse of the cumulative density function of the standard normal distribution, respectively^{84,85}.

With respect to case/control ascertainment, we note that the power of logistic regression is approximately proportional to $\sqrt{P^*(1 - P^*)}$, where P^* is the probability $Y = 1$ at the mean liability in the sample⁸⁶. Given a symmetric distribution of liability, $P^* \approx P$ where P is the sample proportion

of cases. Therefore, power for the analysis in an ascertained sample differs from power in a population sample with the sample size by a factor of $\sqrt{P(1-P)}/\sqrt{K(1-K)}$.

Putting together these adjustments for dichotomization and ascertainment, we adopt the approximation

$$Z_{1j} \sim N\left(\frac{\phi(\Phi^{-1}[K])\sqrt{P(1-P)}}{\sqrt{K(1-K)}\sqrt{K(1-K)}}\sqrt{N_{1j}}\beta_{1j}, 1\right)$$

Importantly, we note that this adjustment mirrors the conversion between observed scale heritability and liability scale heritability derived by Lee et al⁸⁷.

$$h_{obs}^2 = h_{liab}^2 \frac{(\Phi^{-1}[K])^2 P(1-P)}{K(1-K) K(1-K)}$$

Indeed, noting that

$$\beta_{1j} = \beta_{1j}^* \sqrt{h_1^2}$$

we can equate

$$\begin{aligned} E(Z_{1j}) &= \frac{\phi(\Phi^{-1}[K])\sqrt{P(1-P)}}{\sqrt{K(1-K)}\sqrt{K(1-K)}}\sqrt{N_{1j}}\beta_{1j} \\ &= \frac{\phi(\Phi^{-1}[K])\sqrt{P(1-P)}}{\sqrt{K(1-K)}\sqrt{K(1-K)}}\sqrt{N_{1j}}\beta_{1j}^*\sqrt{h_1^2} \\ &= \sqrt{N_{1j}}\sqrt{h_1^2 \frac{(\Phi^{-1}[K])^2 P(1-P)}{K(1-K) K(1-K)}}\beta_{1j}^* \\ &= \sqrt{N_{1j}}\sqrt{h_{1,obs}^2}\beta_{1j}^* \end{aligned}$$

which highlights the parallel between the observed Z_{1j} , the test the effect of variant j on the observed scale, and the corresponding $Z_{\theta,1j}$ testing on the latent scale.

Finally, we can define an effective sample size adjustment for N_1

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

so that the approximate distribution of Z_{1j} from logistic regression of the dichotomous phenotype can be expressed as

$$Z_{1j} \sim N\left(\sqrt{\tilde{N}_{1j}} \beta_{1j}, 1\right)$$

This formulation, with Z_{1j} following a standard normal distribution conditional on (adjusted) sample size and the effect of interest (β_{1j}), is central to allowing the desired meta-analysis.

Test of β_{ij} in GWAS of continuous Y_2

For the continuous phenotype Y_2 , the observed Z scores are computed from conventional linear regression of the observed phenotype. The observed phenotype only differs from the latent θ_2 by a linear transformation to center and scale to unit variance. Thus, since Z scores are invariant linear transformations of the phenotype,

$$Z_{2j} = Z_{\theta,2j}$$

But unlike the dichotomous phenotype analysis, the GWAS of Y_2 is a test of $\beta_{2j} = 0$ rather than $\beta_{1j} = 0$. For the current analysis, our primary interest is in the latter test to identify effects β_{1j} of each variant on the dichotomous phenotype of ADHD diagnosis. The residual effects from f_2 in β_{2j}^* that are unique to the continuous trait (population measures of ADHD-related behavior) are of less relevance, and given a high r_g are anticipated to have a limited contribution even to the continuous measure (proportional to $\sqrt{1 - r_g^2}$).

For that reason, we adopt a random effects framework for β_{2j}^* , and treat them as nuisance parameters. Specifically, we assume that the causal effect sizes

$$\gamma_{2j}^* \sim N\left(0, \frac{1}{M}\right)$$

where M is the total number of variants j . This is equivalent to the standard infinitesimal random effects model used by^{45,88}, but specified here in terms of γ_{2j}^* on the latent scale of f_2 prior to scaling by h_2^2 for the phenotype. For the marginal effects

$$\beta_{2j}^* = \sum_k \gamma_{2k}^* r_{jk}$$

we additionally assume that γ_{2k}^* and r_{jk} are independent, meaning that the causal effect size of variant k is independent of its LD with other variants. This leads to

$$\beta_{2j}^* \sim N\left(0, \sum_k \frac{1}{M} r_{jk}^2\right)$$

$$\beta_{2j}^* \sim N\left(0, \frac{1}{M} l_j\right)$$

where we note that

$$l_j = \sum_k r_{jk}^2$$

is the LD score of variant j as defined by Bulik-Sullivan et al⁴⁵.

Returning to the Z score, we note that the previous expression for Z_{2j}^* is now a conditional distribution given a particular value of β_{2j}^* .

$$Z_{2j} | \beta_{2j}^* \sim N\left(\sqrt{N_{2j} h_2^2} \left[r_g \beta_{1j}^* + \sqrt{1 - r_g^2} \beta_{2j}^* \right], 1\right)$$

Substituting $\beta_{1j} = \beta_{1j}^* \sqrt{h_1^2}$ as the actual parameter of interest and marginalizing over β_{2j}^* as a random effect yields

$$Z_{2j} \sim N\left(r_g \sqrt{N_{2j} h_1^2 h_2^2} \beta_{1j}, 1 + (1 - r_g^2) \frac{N_{2j} h_2^2}{M} l_j\right)$$

To construct the intended meta-analysis for genetics effects on the diagnosis of ADHD, we want to have a test statistic that has a standard normal distribution under the null hypothesis $\beta_{1j} = 0$. Although Z_{2j} is normally distributed with mean zero under that null hypothesis, it's variance is inflated if Y_2 is heritable (i.e. $h_2^2 > 0$) and is not completely genetically correlated with Y_1 (i.e. $r_g < 1$). We note that inflation proportional to $N_{2j}h_2^2l_j/M$ is consistent the expected values derived by Bulik-Sullivan et al.⁴⁵ To obtain a statistic with the desired distribution, we define a new modified Z score

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

which is distributed as

$$\tilde{Z}_{2j} \sim N \left(\frac{|r_g| \sqrt{N_{2j}h_2^2/h_1^2}}{\sqrt{1 + (1 - r_g^2) \frac{N_{2j}h_2^2}{M} l_j}} \beta_{1j}, 1 \right)$$

Including $\text{sign}(r_g)$ in this definition of \tilde{Z}_{2j} ensures that its expected value has the same sign as β_{1j} . We note that $|\tilde{Z}_{2j}| \leq |Z_{2j}|$ since $[1 + (1 - r_g^2) N_{2j}h_2^2l_j/M] \geq 1$ by definition for the involved quantities. Therefore \tilde{Z}_{2j} can be interpreted as a more conservative estimate for inference about β_{1j} that has been attenuated from the raw observation proportional to the potential for genetic effects on the continuous phenotype to be unique to that phenotype rather than shared with the dichotomous outcome (i.e. $[1 - r_g^2]h_2^2$).

Lastly, if we define adjusted N_2 as

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

then the distribution of the modified test statistic can be expressed as

$$\tilde{Z}_{2j} \sim N\left(\sqrt{\tilde{N}_{2j}}\beta_{1j}, 1\right)$$

We note that the multiplication by $sign(r_g)$ in defining \tilde{Z}_{2j} ensures that it's distribution involves the positive root of \tilde{N}_{2j} and matches the sign of β_{1j} in expectation. The resulting statistic provides a clear parallel to Z_{1j} for the dichotomous phenotype, and sets up the meta-analysis of the two Z scores.

Meta-analysis for β_{1j} from GWAS of dichotomous Y_1 and continuous Y_2

Summarizing the above derivations, we have now established a framework for the GWAS results of dichotomous phenotype Y_1 and continuous phenotype Y_2 , respectively, where we can approximate

$$Z_{1j} \sim N\left(\sqrt{\tilde{N}_{1j}}\beta_{1j}, 1\right)$$

$$\tilde{Z}_{2j} \sim N\left(\sqrt{\tilde{N}_{2j}}\beta_{1j}, 1\right)$$

where

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

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

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

From this form, we have a pair of Z statistics for β_{1j} with corresponding (adjusted) sample sizes.

This is sufficient to then proceed with a conventional sample size-weighted meta-analysis⁸⁹.

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

Expanding the numerator makes it clear that the adjusted sample size weights corresponds to weighting each Z proportional to the observed heritability and then balancing by genetic correlation.

$$Z_{j,meta} = \frac{\sqrt{N_{1j} \frac{P(1-P)\phi(\Phi^{-1}[K])^2}{[K(1-K)]^2}} Z_{1j} + \sqrt{N_{2j} \frac{r_g^2 h_2^2 / h_1^2}{1 + (1-r_g^2)N_{2j}h_2^2 l_j / M}} \tilde{Z}_{2j}}{\sqrt{\tilde{N}_{1j} + \tilde{N}_{2j}}}$$

$$Z_{j,meta} \propto \sqrt{h_{1,obs}^2 \tilde{N}_{1j}} Z_{1j} + |r_g| \sqrt{h_2^2 \tilde{N}_{2j}} \frac{\tilde{Z}_{2j}}{\sqrt{1 + (1-r_g^2)N_{2j}h_2^2 l_j / M}}$$

We use these weights to implement the meta-analysis of the GWAS of ADHD-related behaviour from EAGLE/QIMR with GWAS of ADHD diagnoses.

Notes on Implementation

It may be noted that \tilde{Z}_{2j} , \tilde{N}_{1j} , and \tilde{N}_{2j} are computed from both observed values (Z_{1j} , Z_{2j} , N_{1j} , N_{2j} , and P) and unknown population parameters (K , r_g^2 , h_1^2 , h_2^2 , l_j , and M). Sensible estimates for each of these population parameters can be obtained as described below and plugged into the expression for the weights. Importantly, estimation error in K , r_g^2 , h_1^2 , h_2^2 , l_j , and M is only expected to affect the efficiency (i.e. power) of the meta-analysis. For a fully null variant (i.e. $\beta_{1j}^* = \beta_{2j}^* = 0$), Z_{1j} and Z_{2j} both have standard normal distributions, and since they are independent any weighted combination of these Z scores will also follow the null distribution. Suboptimal weights will only affect the power of the meta-analysis when the null hypothesis does not hold. In addition, since we use \tilde{Z}_{2j} in place of

\tilde{Z}_{2j} and $|\tilde{Z}_{2j}| \leq |Z_{2j}|$, the test of $Z_{j,meta}$ will be conservative when there is no genetic effect specific to the continuous phenotype (i.e. $\beta_{2j}^* = 0$).

With this reassurance, we obtain estimates of K , r_g^2 , h_1^2 , h_2^2 , l_j , and M as follows:

- Estimates of K can be derived from the literature. We use $K = .05$ for the prevalence of ADHD here and throughout this paper⁹⁰.
- Estimates of r_g^2 , h_1^2 and h_2^2 are computed from the GWAS results using LD score regression⁴⁵. For meta-analysis with the ADHD GWAS, we use the European ADHD GWAS to estimate these parameters in order to ensure that the population ancestry is matched between the input GWAS for ADHD, EAGLE/QIMR, and the European reference panel used for computing LD scores.
- For M , we utilize the value $M = 5,961,159$ corresponding to LD scores computed from 1000 Genomes Project Phase 3 data on individuals of European ancestry³⁹. These LD scores have been described previously⁶⁹ and are publically available for download at: <http://data.broadinstitute.org/alkesgroup/LDSCORE/>
- We estimate $l_j = 124.718$, the mean LD score of common HapMap3 SNPs as computed from 1000 Genomes Project Phase 3 data on individuals of European ancestry (as above). We use this value for two reasons: (a) using a single value is convenient and allows meta-analysis for all variants in the GWAS, including variants that may not be present in the precomputed LD scores from the 1000 Genomes Project reference data; and (b) we can demonstrate that the value of l_j has a trivial impact on the derived meta-analysis weights when the sample sizes, heritabilities, and genetic correlation are at the levels observed in the current study.

To demonstrate this final point, we consider the relative difference in \tilde{Z}_{2j} and \tilde{N}_{2j} conditional on a range of possible values of l_j with fixed values of N_{1j} , N_{2j} , and P and fixed estimates of K , r_g^2 , h_1^2 , h_2^2 , and M . First, we observe that the >99.5% of 1000 Genomes LD scores have values between 0 and 1000 (Supplementary Figure 23); variants with higher LD scores are predominantly from known regions of long-range LD (e.g. the MHC region⁴⁰). Then we show that for LD scores in this range the value of $\sqrt{1 + (1 - r_g^2)N_{2j}h_2^2 l_j / M}$, the term used to adjust the magnitude of Z_{2j} to account for polygenic effects specific to the second phenotype, is minimally affected by the value of l_j conditional on the estimates of r_g^2 , N_{2j} , and h_2^2 , observed in the current study (Supplementary Figure 24). The impact of l_j on the relative effective sample size \tilde{N}_{2j} is also limited, with the weight effectively unchanged for values of l_j between 0 and 1000 (Supplementary Figure 25).

We note however that the limited effect of l_j on the shrinkage parameter and the relative effective sample size is conditional on the estimated values for the other parameters in the current study.

Specifically, l_j contributes to the meta-analysis weights through the term $\sqrt{1 + (1 - r_g^2)N_{2j}h_2^2 l_j / M}$.

When r_g is large (e.g. $r_g = .97$ for the ADHD GWAS and EAGLE/QIMR) and N_{2j} and h_2^2 are modest (e.g. $N_{2j} = 20,464$ and $h_2^2 = .064$ for EAGLE/QIMR) the potential contribution of l_j to the value of this term is limited. On the other hand, if the genetic correlation is further from 1 (e.g. $r_g = .7$) or the GWAS of the continuous phenotype is better powered (e.g. $N_{2j} = 40,000$ and $h_2^2 = .16$) then the influence of l_j on the meta-analysis weights becomes non-negligible (Supplementary Figures 24-25). Thus, although the parameters for the current study enable the convenient use of mean l_j for all variants, this simplification cannot be expected to hold for all studies.

Supplementary Tables

Supplementary Table 1. Samples included in the GWAS meta-analyses of diagnosed ADHD

Samples marked in bold are included in the GWAS meta-analysis. The European GWAS meta-analysis excludes samples marked with “*” and PUWMa (strict) is used in place of PUWMa. For each cohort, the number of cases and controls and the percentage of females (“% F”) among each group is reported. The age group of ADHD cases (children or adults) is given. Principal components (“PCs”) and other covariates used in each GWAS are indicated. “Literature” lists previously published studies including the ADHD cohort.

Sample	Cases	% F cases	Controls	% F controls	Age Group	Sample design	Ancestry	Covariates	Genotyping chip	Literature
Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH)										
iPSYCH-ADHD, Denmark	14584	26.6%	22492	49.2%	Children & adults	Case-control	European	PCs	PsychChip	New
Psychiatric Genomics Consortium (PGC)										
Barcelona, Spain	572	30.20%	425	23.50%	Adults	Case-control	European	PCs	Illumina Omni1-Quad	Ribases et al. ⁹¹ , Sanchez-Mora et al. ¹³
Beijing, China*	1012	15.70%	925	37.80%	Children	Case-control	Han Chinese	PCs	Affymetrix 6.0	Yang et al. ¹⁴
Bergen, Norway	295	53.60%	202	60.90%	Adults	Case-control	European	PCs	Illumina OmniExpress-12v1	Zayats et al. ¹⁵
Cardiff, UK	721	12.90%	5081	49.40%	Children	Case-control	European	PCs	Illumina 660K (cases) & Illumina 1.2M (controls)	Stergiakouli et al. ¹⁶
CHOP, USA	262	24.40%	262	24.40%	Children	Trios	European	None	Illumina 550K	Elia et al. ⁸ , Neale et al. ¹⁰
Germany	487	19.30%	1290	49.10%	Children	Case-control	European	PCs	Illumina 660K (cases) & Illumina 550v3 (controls)	Hinney et al. ¹⁷
IMAGE-I	700	12.10%	700	12.10%	Children	Trios	European	Site	Perlegen 600K	Neale et al. ^{9,10}

IMAGE-II	624	18.60%	1755	50.00%	Children	Case-control	European	PCs	Affymetrix 5.0 & Affymetrix 6.0	Neale et al. ¹⁰
PUWMa*	635	35.70%	635	35.70%	Children	Trios	Diverse (USA)	PCs	Illumina 1M-Duo	Mick et al. ¹¹ , Neale et al. ¹⁰
PUWMa (strict)	563	35.90%	563	35.90%	Children	Trios	European	PCs	Illumina 1M-Duo	
Toronto, Canada	109	24.80%	109	24.80%	Children	Trios	European	None	Affymetrix 6.0	Lionel et al. ¹²
Yale-Penn	182	30.20%	1315	42.20%	Adults	Case-control	European	PCs, alcohol dependence diagnosis	Illumina HumanOmni1-Quad & Illumina Infinium Human Core Exome	Gelernter et al. ²¹⁻²³ (studies of substance use disorders)

Supplementary Table 2. Conditional Analysis of Secondary GWAS Signals

Linkage disequilibrium (r^2) computed between the putative secondary effect variant and the index variant in individuals of European ancestry from the 1000 Genomes Project (1KG), and imputed genotypes in the merged PGC and iPSYCH cohorts. Odds ratio (OR) and standard error (SE) of the secondary variant reported for the primary GWAS and conditional on the corresponding index variant.

Variant	CHR	BP	Index Var.	r^2 with Index Variant			Marginal Association			Conditional Association		
				1KG	PGC	iPSYCH	OR	SE	P	OR	SE	P
rs3952787	1	44323244	rs11420276	0.054	0.046	0.064	1.085	0.015	3.49×10^{-8}	1.063	0.015	6.02×10^{-5}
rs304132	5	88215594	rs4916723	0.051	0.059	0.091	0.925	0.014	4.23×10^{-8}	0.939	0.015	2.03×10^{-5}

Supplementary Table 3. Summary of Bayesian Credible Set Results

For each genome-wide significant locus (denoted with the chromosome [CHR] and base pair [BP] position of the index variant), the number of variants that are in the 99% credible set computed based on linkage disequilibrium (LD) in the PGC cohorts, the 99% credible set computed based on LD in the iPSYCH data, and their overlap.

CHR	Index Variant	BP	Number of Variants in Credible Set		
			PGC only	iPSYCH only	Both sets
1	rs11420276	44184192	0	0	96
12	rs1427829	89760744	0	0	13
16	rs212178	72578131	0	3	21
4	rs28411770	31151456	0	0	53
10	rs11591402	106747354	0	0	87
8	rs74760947	34352610	0	1	13
2	rs9677504	215181889	0	0	23
5	rs4916723	87854395	0	0	67
7	rs5886709	114086133	5	0	63
3	rs4858241	20669071	1	5	44
15	rs281324	47754018	0	0	43
1	rs1222063	96602440	0	4	4

Supplementary Table 4. Biological function of potential ADHD risk genes located in genome-wide significantly associated loci

Literature review of biological function, mutational constraint, tissue-specific expression, and relevant phenotypic associations for genes affiliated with the 12 loci significantly associated with ADHD. Mutational constraint is indexed by the estimated probability of loss-of-function intolerance (pLI) reported by the Exome Aggregation Consortium (ExAC; release 0.3.1)⁶³.

Chr	Gene	Index SNP	P-value	Function of encoded product	pLI	Tissue specificity	ADHD-related phenotype associations
1	<i>ST3GAL3</i>	rs11420276	2.14 x 10 ⁻¹³	ST3 Beta-Galactoside Alpha-2,3-Sialyltransferase 3 (<i>ST3GAL3</i>), encodes a membrane protein (ST3Gal III) that adds sialic acid to the terminal site of glycolipids or glycoproteins. ST3Gal III may play an important role in brain development as the human brain is especially enriched in sialic acid-containing glycolipids (termed gangliosides) ⁹²⁻⁹⁴ and in mice <i>St3gal2</i> and <i>St3gal3</i> were found to be responsible for nearly all the terminal sialylation of brain gangliosides ⁹⁵ as well as playing an important role for normal cognition ⁹³ . Gangliosides are known to modulate Ca(2+) homeostasis and signal transduction in neurons ^{96,97} .	0.57	This gene is expressed in several tissues including neurons ⁹³ .	Mutations in this gene have been associated with autosomal recessive mental retardation ⁹⁴ and early infantile epileptic encephalopathy ⁹⁸ . DNA methylation at sites annotated to <i>ST3GAL3</i> were reported capable to differentiate individuals with high and low ADHD symptomatology ratings ⁹⁹ . Variants in <i>ST3GAL3</i> have also been associated with educational attainment ⁸² .
1	<i>PTPRF</i>	rs3001723	3.62 x 10 ⁻¹⁰	Homo sapiens protein tyrosine phosphatase, receptor type, F (<i>PTPRF</i>). <i>PTPRF</i> encodes the leukocyte common antigen-related (LAR) receptor PTP which is present in neurons expressing TrkB, and LAR is associated with caveolae and regulates survival and	1.00	This gene is expressed in several tissues including neurons ¹⁰¹ .	Gws association of genetic markers in <i>PTPRF</i> with schizophrenia has been found ⁵⁵ . Gws association of genetic markers in <i>PTPRF</i> with educational attainment has been found in a study of individuals from the UK Biobank (N=112,151) ⁷⁶ . Overexpression of the

				neurite outgrowth ¹⁰⁰ . The LAR receptor is present in neurons expressing TrkB, which is receptor for the brain-derived neurotrophic factor (BDNF), and it has been demonstrated that LAR, through its interaction with TrkB can affect the neurotrophic activity of BDNF ¹⁰¹ .			LAR receptor encoded by <i>PTPRF</i> may contribute to insulin resistance ¹⁰² .
1	Intergenic	rs1222063	3.07 x 10 ⁻⁸	Not applicable.	n/a	Not applicable.	Not applicable.
2	<i>SPAG16</i>	rs9677504	1.39 x 10 ⁻⁸	Homo sapiens sperm associated antigen 16 (<i>SPAG16</i>). <i>SPAG16</i> encodes two major proteins that associate with the microtubular backbone of sperm tails and the nucleus of postmeiotic germ cells ^{103,104} .	0.00	Highly expressed in testis, but also detected throughout many tissues including brain, spinal cord, pituitary ovary, esophagus, thyroid, vagina, tibial nerve, bladder (http://www.gtexportal.org/home/gene/ENSG00000144451.14)	Studies have suggested that <i>SPAG16</i> may play a role in multiple sclerosis ^{105,106} .
3	Intergenic	rs4858241	1.74 x 10 ⁻⁸	Not applicable.	n/a	Not applicable.	Not applicable.
4	<i>PCDH7</i>	rs28411770	1.15 x 10 ⁻⁸	Homo sapiens protocadherin 7 (<i>PCDH7</i>). This gene belongs to the protocadherin gene family, a subfamily of the cadherin superfamily. It encodes an integral membrane protein that is thought to function in cell-cell recognition and calcium-dependent adhesion ^{107,108} and plays an important role in neuron development ¹⁰⁹ .	n/a	The gene is expressed in several brain regions especially the thalamus, cerebral cortex and brainstem circuits ¹¹⁰ .	Variants in <i>PCDH7</i> have been significantly associated with generalised epilepsy in GWAS ¹¹¹ . <i>PCDH7</i> is a target gene for <i>MECP2</i> ¹¹² and <i>MECP2</i> mutations causes Rett syndrome, which is a neurodevelopmental disorder characterized by loss of speech, microcephaly, seizures, and mental retardation (http://omim.org/entry/312750).
5	<i>LINC00461</i>	rs4916723	1.58 x 10 ⁻⁸	Homo sapiens long intergenic noncoding RNA 461 (<i>LINC00461</i>), also known as visual cortex-expressed gene	n/a	Primarily expressed in the brain ⁵² (https://www.gtexportal.org)	Variants in <i>LINC00461</i> have been associated with educational attainment ⁸² .

				(Visc). The locus is conserved across diverse mammals, but mouse knockouts of the Visc-2 transcript do not exhibit any clear anatomical phenotype ¹¹³ .		/home/gene/LINC00461). In mice, it's strongly localized in the cortex and sites of neurogenesis during neurodevelopment and continuing into adulthood ¹¹³ .	
5	<i>MEF2C</i> / <i>MEF2C-ASI</i>	rs304132	4.22 x 10 ⁻⁸	Homo sapiens myocyte enhancer factor 2C (<i>MEF2C</i>). <i>MEF2C</i> encodes a member of the MADS box transcription factors, which binds to the conserved MADS box sequence motif ¹¹⁴ . <i>MEF2C</i> is important for normal neuronal function by regulating neuronal proliferation, differentiation, survival and synapse development ¹¹⁵⁻¹¹⁷ . Plays a role in hippocampal-dependent learning and memory, possibly by controlling the number of excitatory synapses ^{116,118} .	0.00	<i>MEF2C</i> is expressed in brain especially the frontal cortex, cortex and skeletal muscle (http://www.gtexportal.org/home/gene/MEF2)	Mutations and deletions in <i>MEF2C</i> have been associated with severe mental retardation, stereotypic movements, epilepsy, lack of speech and cerebral malformation (http://omim.org/entry/613443). GWAS studies have identified genome-wide significant association of variants in loci implicating <i>MEF2C</i> with Alzheimer's disease ¹¹⁹ , depression ¹²⁰ and schizophrenia ³⁵ . <i>Mef2c</i> knockout mice have demonstrated autism like behaviours ^{115,117} , and individuals with <i>MEF2C</i> deletions have been found to display autism like traits ^{121,122} .
7	<i>FOXP2</i>	rs5886709	1.66 x 10 ⁻⁸	Homo sapiens forkhead box P2 (<i>FOXP2</i>). This gene encodes a member of the forkhead/winged-helix (FOX) family of transcription factors. <i>FOXP2</i> is involved in e.g. synapse formation and neural mechanisms mediating the development of speech and language and learning related to linguistic issues ¹²³⁻¹²⁵ . It influences a large number of downstream gene targets ^{62,123,126} with potential regional or tissue-specific differences in activity ⁶² .	1.00	<i>FOXP2</i> is expressed in both fetal and adult human brain ^{127,128} .	Deletions in <i>FOXP2</i> may cause speech-language disorder 1 (SPCH1) inherited in an autosomal dominant manner. The disorder is characterized by abnormal development of several brain areas critical for both orofacial movements and sequential articulation ^{126,129} (http://omim.org/entry/602081). Candidate gene analysis previously suggested tentative evidence for association between <i>FOXP2</i> and ADHD ¹³⁰ . Variants in <i>FOXP2</i> have also been associated with educational attainment ⁸² , and age of having a child ¹³¹ .

8	<i>LINC01288</i>	rs74760947	1.35 x 10 ⁻⁸	Homo sapiens long intergenic non-protein coding RNA 1288 (<i>LINC01288</i>). No additional information available.	n/a	No information available.	No information available.
10	<i>SORCS3</i>	rs11591402	1.34 x 10 ⁻⁸	Homo sapiens sortilin-related VPS10 domain containing receptor 3 (<i>SORCS3</i>). This gene encodes a transmembrane receptor that is a member of the vacuolar protein sorting 10 receptor family ¹³² . <i>SORCS3</i> is involved in signalling and intracellular sorting ¹³³ ; important for neuronal development and synaptic plasticity ^{134,135} .	0.33	Expressed in both the prenatal and adult brain regions ¹³⁵ (http://www.gtexportal.org/home/gene/SORCS3)	Rare CNVs overlapping <i>SORCS3</i> have been suggested to be involved in ADHD ¹² . Decreased expression of <i>SORCS3</i> in brains from patients with Alzheimer's disease compared to controls has been found ¹³⁶ . GWAS studies have demonstrated strong association of variants in <i>SORCS3</i> with schizophrenia (however not gws) ³⁵ and gws association with depression ¹²⁰ .
12	<i>DUSP6</i>	rs1427829	1.82 x 10 ⁻⁹	Homo sapiens dual specificity phosphatase 6 (<i>DUSP6</i>). The protein encoded by <i>DUSP6</i> is a member of the dual specificity protein phosphatase subfamily ¹³⁷ . <i>DUSP6</i> (also referred to as mitogen kinase phosphatase 3 (<i>MKP-3</i>)) is involved in negative regulation of mitogen-activated protein kinases (MAPKs) by acting as a dual phosphatase that dephosphorylate MAPKs at both threonine and tyrosine residues and thereby inactivate them ¹³⁸ . <i>DUSP6</i> is a cytoplasmic enzyme which has preference for extracellular signal-regulated MAPKs ¹³⁸⁻¹⁴⁰ . MAPKs are components of highly conserved signal transduction pathways, responding to a wide variety of extracellular and intracellular stimuli, and they are involved in e.g. embryogenesis, cellular proliferation and differentiation ¹⁴¹⁻¹⁴³ . Additionally <i>MKP-3</i> has been suggested to play a role in regulating	0.91	Generally expressed at low levels in the brain (http://gtexportal.org/home/gene/DUSP6) and is strongly regulated during development ¹⁴³ .	Rare mutations in <i>DUSP6</i> may lead to congenital hypogonadotropic hypogonadism ¹⁴⁷ (http://omim.org/entry/602748). <i>DUSP6</i> may play a role in Hirschsprung's disease, due to decreased expression ¹⁴⁸ . <i>DUSP6</i> have been found in reduced levels in Alzheimer brains ¹⁴⁹ . Additionally, <i>MKP-3</i> may play a critical role in cancer development ^{150,151} .

				neurotransmitter homeostasis, as increased <i>MKP-3</i> was found to reduce depolarization-dependent release of dopamine in rat PC12 cells, possibly through a down regulation of <i>Cacnalc</i> ¹⁴⁴ . It has been suggested that <i>MKP-3</i> stabilizes the dopamine transporter in the presynaptic dopaminergic neuron ¹⁴⁵ . <i>MKP-3</i> is upregulated by methamphetamine ¹⁴⁶ .			
15	<i>SEMA6D</i>	rs281324	2.68 x 10 ⁻⁸	Homo sapiens sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6D (<i>SEMA6D</i>). The product encoded by this gene is a transmembrane semaphoring which play role in maintenance and remodelling of neuronal connections ¹⁵² . <i>Sema6D</i> acts as ligand for PlexinA1 which is involved in critical steps of neuronal development in the spinal cord ¹⁵³ as well as cardiac development ^{154,155} .	1.00	Expressed in adult brain, spinal cord, and fetal brains ¹⁵² .	Variants in <i>SEMA6D</i> have been associated with educational attainment ⁸² .
16	<i>LINC01572</i>	rs212178	7.68 x 10 ⁻⁹	Homo sapiens long intergenic non-protein coding RNA 1572 (<i>LINC01572</i>). No additional information available.	n/a	No information available.	No information available.

Supplementary Table 5. Results from MAGMA gene-based association with ADHD

Genes demonstrating significant gene-wise association with ADHD after Bonferroni correction in the MAGMA⁴⁶ analysis. Chromosome (CHR), number of SNPs in the genes (N SNPS) and number of relevant parameters used in the model (N PARAM) are shown. Chromosome band location of the associated gene and the chromosome band location of the nearest gws single marker are shown. Genes marked in bold are not overlapping with loci with gws single markers.

Gene	CHR	N SNPS	N PARAM	P	Location of Gene	Location of nearest gws marker
<i>ST3GAL3</i>	1	483	57	7.38 x 10 ⁻¹²	1p34.1	1p34.1
<i>KDM4A</i>	1	71	28	2.15 x 10 ⁻¹¹	1p34.1	1p34.1
<i>PTPRF</i>	1	226	60	5.68 x 10 ⁻¹⁰	1p34.2	1p34.2
<i>SZT2</i>	1	87	25	8.47 x 10 ⁻⁹	1p34.2	1p34.2
<i>TIE1</i>	1	30	15	2.01 x 10 ⁻⁸	1p34.2	1p34.2
<i>MPL</i>	1	13	6	3.33 x 10 ⁻⁸	1p34.2	1p34.2
<i>CDC20</i>	1	5	5	6.34 x 10 ⁻⁸	1p34.2	1p34.2
<i>HYI</i>	1	5	4	3.28 x 10 ⁻⁷	1p34.2	1p34.2
<i>SLC6A9</i>	1	60	31	7.58 x 10 ⁻⁷	1p34.1	1p34.1
<i>ELOVL1</i>	1	3	3	1.26 x 10 ⁻⁶	1p34.2	1p34.2
<i>CCDC24</i>	1	6	5	2.12 x 10 ⁻⁶	1p34.1	1p34.1
<i>MANBA</i>	4	203	55	6.00 x 10⁻⁸	4q24	4p15.1(PCDH7)
<i>MEF2C</i>	5	320	54	3.19 x 10 ⁻⁸	5q14.3	5q14.3
<i>FOXP2</i>	7	812	110	5.50 x 10 ⁻⁷	7q31.1	7q31.1
<i>SORCS3</i>	10	1823	106	2.18 x 10 ⁻⁹	10q25.1	10q25.1
<i>CUBN</i>	10	1172	167	1.59 x 10⁻⁷	10p13	10q25.1 (SORCS3)
<i>PIDD1</i>	11	27	12	5.30 x 10⁻⁷	11p15.5	NA
<i>DUSP6</i>	12	20	8	2.24 x 10 ⁻⁹	12q21.33	12q21.33
<i>SEMA6D</i>	15	1458	138	2.63 x 10 ⁻¹⁰	15q21.1	15q21.1
<i>CDH8</i>	16	764	79	4.67 x 10⁻⁸	16q21	16q22.2 (LINC01572)

Supplementary Table 6. Results from MAGMA gene-based association of ADHD candidate genes

MAGMA⁴⁶ analysis of previously reported candidate genes for ADHD from a recent review⁵⁸.

Number of SNPs in the genes (N SNPS), number of relevant parameters used in the model (N PARAM) are shown.

Gene symbol	Entrez ID	N SNPS	N PARAM	Z	P
<i>SLC9A9</i>	285195	1609	129	3.395	3.40 x10 ⁻⁴
<i>DRD5</i>	1816	4	2	-1.374	0.92
<i>SLC6A3</i>	6531	101	12	-0.975	0.84
<i>HTR1B</i>	3351	3	1	2.246	0.012
<i>DRD4</i>	1815	5	2	-0.192	0.58
<i>NOS1</i>	4842	410	30	1.088	0.14
<i>GIT1</i>	28964	21	5	0.77	0.22
<i>SLC6A4</i>	6532	67	12	-0.021	0.51
<i>SNAP25</i>	6616	180	25	-0.512	0.7

Supplementary Table 7. Enrichment analysis of gene sets related to *FOXP2* downstream target genes

Competitive gene set analysis of each set of *FOXP2* target genes performed using MAGMA⁴⁶. For each gene set, the number of genes (N Genes), raw and semi-standardized (Std.) regression coefficients, and corresponding standard error (SE) are reported.

Gene Set	N Genes	Beta	Beta (Std.)	SE	P
Mouse brain (ChIP-chip)	219	0.016	0.002	0.06	0.39
Mouse brain (knockout)	243	0.034	0.004	0.055	0.27
Human brain (ChIP-chip)	258	-0.094	-0.011	0.053	0.96

Supplementary Table 8. Enrichment analysis for a set of highly constrained genes

Competitive gene set analysis of highly constrained genes ($pLI > 0.9$) performed using MAGMA⁴⁶.

The number of genes (N Genes), raw and semi-standardized (Std.) regression coefficients, and corresponding standard error (SE) are reported.

Gene Set	N Genes	Beta	Beta (Std.)	SE	P
Highly constrained genes	2932	0.062	0.023	0.018	2.60×10^{-4}

Supplementary Table 9. SNP heritability of ADHD

Estimated SNP heritability (h^2) and standard error (SE) using LD score regression (ldsc) and univariate GREML analysis in GCTA. Estimates are reported on the liability scale assuming a 5% population prevalence of ADHD. Only PGC European samples (Eur samples) were included. NA = not available.

Sample	h^2 (ldsc)	SE (ldsc)	h^2 (GCTA)	SE (GCTA)
iPSYCH	0.26	0.02	0.187	0.008
PGC (Eur samples)	0.12	0.03	0.104	0.013
iPSYCH+PGC (Eur samples)	0.22	0.01	NA	NA

Supplementary Table 10. Heritability and genetic correlations for PGC ADHD samples

Univariate and bivariate heritability estimates for PGC ADHD samples stratified by study design type (case-control vs. parent-offspring trios). The estimated SNP heritability (SNP- h^2) of each subset and the genetic correlation (r_g) between the two sets are reported with their respective standard errors (SE). Heritability estimates are reported on the liability scale assuming a 5% population prevalence of ADHD.

Cohort	N cases	N controls	SNP-h^2 (SE)	r_g (SE)
PGC case-control	2871	9983	0.138 (0.019)	1.02 (0.32)
PGC trios	1628	1629	0.081 (0.045)	

Supplementary Table 11. Genetic correlations of ADHD with other selected traits

Genetic correlation (r_g) and its standard error (SE) is estimated using LD score regression and summary statistics from the European GWAS meta-analysis (iPSYCH + PGC European samples (Eur samples)) and summary statistics from published GWASs of other selected traits. Genetic correlations (based on analyses of European populations) significant after Bonferroni correction are presented in the table (correction for 219 tests). Extended table with results for all 219 phenotypes can be found in Supplementary eTable 5). Four significant results are omitted in this table as they were based on analyses of populations with mixed ancestry (Body fat and Coronary artery disease) and large overlap in samples already represented in other analyses (Years of schooling 2013 and 2014), results for correlation with these traits can be found in supplementary eTable 5.

Trait	Trait type	N	r_g	SE	P-value
Childhood IQ ¹⁵⁶	Cognition/education	17,989	-0.411	0.082	5.09 x 10 ⁻⁷
Years of schooling ⁸²	Cognition/education	328,917	-0.535	0.028	1.44 x 10 ⁻⁸⁰
College completion ¹⁵⁷	Cognition/education	126,559	-0.538	0.046	3.30 x 10 ⁻³¹
Human intelligence ⁷⁵	Cognition/education	78,308	-0.407	0.039	7.03 x 10 ⁻²⁶
UKB college/university degree ⁷⁶	Cognition/education	111,114	-0.520	0.038	5.60 x 10 ⁻⁴²
UKB verbal-numerical reasoning ⁷⁶	Cognition/education	36,035	-0.360	0.050	5.58 x 10 ⁻¹³
Neuroticism ¹⁵⁸	Personality	170,911	0.264	0.046	1.02 x 10 ⁻⁸
Depressive symptoms ¹⁵⁸	Psychiatric	161,460	0.446	0.050	7.00 x 10 ⁻¹⁹
Subjective well being ¹⁵⁸	Psychiatric	298,420	-0.283	0.048	3.73 x 10 ⁻⁹
Major depressive disorder ⁷⁹	Psychiatric	461,134	0.424	0.033	7.38 x 10 ⁻³⁸
PGC cross-disorder analysis ¹⁵⁹	Psychiatric	61,220	0.266	0.046	5.58 x 10 ⁻⁹
Anorexia nervosa ⁷⁸	Psychiatric	14,477	-0.244	0.065	1.62 x 10 ⁻⁴
Body mass index ¹⁶⁰	Weight related	123,865	0.258	0.032	1.68 x 10 ⁻¹⁵
Waist circumference ¹⁶¹	Weight related	224,459	0.269	0.034	2.20 x 10 ⁻¹⁵
Hip circumference ¹⁶¹	Weight related	254,459	0.160	0.034	2.13 x 10 ⁻⁶
Waist-to-hip ratio ¹⁶²	Weight related	254,459	0.304	0.036	1.16 x 10 ⁻¹⁷
Overweight ¹⁶¹	Weight related	158,855	0.275	0.036	1.73 x 10 ⁻¹⁴
Obesity class 1 ¹⁶¹	Weight related	98,697	0.285	0.036	1.81 x 10 ⁻¹⁵
Obesity class 2 ¹⁶¹	Weight related	75,729	0.320	0.046	5.10 x 10 ⁻¹²
Obesity class 3 ¹⁶¹	Weight related	50,364	0.338	0.067	4.05 x 10 ⁻⁷
Extreme BMI ¹⁶¹	Weight related	16,068	0.254	0.052	9.31 x 10 ⁻⁷
Childhood obesity ¹⁶³	Weight related	13,848	0.216	0.046	3.29 x 10 ⁻⁶
Type 2 Diabetes ¹⁶⁴	Glycemic	149,821	0.185	0.047	7.80 x 10 ⁻⁵
HDL cholesterol ¹⁶⁵	Lipids	99,900	-0.217	0.042	2.44 x 10 ⁻⁷

Triglycerides ¹⁶⁵	Lipids	96,598	0.159	0.040	6.49 x 10 ⁻⁵
Ever vs never smoked ¹⁶⁶	Smoking behaviour	74,035	0.478	0.059	4.33 x 10 ⁻¹⁶
Cigarettes smoked per day ¹⁶⁶	Smoking behaviour	68,028	0.451	0.103	1.07 x 10 ⁻⁵
Former vs Current smoker ¹⁶⁶	Smoking behaviour	41,969	-0.344	0.086	6.74 x 10 ⁻⁵
Lung cancer ¹⁶⁷	Cancer	56,697	0.390	0.063	6.35 x 10 ⁻¹⁰
Lung cancer (all) ¹⁶⁸	Cancer	27,209	0.368	0.071	2.53 x 10 ⁻⁷
Squamous cell lung cancer ¹⁶⁷	Cancer	56,697	0.549	0.135	4.57 x 10 ⁻⁵
Age of first birth ¹³¹	Reproductive	251,151	-0.612	0.037	3.70 x 10 ⁻⁶¹
Number of children ever born ¹³¹	Reproductive	343,072	0.421	0.051	8.51 x 10 ⁻¹⁷
Age at Menopause ⁵⁹	Reproductive	69,360	-0.161	0.042	1.50 x 10 ⁻⁴
Mothers age at death ¹⁶⁹	Aging	52,776	-0.432	0.087	6.48 x 10 ⁻⁷
Fathers age at death ¹⁶⁹	Aging	63,775	-0.298	0.066	7.19 x 10 ⁻⁶
Parents age at death ¹⁶⁹	Aging	45,627	-0.376	0.091	3.51 x 10 ⁻⁵
Insomnia ⁷⁷	Sleep	113,006	0.421	0.064	3.85 x 10 ⁻¹¹
Rheumatoid Arthritis ¹⁷⁰	Autoimmune	103,638	0.162	0.042	1.32 x 10 ⁻⁴

Supplementary Table 12. Sign test results for each replication cohort

Test of whether the proportion of loci with estimated effects in the same direction as the ADHD

GWAS (π) is greater than expected by chance. Values in bold are nominally significant after

Bonferroni correction for testing in 3 replication cohorts at a given P-value threshold ($P < 0.0167$).

<i>P</i> Threshold	deCODE Concordance			23andMe Concordance			EAGLE/QIMR Concordance		
	# Loci	π	<i>P</i>	# Loci	π	<i>P</i>	# Loci	π	<i>P</i>
5.00E-08	12	0.833	0.0193	11	0.727	0.1133	12	1.000	2.44E-04
1.00E-07	14	0.786	0.0287	13	0.615	0.2905	14	1.000	6.10E-05
1.00E-06	37	0.73	0.00382	36	0.722	0.00567	34	0.824	9.76E-05
1.00E-05	99	0.636	0.00432	99	0.747	4.25E-07	98	0.694	7.81E-05
1.00E-04	--	--	--	326	0.607	6.28E-05	301	0.641	5.51E-07

Supplementary Table 13. Results for significant ADHD loci in each replication cohort

Comentado [DD1]: I have added a column with chromosome info to this table

GWAS results from deCODE, 23andMe, and EAGLE/QIMR for the genome-wide significant loci identified in the ADHD GWAS. Replication is tested for the index variant from the ADHD GWAS, or for a proxy variant when the index variant is not present in the replication cohort.

Proxy variants are identified by linkage disequilibrium (LD) clumping of the ADHD GWAS results using European ancestry samples from the 1000 Genomes Project after restricting to variants present in the replication cohort. No proxy variant is available for rs28411770. Effects (Z or odds ratio [OR]) that are sign concordant with the ADHD GWAS are indicated in bold.

Variant		chr	Effect Allele	LD to Index	iPSYCH/PGC		DeCODE		23andMe		EAGLE/QIMR	
Index	Proxy				OR	P	OR	P	OR	P	Z	P
rs11420276	--	1	G	--	1.113	2.14E-13	1.006	8.23E-01	--	--	--	--
rs11420276	rs112984125		A	0.98	0.899	3.58E-13	--	--	1.044	4.01E-02	-1.540	1.24E-01
rs1222063	--	1	A	--	1.101	3.07E-08	1.006	7.97E-01	--	--	0.522	6.02E-01
rs1222063	rs2391769		A	0.093	0.927	3.96E-08	1.013	6.01E-01	0.957	3.20E-02	-0.501	6.16E-01
rs9677504	--	2	A	--	1.124	1.39E-08	0.938	1.33E-01	1.068	3.52E-02	0.160	8.73E-01
rs4858241	--	3	T	--	1.082	1.74E-08	1.017	4.75E-01	1.016	4.44E-01	1.259	2.08E-01
rs28411770	--	4	T	--	1.090	1.15E-08	1.017	7.05E-01	--	--	0.398	6.91E-01
rs4916723	--	5	A	--	0.926	1.58E-08	0.976	2.92E-01	0.989	5.88E-01	-2.752	5.93E-03
rs5886709	--	7	G	--	1.079	1.66E-08	1.049	4.06E-02	1.045	2.62E-02	--	--
rs5886709	rs10262192		A	0.955	1.076	2.89E-08	--	--	1.045	2.57E-02	1.639	1.01E-01
rs74760947	--	8	A	--	0.835	1.35E-08	0.996	9.28E-01	0.955	3.90E-01	-0.870	3.84E-01
rs11591402	--	10	A	--	0.911	1.34E-08	0.955	9.40E-02	0.957	6.27E-02	-1.175	2.40E-01
rs1427829	--	12	A	--	1.083	1.82E-09	1.021	3.63E-01	1.036	6.97E-02	0.418	6.76E-01
rs281324	--	15	T	--	0.928	2.68E-08	0.988	6.06E-01	1.007	7.05E-01	-1.659	9.72E-02
rs212178	--	16	A	--	0.891	7.68E-09	1.012	7.32E-01	1.030	3.80E-01	-1.044	2.96E-01

Supplementary Table 14. Comparison of profile of genetic correlations for the ADHD GWAS and 23andMe

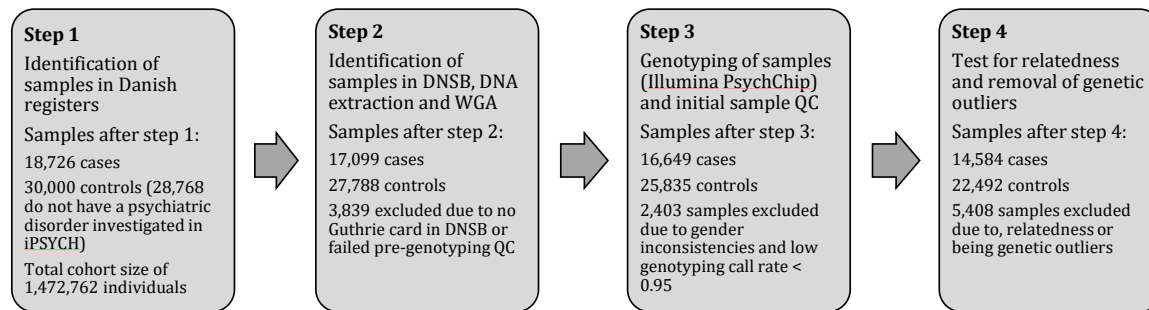
Comparison of genetic correlation (r_g) results, including block jack-knife standard errors (SE), of each GWAS with 28 selected traits from LD Hub (<http://ldsc.broadinstitute.org/ldhub/>). For each trait, the absolute difference between the 23andMe and ADHD GWAS r_g estimates is reported, along with the approximate pooled standard error and corresponding approximate P value. P values for the difference reaching nominal significance ($P < .05$) are indicated in italics, and Bonferroni-corrected significant P values ($P < 0.05/28 = 1.8 \times 10^{-3}$) are indicated in bold.

References for each trait are the same as Supplementary Table 11.

Trait	Category	ADHD GWAS			23andMe			Difference		
		r_g	SE	P	r_g	SE	P	$ r_1-r_2 $	SE	P
College completion	Cognition/education	-0.538	0.046	3.30E-31	0.056	0.086	5.13E-01	0.594	0.097	1.07E-09
Years of schooling 2016	Cognition/education	-0.535	0.028	1.44E-80	-0.199	0.058	6.00E-04	0.336	0.064	1.84E-07
Childhood IQ	Cognition/education	-0.411	0.082	5.09E-07	-0.219	0.144	1.27E-01	0.192	0.165	2.46E-01
Neuroticism	Personality	0.264	0.046	1.02E-08	0.145	0.090	1.08E-01	0.120	0.101	2.38E-01
Schizophrenia	Psychiatric	0.122	0.036	7.00E-04	0.275	0.070	8.95E-05	0.153	0.079	5.30E-02
Bipolar disorder	Psychiatric	0.095	0.055	8.05E-02	0.293	0.103	4.60E-03	0.198	0.117	9.10E-02
Subjective well being	Psychiatric	-0.283	0.048	3.73E-09	-0.146	0.102	1.53E-01	0.137	0.113	2.26E-01
Depressive symptoms	Psychiatric	0.446	0.050	7.00E-19	0.300	0.114	8.80E-03	0.146	0.125	2.44E-01
PGC cross-disorder analysis	Psychiatric	0.266	0.046	5.58E-09	0.286	0.104	6.00E-03	0.020	0.113	8.61E-01
Obesity class 1	Weight related	0.285	0.036	1.81E-15	0.108	0.070	1.26E-01	0.178	0.079	2.43E-02
Childhood obesity	Weight related	0.216	0.046	3.29E-06	-0.025	0.105	8.10E-01	0.241	0.115	3.61E-02
Waist-to-hip ratio	Weight related	0.304	0.036	1.16E-17	0.154	0.073	3.38E-02	0.150	0.081	6.30E-02
Birth weight	Weight related	-0.132	0.039	8.00E-04	0.012	0.079	8.79E-01	0.144	0.088	1.03E-01

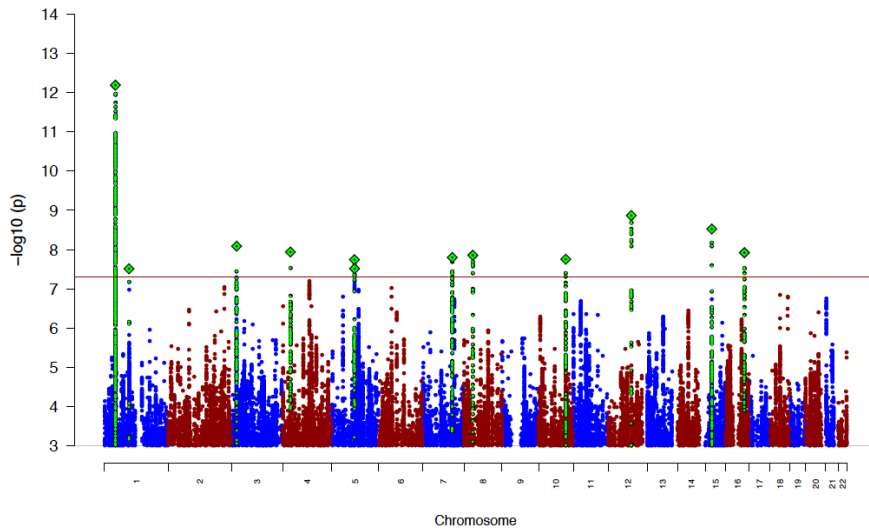
Body mass index	Weight related	0.258	0.032	1.68E-15	0.132	0.075	8.06E-02	0.126	0.082	1.25E-01
Overweight	Weight related	0.275	0.036	1.73E-14	0.183	0.077	1.72E-02	0.092	0.085	2.81E-01
Type 2 Diabetes	Glycemic	0.185	0.047	7.80E-05	0.058	0.100	5.64E-01	0.127	0.110	2.51E-01
Triglycerides	Lipids	0.159	0.040	6.49E-05	0.048	0.065	4.58E-01	0.111	0.076	1.47E-01
HDL cholesterol	Lipids	-0.217	0.042	2.44E-07	-0.131	0.082	1.11E-01	0.086	0.092	3.54E-01
Former vs Current smoker	Smoking behaviour	-0.344	0.086	6.74E-05	-0.132	0.157	4.00E-01	0.212	0.179	2.37E-01
Ever vs never smoked	Smoking behaviour	0.478	0.059	4.33E-16	0.340	0.108	1.60E-03	0.139	0.123	2.58E-01
Cigarettes smoked per day	Smoking behaviour	0.451	0.103	1.07E-05	0.343	0.149	2.08E-02	0.108	0.181	5.50E-01
Lung cancer	Cancer	0.390	0.063	6.35E-10	0.139	0.113	2.18E-01	0.251	0.129	5.15E-02
Squamous cell lung cancer	Cancer	0.549	0.135	4.57E-05	0.393	0.180	2.86E-02	0.156	0.224	4.88E-01
Age of first birth	Reproductive	-0.612	0.037	3.70E-61	-0.331	0.074	6.66E-06	0.281	0.082	6.34E-04
Number of children ever born	Reproductive	0.421	0.051	8.51E-17	0.310	0.094	1.00E-03	0.112	0.106	2.94E-01
Age at Menopause	Reproductive	-0.161	0.042	1.00E-04	-0.129	0.087	1.40E-01	0.032	0.097	7.39E-01
Parents age at death	Aging	-0.376	0.091	3.51E-05	-0.339	0.177	5.59E-02	0.038	0.199	8.50E-01
Rheumatoid Arthritis	Autoimmune	0.162	0.042	1.00E-04	-0.040	0.093	6.67E-01	0.202	0.102	4.88E-02

Supplementary Figures



Supplementary Figure 1. Genotyping iPSYCH-ADHD sample, main steps and sample loss

Flowchart demonstrating the main steps and sample loss during the process of obtaining high quality genotypes for the iPSYCH sample. Detailed description of the Danish registers, DNA generation, genotyping and QC is described under the detailed description of the iPSYCH sample.



Supplementary Figure 2. Manhattan plot from ADHD European GWAS meta-analysis
Results from GWAS meta-analysis of iPSYCH and PGC European samples.

Supplementary Figure 3.A1 - M1. Forest plots for index SNPs in gws loci

Forest plots for the index SNP in each of the gws associated regions in the GWAS meta-analysis. Each plot provides a visualization of the effect size estimates for each wave in the iPSYCH sample and each PGC sample and for the summary meta-analysis in addition the 95% confidence intervals are included for the estimates.

Supplementary Figure 3.A2 - M2. Regional association plots for index SNPs in gws loci

Regional association plots of the local association results. Each plot includes information about the gws locus, the location and orientation of the genes in the region, LD estimates of surrounding SNPs with the index SNP (r^2 values estimated based on 1KGP3) is indicated by colour (colour bar in upper left corner indicates r^2 values), if multiple index SNPs then different colour scheme for each index SNP. Additionally, the local estimates of recombination rate are indicated in light blue (legend on vertical axis at right). Detailed SNP info in upper right corner (blue letters): SNP name (rsid), P-value (p), odds ratio (or), minor allele frequency (maf), imputation INFO score (info), directions in the analysed samples/waves (risk increasing - decreasing - missing). Gene lists were downloaded from <ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/refGene.txt.gz>. Previously reported gws regions were downloaded from the NHGRI GWAS catalogue available from <http://www.ebi.ac.uk/gwas>.

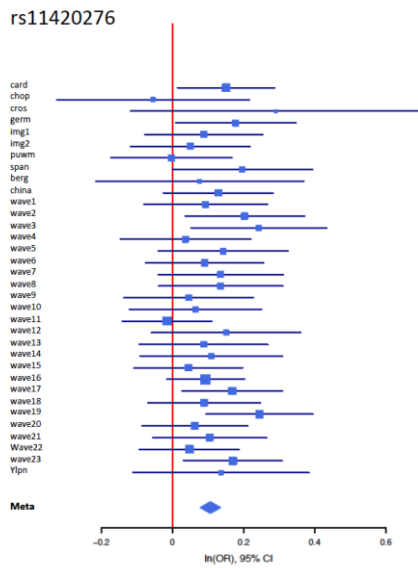


Figure 3.A1. Forest plot for rs11420276

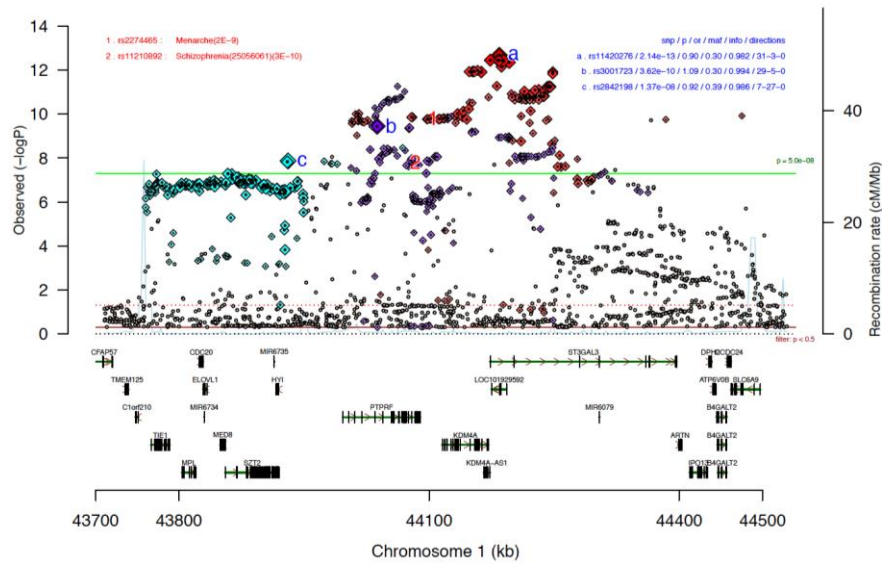


Figure 3.A2. Regional association plot for rs11420276

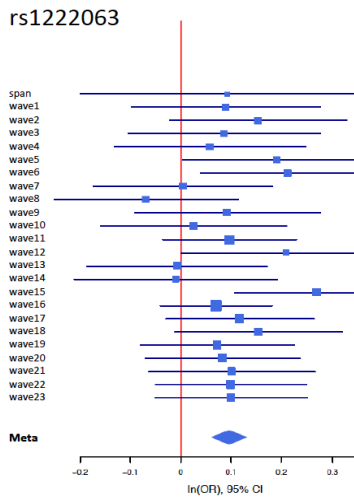


Figure 3.B1. Forest plot for rs1222063

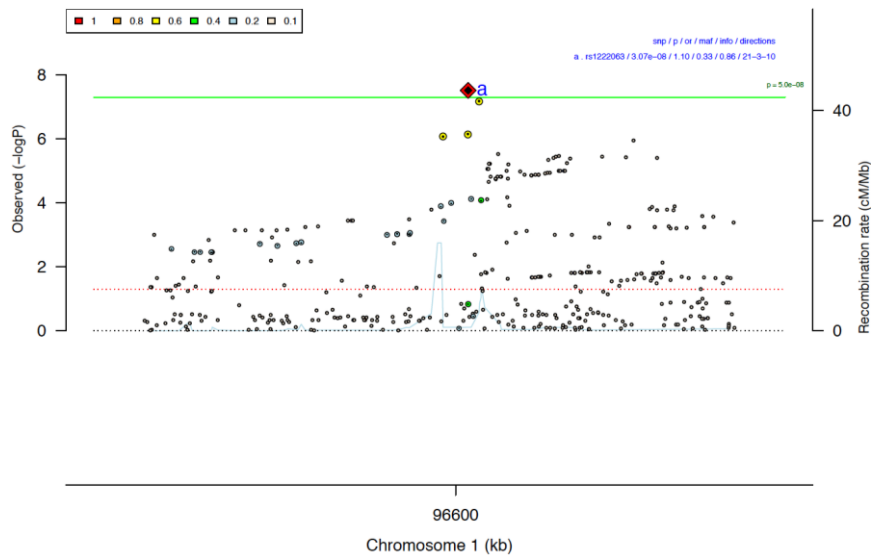


Figure 3.B2. Regional association plot for rs1222063

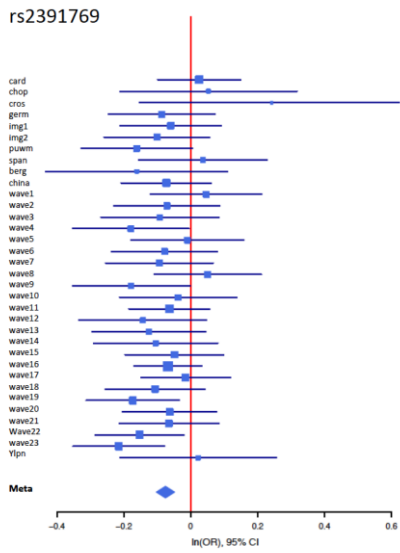


Figure 3.C1. Forest plot for rs2391769

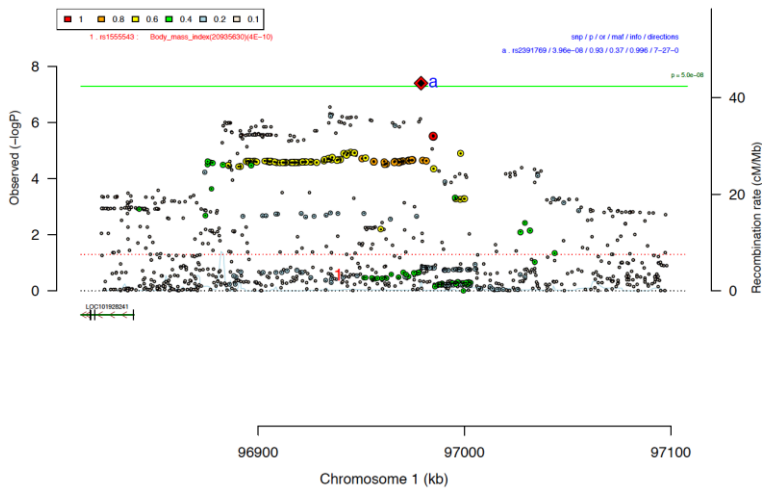


Figure 3.C2. Regional association plot for rs2391769

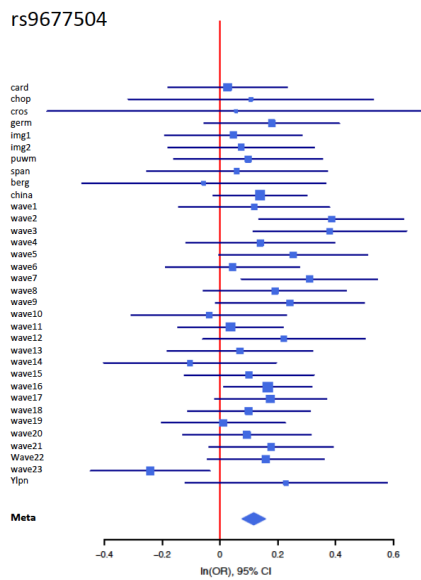


Figure 3.D1. Forest plot for rs9677504

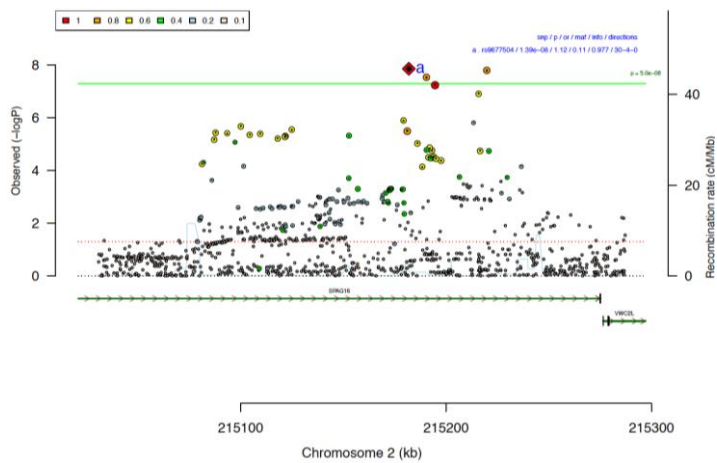


Figure 3.D2. Regional association plot for rs9677504

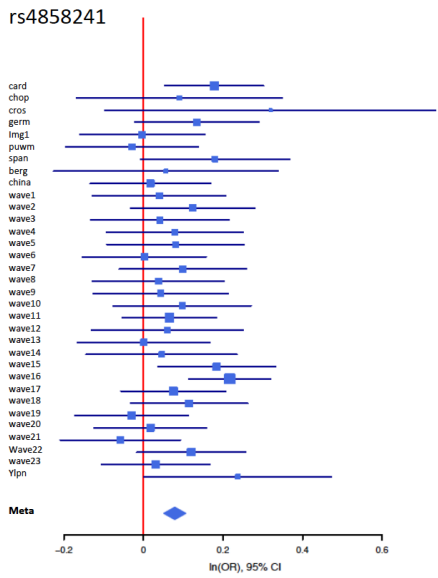


Figure 3.E1. Forest plot for rs4858241

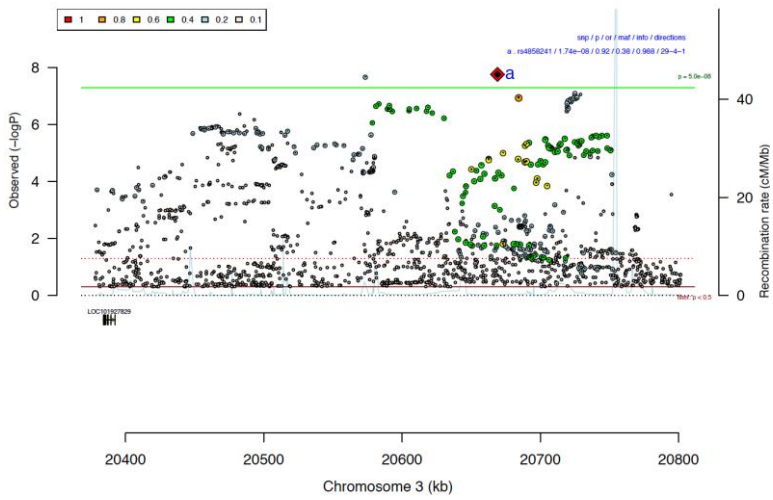


Figure 3.E2. Regional association plot for rs4858241

rs28411770

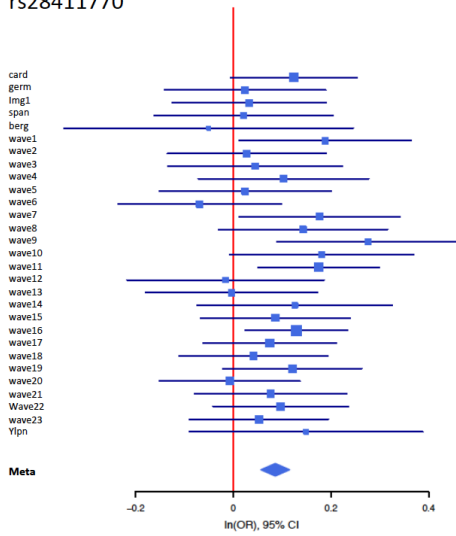


Figure 3.F1. Forest plot for rs28411770

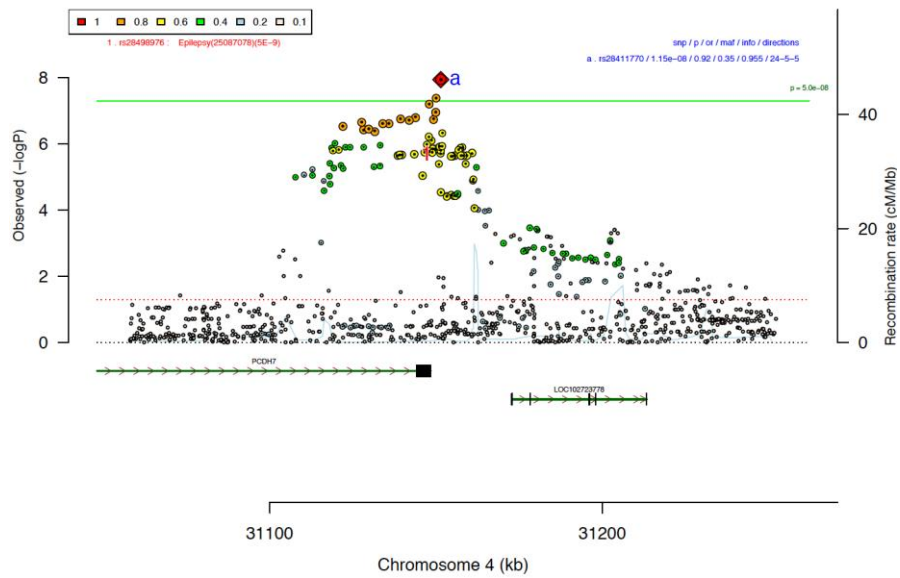


Figure 3.F2. Regional association plot for rs28411770

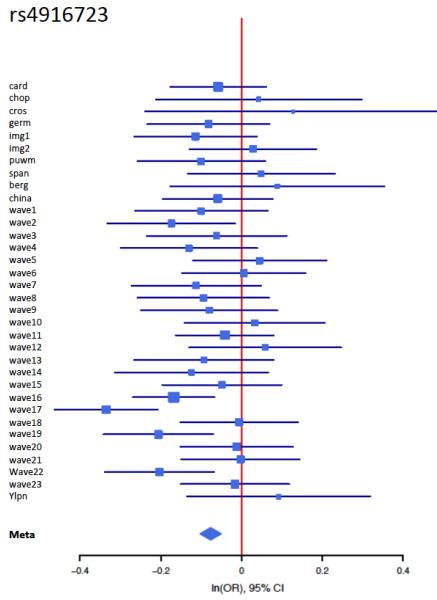


Figure 3.G1. Forest plot for rs4916723

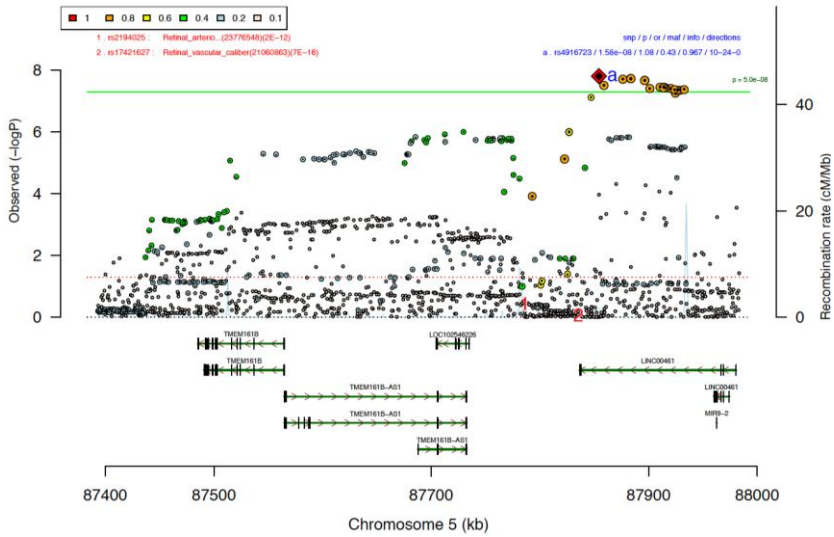


Figure 3.G2. Regional association plot for rs4916723

rs304132

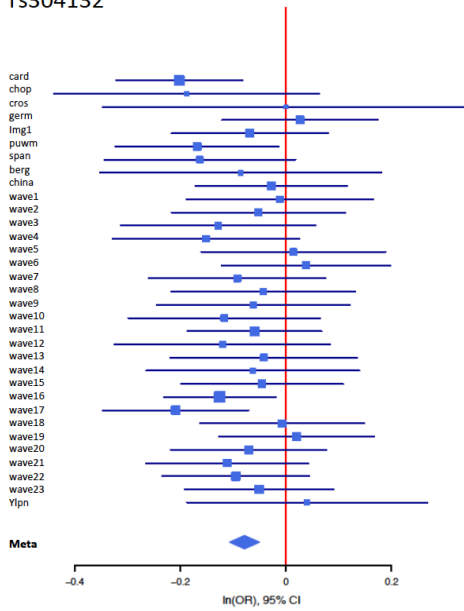


Figure 3.H1. Forest plot for rs304132

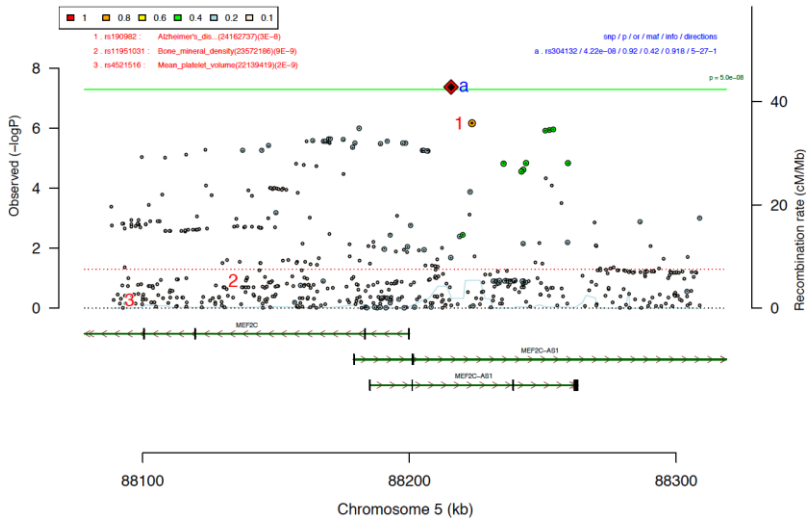


Figure 3.H2. Regional association plot for rs304132

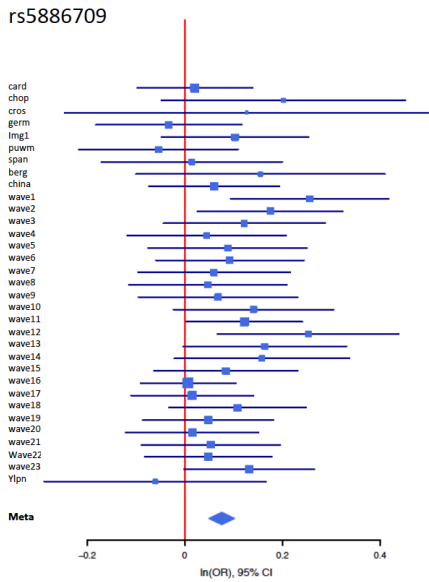


Figure 3.I1. Forest plot for rs5886709

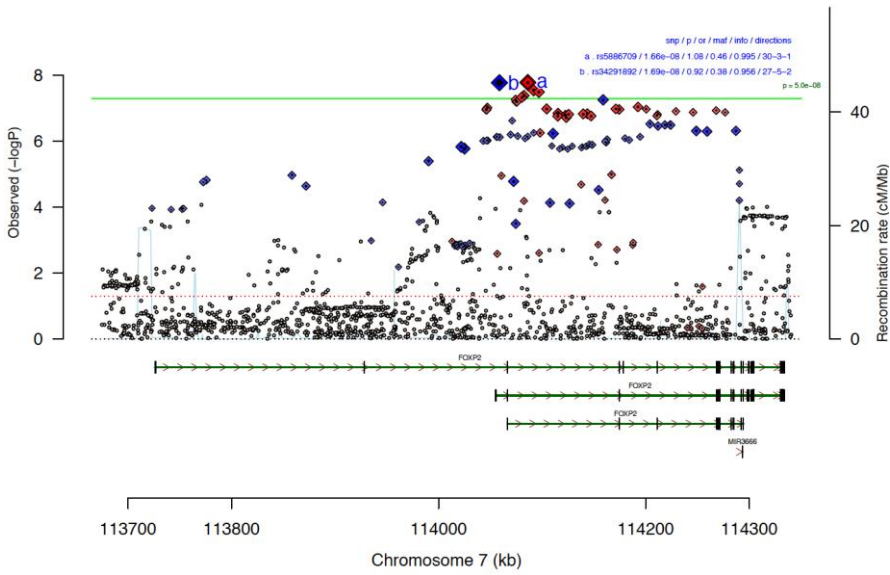


Figure 3.I2. Regional association plot for rs5886709

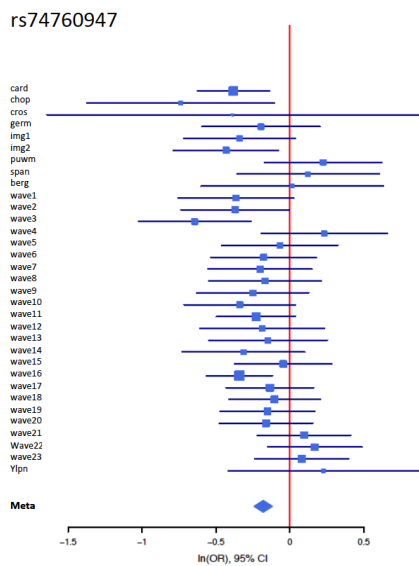


Figure 3.J1. Forest plot for rs74760947

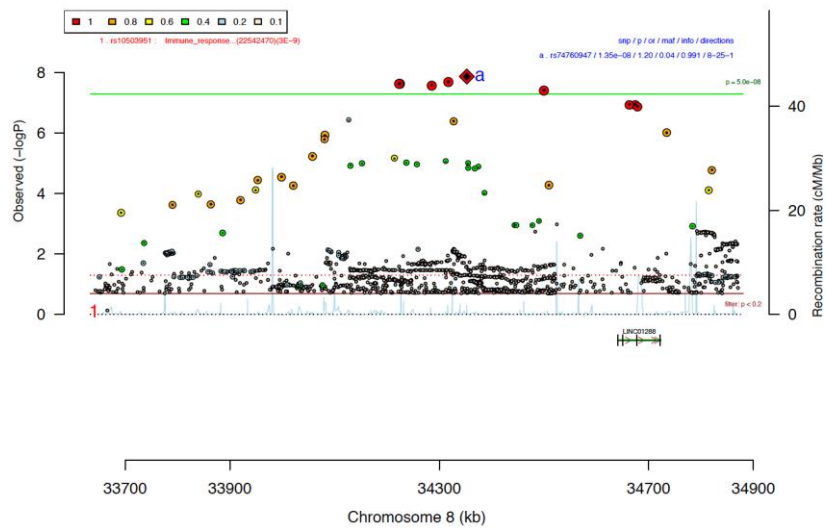


Figure 3.J2. Regional association plot for rs74760947

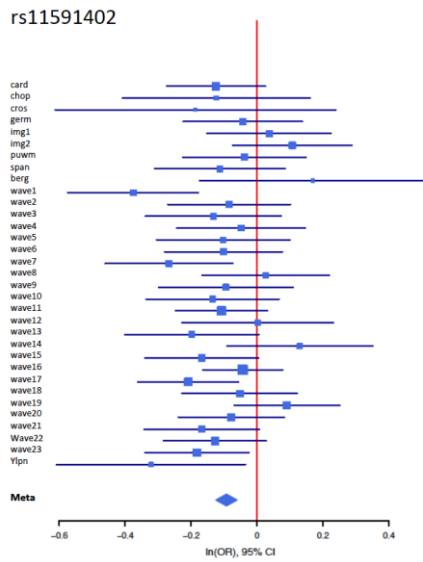


Figure 3.K1. Forest plot for rs11591402

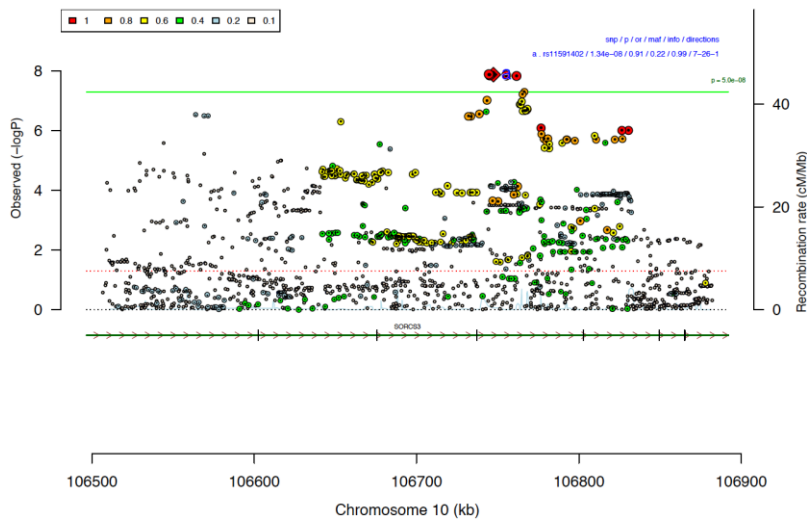


Figure 3.K2. Regional association plot for rs11591402

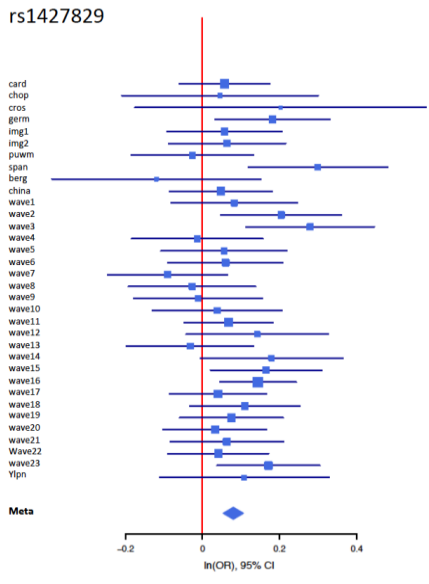


Figure 3.L1. Forest plot for rs1427829

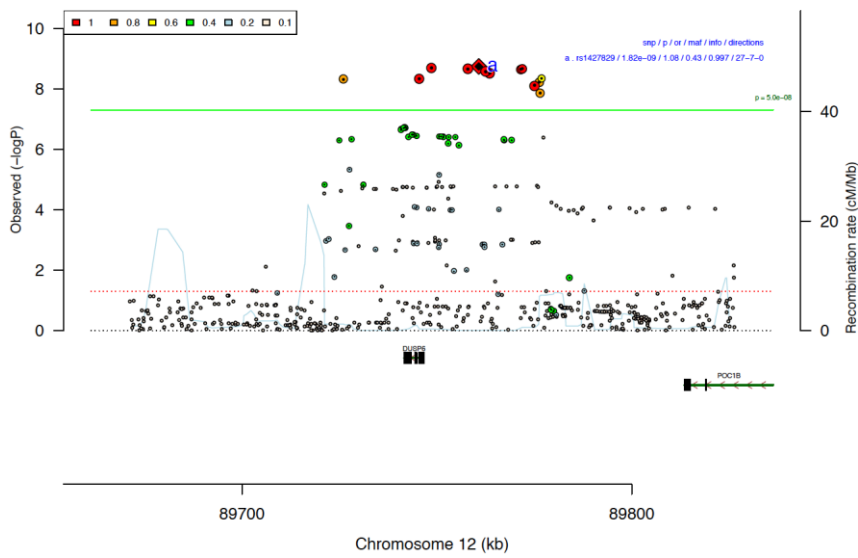


Figure 3.L2. Regional association plot for rs1427829

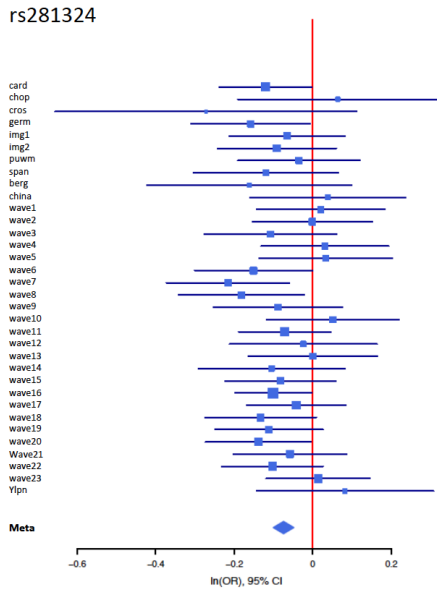


Figure 3.M1. Forest plot for rs281324

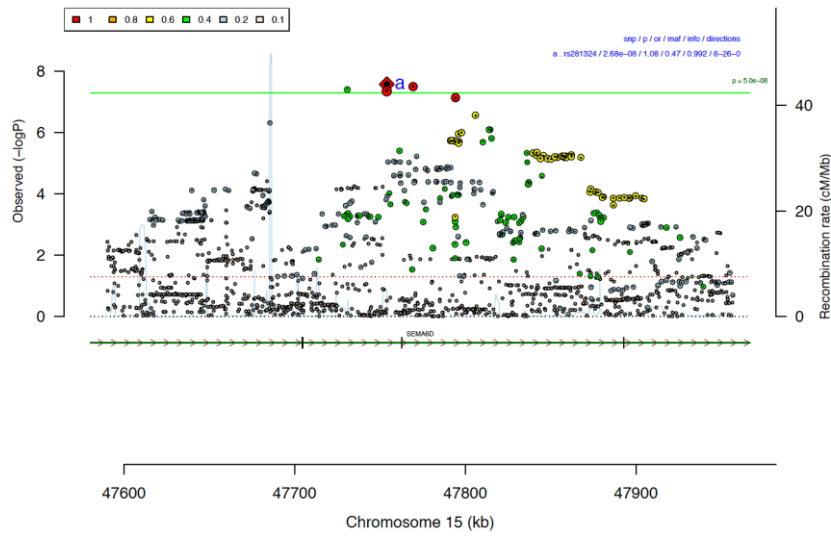


Figure 3.M2. Regional association plot for rs281324

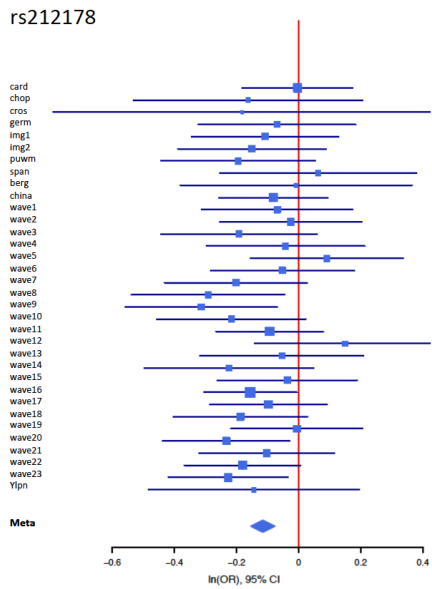


Figure 3.N1. Forest plot for rs212178

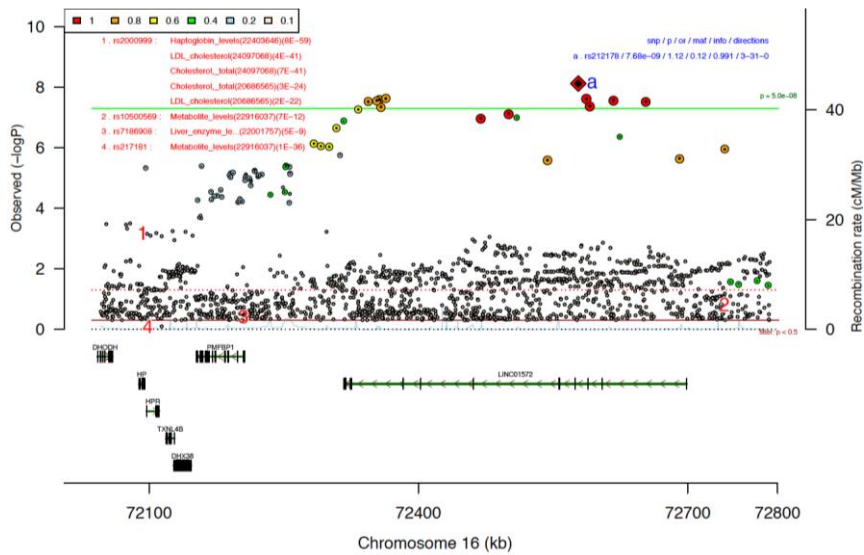
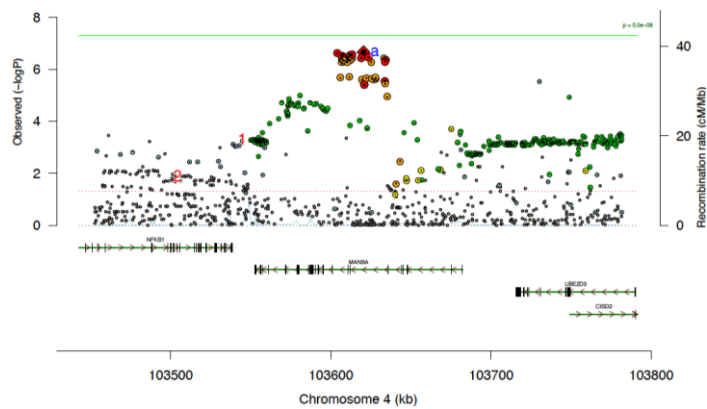


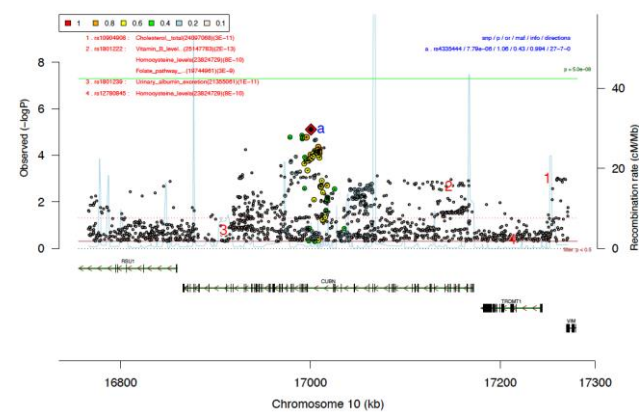
Figure 3.N2. Regional association plot for rs212178

Supplementary Figure 4.A – 4.D. Gene-based association, regional association plots

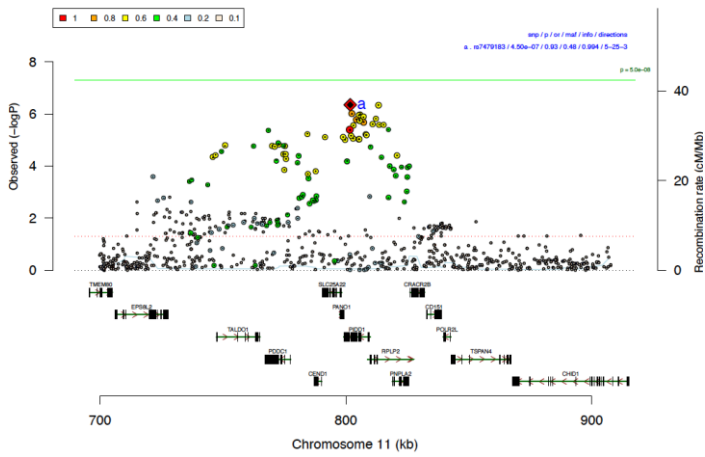
LD structure in the region around the four new genes (genes not overlapping with loci being gws in the single marker GWAS meta-analysis) significantly association with ADHD in the MAGMA gene-based association analysis.



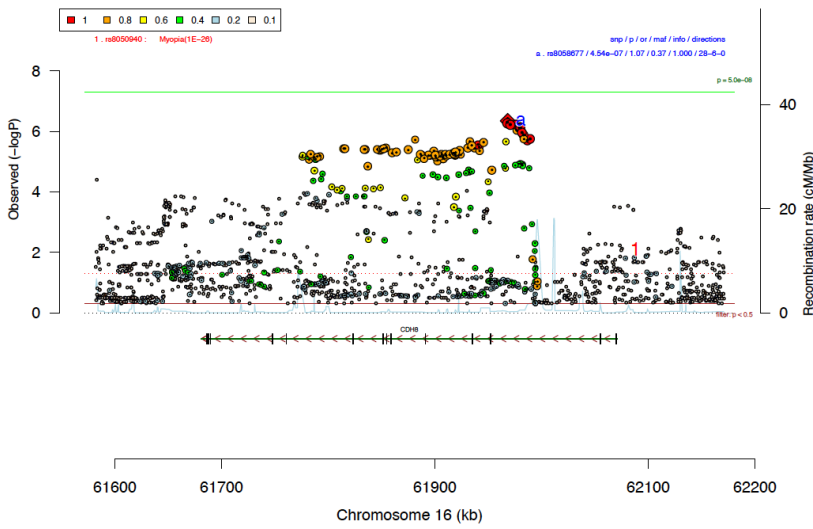
Supplementary Figure 4.A. Regional association plot for *MANBA* (+/- 100,000 bp up- and downstream of the gene).



Supplementary Figure 4.B. Regional association plot for *CUBN* (+/- 100,000 bp up- and downstream of the gene).

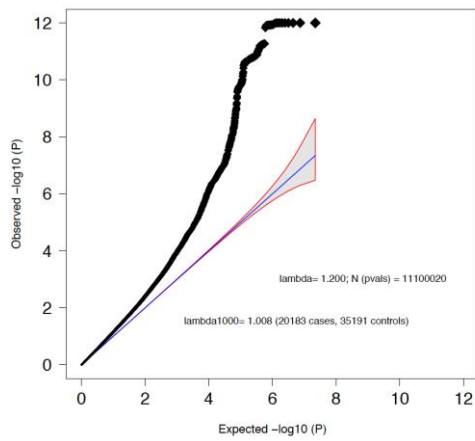


Supplementary Figure 4.C. Regional association plot for *PIDD1* (+/- 100,000 bp up- and down-stream of the gene).

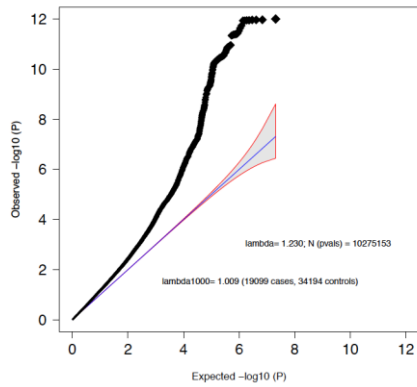


Supplementary Figure 4.D. Regional association plot for *CDH8* (+/- 100,000 bp up- and down-stream of the gene).

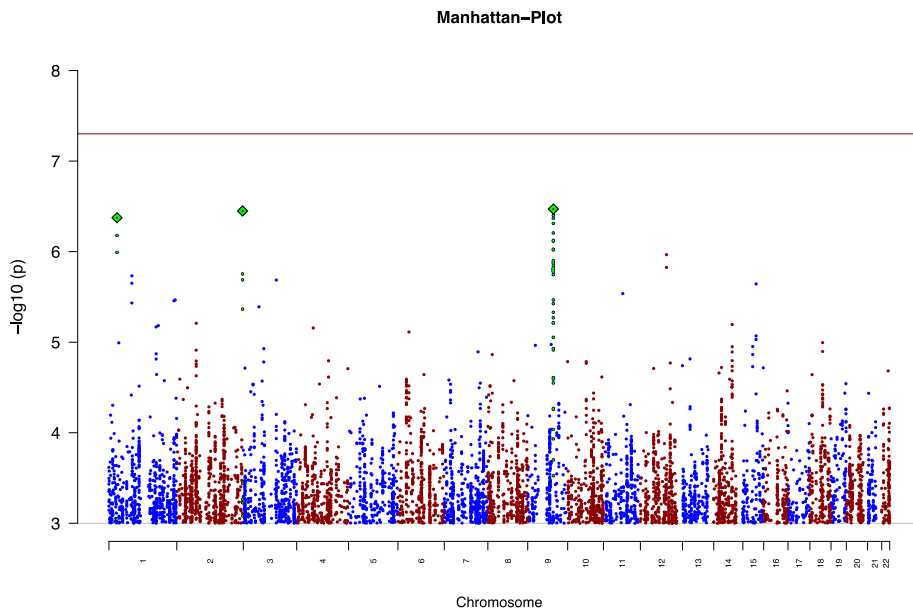
Supplementary Figure 5.A. – 5.B. Q-Q plot from GWAS meta-analyses
Quantile-quantile plot of the $-\log_{10}$ P-values from GWAS meta-analyses.



Supplementary Figure 5.A. Q-Q plot from GWAS meta-analysis

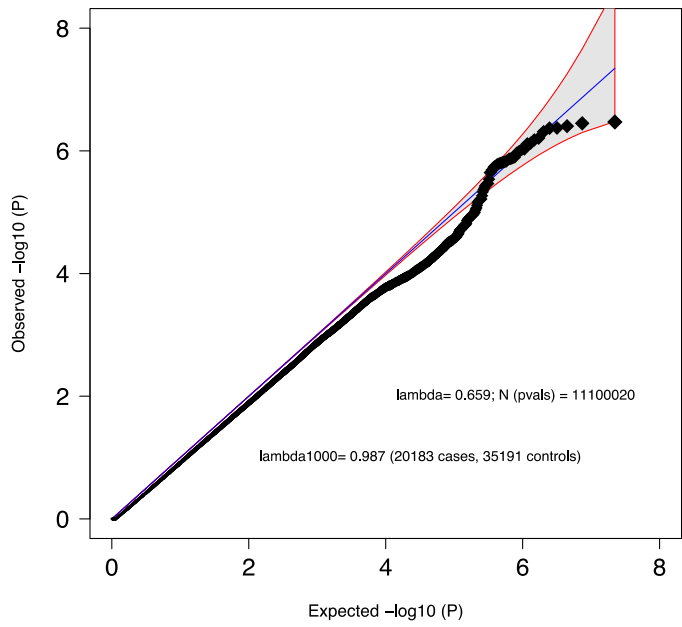


Supplementary Figure 5.B. Q-Q plot from European GWAS meta-analysis



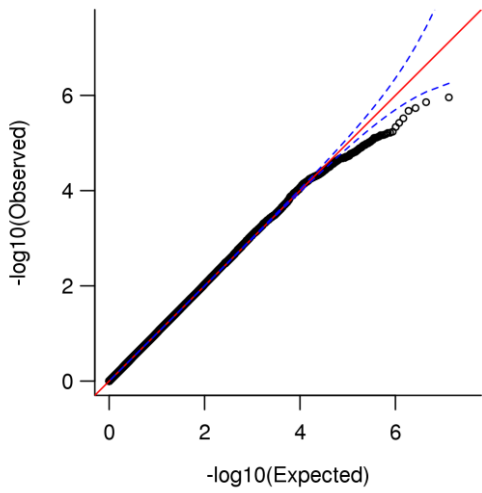
Supplementary Figure 6. Manhattan plot from test for heterogeneity between studies/waves in the ADHD GWAS meta-analysis

Omnibus test of heterogeneity across cohorts. Red reference line indicates genome-wide significance threshold (5×10^{-8}).



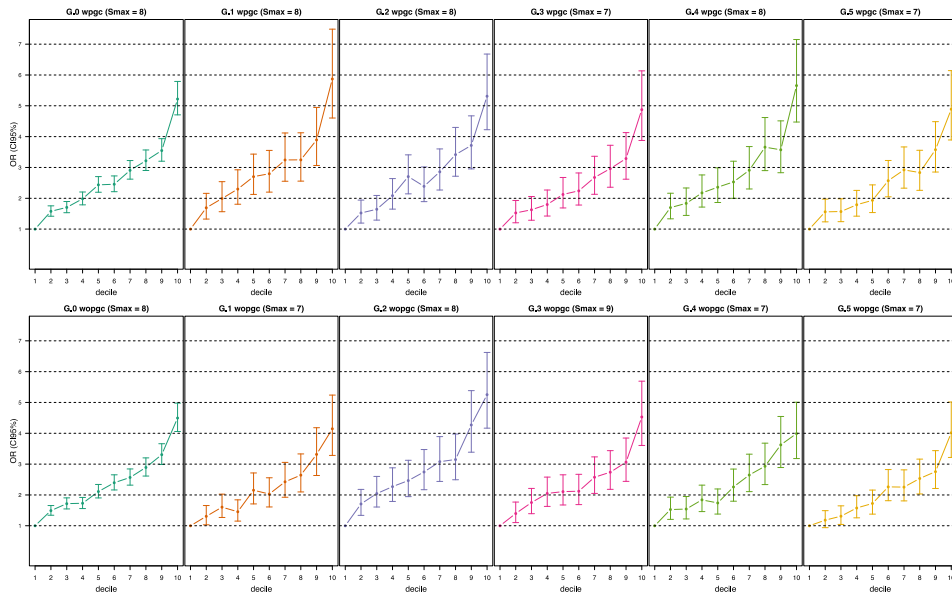
Supplementary Figure 7.A. Q-Q plot from test for heterogeneity between all samples/waves in the ADHD GWAS meta-analysis

Quantile-quantile plot of P -values from the omnibus test of heterogeneity.



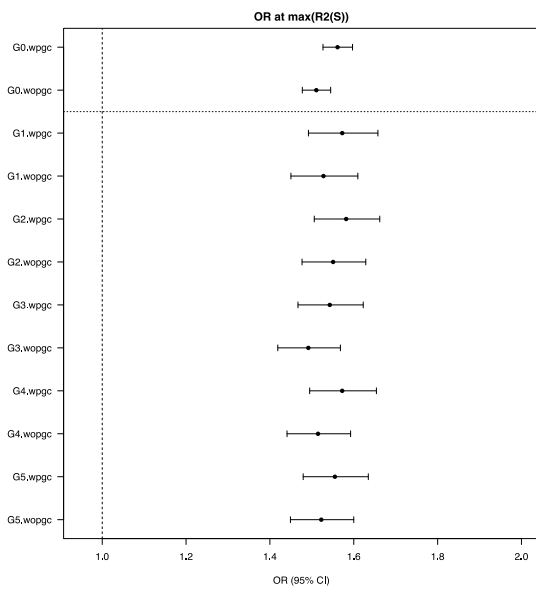
Supplementary Figure 7.B. Q-Q plot from test for heterogeneity between Chinese and European ancestry cohorts in the ADHD GWAS meta-analysis

Quantile-quantile plot of P -values from the 1 degree of freedom test of heterogeneity between the Beijing, China cohort and all European ancestry cohorts.



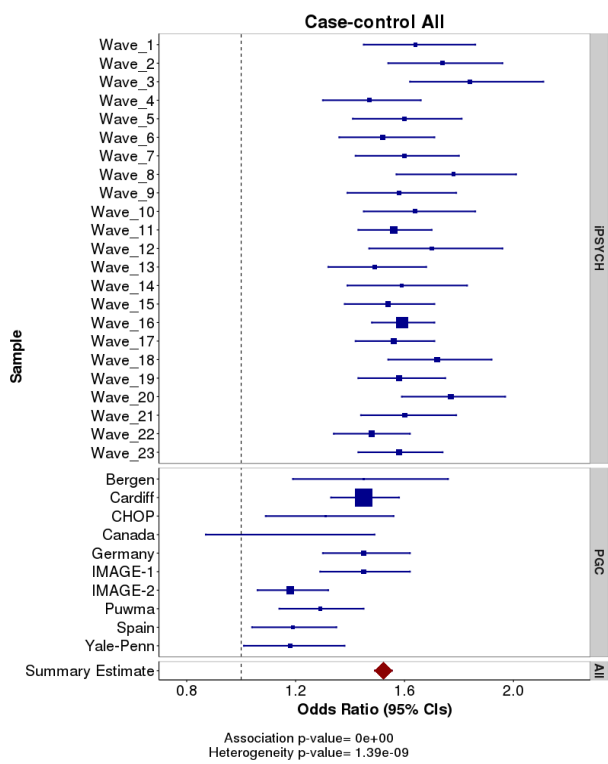
Supplementary Figure 8. Odds ratios by PRS within deciles in target groups

Odds ratio by PRS within each decile for each of the five target groups (G1-G5) and for the pooled (G0) analysis with (upper panels) and without (lower panels) PGC European samples included among the training data sets. Plots are shown for the P-value threshold with the highest Nagelkerke's R^2 (Smax). Error bars indicate 95% confidence limits.



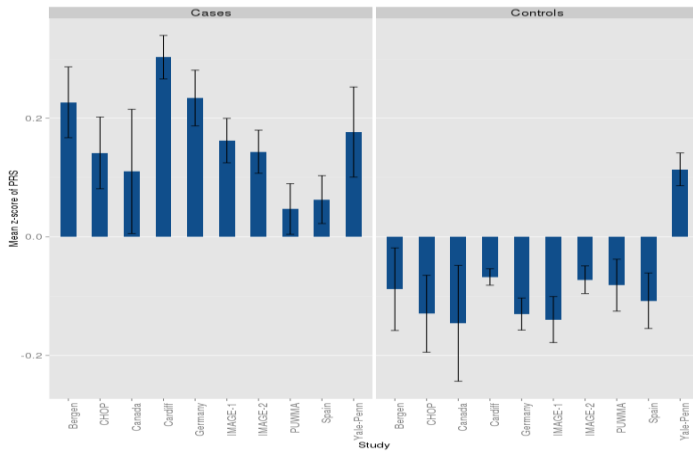
Supplementary Figure 9. Odds ratios within target groups in iPSYCH

PRS-based odds ratio and 95% confidence limits from logistic regression of continuous PRS (normalized by target group) for each target group considered separately (G1-G5) and pooled (G0). PRS estimated using iPSYCH waves alone as training sample (wopgc) or iPSYCH waves together with PGC European samples (wpgc).



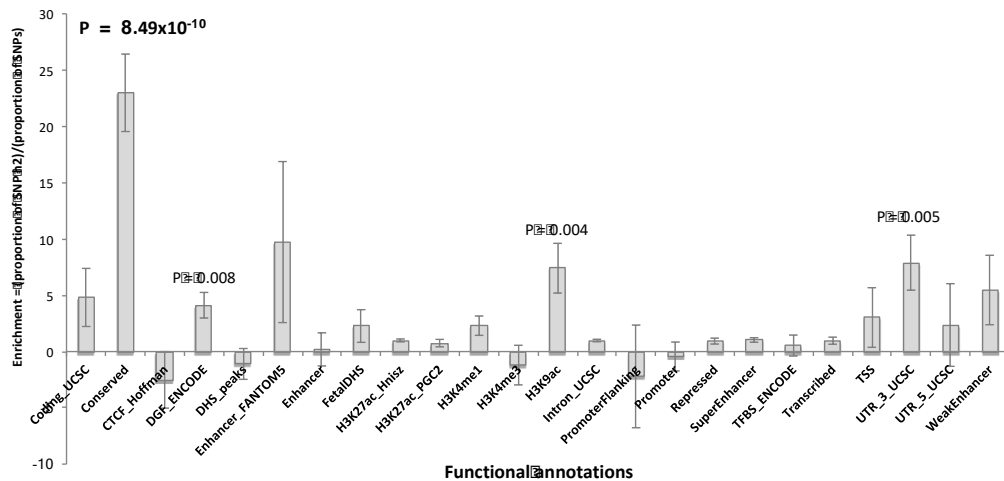
Supplementary Figure 10. PRS-based odds ratios within each study/wave

PRS-based odds ratio and 95% confidence limits from logistic regression of standardised PRS for each target study/wave.



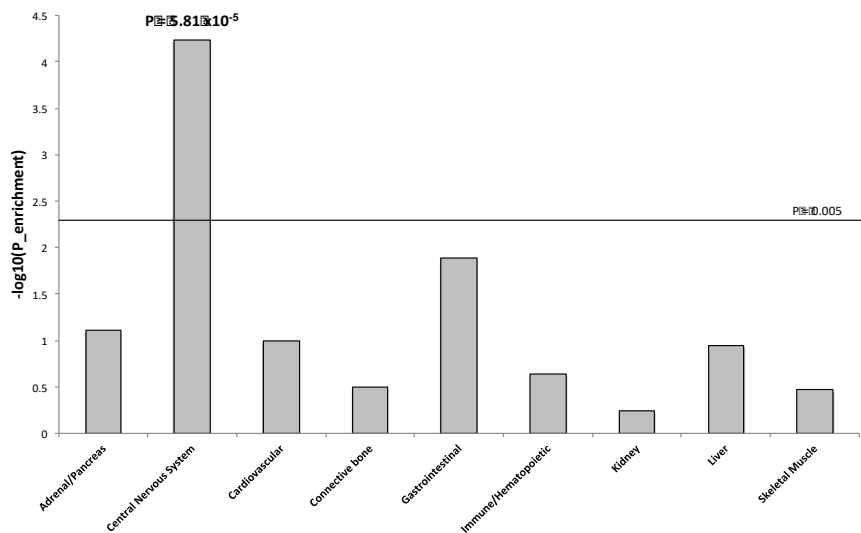
Supplementary Figure 11. ADHD PRS stratified by case-control status and PGC study

Mean PRS z-score (+/- standard error) plotted stratified by case status and PGC study.



Supplementary Figure 12. Partitioning of h^2 by functional annotations

Enrichment of heritability per SNP in 24 functional annotations defined by Finucane et al.⁶⁹ Error bars represent 95% confidence intervals. P-values for annotation categories with nominal significant enrichment are shown and values on bold indicate significance after Bonferroni correction.



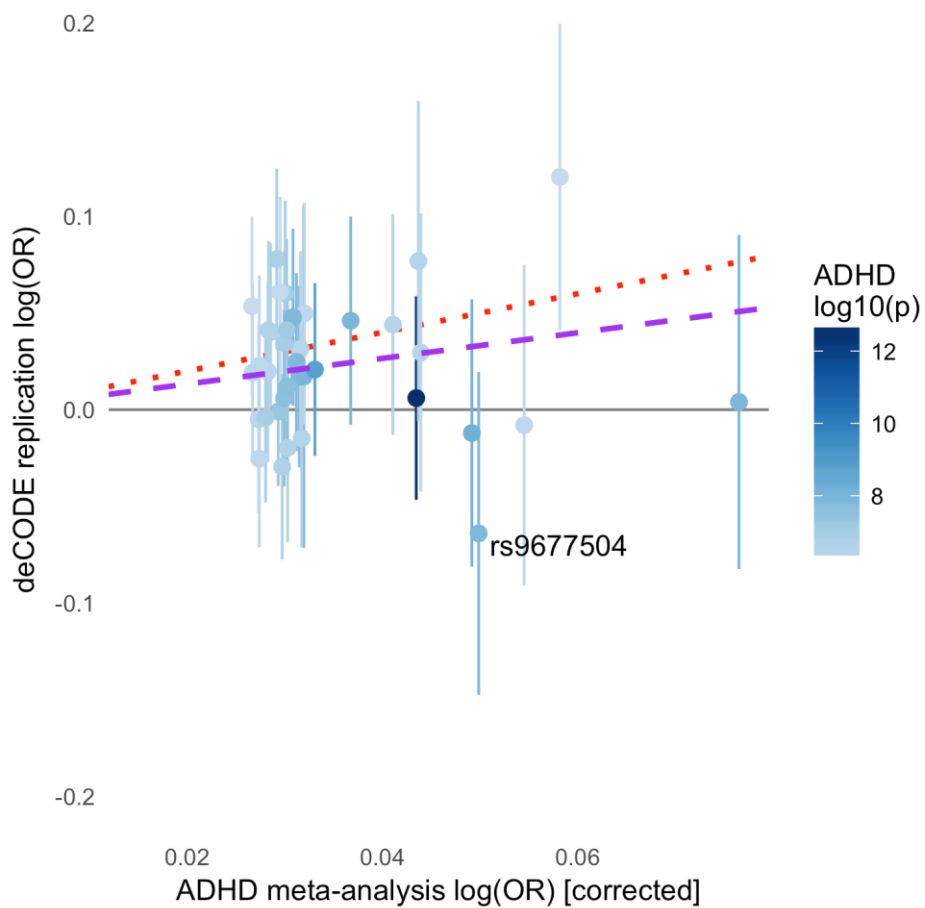
Supplementary Figure 13. Partitioning of h^2 by tissue-group annotations

Results from partitioning heritability by SNPs located in cell-group specific regulatory elements. The line indicate significance after Bonferroni correction ($P = 0.005$).



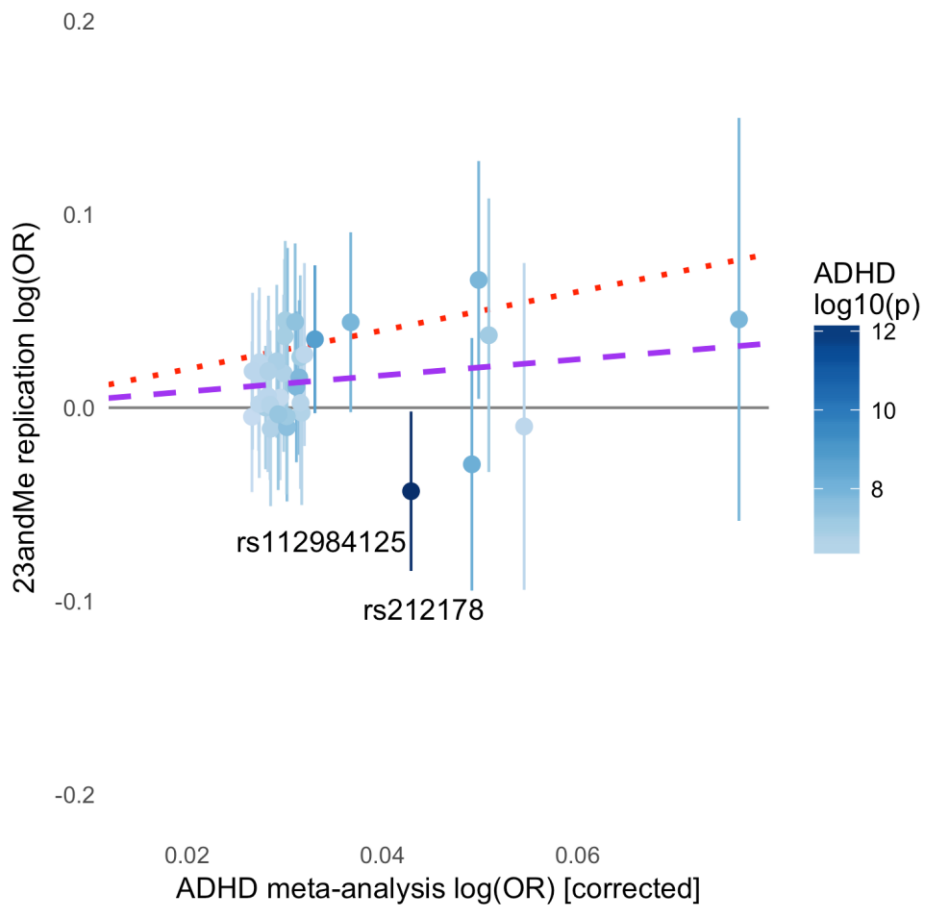
Supplementary Figure 14. Partitioning of h^2 by tissue-specific H3K4Me1 annotations

P-values for enrichment in the SNP heritability of ADHD by variants located within regulatory regions (H3K4Me1 peaks) of various cells and tissues (annotations from the Roadmap Epigenomics Mapping Consortium⁷²). Dashed line: threshold for nominal significance. Full line: threshold for significance after Bonferroni correction.



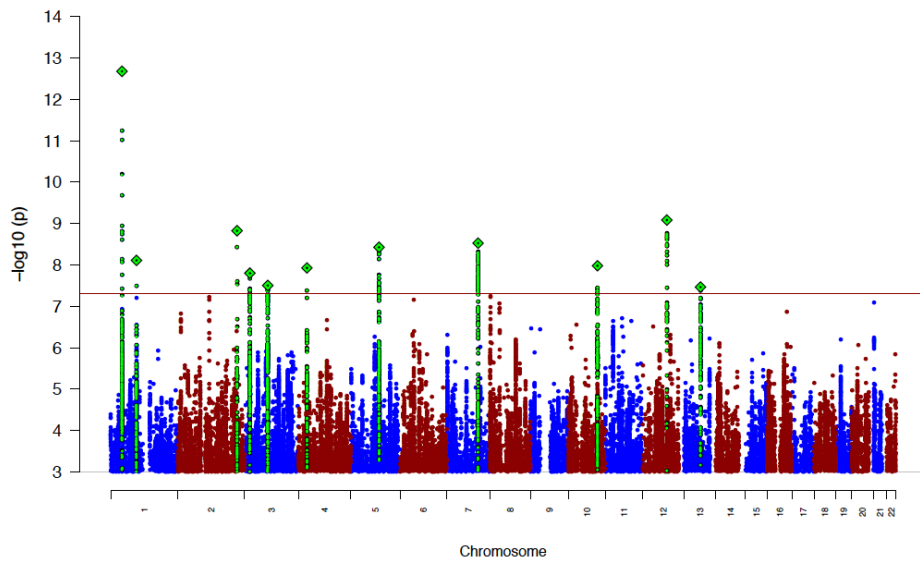
Supplementary Figure 15. Comparison of estimated effect sizes from deCODE

Regression of deCODE effect size estimates on winner's curse-adjusted effect size estimates for top loci ($P < 1 \times 10^{-6}$) in the ADHD GWAS. All variants oriented to the allele estimated to increase risk in the ADHD GWAS. Dotted red reference line indicates a slope of 1, the target result for an "ideal" replication. Dashed purple line indicates observed slope of linear regression of log odds ratio (OR) weighted by inverse standard error.

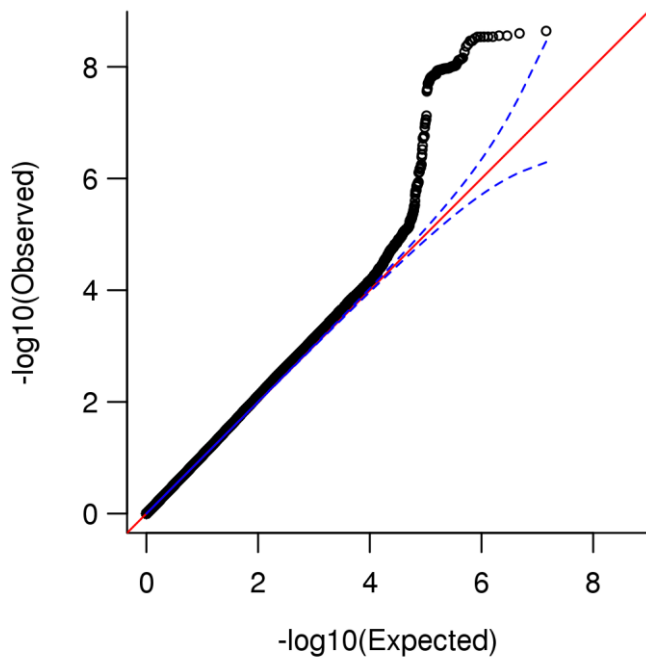


Supplementary Figure 16. Comparison of estimated effect sizes from 23andMe

Regression of 23andMe effect size estimates on winner’s curse-adjusted effect size estimates for top loci ($p < 1e-6$) in the ADHD GWAS. All variants oriented to the allele estimated to increase risk in the ADHD GWAS. Dotted red reference line indicates a slope of 1, the target result for an “ideal” replication. Dashed purple line indicates observed slope of linear regression of log odds ratio (OR) weighted by inverse standard error.

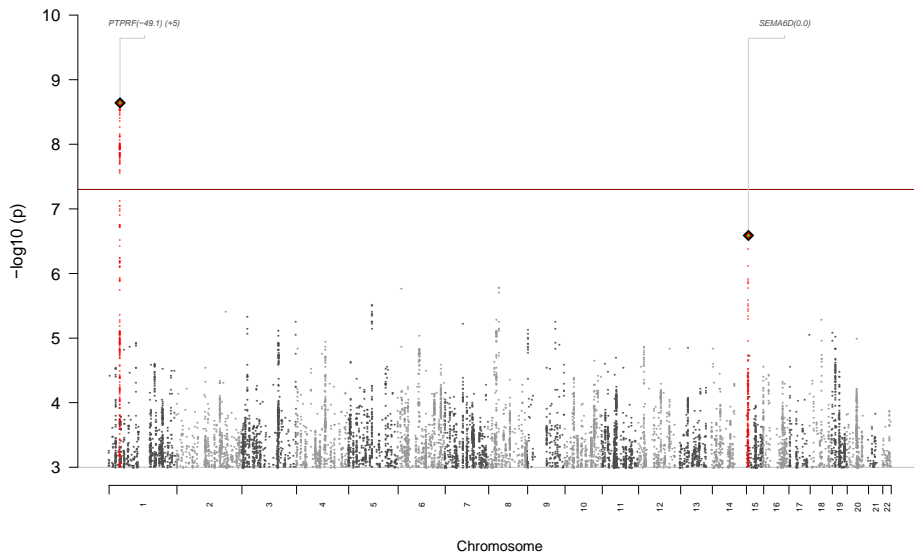


Supplementary Figure 17. Manhattan plot of results from meta-analysis of ADHD+23andMe
Genome-wide results from the replication meta-analysis, performed using the inverse-variance weighted fixed effects model. The genome-wide significant locus on chromosome four is only driven by iPSYCH/PGC since there is no information about this locus in 23andMe.



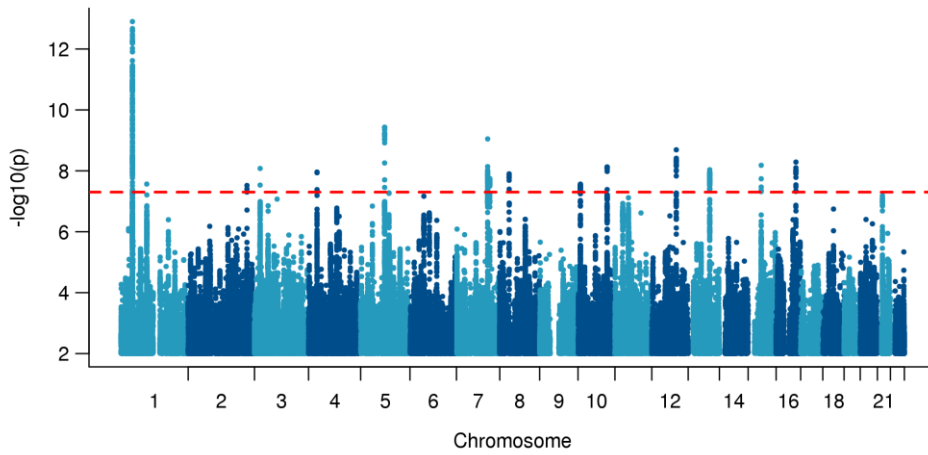
Supplementary Figure 18. Q-Q plot from test for heterogeneity between ADHD GWAS meta-analysis and 23andMe

Quantile-quantile plot of P -values for the 1 degree of freedom test of heterogeneity between 23andMe and the ADHD GWAS for genome-wide markers.



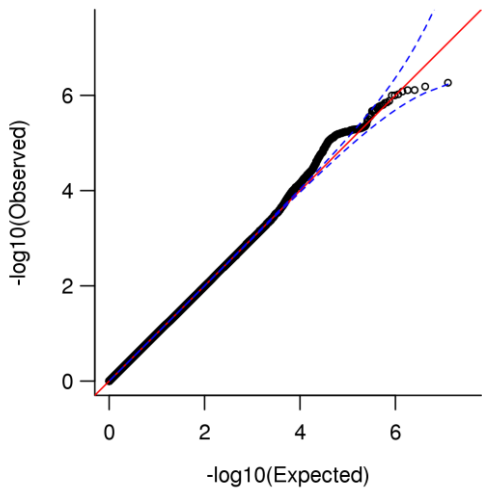
Supplementary Figure 19. Manhattan plot from test for heterogeneity between ADHD GWAS meta-analysis and 23andMe

Genome-wide results for the 1 degree of freedom test of heterogeneity between 23andMe and the ADHD GWAS. Red reference line indicates genome-wide significance (5×10^{-8}).



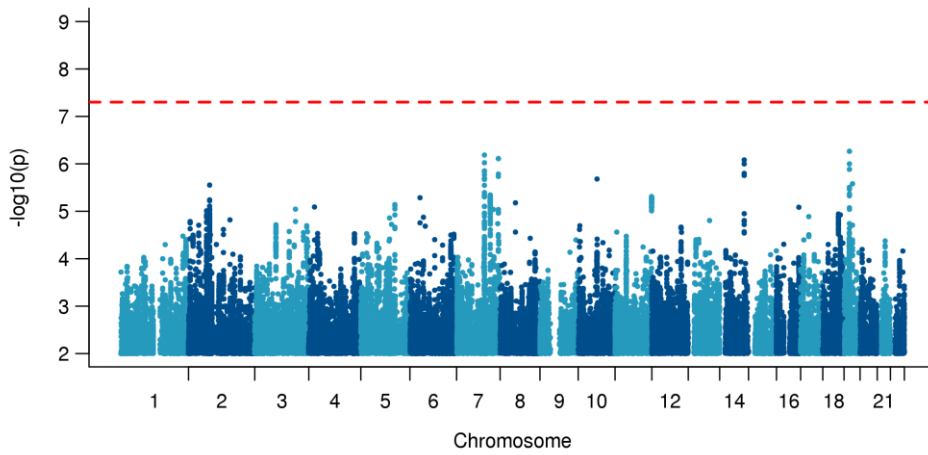
Supplementary Figure 20. Manhattan plot of results from meta-analysis of **ADHD+EAGLE/QIMR**

Genome-wide results for meta-analysis of the ADHD GWAS and EAGLE/QIMR using the modified sample size-based weights calibrated by estimates of heritability and genetic correlation.



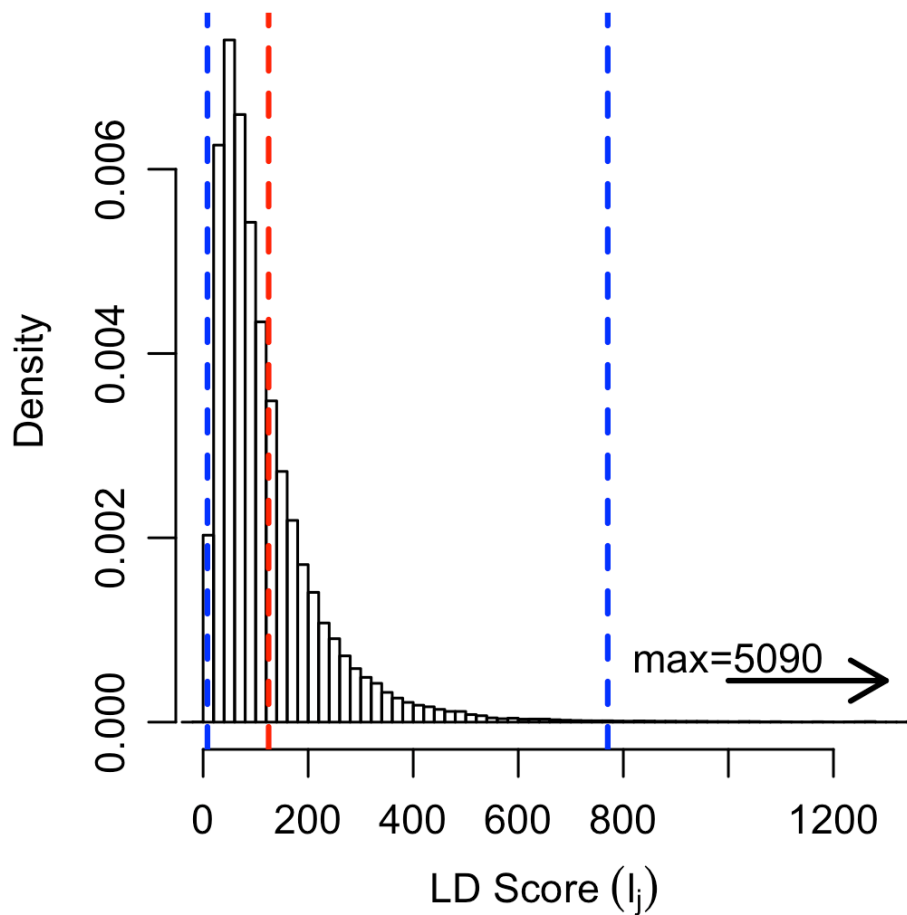
Supplementary Figure 21. Q-Q plot from test for heterogeneity between ADHD GWAS meta-analysis and EAGLE/QIMR

Quantile-quantile plot of P -values for the 1 degree of freedom test of heterogeneity between EAGLE/QIMR and the ADHD GWAS for genome-wide markers.



Supplementary Figure 22. Manhattan plot from test for heterogeneity between ADHD GWAS meta-analysis and EAGLE/QIMR

Genome-wide results for the 1 degree of freedom test of heterogeneity between 23andMe and the ADHD GWAS for genome-wide markers. The dashed red reference line indicates genome-wide significance (5×10^{-8}).

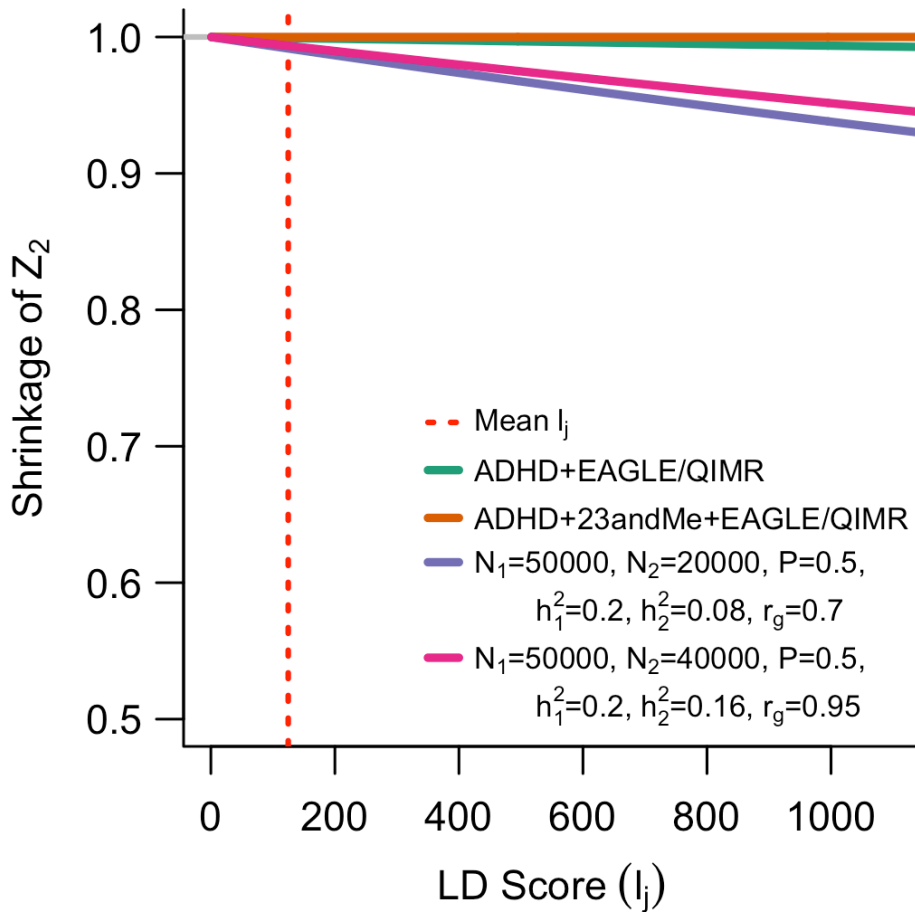


Supplementary Figure 23. Distribution of 1000 Genomes Phase 3 European LD Scores

Distribution of LD scores l_j for common HapMap3 SNPs estimated in 1000 Genomes Phase 3 data using individuals of European ancestry. LD scores downloaded from

<http://data.broadinstitute.org/alkesgroup/LDSCORE/>. Red reference line indicates mean LD score.

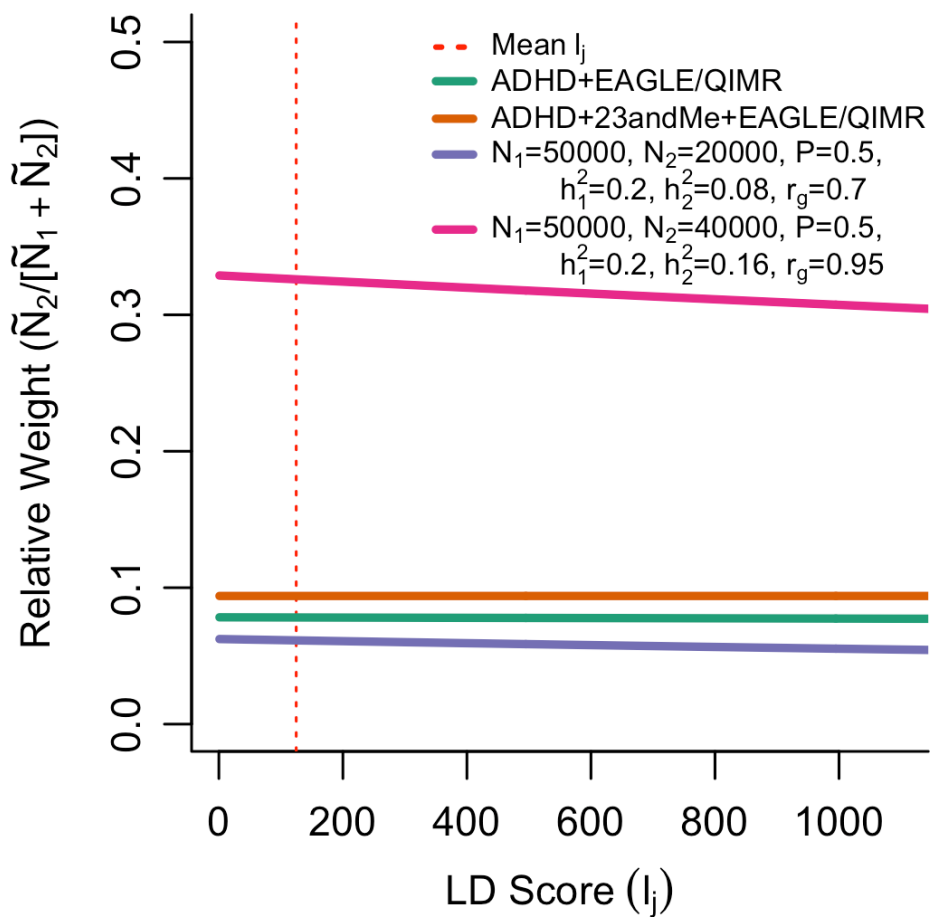
Blue reference lines indicate 0.5% and 99.5% quantiles of the distribution.



Supplementary Figure 24. Shrinkage factor for \tilde{Z}_{2j} with varying l_j

Value of $1/\sqrt{1 + (1 - r_g^2)N_{2j}h_2^2 l_j/M}$, the reduction in Z_{2j} to account for polygenic effects specific to the second phenotype, across the range of observed values for l_j . We compare the value of this term at the estimates of r_g^2, N_2, h_2^2 , and M observed in the current study, as well as with example values for scenarios with lower r_g^2 or a more highly powered GWAS of the second phenotype (i.e.

increased N_2 and h_2^2). The red reference line indicates the fixed value of $l_j = 124.718$ used for the current study.



Supplementary Figure 25. Relative effective sample size \tilde{N}_{2j} with varying l_j

Value of $\tilde{N}_{2j} / (\tilde{N}_{1j} + \tilde{N}_{2j})$, the relative effective sample size for the second phenotype, across the range of observed values for l_j . We compare the value of this term at the estimates of

$N_1, N_2, r_g^2, h_1^2, h_2^2, K, P$, and M observed in the current study, as well as with example values for scenarios with lower r_g^2 or a more highly powered GWAS of the second phenotype (i.e. increased

N_2 and h_2^2). The red reference line indicates the fixed value of $l_j = 124.718$ used for the current study.

References

- 1 Pedersen, C. B., Gotzsche, H., Moller, J. O. & Mortensen, P. B. The Danish Civil Registration System. A cohort of eight million persons. *Danish Medical Bulletin* **53**, 441-449 (2006).
- 2 Mors, O., Pertou, G. P. & Mortensen, P. B. The Danish Psychiatric Central Research Register. *Scand J Public Health* **39**, 54-57, doi:10.1177/1403494810395825 (2011).
- 3 Borglum, A. D. *et al.* Genome-wide study of association and interaction with maternal cytomegalovirus infection suggests new schizophrenia loci. *Molecular psychiatry* **19**, 325-333, doi:10.1038/mp.2013.2 (2014).
- 4 Hollegaard, M. V. *et al.* Robustness of genome-wide scanning using archived dried blood spot samples as a DNA source. *BMC Genet* **12**, 58, doi:10.1186/1471-2156-12-58 (2011).
- 5 Illumina. illumina GenCall Data Analysis Software. *Illumina Tech Note* (2005).
- 6 Korn, J. M. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nature genetics* **40**, 1253-1260, doi:10.1038/ng.237 (2008).
- 7 Goldstein, J. I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* **28**, 2543-2545, doi:10.1093/bioinformatics/bts479 (2012).
- 8 Elia, J. *et al.* Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. *Molecular psychiatry* **15**, 637-646, doi:10.1038/mp.2009.57 (2010).
- 9 Neale, B. M. *et al.* Genome-wide association scan of attention deficit hyperactivity disorder. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **147B**, 1337-1344, doi:10.1002/ajmg.b.30866 (2008).
- 10 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).
- 11 Mick, E. *et al.* Family-based genome-wide association scan of attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry* **49**, 898-905 e893, doi:10.1016/j.jaac.2010.02.014 (2010).
- 12 Lionel, A. C. *et al.* Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. *Science translational medicine* **3**, 95ra75, doi:10.1126/scitranslmed.3002464 (2011).
- 13 Sanchez-Mora, C. *et al.* Case-control genome-wide association study of persistent attention-deficit hyperactivity disorder identifies FBX033 as a novel susceptibility gene for the disorder. *Neuropsychopharmacology* **40**, 915-926, doi:10.1038/npp.2014.267 (2015).
- 14 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).

- 15 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).
- 16 Stergiakouli, E. *et al.* Investigating the contribution of common genetic variants to the risk and pathogenesis of ADHD. *The American journal of psychiatry* **169**, 186-194, doi:10.1176/appi.ajp.2011.11040551 (2012).
- 17 Hinney, A. *et al.* Genome-wide association study in German patients with attention deficit/hyperactivity disorder. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **156B**, 888-897, doi:10.1002/ajmg.b.31246 (2011).
- 18 Neale, B. M. *et al.* Case-control genome-wide association study of attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry* **49**, 906-920, doi:10.1016/j.jaac.2010.06.007 (2010).
- 19 Bakker, S. C. *et al.* A whole-genome scan in 164 Dutch sib pairs with attention-deficit/hyperactivity disorder: suggestive evidence for linkage on chromosomes 7p and 15q. *American journal of human genetics* **72**, 1251-1260 (2003).
- 20 Myocardial Infarction Genetics, C. *et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nature genetics* **41**, 334-341, doi:10.1038/ng.327 (2009).
- 21 Gelernter, J. *et al.* Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. *Molecular psychiatry* **19**, 41-49, doi:10.1038/mp.2013.145 (2014).
- 22 Gelernter, J. *et al.* Genome-wide association study of opioid dependence: multiple associations mapped to calcium and potassium pathways. *Biological psychiatry* **76**, 66-74, doi:10.1016/j.biopsych.2013.08.034 (2014).
- 23 Gelernter, J. *et al.* Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. *Molecular psychiatry* **19**, 717-723, doi:10.1038/mp.2013.99 (2014).
- 24 Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nature genetics* **47**, 435-444, doi:10.1038/ng.3247 (2015).
- 25 Pappa, I. *et al.* A genome-wide approach to children's aggressive behavior: The EAGLE consortium. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, doi:10.1002/ajmg.b.32333 (2015).
- 26 Paternoster, L. *et al.* Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. *Nature genetics* **44**, 187-192, doi:10.1038/ng.1017 (2011).
- 27 van der Valk, R. J. *et al.* A novel common variant in DCST2 is associated with length in early life and height in adulthood. *Hum Mol Genet* **24**, 1155-1168, doi:10.1093/hmg/ddu510 (2015).
- 28 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).
- 29 Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65, doi:10.1038/nature11632 (2012).

- 30 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of
genome-wide association scans. *Bioinformatics* **26**, 2190-2191,
doi:10.1093/bioinformatics/btq340 (2010).
- 31 Wright, M. J. & Martin, N. G. Brisbane Adolescent Twin Study: Outline of study methods
and research projects. *Australian Journal of Psychology* **56**, 65-78,
doi:10.1080/00049530410001734865 (2004).
- 32 Swanson, J. M. *et al.* Categorical and Dimensional Definitions and Evaluations of
Symptoms of ADHD: History of the SNAP and the SWAN Rating Scales. *Int J Educ
Psychol Assess* **10**, 51-70 (2012).
- 33 Das, S. *et al.* Next-generation genotype imputation service and methods. *Nature
genetics* **48**, 1284-1287, doi:10.1038/ng.3656 (2016).
- 34 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).
- 35 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).
- 36 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).
- 37 Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes.
G3 **1**, 457-470, doi:10.1534/g3.111.001198 (2011).
- 38 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).
- 39 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**,
68-74, doi:10.1038/nature15393 (2015).
- 40 Price, A. L. *et al.* Long-range LD can confound genome scans in admixed populations.
American journal of human genetics **83**, 132-135; author reply 135-139,
doi:10.1016/j.ajhg.2008.06.005 (2008).
- 41 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).
- 42 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).
- 43 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).
- 44 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).
- 45 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).
- 46 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).
- 47 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).

- 48 Gormley, P. *et al.* Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nature genetics* **48**, 856-866, doi:10.1038/ng.3598 (2016).
- 49 McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122, doi:10.1186/s13059-016-0974-4 (2016).
- 50 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).
- 51 Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523-527, doi:10.1038/nature19847 (2016).
- 52 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).
- 53 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).
- 54 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).
- 55 Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics* **46**, 310-315, doi:10.1038/ng.2892 (2014).
- 56 Hou, C. A simple approximation for the distribution of the weighted combination of non-independent or independent probabilities. *Statistics and Probability Letters* **73**, 179-187 (2005).
- 57 Farrell, M. S. *et al.* Evaluating historical candidate genes for schizophrenia. *Molecular psychiatry* **20**, 555-562, doi:10.1038/mp.2015.16 (2015).
- 58 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).
- 59 Day, F. R. *et al.* Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nature genetics* **47**, 1294-1303, doi:10.1038/ng.3412 (2015).
- 60 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).
- 61 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).
- 62 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).
- 63 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).
- 64 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).
- 65 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).
- 66 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).

- 67 Sherva, R. *et al.* Genome-wide Association Study of Cannabis Dependence Severity, Novel Risk Variants, and Shared Genetic Risks. *JAMA Psychiatry* **73**, 472-480, doi:10.1001/jamapsychiatry.2016.0036 (2016).
- 68 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).
- 69 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).
- 70 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics* **47**, 1228-1235 (2015).
- 71 Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nature genetics* **45**, 124-130, doi:10.1038/ng.2504 (2013).
- 72 Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).
- 73 Ernst, J. & Kellis, M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nature biotechnology* **33**, 364-376 (2015).
- 74 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).
- 75 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).
- 76 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).
- 77 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).
- 78 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).
- 79 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).
- 80 Yates, F. Contingency Tables Involving Small Numbers and the χ^2 Test. *Supplement to the Journal of the Royal Statistical Society* **1**, 217-235, doi:10.2307/2983604 (1934).
- 81 R Core Team. *R: A language and environment for statistical computing.* , <<http://www.r-project.org/>> (2014).
- 82 Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539-542, doi:10.1038/nature17671 (2016).
- 83 Pearson K, L. A. On the inheritance of characters not capable of exact quantitative measurement. *Philosophical Transactions of the Royal Society of London Series A*, 79-150 (1901).
- 84 Cohen, J. The cost of dichotomization. *Applied Psychological Measurement* **7**, 249-253, doi:10.1177/014662168300700301 (1983).

- 85 Hunter, J. & Schmidt, F. Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology* **75**, 334-349, doi:<http://dx.doi.org/10.1037/0021-9010.75.3.334> (1990).
- 86 Hsieh, F. Y., Bloch, D. A. & Larsen, M. D. A simple method of sample size calculation for linear and logistic regression. *Stat Med* **17**, 1623-1634 (1998).
- 87 Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *American journal of human genetics* **88**, 294-305, doi:10.1016/j.ajhg.2011.02.002 (2011).
- 88 Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature genetics* **42**, 565-569, doi:10.1038/ng.608 (2010).
- 89 Stouffer, S., DeVinney, L. & Suchmen, E. *The American soldier: Adjustment during army life.*, Vol. 1 (Princeton University Press, 1949).
- 90 Faraone, S. V. *et al.* Attention-deficit/hyperactivity disorder. *Nature Reviews Disease Primers*, 15020, doi:10.1038/nrdp.2015.20 (2015).
- 91 Ribases, M. *et al.* Exploration of 19 serotonergic candidate genes in adults and children with attention-deficit/hyperactivity disorder identifies association for 5HT2A, DDC and MAOB. *Molecular psychiatry* **14**, 71-85, doi:10.1038/sj.mp.4002100 (2009).
- 92 Wang, B. Molecular mechanism underlying sialic acid as an essential nutrient for brain development and cognition. *Adv Nutr* **3**, 465S-472S, doi:10.3945/an.112.001875 (2012).
- 93 Yoo, S. W. *et al.* Sialylation regulates brain structure and function. *FASEB J* **29**, 3040-3053, doi:10.1096/fj.15-270983 (2015).
- 94 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).
- 95 Sturgill, E. R. *et al.* Biosynthesis of the major brain gangliosides GD1a and GT1b. *Glycobiology* **22**, 1289-1301, doi:10.1093/glycob/cws103 (2012).
- 96 Ledeen, R. W. & Wu, G. Ganglioside function in calcium homeostasis and signaling. *Neurochemical research* **27**, 637-647 (2002).
- 97 Wu, G., Xie, X., Lu, Z. H. & Ledeen, R. W. Cerebellar neurons lacking complex gangliosides degenerate in the presence of depolarizing levels of potassium. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 307-312, doi:10.1073/pnas.011523698 (2001).
- 98 Edvardson, S. *et al.* West syndrome caused by ST3Gal-III deficiency. *Epilepsia* **54**, e24-27, doi:10.1111/epi.12050 (2013).
- 99 Walton, E. *et al.* Epigenetic profiling of ADHD symptoms trajectories: a prospective, methylome-wide study. *Molecular psychiatry*, doi:10.1038/mp.2016.85 (2016).
- 100 Woo, J. *et al.* Trans-synaptic adhesion between NGL-3 and LAR regulates the formation of excitatory synapses. *Nature neuroscience* **12**, 428-437, doi:10.1038/nn.2279 (2009).
- 101 Yang, T., Massa, S. M. & Longo, F. M. LAR protein tyrosine phosphatase receptor associates with TrkB and modulates neurotrophic signaling pathways. *J Neurobiol* **66**, 1420-1436, doi:10.1002/neu.20291 (2006).
- 102 Zabolotny, J. M. *et al.* Overexpression of the LAR (leukocyte antigen-related) protein-tyrosine phosphatase in muscle causes insulin resistance. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5187-5192, doi:10.1073/pnas.071050398 (2001).

- 103 Zhang, Z. *et al.* A heterozygous mutation disrupting the SPAG16 gene results in biochemical instability of central apparatus components of the human sperm axoneme. *Biol Reprod* **77**, 864-871, doi:10.1095/biolreprod.107.063206 (2007).
- 104 Zhang, Z. *et al.* A sperm-associated WD repeat protein orthologous to Chlamydomonas PF20 associates with Spag6, the mammalian orthologue of Chlamydomonas PF16. *Mol Cell Biol* **22**, 7993-8004 (2002).
- 105 Somers, V. *et al.* Autoantibody profiling in multiple sclerosis reveals novel antigenic candidates. *J Immunol* **180**, 3957-3963 (2008).
- 106 de Bock, L. *et al.* Anti-SPAG16 antibodies in primary progressive multiple sclerosis are associated with an elevated progression index. *Eur J Neurol* **23**, 722-728, doi:10.1111/ene.12925 (2016).
- 107 Leung, L. C. *et al.* Coupling of NF-protocadherin signaling to axon guidance by cue-induced translation. *Nature neuroscience* **16**, 166-173, doi:10.1038/nn.3290 (2013).
- 108 Blevins, C. J., Emond, M. R., Biswas, S. & Jontes, J. D. Differential expression, alternative splicing, and adhesive properties of the zebrafish delta1-protocadherins. *Neuroscience* **199**, 523-534, doi:10.1016/j.neuroscience.2011.09.061 (2011).
- 109 Krishna, K. K., Hertel, N. & Redies, C. Cadherin expression in the somatosensory cortex: evidence for a combinatorial molecular code at the single-cell level. *Neuroscience* **175**, 37-48, doi:10.1016/j.neuroscience.2010.11.056 (2011).
- 110 Kim, S. Y., Chung, H. S., Sun, W. & Kim, H. Spatiotemporal expression pattern of non-clustered protocadherin family members in the developing rat brain. *Neuroscience* **147**, 996-1021, doi:10.1016/j.neuroscience.2007.03.052 (2007).
- 111 International League Against Epilepsy Consortium on Complex Epilepsies. Electronic address, e.-a. u. e. a. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. *Lancet Neurol* **13**, 893-903, doi:10.1016/S1474-4422(14)70171-1 (2014).
- 112 Miyake, K. *et al.* The protocadherins, PCDHB1 and PCDH7, are regulated by MeCP2 in neuronal cells and brain tissues: implication for pathogenesis of Rett syndrome. *BMC Neurosci* **12**, 81, doi:10.1186/1471-2202-12-81 (2011).
- 113 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).
- 114 Janson, C. G., Chen, Y., Li, Y. & Leifer, D. Functional regulatory regions of human transcription factor MEF2C. *Brain Res Mol Brain Res* **97**, 70-82 (2001).
- 115 Harrington, A. J. *et al.* MEF2C regulates cortical inhibitory and excitatory synapses and behaviors relevant to neurodevelopmental disorders. *Elife* **5**, doi:10.7554/eLife.20059 (2016).
- 116 Adachi, M., Lin, P. Y., Pranav, H. & Monteggia, L. M. Postnatal Loss of Mef2c Results in Dissociation of Effects on Synapse Number and Learning and Memory. *Biological psychiatry* **80**, 140-148, doi:10.1016/j.biopsych.2015.09.018 (2016).
- 117 Li, H. *et al.* Transcription factor MEF2C influences neural stem/progenitor cell differentiation and maturation in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 9397-9402, doi:10.1073/pnas.0802876105 (2008).
- 118 Barbosa, A. C. *et al.* MEF2C, a transcription factor that facilitates learning and memory by negative regulation of synapse numbers and function. *Proceedings of the National*

- Academy of Sciences of the United States of America* **105**, 9391-9396, doi:10.1073/pnas.0802679105 (2008).
- 119 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).
- 120 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).
- 121 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).
- 122 Mikhail, F. M. *et al.* Clinically relevant single gene or intragenic deletions encompassing critical neurodevelopmental genes in patients with developmental delay, mental retardation, and/or autism spectrum disorders. *Am J Med Genet A* **155A**, 2386-2396, doi:10.1002/ajmg.a.34177 (2011).
- 123 Sia, G. M., Clem, R. L. & Haganir, 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).
- 124 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).
- 125 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).
- 126 Vernes, S. C. *et al.* A functional genetic link between distinct developmental language disorders. *N Engl J Med* **359**, 2337-2345, doi:10.1056/NEJMoa0802828 (2008).
- 127 Lai, C. S., Gerrelli, D., Monaco, A. P., Fisher, S. E. & Copp, A. J. FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain* **126**, 2455-2462, doi:10.1093/brain/awg247 (2003).
- 128 Wilcke, A. *et al.* Imaging genetics of FOXP2 in dyslexia. *Eur J Hum Genet* **20**, 224-229, doi:10.1038/ejhg.2011.160 (2012).
- 129 Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F. & Monaco, A. P. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519-523, doi:10.1038/35097076 (2001).
- 130 Ribases, M. *et al.* An association study of sequence variants in the forkhead box P2 (FOXP2) gene and adulthood attention-deficit/hyperactivity disorder in two European samples. *Psychiatric genetics* **22**, 155-160, doi:10.1097/YPG.0b013e328353957e (2012).
- 131 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).
- 132 Willnow, T. E., Petersen, C. M. & Nykjaer, A. VPS10P-domain receptors - regulators of neuronal viability and function. *Nature reviews. Neuroscience* **9**, 899-909, doi:10.1038/nrn2516 (2008).
- 133 Hermey, G. *et al.* The three sorCS genes are differentially expressed and regulated by synaptic activity. *J Neurochem* **88**, 1470-1476 (2004).

- 134 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).
- 135 Oetjen, S., Mahlke, C., Hermans-Borgmeyer, I. & Hermey, G. Spatiotemporal expression analysis of the growth factor receptor SorCS3. *J Comp Neurol* **522**, 3386-3402, doi:10.1002/cne.23606 (2014).
- 136 Reitz, C. *et al.* Independent and epistatic effects of variants in VPS10-d receptors on Alzheimer disease risk and processing of the amyloid precursor protein (APP). *Transl Psychiatry* **3**, e256, doi:10.1038/tp.2013.13 (2013).
- 137 Muda, M. *et al.* MKP-3, a novel cytosolic protein-tyrosine phosphatase that exemplifies a new class of mitogen-activated protein kinase phosphatase. *The Journal of biological chemistry* **271**, 4319-4326 (1996).
- 138 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).
- 139 Owens, D. M. & Keyse, S. M. Differential regulation of MAP kinase signalling by dual-specificity protein phosphatases. *Oncogene* **26**, 3203-3213, doi:10.1038/sj.onc.1210412 (2007).
- 140 Stewart, A. E., Dowd, S., Keyse, S. M. & McDonald, N. Q. Crystal structure of the MAPK phosphatase Pyst1 catalytic domain and implications for regulated activation. *Nat Struct Biol* **6**, 174-181, doi:10.1038/5861 (1999).
- 141 Bermudez, O., Pages, G. & Gimond, C. The dual-specificity MAP kinase phosphatases: critical roles in development and cancer. *Am J Physiol Cell Physiol* **299**, C189-202, doi:10.1152/ajpcell.00347.2009 (2010).
- 142 Dickinson, R. J., Eblaghie, M. C., Keyse, S. M. & Morriss-Kay, G. M. Expression of the ERK-specific MAP kinase phosphatase PYST1/MKP3 in mouse embryos during morphogenesis and early organogenesis. *Mech Dev* **113**, 193-196 (2002).
- 143 Li, C., Scott, D. A., Hatch, E., Tian, X. & Mansour, S. L. Dusp6 (Mkp3) is a negative feedback regulator of FGF-stimulated ERK signaling during mouse development. *Development* **134**, 167-176, doi:10.1242/dev.02701 (2007).
- 144 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).
- 145 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).
- 146 Takaki, M. *et al.* Two kinds of mitogen-activated protein kinase phosphatases, MKP-1 and MKP-3, are differentially activated by acute and chronic methamphetamine treatment in the rat brain. *J Neurochem* **79**, 679-688 (2001).
- 147 Miraoui, H. *et al.* Mutations in FGF17, IL17RD, DUSP6, SPRY4, and FLRT3 are identified in individuals with congenital hypogonadotropic hypogonadism. *American journal of human genetics* **92**, 725-743, doi:10.1016/j.ajhg.2013.04.008 (2013).
- 148 Luo, Y. *et al.* Differential expression of FOXA1, DUSP6, and HA117 in colon segments of Hirschsprung's disease. *Int J Clin Exp Pathol* **8**, 3979-3986 (2015).

- 149 Banzhaf-Strathmann, J. *et al.* MicroRNA-125b induces tau hyperphosphorylation and cognitive deficits in Alzheimer's disease. *EMBO J* **33**, 1667-1680, doi:10.15252/embj.201387576 (2014).
- 150 Keyse, S. M. Dual-specificity MAP kinase phosphatases (MKPs) and cancer. *Cancer Metastasis Rev* **27**, 253-261, doi:10.1007/s10555-008-9123-1 (2008).
- 151 Kidger, A. M. & Keyse, S. M. The regulation of oncogenic Ras/ERK signalling by dual-specificity mitogen activated protein kinase phosphatases (MKPs). *Semin Cell Dev Biol* **50**, 125-132, doi:10.1016/j.semcdb.2016.01.009 (2016).
- 152 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).
- 153 Yoshida, Y., Han, B., Mendelsohn, M. & Jessell, T. M. PlexinA1 signaling directs the segregation of proprioceptive sensory axons in the developing spinal cord. *Neuron* **52**, 775-788, doi:10.1016/j.neuron.2006.10.032 (2006).
- 154 Toyofuku, T. *et al.* Dual roles of Sema6D in cardiac morphogenesis through region-specific association of its receptor, Plexin-A1, with off-track and vascular endothelial growth factor receptor type 2. *Genes Dev* **18**, 435-447, doi:10.1101/gad.1167304 (2004).
- 155 Toyofuku, T. *et al.* Guidance of myocardial patterning in cardiac development by Sema6D reverse signalling. *Nat Cell Biol* **6**, 1204-1211, doi:10.1038/ncb1193 (2004).
- 156 Benyamin, B. *et al.* Childhood intelligence is heritable, highly polygenic and associated with FBNP1L. *Molecular psychiatry* **19**, 253-258, doi:10.1038/mp.2012.184 (2014).
- 157 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).
- 158 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).
- 159 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).
- 160 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).
- 161 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).
- 162 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).
- 163 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).
- 164 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).
- 165 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).

- 166 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).
- 167 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).
- 168 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).
- 169 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).
- 170 Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376-381, doi:10.1038/nature12873 (2014).

Supplementary Notes

Consortium Members

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

ADHD Working Group of the Psychiatric Genomics Consortium (PGC)

Özgür Albayrak,^{1,2} Richard J.L. Anney,³ Alejandro Arias Vasquez,⁴ Maria Jesús Arranz,⁵ Philip Asherson,⁶ Tobias J. Banaschewski,⁷ Claiton Bau,^{8,9} Joseph Biederman,^{10,11} Preben Bo Mortensen,¹² Anders Børglum,¹² Jan K. Buitelaar,¹³ Miguel Casas,^{13,14,15} Alice Charach,¹⁶ Bru Cormand,^{17,18,19,20} Jennifer Crosbie,¹⁶ Soeren Dalsgaard,¹² Mark J. Daly,²¹ Ditte Demontis,¹² Astrid Dempfle,²² Alysa E. Doyle,²³ Richard P. Ebstein,²⁴ Josephine Elia,²⁵ Stephen V. Faraone,^{26,27} Manuel Föcker,¹ Barbara Franke,²⁸ Christine Freitag,²⁹ Joel Gelernter,^{30,31} Michael Gill,³² Eugenio Grevet,^{8,33} Jan Haavik,³⁴ Hakon Hakonarson,^{35,36} Ziarh Hawi,³⁷ Johannes Hebebrand,¹ Beate Herpertz-Dahlmann,³⁸ Amaia Hervas,⁵ Anke Hinney,¹ Sarah Hohmann,⁷ Peter Holmans,³ Mara Hutz,⁹ Abel Ickowitz,¹⁶ Stefan Johansson,³⁹ Lindsey Kent,⁴⁰ Sarah Kittel-Schneider,⁴¹ Henry Kranzler,^{42,43} Jonna Kuntsi,⁶ Nanda Lambregts-Rommelse,⁴⁴ Kate Langley,³ Gerd Lehmkuhl,⁴⁵ Klaus-Peter Lesch,^{46,47} Sandra K. Loo,⁴⁸ Joanna Martin,^{3,49,50} James J. McGough,⁴⁸ Sarah E. Medland,⁵¹ Jobst Meyer,⁵² Eric Mick,⁵³ Frank Middleton,²⁷ Ana Miranda,⁵⁴ Fernando Mulas,⁵⁵ Aisling Mulligan,⁵⁶ Benjamin M. Neale,²¹ Stan F. Nelson,⁴⁸ T Trang Nguyen,⁵⁷ Michael C. O'Donovan,³ Robert D. Oades,⁵⁸ Michael J. Owen,³ Haukur Palmason,⁵² Josep Antoni Ramos-Quiroga,^{13,14,15} Andreas Reif,⁴¹ Tobias J. Renner,^{46,59} Luis Rhode,^{8,33} Marta Ribasés,^{13,14,60} Marcella Rietschel,⁶¹ Stephan Ripke,²¹ Olga Rivero,^{46,47} Herbert Roeyers,⁶² Marcel Romanos,⁶³ Jasmin Romanos,⁶³ Nina Roth Mota,^{8,64} Aribert Rothenberger,⁶⁵ Cristina Sánchez-Mora,^{13,14,60} Russell Schachar,¹⁶ Helmut Schäfer,⁵⁷ André Scherag,^{66,67} Benno G. Schimmelmann,⁶⁸ Joseph Sergeant,⁶⁹ Judith Sinzig,^{45,70} Susan L. Smalley,⁴⁸ Edmund J.S. Sonuga-Barke,^{62,71} Hans-Christoph Steinhausen,^{72,73,74} Patrick F. Sullivan,^{50,75} Anita Thapar,^{3,76} Margaret Thompson,⁷¹ Alexandre Todorov,⁷⁷ Irwin Waldman,⁷⁸ Susanne Walitza,^{63,79} Raymond K. Walters,^{21,49} Yufeng Wang,⁸⁰ Andreas Warnke,⁶³ Nigel Williams,³ Stephanie H. Witt,⁶¹ Li Yang,⁸⁰ Tetyana Zayats,³⁴ & Yanli Zhang-James²⁶

Affiliations

1. Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Essen, University of Duisburg-Essen, Essen, Germany
2. Department of Psychosomatic Medicine and Psychotherapy, Hannover Medical School (MHH), Hannover, Germany
3. MRC Centre for Neuropsychiatric Genetics and Genomics, Department of Psychological Medicine and Neurology, School of Medicine, Cardiff University, Cardiff, Wales, UK
4. Departments of Human Genetics, Psychiatry, and Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, The Netherlands
5. University Hospital Mutua Terrassa, Barcelona, Spain

6. MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
7. Department of Child and Adolescent Psychiatry, Central Institute of Mental Health and Mannheim Medical Faculty, University of Heidelberg, Heidelberg, Germany
8. ADHD Outpatient Clinic, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil
9. Department of Genetics, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
10. Pediatric Psychopharmacology Unit, Massachusetts General Hospital, Boston, MA, USA
11. Department of Psychiatry, Harvard Medical School, Boston, MA, USA
12. National Centre for Register-based Research, Aarhus University, Aarhus, Denmark
13. Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain
14. Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands
15. Biomedical Network Research Centre on Mental Health (CIBERSAM), Barcelona, Spain
16. Department of Psychiatry and Legal Medicine, Universitat Autònoma de Barcelona, Catalonia, Spain
17. The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada
18. Department de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain
19. Biomedical Network Research Centre on Rare Diseases (CIBERER), Barcelona, Spain
20. Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Spain
21. Institut de Recerca Sant Joan de Déu (IR-SJD), Barcelona, Catalonia, Spain
22. Analytical and Translational Genetics Unit (ATGU), Massachusetts General Hospital, Boston, MA
23. Institute for Medical Informatics and Statistics, University Medical Center Schleswig-Holstein, Campus Kiel, Kiel, Germany
24. Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA
25. National University of Singapore, Singapore
26. duPont Children's Hospital, Wilmington, DE, USA
27. Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY, USA
28. Department of Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, NY, USA
29. Departments of Human Genetics (855) and Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, The Netherlands
30. Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, JW Goethe University, Frankfurt, Germany
31. Department of Psychiatry, Genetics, and Neuroscience, Yale University School of Medicine, New Haven, Connecticut, USA
32. Veterans Affairs Connecticut Healthcare Center, West Haven, Connecticut, USA
33. Department of Psychiatry, Trinity College Dublin, Trinity Centre for Health Sciences, St. James's Hospital, Dublin, Ireland
34. Department of Psychiatry, Faculdade de Medicina, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
35. K.G.Jebsen Centre for Neuropsychiatric Disorders, Department of Biomedicine, University of Bergen, Bergen, Norway
36. The Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA
37. Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

38. School of Psychological Sciences and Monash Institute for Cognitive and Clinical Neurosciences, Monash University, Clayton, Australia
39. Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, RWTH Aachen, University Clinics, Aachen, Germany
40. K.G.Jebens Centre for Neuropsychiatric Disorders, Department of Clinical Science, University of Bergen, Bergen, Norway
41. University of St Andrews, St Andrews, UK
42. Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital, Frankfurt, Germany
43. Department of Psychiatry, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
44. Veterans Integrated Service Network (VISN4) Mental Illness Research, Education, and Clinical Center (MIRECC), Crescenz VA Medical Center, Philadelphia, PA, USA
45. Karakter Child and Adolescent Psychiatry University Center and Department of Psychiatry, Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
46. Department of Child and Adolescent Psychiatry, University of Cologne, Cologne, Germany
47. Division of Molecular Psychiatry, ADHD Clinical Research Unit, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany
48. Department of Neuroscience, School for Mental Health and Neuroscience (MHENS), Maastricht University, Maastricht, The Netherlands
49. University of California Los Angeles, Los Angeles, CA
50. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, USA
51. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
52. Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia
53. Institute of Psychobiology, Department of Neurobehavioral Genetics, University of Trier, Trier, Germany
54. Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, USA
55. Department of Developmental and Educational Psychology, University of Valencia, Spain
56. Instituto Valenciano de Neurologia Pediátrica (INVANEP), Valencia, Spain
57. Child and Adolescent Psychiatry, University College Dublin, Dublin, Ireland
58. University of Marburg, Marburg, Germany
59. University of Duisburg-Essen, Duisburg, Germany
60. Department of Child and Adolescent Psychiatry, Universitätsklinikum Tübingen, Tübingen, Germany
61. Psychiatric Genetics Unit, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain
62. Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, University of Heidelberg, Mannheim, Germany
63. Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium
64. Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany
65. Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands

66. Department of Child and Adolescent Psychiatry, University Medicine Goettingen, Goettingen, Germany
67. Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), University of Duisburg-Essen, Essen, Germany
68. Clinical Epidemiology, Integrated Research and Treatment Center, Center for Sepsis Control and Care (CSCC), Jena University Hospital, Jena, Germany
69. University Hospital of Child- and Adolescent Psychiatry, University of Bern, Bern, Switzerland
70. Department of Clinical Neuropsychology, Vrije Universiteit, Amsterdam, The Netherlands
71. Department of Child and Adolescent Psychiatry and Psychotherapy, LVR – Clinic Bonn, Bonn, Germany
72. School of Psychology, University of Southampton, Southampton, UK
73. University of Zurich, Zurich, Switzerland
74. University of Basel, Basel, Switzerland
75. Child and Adolescent Mental Health Centre, Capital Region Psychiatry, Copenhagen, Denmark
76. Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC, USA
77. School of Psychology, Cardiff University, Cardiff, UK
78. Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA
79. Department of Psychology, Emory University, Atlanta, Georgia, USA
80. Department of Child and Adolescent Psychiatry, University of Zurich, Zurich, Switzerland
81. Institute of Mental Health, Peking University Sixth Hospital, Beijing, China

Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium

Dorret I. Boomsma,^{1,2} George Davey Smith,³ Gareth E. Davies,^{1,4} Erik A. Ehli,⁴ David M. Evans,^{3,5} Iryna O. Fedko,¹ Corina U. Greven,^{6,7,8} Maria M. Groen-Blokhuis,⁹ Monica Guxens,^{10,11,12,13} Anke R. Hammerschlag,¹⁴ Catharina A. Hartman,¹⁵ Joachim Heinrich,^{16,17} Jouke- Jan Hottenga,¹⁸ James Hudziak,^{19,20,21,22} Astanand Jugessur,^{23,24} John P. Kemp,^{3,5} Eva Krapohl,⁸ Nicholas G. Martin,²⁵ Sarah E. Medland,²⁶ Christel M. Middeldorp,^{1,27,28} Mario Murcia,^{12,29} Ronny Myhre,³⁰ Ilja M. Nolte,³¹ Dale R. Nyholt,³² Johan Ormel,¹⁵ Klaasjan G. Ouwens,¹ Irene Pappa,^{13,33} Craig E. Pennell,³⁴ Robert Plomin,⁸ Susan Ring,^{3,35} Marie Standl,¹⁶ Evie Stergiakouli,^{3,5} Beate St Pourcain,^{3,36} Camilla Stoltenberg,³⁷ Jordi Sunyer,^{11,38,39} Elisabeth Thiering,^{16,40} Henning Tiemeier,²¹ Carla M.T. Tiesler,^{16,40} Nicholas J. Timpson,³ Maciej Trzaskowski,⁴¹ Peter Johannes van der Most,³¹ Natalia Vilor-Tejedor,^{10,11,38} Carol A. Wang,³⁴ Andrew J.O. Whitehouse,⁴² & Huiying Zhao³²

Affiliations

1. Department of Biological Psychology, Neuroscience Campus Amsterdam, VU University, Amsterdam, The Netherlands
2. EMGO Institute for Health and Care Research, Amsterdam, The Netherlands
3. Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK.
4. Avera Institute for Human Genetics, Sioux Falls, SD, USA
5. University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia

6. Radboud University Medical Center, Donders Institute for Brain, Cognition and Behaviour, Department of Cognitive Neuroscience, Nijmegen, The Netherlands
7. Karakter Child and Adolescent Psychiatry University Center, Nijmegen, The Netherlands
8. Medical Research Council Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK
9. GGZ inGeest, Amsterdam, the Netherlands
10. ISGlobal - Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain
11. Universitat Pompeu Fabra (UPF), Barcelona, Spain
12. Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Instituto de Salud Carlos III, Madrid, Spain
13. Department of Child and Adolescent Psychiatry/Psychology, Erasmus University Medical Centre–Sophia Children's Hospital, Rotterdam, The Netherlands
14. Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam, The Netherlands
15. Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
16. Institute of Epidemiology I, Helmholtz Zentrum München – German Research Centre for Environmental Health, Neuherberg, Germany
17. Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Inner City Clinic, University Hospital of Munich (LMU), Munich, Germany
18. Biological Psychology, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
19. Vermont Center for Children Youth and Families and University of Vermont Medical Center, University of Vermont, Burlington, VT, USA
20. Child Psychiatry, School of Medicine, Washington University, St. Louis, MO, USA
21. Erasmus University Medical Centre–Sophia Children's Hospital, Rotterdam, The Netherlands
22. Geisel School of Medicine, Dartmouth, Hanover, NH, USA
23. Department of Genetic Research and Bioinformatics, Norwegian Institute of Public Health, Oslo, Norway
24. Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway
25. Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia
26. Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia
27. Child Health Research Centre, University of Queensland, Brisbane, Queensland, Australia
28. Child and Youth Mental Health Service, Children's Health Queensland Hospital and Health Service, Queensland, Australia
29. Epidemiology and Environmental Health Joint Research Unit, FISABIO-Universitat Jaume I-Universitat de València, Valencia, Spain.
30. Department of Genes and Environment, Norwegian Institute of Public Health, Oslo, Norway
31. Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
32. Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia
33. Generation R Study Group, Erasmus Medical Center, Rotterdam, The Netherlands
34. School of Women's and Infants' Health, The University of Western Australia, Crawley, Western Australia, Australia
35. School of Social and Community Medicine, University of Bristol, Bristol, UK

36. Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
37. Norwegian Institute of Public Health, Oslo, Norway
38. CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain
39. ISGlobal Barcelona Institute for Global Health, Barcelona, Spain
40. Division of Metabolic Diseases and Nutritional Medicine, Dr. von Hauner Children's Hospital, Ludwig-Maximilians-University of Munich, Munich, Germany
41. Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia
42. Telethon Kids Institute, University of Western Australia, West Perth, Western Australia, Australia

23andMe Research Team

Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, David A. Hinds, Bethann S. Hromatka, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Matthew H. McIntyre, Joanna L. Mountain, Carrie A.M. Northover, Steven J. Pitts, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, & Catherine H. Wilson