

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a preprint version which may differ from the publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/201032>

Please be advised that this information was generated on 2021-09-23 and may be subject to change.

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

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

† Equal contributions. \* Co-last authors.

Correspond with: Benjamin M. Neale ([bneale@broadinstitute.org](mailto:bneale@broadinstitute.org)), Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. Anders D. Børglum ([anders@biomed.au.dk](mailto:anders@biomed.au.dk)) Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark. Stephen V. Faraone ([sfaraone@childpsychresearch.org](mailto: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. Department of Biological Psychology, Neuroscience Campus Amsterdam, VU University, Amsterdam, The Netherlands
27. EMGO Institute for Health and Care Research, Amsterdam, The Netherlands
28. Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain
29. Biomedical Network Research Centre on Rare Diseases (CIBERER), Barcelona, Spain
30. Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Spain
31. Institut de Recerca Pediàtrica Hospital Sant Joan de Déu, Barcelona, Catalonia, Spain
32. Departments of Human Genetics (855) and Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, The Netherlands
33. Department of Psychiatry, Genetics, and Neuroscience, Yale University School of

Medicine, New Haven, Connecticut, USA

34. Veterans Affairs Connecticut Healthcare Center, West Haven, Connecticut, USA
35. The Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA
36. Department of Psychiatry, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
37. Veterans Integrated Service Network (VISN4) Mental Illness Research, Education, and Clinical Center (MIRECC), Crescenz VA Medical Center, Philadelphia, PA, USA
38. MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
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. Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Copenhagen, Denmark
56. Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
57. Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark
58. 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 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. These findings, based on clinical interviews and/or medical records are supported by additional analyses of a self-reported ADHD sample and a study of quantitative measures of ADHD symptoms in the population. Meta-analyzing these data with our primary scan yielded a total of 16 genome-wide significant loci. The results support 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 worldwide<sup>1</sup>. ADHD is often persistent and markedly impairing with increased risk of harmful outcomes such as injuries<sup>2</sup>, traffic accidents<sup>3</sup>, increased health care utilization<sup>4,5</sup>, substance abuse<sup>6</sup>, criminality<sup>7</sup>, unemployment<sup>8</sup>, divorce<sup>4</sup>, suicide<sup>9</sup>, AIDS risk behaviors<sup>8</sup>, and premature mortality<sup>10</sup>. Epidemiologic and clinical studies implicate genetic and environmental risk factors that affect

the structure and functional capacity of brain networks involved in behavior and cognition<sup>1</sup>, in the etiology of ADHD.

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

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

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

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 a national research register 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<sup>24,25,31,32,22</sup> (see Supplementary Information for detailed cohort descriptions).

Prior to analysis, stringent quality control procedures were performed on the genotyped markers and individuals in each cohort using a standardized pipeline<sup>33</sup> (see Online Methods). Related individuals were removed, and genetic outliers were excluded based on principal component analysis. Non-genotyped markers were imputed using the 1000 Genomes Project Phase 3 reference panel<sup>34</sup> (see Online Methods).

GWAS was conducted in each cohort using logistic regression with the imputed additive genotype dosages. Principal components were included as covariates to correct for population stratification<sup>35</sup> (see Supplementary Information), and variants with imputation INFO score < 0.8 or minor allele frequency (MAF) < 0.01 were excluded. The GWAS were then meta-analyzed using an inverse-variance weighted fixed effects model<sup>36</sup>. 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.M2). No marker demonstrated significant heterogeneity between studies (Supplementary Figure 6 and 7). Conditional analysis within each locus did not identify any independent secondary signals meeting genome-wide significance (see Online Methods, Supplementary Table 2).

### **Homogeneity of effects between cohorts**

No genome-wide significant heterogeneity was observed in the ADHD GWAS meta-analysis (see Supplementary Information). Genetic correlation analysis (see Online Methods) provided further evidence that effects were consistent across different cohort study designs. The estimated genetic correlation between the European ancestry PGC samples and the iPSYCH sample from LD score regression<sup>37</sup> was not significantly less than 1 ( $r_g = 1.17$ , SE = 0.2). The correlation



between European ancestry PGC case/control and trio cohorts estimated with bivariate GREML was close to one ( $r_g = 1.02$ ; SE = 0.32).

Polygenic risk scores (PRS)<sup>38</sup> also show consistency over 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 (see 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 sample, with PRS for each subset computed from the other four iPSYCH subsets and the PGC samples used as training samples (see 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 the risk is significantly associated with the individual PRS burden in a dose-dependent manner.

### **Polygenic Architecture of ADHD**

To assess the proportion of phenotypic variance explained by common variants we applied LD score regression<sup>37</sup> in the European ancestry meta-analysis (Online Methods). Assuming a population prevalence of 5% for ADHD<sup>39</sup>, 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; 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 performed partitioning based on the functional annotations described in Finucane et al.<sup>40</sup> (see Online Methods). The analysis revealed significant enrichment in the heritability by SNPs located in conserved regions ( $P = 8.49 \times 10^{-10}$ ) (Supplementary Figure 12), supporting the general biological importance of conserved regions and the potential impact on ADHD of variants located in these regions. Additionally enrichment in the heritability of SNPs located in cell-type-specific regulatory elements was evaluated by using the cell-type-specific group annotations described in Finucane et al.<sup>40</sup>. The analysis revealed 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 220 phenotypes using LD score regression<sup>41,42</sup> (Online Methods, Supplementary eTable 5). Thirty-eight phenotypes demonstrated significant genetic overlap with ADHD ( $P < 2.27 \times 10^{-4}$ ), including major depressive disorder (submitted), educational outcomes<sup>43-46</sup>, obesity-related phenotypes<sup>47-52</sup>, smoking<sup>53-55</sup>, reproductive success<sup>56</sup>, and mortality<sup>57</sup> (Figure 3; Supplementary Table 11). In each case the genetic correlation is supported in 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 smoking ( $r_g = 0.48$ ,  $P = 4.33 \times 10^{-16}$ ) 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 childhood IQ ( $r_g = -0.41$ ,  $P = 1.24 \times 10^{-6}$ ).

### **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 causal effect (see 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 (see Online Methods, Supplementary eTable 2). Broadly, the significant loci do not coincide with candidate genes proposed to play a role in ADHD<sup>58</sup>.

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 learning<sup>59-61</sup>. Comorbidity of ADHD with specific developmental disorders of language and learning is common (7-11%)<sup>62,63</sup>, and poor language skills have been associated with higher inattention/hyperactivity symptoms in primary school<sup>64</sup>.

Genome-wide significant loci on chromosomes 12 and 15 have 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). The chromosome 15 locus is located in *SEMA6D*, and the majority of variants in the credible set are strongly associated with expression of *SEMA6D* in fibroblasts<sup>65</sup>. *DUSP6* encodes a dual specificity phosphatase<sup>66</sup>, and may play a role in regulating neurotransmitter homeostasis by affecting dopamine levels in the synapses<sup>67,68</sup>. Regulation of dopamine levels is likely to be relevant to ADHD since widely used ADHD medications have dopaminergic targets<sup>69,70</sup> that increases the availability of synaptic dopamine. *SEMA6D* is active in the brain during embryonic development, and may play a role in neuronal wiring<sup>71</sup>. Variants in *SEMA6D* have previously been associated with educational attainment<sup>72</sup>.

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 blood<sup>73</sup> (Supplementary eTable 2). Missense mutations in *ST3GAL3* have been shown to cause

autosomal recessive intellectual disability<sup>74</sup>. Hi-C and eQTL annotations suggest multiple alternative genes however (Supplementary eTable 3). The locus also includes an intergenic variant, rs11210892, that has previously been associated with schizophrenia<sup>33</sup>.

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 brain<sup>75</sup> and the genome-wide significant locus overlaps with variants in *LINC00461* associated with educational attainment<sup>72</sup>. Alternatively, a genome-wide significant SNP in this locus (rs304132) is located in *MEF2C-AS1*, of strong interest given previous associations between *MEF2C* and severe intellectual disability,<sup>76-78</sup> cerebral malformation<sup>77</sup>, depression<sup>79</sup>, schizophrenia<sup>33</sup> and Alzheimer's disease<sup>80</sup>, but the corresponding variant is not supported by the credible set analysis. Credible set annotations for other significant loci are similarly cryptic.

### **Analysis of gene sets**

Competitive gene based tests were performed for *FOXP2* target genes, highly constrained genes, and for all Gene Ontology terms<sup>81</sup> from MsigDB 6.0<sup>82</sup> using MAGMA<sup>83</sup> (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<sup>84,85</sup> were tested (Online Methods), none of which demonstrated significant association to ADHD (Supplementary Table 7). The lack of association, might be caused by unknown functions of *FOXP2* driving ADHD risk; insufficient power to detect relevant downstream genes

or that only a small subset of biological functions regulated by FOXP2 are relevant to ADHD pathogenesis.

Consistent with the partitioning of heritability, a set of 2,932 genes that are highly constrained and show high intolerance to loss of function<sup>86</sup> 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 ADHD<sup>58</sup>; 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 correction for multiple testing, but the top pathways did include interesting and nominally significant pathways such as “dopamine receptor binding” ( $p=0.0010$ ) and “Excitatory Synapse” ( $P = 0.0088$ ). (Supplementary eTable 4).

### **Replication of GWAS loci**

Here we describe the comparison of the GWAS meta-analysis of ADHD with two other ADHD-related GWASs: a 23andMe self-report cohort (5,857 cases and 70,393 controls) and a meta-analysis of childhood rating scales of ADHD symptoms performed by the EAGLE consortium (17,666 children < 13 years of age)<sup>30</sup>. We observed moderate concordance of genome-wide results between the ADHD GWAS and the cohort with a self-reported history of diagnosis for ADHD or Attention Deficit Disorder genotyped by 23andMe (see Supplementary Information). The estimated genetic correlation between the two analyses was strong ( $r_g=0.653$ ,  $SE=0.114$ ), but significantly less than 1 ( $P=2.35 \times 10^{-3}$ ). Of 94 clumped loci with  $P < 1 \times 10^{-5}$  in the ADHD GWAS meta-analysis, 71 had effects in the same direction in the 23andMe GWAS, significantly more concordant than expected by chance ( $P=1.25 \times 10^{-6}$ ; Online Methods; Supplementary Table

12). We also observe a striking genetic correlation between the ADHD GWAS meta-analysis and quantitative measures of ADHD-related behavior in the general population as assembled by the EAGLE Consortium ( $r_g=0.943$ ,  $SE = 0.204$ ). The direction of effect in EAGLE was concordant for 65 of 94 clumped loci with  $P < 1 \times 10^{-5}$  in the ADHD GWAS ( $P=3.06 \times 10^{-4}$ ; Supplementary Table 12). For the 12 genome-wide significant loci in the ADHD GWAS meta-analysis, seven of the 11 loci present in the 23andMe GWAS had effects in the same direction ( $P=0.386$  for sign concordance; Supplementary Table 13). In EAGLE, the direction of effect is concordant for 10 of 11 genome-wide significant loci from the ADHD GWAS meta-analysis ( $P = 0.0159$ ).

We then meta-analyzed the ADHD GWAS with the 23andMe and EAGLE results. For EAGLE, we developed a 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 based on the heritability and effective sample size for each phenotype (detailed description in Supplementary Information).

All 12 genome-wide significant loci from the ADHD GWAS meta-analysis maintained genome-wide significance after meta-analysis with EAGLE (Supplementary Figure 18), though the EAGLE data does not contribute to significance at two of those loci. Meta-analysis with EAGLE also yielded three additional loci surpassing genome-wide significance (Supplementary eTable 6). Eight of the 12 ADHD loci were significant after inverse variance-weighted meta-analysis of the ADHD GWAS with 23andMe (Supplementary Figure 15), including two loci without 23andMe results. Joint meta-analysis of the ADHD, 23andMe, and EAGLE GWASs yielded

significance for 10 of the 12 genome-wide significant ADHD loci and a total of 16 significant loci (Supplementary eTable 6, Supplementary Figure 21).

Consistent with the genetic correlation results, there is evidence of heterogeneity between the ADHD GWAS meta-analysis results and 23andMe GWAS. Genome-wide significant heterogeneity was observed at the lead chromosome 1 locus from the ADHD GWAS meta-analysis (rs12410155:  $I^2=97.2$ ,  $P=2.29\times 10^{-9}$ ; Supplementary Figure 17). No genome-wide significant heterogeneity was observed in the meta-analysis of ADHD and EAGLE (Supplementary Figure 20).

## 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* 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. schizophrenia<sup>33</sup>. A considerably larger proportion of the heritability of ADHD can be explained by all common variants ( $h^2_{\text{snp}} = 0.22$ , SE = 0.01). This is consistent with previous estimates of  $h^2_{\text{snp}}$  for ADHD in smaller studies ( $h^2_{\text{snp}}$ : 0.1-0.28)<sup>23,24</sup>, and also comparable to what has been found for schizophrenia ( $h^2_{\text{snp}}$  0.23 - 0.26)<sup>23,24</sup>. 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 disorder<sup>40</sup>.

Along with polygenicity, selection and evolutionary pressures may be an important feature of the architecture of ADHD genetics. We observe that ADHD risk variants are strongly enriched in genomic regions conserved in mammals<sup>87</sup>, and constrained genes likely to be intolerant of loss-of-function mutations<sup>86</sup> are associated with ADHD. The directionality of potential selective effect is unclear: we find that common variant risk for ADHD is genetically correlated with having children younger and having more children, but is also correlated with a family history of parental mortality at a younger age. Given the documented association between ADHD and educational underachievement<sup>88,89</sup>, reinforced by genetic correlation of ADHD with educational attainment and childhood IQ<sup>43</sup> observed in this study, selective pressure on the genetics of ADHD would be consistent with recent work suggesting that variants associated with educational attainment are under negative selection in Iceland<sup>90</sup>. Future studies of rare and *de novo* variants 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<sup>24,42</sup>, as well as the broader genetic overlap of psychiatric disorders<sup>23,24</sup>. Positive genetic correlations between ADHD and health risk behaviors such as smoking and obesity are consistent with the observed increase in those behaviors among individuals with ADHD<sup>91-94</sup> and indicates a shared genetic basis across these traits.

The current results further support the hypothesis that ADHD is the extreme expression of one or more heritable quantitative traits. We observe strong concordance between the GWAS of ADHD and previous GWAS of ADHD-related traits in the population from the EAGLE Consortium<sup>30</sup>, both in terms of genome-wide genetic correlation and concordance at individual loci. Polygenic risk for ADHD has previously been associated with inattentive and hyperactive/impulsive trait variation below clinical thresholds in the population<sup>29</sup>. Shared genetic risk with health risk behaviors may similarly be hypothesized to reflect an impaired ability to self-regulate and inhibit impulsive behavior<sup>95,96</sup>.

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*, *DUSP6*, and other 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.

## 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)<sup>33</sup>. 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 (see 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 (see Supplementary Information). Each cohort was then phased and imputed using the 1000 Genomes Project phase 3 (1KGP3)<sup>34,97</sup> imputation reference panel using SHAPEIT<sup>98</sup> and IMPUTE2<sup>99</sup>, 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<sup>100,101</sup> 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<sup>35,102</sup>. This was done separately for each of the PGC samples and on a merged set of genotypes for the iPSYCH sample (see 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<sup>100,101</sup>. Principal components were included as covariates to control for population stratification<sup>35,102</sup>, along with relevant study-specific covariates where applicable (see Supplementary Information). Subsequently the results were meta-analysed using an inverse-variance weighted fixed effects model, implemented in METAL (version 2011-03-25)<sup>36</sup>. Variants were filtered and included for imputation quality (INFO score) > 0.8 and MAF > 0.01. Only markers supported by an effective sample size  $N_{\text{eff}} = 2/(1/N_{\text{cases}} + 1/N_{\text{controls}})$ <sup>103</sup> 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 (see 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. principle components) were also included. The conditional association results were then combined in an inverse-variance weighted meta-analysis.

### **Genetic correlations between ADHD samples**

Genetic correlation between the European-ancestry PGC and iPSYCH GWAS results was calculated using LD Score regression<sup>37</sup>. The regression was performed using pre-computed LD

scores for HapMap3 SNPs calculated based on 378 European-ancestry individuals from the 1000 Genomes Project (available on <https://github.com/bulik/ldsc>). Only results for markers with an imputation INFO score  $> 0.90$  were included in the analysis. In addition, a bivariate GREML analysis was conducted using GCTA<sup>104</sup> in order to estimate the genetic correlation between PGC case/control and trio study designs.

### **Polygenic Risk Scores for ADHD**

The iPSYCH sample were split into five groups, and subsequently five leave-one-out association analyses were conducted, using four out of five groups and the PGC samples as training datasets<sup>38</sup>. 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 principle 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 (see Supplementary Information).

### **SNP heritability and intercept evaluation**

LD score regression<sup>37</sup> 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  $\chi^2$  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, its expected value under the null hypothesis of no confounding from e.g. population stratification. The ratio between this deviation and the deviation of the mean  $\chi^2$  from one (i.e. its expected value under the null hypothesis of no association) was used to estimate the proportion of inflation in  $\chi^2$  attributable to confounding as opposed to true polygenic effects (ratio = (intercept-1)/(mean  $\chi^2$ -1)). SNP heritability was estimated based on the slope of the LD score regression, with heritability on the liability scale calculated assuming a 5% population prevalence of ADHD<sup>39</sup>.

### **Partitioning of the heritability**

SNP heritability was partitioned by functional category and tissue association using LD score regression<sup>40</sup>. 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.<sup>40</sup>.

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<sup>105,106</sup> (see Supplementary Information). The analyses were restricted to the European GWAS meta-analysis results to ensure matching of population LD structure. Results for each functional category were evaluated based on marginal enrichment, defined as the proportion of SNP heritability explained by SNPs in the annotation divided by the proportion of genome-wide SNPs in the annotation<sup>40</sup>. 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 correlation of ADHD with other traits were evaluated using LD Score regression<sup>42</sup>. 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 219 phenotypes with publically available GWAS summary statistics using LD Hub<sup>41</sup> (see Supplementary Information). 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 Consortium (submitted). 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.<sup>107</sup> (see Supplementary Information), implemented by a freely available R script (<https://github.com/hailianghuang/FM-summary>). Under the assumption that (a) there is one causal variant in each locus, and (b) the causal variant is observed in the genotype data, the credible set can be considered to have a 99% probability of containing the causal variant. For each the 12 genome-wide significant loci, variants within 1MB and in LD with correlation  $r^2 > 0.4$  to the index variant were considered for inclusion in the credible set analysis. The credible set analysis was done using the European GWAS meta-analysis to ensure consistent LD structure in the analyzed cohorts.

### **Biological annotation of variants in credible set**

The variants in the credible set for each locus, were annotated based on external reference data in order to evaluate potential functional consequences. In particular, we identify: (a) Gene and regulatory consequences annotated by Variant Effect Predictor (VEP) using Ensembl with genome build GRCh37<sup>108</sup>. We exclude upstream and downstream consequences, and consequences for transcripts that lack a HGNC gene symbol (e.g. vega genes). (b) Variants within 2kb upstream of the transcription start site (TSS) of at least one gene isoform based on Gencode v19<sup>109</sup>. (c) Variants annotated as interacting with a given gene in Hi-C data from samples of developing human cerebral cortex during neurogenesis and migration<sup>110</sup>. Annotations



are considered for both the germinal zone (GZ), primarily consisting of actively dividing neural progenitors, and the cortical and subcortical plate (CP), primarily consisting of post-mitotic neurons. (d) Variants identified as eQTLs based on gene expression in GTEx<sup>111</sup> or BIOS<sup>73</sup>. 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 Roadmap<sup>112</sup>. 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), 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 1.05<sup>83</sup> using the summary statistics from the European GWAS meta-analysis ( $N_{\text{cases}} = 19,099$ ;  $N_{\text{controls}} = 34,194$ ; See 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 test statistic. MAGMA accounts for gene-size, number of SNPs in a gene and LD between markers when estimating gene-based P-values. LD correction was based on estimates from the 1000 genome phase 3 European ancestry samples<sup>34</sup>.

The generated gene-based P-values were used to analyze sets of genes in order to test for enrichment of association signals in genes belonging to specific biological pathways or processes. In the analysis only genes on autosomes, and genes located outside the broad MHC region (hg19:chr6:25-35M) were included. We used the gene names and locations and the European genotype reference panel provided with MAGMA. For gene sets we used sets with 10-1000 genes from the Gene Ontology sets<sup>81</sup> curated from MsigDB 6.0<sup>82</sup>.

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.<sup>84</sup> and Spiteri et al<sup>85</sup>.

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 (the Exome Aggregation Consortium; ExAC)<sup>86</sup>. 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 23andMe and EAGLE. We evaluated evidence for replication based on: (a) genetic

correlation between the ADHD GWAS and each replication cohort; (b) sign tests of concordance between the ADHD GWAS meta-analysis and each replication cohort; (c) meta-analysis of the ADHD GWAS meta-analysis results with the results from analyses of the replication cohorts; and (d) tests of heterogeneity in the meta-analyses of the ADHD GWAS meta-analysis with the replication cohorts.

Genetic correlations were calculated using LD score regression<sup>37</sup> with the same procedure as described above. For the sign test, LD clumping was performed for all variants with  $P < 1 \times 10^{-4}$  in the ADHD GWAS meta-analysis using LD estimated from European ancestry individuals from 1000 Genomes Phase 3 data. The proportion of variants with a concordant direction of effect in the two replication samples ( $\pi$ ) was evaluated using a one sample test of the proportion with Yates' continuity correction against a null hypothesis of  $\pi = 0.50$  (i.e. the signs are concordant between the two analyses by chance). This test was done for loci passing P-value thresholds of  $P < 5 \times 10^{-8}$ ,  $P < 1 \times 10^{-7}$ ,  $P < 1 \times 10^{-6}$ ,  $P < 1 \times 10^{-5}$ , and  $P < 1 \times 10^{-4}$  in the ADHD GWAS meta-analysis (see Supplementary Information).

We performed three meta-analyses based on the ADHD GWAS meta-analysis result and the results from the two replication cohorts. First, we performed an inverse variance-weighted meta-analysis of the ADHD GWAS meta-analysis with the results of the 23andMe GWAS of self-reported ADHD case status. Second, we performed a meta-analysis combining the results from clinically ascertained ADHD with results from GWAS of ADHD-related behavior in childhood population samples (the EAGLE data). This was done using a modified sample size-based weighting method (see below). Third, we applied the modified sample size-based weighting method to meta-analyze the EAGLE GWAS with the ADHD+23andMe GWAS meta-analysis.

For meta-analyses including the EAGLE cohort, 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

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

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

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

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 traits<sup>113</sup>.

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

## 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 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).
- 44 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).
- 45 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).
- 46 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).
- 47 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).
- 48 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).
- 49 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).
- 50 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).
- 51 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).
- 52 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).
- 53 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).
- 54 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).
- 55 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).



- 56 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).
- 57 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).
- 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 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).
- 60 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).
- 61 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).
- 62 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).
- 63 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).
- 64 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).
- 65 Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nature genetics* **45**, 580-585, doi:10.1038/ng.2653 (2013).
- 66 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).
- 67 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).
- 68 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).
- 69 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).
- 70 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).
- 71 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).
- 72 Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539-542, doi:10.1038/nature17671 (2016).
- 73 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).

- 74 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).
- 75 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).
- 76 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).
- 77 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).
- 78 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).
- 79 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).
- 80 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).
- 81 Gene Ontology, C. Gene Ontology Consortium: going forward. *Nucleic acids research* **43**, D1049-1056, doi:10.1093/nar/gku1179 (2015).
- 82 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).
- 83 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).
- 84 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).
- 85 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).
- 86 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).
- 87 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).
- 88 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).
- 89 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).
- 90 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).
- 91 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).
- 92 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).

- 93 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).
- 94 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).
- 95 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).
- 96 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).
- 97 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).
- 98 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).
- 99 Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3* **1**, 457-470, doi:10.1534/g3.111.001198 (2011).
- 100 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).
- 101 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).
- 102 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).
- 103 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).
- 104 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).
- 105 Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).
- 106 Ernst, J. & Kellis, M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nature biotechnology* **33**, 364-376 (2015).
- 107 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).
- 108 McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122, doi:10.1186/s13059-016-0974-4 (2016).
- 109 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).
- 110 Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523-527, doi:10.1038/nature19847 (2016).
- 111 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).
- 112 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).
- 113 Turley, P. *et al.* MTAG: Multi-Trait Analysis of GWAS. *bioRxiv*, doi:10.1101/118810 (2017).

## 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, Aggessotype), 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 Aggessotype); Action Medical Research (GN2080 and GN2315); 4 Medical Research Council and SGDP Centre PhD studentships; and by the ECNP Network ADHD Across the Lifespan.

Dr. Langley was funded by Wellcome Trust (Grant No: 079711)

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

Barbara Franke's research is supported by funding from a personal Vici grant of the Netherlands Organisation for Scientific Research (NWO; grant 016-130-669, to BF), from the European Community's Seventh Framework Programme (FP7/2007 – 2013) under grant agreements n° 602805 (Aggrosotype), 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), IMSGC (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

This section will be filled in the review process

## 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](http://www.adhdinadults.com).

Dr. Neale is a member of Deep Genomics Scientific Advisory Board and has received travel expenses from Illumina. He also serves as a consultant for Avanir and 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: have 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.

**Correspondence and requests for materials should be addressed to:**

Benjamin Neale, [bneale@broadinstitute.org](mailto:bneale@broadinstitute.org)  
Anders D. Børglum, [anders@biomed.au.dk](mailto:anders@biomed.au.dk)  
Stephen Faraone, [sfaraone@childpsychresearch.org](mailto:sfaraone@childpsychresearch.org)

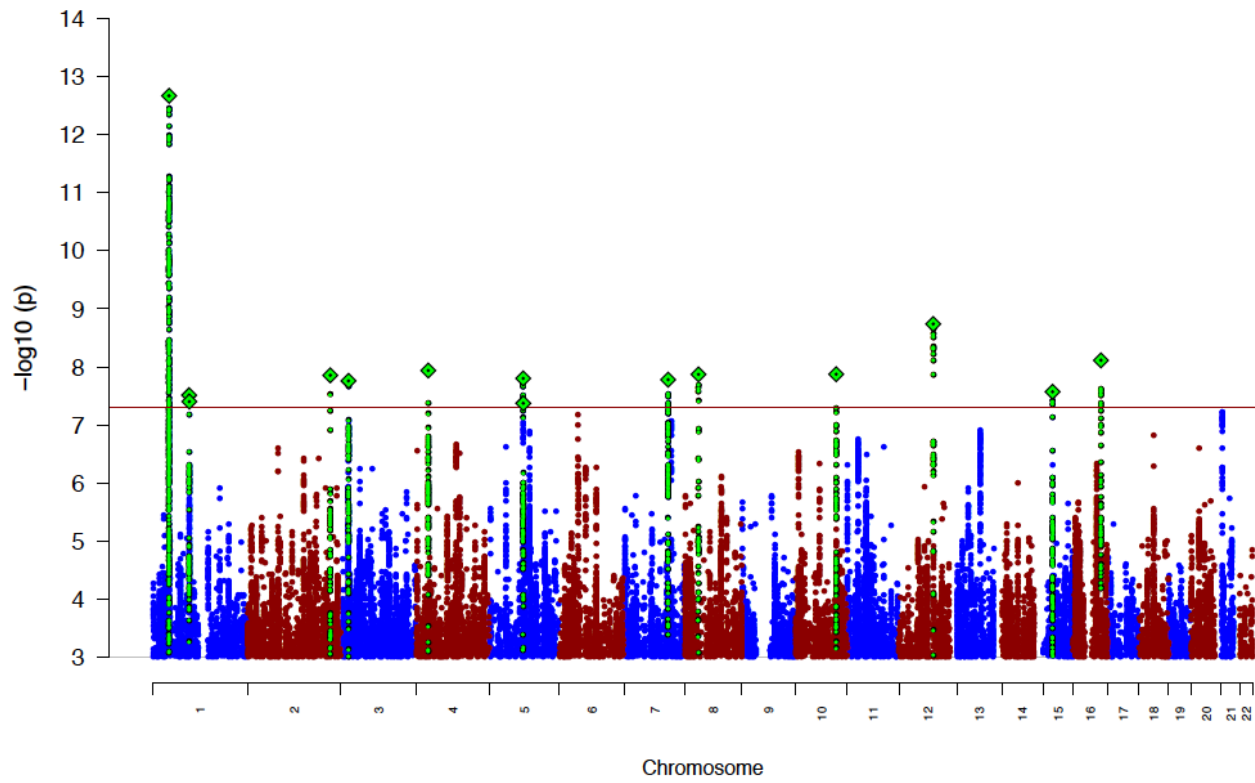
## Figure legends

**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.

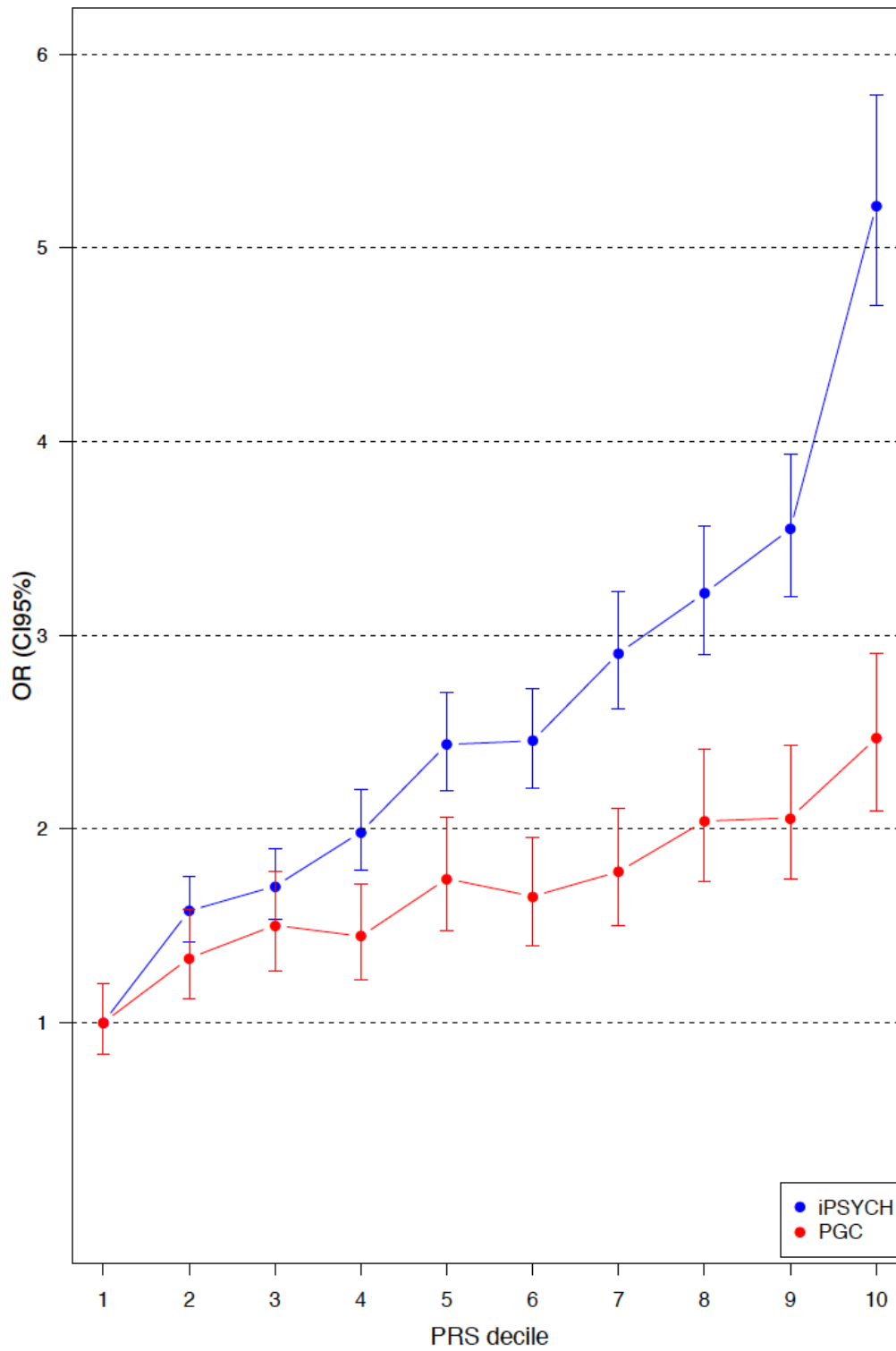
**Figure 3.** Significant genetic correlations between ADHD and other traits reveal overlap of genetic risk factors for ADHD across several groups of traits (grouping indicated by a horizontal line): educational, psychiatric/personality, weight (and possible weight related traits), smoking behaviour/smoking-related 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.**

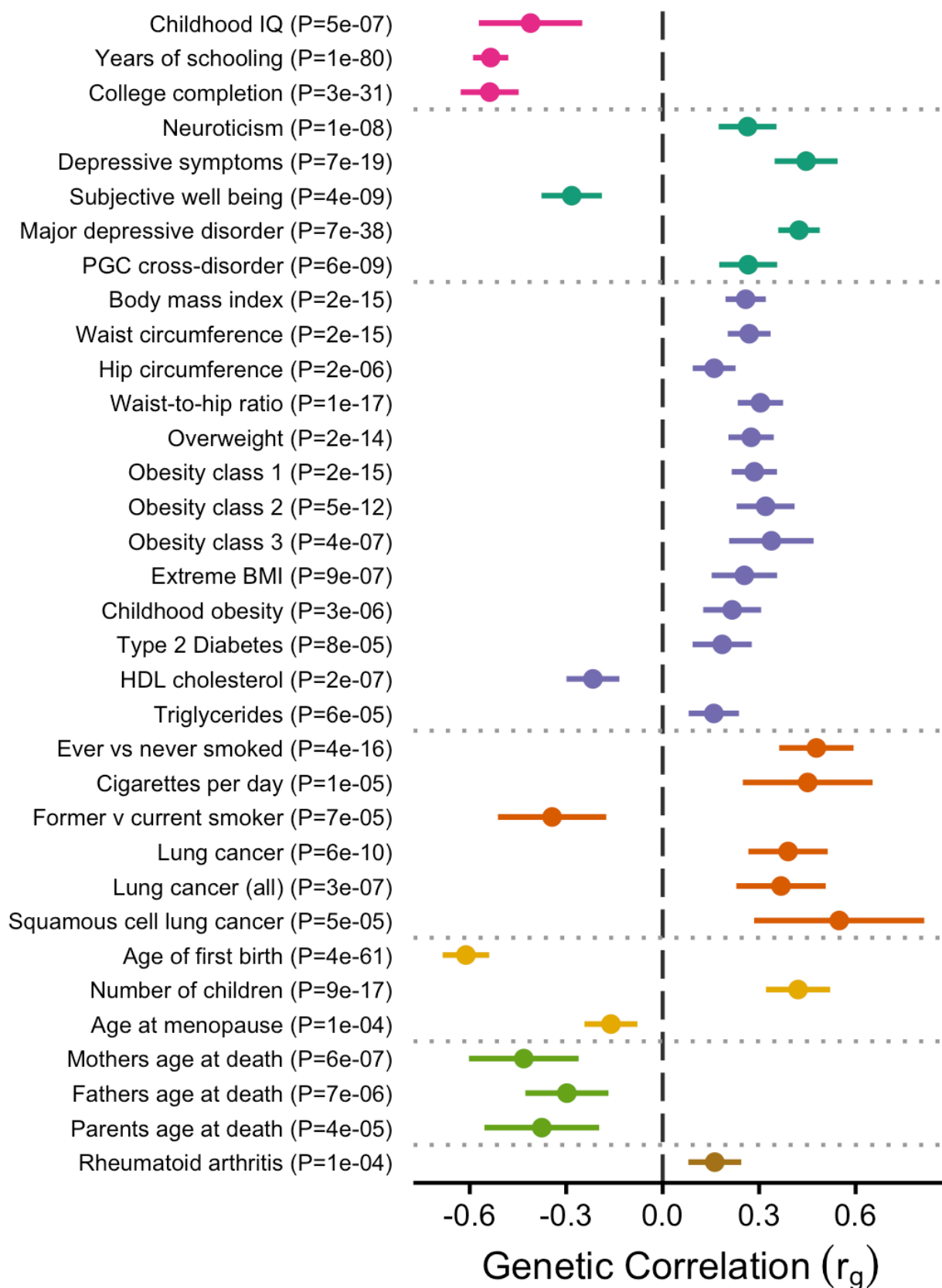




**Figure 2.**



**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	$2.14 \times 10^{-13}$
2	1	96602440	rs1222063	Intergenic	A	G	0.328	1.101	$3.07 \times 10^{-8}$
3	2	215181889	rs9677504	<i>SPAG16</i>	A	G	0.109	1.124	$1.39 \times 10^{-8}$
4	3	20669071	rs4858241	Intergenic	T	G	0.622	1.082	$1.74 \times 10^{-8}$
5	4	31151456	rs28411770	<i>PCDH7, LINC02497</i>	T	C	0.651	1.09	$1.15 \times 10^{-8}$
6	5	87854395	rs4916723	<i>LINC00461, MIR9-2, LINC02060, TMEM161B-AS1</i>	A	C	0.573	0.926	$1.58 \times 10^{-8}$
7	7	114086133	rs5886709	<i>FOXP2, MIR3666</i>	G	GTC	0.463	1.079	$1.66 \times 10^{-8}$
8	8	34352610	rs74760947	<i>LINC01288</i>	A	G	0.957	0.835	$1.35 \times 10^{-8}$
9	10	106747354	rs11591402	<i>SORCS3</i>	A	T	0.224	0.911	$1.34 \times 10^{-8}$
10	12	89760744	rs1427829	<i>DUSP6, POC1B</i>	A	G	0.434	1.083	$1.82 \times 10^{-9}$
11	15	47754018	rs281324	<i>SEMA6D</i>	T	C	0.531	0.928	$2.68 \times 10^{-8}$
12	16	72578131	rs212178	<i>LINC01572</i>	A	G	0.883	0.891	$7.68 \times 10^{-9}$

## **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 220 phenotypes

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