

## Polygenic profiles define aspects of clinical heterogeneity in ADHD

Sonja LaBianca<sup>a,b</sup>, Isabell Brikell<sup>b,c#</sup>, Dorte Helenius<sup>a,b#</sup>, Robert Loughnan<sup>d#</sup>, Joel Mefford<sup>e#</sup>, Clare E. Palmer<sup>f#</sup>, Rebecca Walker<sup>g,h#</sup>, Jesper R. Gådin<sup>a,b</sup>, Morten Krebs<sup>a,b</sup>, Vivek Appadurai<sup>a,b</sup>, Morteza Vaez<sup>a</sup>, Esben Agerbo<sup>b,c,i</sup>, Marianne Gørtz Pedersen<sup>b,c,i</sup>, Anders D. Børghlum<sup>b,j,k</sup>, David M. Hougaard<sup>b,l</sup>, Ole Mors<sup>b,m</sup>, Merete Nordentoft<sup>b,n,o</sup>, Preben Bo Mortensen<sup>b,c,i,k</sup>, Kenneth S. Kendler<sup>p</sup>, Terry L. Jernigan<sup>f,d,q,r</sup>, Daniel H. Geschwind<sup>g,h,s,t</sup>, Andrés Ingason<sup>a,b</sup>, Andrew W. Dahl<sup>u</sup>, Noah Zaitlen<sup>s</sup>, Søren Dalsgaard<sup>b,c,i</sup>, Thomas M. Werge<sup>a,b,o,\*</sup>, Andrew J. Schork<sup>a,b,v,\*</sup>.

#These authors contributed equally to this work

\* Please address correspondence to: [andrew.joseph.schork@regionh.dk](mailto:andrew.joseph.schork@regionh.dk) or [thomas.werge@regionh.dk](mailto:thomas.werge@regionh.dk)

<sup>a</sup> Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Mental Health Services Copenhagen, Roskilde, Denmark.

<sup>b</sup> The Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), Copenhagen, Denmark.

<sup>c</sup> National Centre for Register-based Research (NCRR), Department of Economics and Business Economics, Aarhus University, Aarhus, Denmark

<sup>d</sup> Department of Cognitive Science, University of California San Diego, La Jolla, CA, USA

<sup>e</sup> Department of Neurology, University of California, Los Angeles, Los Angeles, CA, USA

<sup>f</sup> Center for Human Development, University of California San Diego, La Jolla, CA, USA

<sup>g</sup> Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA, USA

<sup>h</sup> Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

<sup>i</sup> Centre for Integrated Register-based Research (CIRRAU), Aarhus University, Aarhus, Denmark

<sup>j</sup> Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark

<sup>k</sup> Centre for Integrative Sequencing (iSEQ), Aarhus University, Aarhus, Denmark

<sup>l</sup> Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark

<sup>m</sup> Psychosis Research Unit, Aarhus University Hospital – Psychiatry, Denmark

<sup>n</sup> Copenhagen Mental Health Center, Mental Health Services Capital Region of Denmark Copenhagen, Copenhagen, Denmark

<sup>o</sup> Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

<sup>p</sup> Department of Psychiatry, Virginia Commonwealth University, Richmond, Virginia, USA

<sup>q</sup> Department of Psychiatry, University of California San Diego, La Jolla, CA, USA

<sup>r</sup> Department of Radiology, University of California San Diego, La Jolla, CA, USA

<sup>s</sup> Department of Neurology, University of California, Los Angeles, Los Angeles, CA, USA

<sup>t</sup> Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

<sup>u</sup> Section of Genetic Medicine, University of Chicago, Chicago, Illinois, USA

<sup>v</sup> Neurogenomics Division, The Translational Genomics Research Institute (TGEN), Phoenix, AZ, USA

Word Counts:

Citations Count:

Abstract:	200
Introduction:	770
Results:	1955
Discussion:	1228
Methods:	2268

Main citations:	1-65
Methods citations:	66-106

## Abstract

Attention deficit hyperactivity disorder (ADHD) is a complex disorder with heterogeneous clinical presentations that manifest variability in long-term outcomes. The genetic contributions to this clinical heterogeneity, however, are not well understood. Here, we study 14 084 individuals diagnosed with ADHD to identify several genetic factors underlying clinical heterogeneity. One genome-wide significant locus was specifically associated with an autism spectrum disorder (ASD) diagnosis among individuals diagnosed with ADHD and it was not previously associated with ASD nor ADHD, individually. We used a novel approach to compare profiles of polygenic scores for groups of individuals diagnosed with ADHD and uncovered robust evidence that biology is an important factor in on-going clinical debates. Specifically, individuals diagnosed with ASD and ADHD, substance use disorder (SUD) and ADHD, or first diagnosed with ADHD in adulthood had different profiles of polygenic scores for ADHD and multiple other psychiatric, cognitive, and socio-behavioral traits. A polygene overlap between an ASD diagnosis in ADHD and cognitive performance was replicated in an independent, typically developing cohort. Our unique approach uncovered evidence of genetic heterogeneity in a widely studied complex disorder, allowing for timely contributions to the understanding of ADHD etiology and providing a model for similar studies of other disorders.

## Introduction

Attention deficit hyperactivity disorder (ADHD) is a multifactorial neurodevelopmental disorder with typical symptom onset in childhood, often persisting into adulthood, and affecting many aspects of life through impaired attention, impulsivity and hyperactivity<sup>1,2</sup>. Clinical presentations are heterogeneous and diagnosed individuals demonstrate substantial variability in symptom severity, predominance and duration, treatment, age at time of diagnosis, need for hospitalization, and other long-term outcomes<sup>3-5</sup>. Some diagnosed individuals experience more negative outcomes such as criminality<sup>6</sup>, premature mortality<sup>7</sup>, poor educational attainment<sup>8</sup>, or lower socioeconomic status<sup>9</sup>. Both psychiatric and somatic comorbidities in ADHD are common but their prevalence and (co)occurrence can vary substantially among individuals and across the life-span<sup>3,10</sup>. Better understanding the etiology of clinical heterogeneity is important for improving long term outcomes and precision care. Although genetic heterogeneity is proposed as a feature of complex disorders<sup>11</sup> and important to understand<sup>12</sup>, insights have proven elusive, especially when considering the extreme polygenic architectures of psychiatric disorders (e.g.,<sup>13,14</sup>).

Family studies show that genetic factors play an important role in ADHD etiology, with estimates of the narrow-sense heritability ( $h^2$ ) around 0.74<sup>15</sup>. This genetic contribution is complex, including copy number variants (CNVs)<sup>16,17</sup>, rare protein truncating variants (PTV)<sup>18</sup>, and polygenes - the numerous common variants with small, independent, additive effects on liability<sup>19</sup>. Estimates of SNP-based heritability ( $h^2_{SNP}$ ) for ADHD imply this polygene contribution is substantial (22%)<sup>19</sup>. ADHD associated polygenes are known to be pleiotropic, shared broadly across several clustered psychiatric, cognitive, and socio-behavioral traits, and concentrated in chromatin with active roles in neural development and functioning<sup>19,20</sup>. These insights come predominantly from consortia-driven meta-analyses<sup>21,22</sup> that aggregate multiple cohorts sampled according to different ascertainment criteria, case-control definitions, and healthcare contexts. Mapping robustly associated individual loci via genome wide association studies (GWAS) does not appear overly sensitive to such differences in contributing cohorts, however, inferences from polygenes may be more susceptible<sup>23-25</sup>. Thus, a polygene contribution to clinical heterogeneity in ADHD is plausible and has important implications for study design, across cohort replication, and providing a more nuanced picture of ADHD etiology<sup>12</sup>.

Recent studies have sought robust evidence of genetic heterogeneity for numerous complex disorders<sup>12</sup>, but many have been limited by conceptual, methodological, and data-related challenges that are exemplified by, but not limited to, ADHD. When considering ADHD, few cohorts have been ascertained with both the requisite statistical power for genetic analysis and the necessary depth and breadth of adjacent phenotyping for systematic investigations of clinical heterogeneity. As a result, analyses targeting such heterogeneity are often conducted *post-hoc* or secondary to other aims (i.e., mapping disorder associated loci) where they may also be limited to a focus on single trait or selected variants. This includes, as examples, the association of previously discovered ADHD risk variants with comorbid substance use

disorder (SUD)<sup>26</sup>, autism spectrum disorder (ASD)<sup>27</sup>, sex-differences<sup>28</sup>, or symptom persistence<sup>29</sup>. More generally, state of the field analytical approaches (e.g.,<sup>30</sup>) often specifically emphasize variants with an *a priori* association to disorder onset and may miss or under prioritize contributions from modifier variants<sup>31,32</sup> that could alter clinical trajectories without such a prior association (e.g., those disrupting drug metabolizing enzymes). Clinical presentations of complex disorders vary on a multitude of phenotypic dimensions and likely due to an equally diverse set of genetic factors, so studies that can overcome some of these prior limitations are poised to have wide-reaching impact.

Here, we use data from iPSYCH2012 case-cohort study<sup>33</sup>, which includes the largest single cohort of individuals clinically diagnosed with ADHD (N=14,084) and is linked with a wealth of adjacent phenotyping from Danish population-wide health and civil registers<sup>34-37</sup>. We first define a collection of *ADHD-adjacent traits*, using this term to refer to plausibly relevant, clinical phenotypes of individuals diagnosed with ADHD that we then assess for etiological relevance. We implement a well understood estimator of  $h^2_{SNP}$  in a novel way to prioritize ADHD-adjacent traits that most strongly associate with genetic differences among diagnosed individuals. Then, we conduct GWAS to identify single variants associated with these prioritized traits and describe plausible biological mechanisms. We then introduce a novel polygenic profiling approach that uses multivariate, multinomial logistic regression. This method simultaneously compares healthy controls and multiple case groups across sets of mutually adjusted polygenic scores (PGS) for cognitive, psychiatric, and socio-behavioral traits, while accounting for primary disorder (e.g., ADHD) PGS and covariates. We use this to identify specific and across PGS evidence of genetic heterogeneity. Finally, as an external validation of our findings, we construct PGS using variant effects from ADHD-adjacent trait GWAS to predict cognitive and behavioral performance in an independent, typically developing cohort. Our study adds robust, new genetic perspectives on existing clinical debates surrounding ADHD etiology and can serve as a model for similar investigations in other complex disorders.

## Results

### *ADHD-adjacent traits associate with genetic variability among diagnosed individuals*

We defined 22 ADHD-adjacent traits from national register data to broadly describe variability in medication use, psychiatric comorbidity, need of care, and sex recorded at birth for 14,084 individuals diagnosed with ADHD (**Methods, Table 1, Supplementary table 1**). We estimated the SNP heritability ( $h^2_{SNP}$ ) of each trait within the ADHD case group to prioritize those most relevant for follow up with more detailed investigation. (**Methods, Figure 1, Supplementary table 2**). We observed significant  $h^2_{SNP}$  ( $p < 0.002$ , adjusted for 22 tests) for three traits: a first ADHD diagnosis as an adult ( $\geq 18$  years of age) ( $h^2_{SNP} = 0.143$ , s.e. = 0.025,  $p = 1.7 \times 10^{-10}$ ), an ADHD-adjacent diagnosis of ASD ( $h^2_{SNP} = 0.091$ , s.e. = 0.024,  $p = 2.7 \times 10^{-5}$ ), and an ADHD-adjacent diagnosis of SUD ( $h^2_{SNP} = 0.085$ ; s.e. = 0.024;  $p = 1.0 \times 10^{-4}$ ). An additional

seven traits had nominally significant p-values ( $p < 0.05$ ), suggesting with additional samples more traits could be considered for follow up. We used hierarchical clustering to identify patterns in tetrachoric correlations of all pairs of traits. There were roughly three groups: male sex-childhood diagnoses, medication traits, and adult first ADHD diagnosis-mood diagnoses (**Supplementary table 3**, clustered heatmap: **Supplementary figure 1**). We note prioritized traits are at least partially independent with a first ADHD diagnosis as an adult positively associated with an SUD diagnosis ( $r_{\text{TET}} = 0.62$ , s.e. = 0.01,  $p < 1 \times 10^{-10}$ ) and negatively associated with an ASD diagnosis ( $r_{\text{TET}} = -0.43$ , s.e. = 0.017,  $p < 1 \times 10^{-10}$ ). SUD diagnoses were negatively associated with ASD diagnoses ( $r_{\text{TET}} = -0.33$ , s.e. = 0.019,  $p < 1 \times 10^{-10}$ ). ADHD-adjacent traits appear widely and robustly associated with genetic differences and further investigations are warranted for a first ADHD diagnosis as an adult, an ADHD-adjacent ASD diagnosis, and an ADHD-adjacent ASD diagnosis.

### *rs8178395 is specifically associated with an ADHD-adjacent ASD diagnosis*

We performed three GWAS to identify single variants associated with the prioritized ADHD-adjacent traits (**Methods**). No individual SNPs were associated ( $p > 5 \times 10^{-8}$ ) with a first ADHD diagnosis as an adult (**Supplementary figure 2**) or an ADHD-adjacent SUD diagnosis (**Supplementary figure 3**). However, one locus on chromosome 17q22 was significantly associated with an ADHD-adjacent ASD diagnosis (hg19: 55,341,733-57,341,733, lead SNP: rs8178395, minor allele (T) frequency (MAF) = 0.14, odds ratio (OR) = 1.30, s.e. on ln scale = 0.05,  $p_{\text{GWAS}} = 1.98 \times 10^{-8}$ ; **Figure 2a**, **Supplementary figure 4**). This locus was not identified in previous ADHD<sup>19</sup> (rs8178395,  $p = 0.44$ ), ASD<sup>38</sup> (rs8178395,  $p = 0.26$ ), or across-psychiatric disorders<sup>39</sup> (rs8178289, a proxy SNP with  $r^2 \text{ LD} = 0.9$ ,  $p = 0.23$ ) GWAS, although these individuals were included in each meta-analysis. PheWAS for rs8178395 and rs8178289 showed evidence for prior associations with physiological measures of blood cell composition, protein levels, metabolites, cardiovascular complications, and sleep behavior (**Supplementary figure 5**, **Supplementary table 4**). Quality checks suggest rs8178395 is reliably imputed (**Supplementary figure 6**, **Supplementary figure 7**, **Supplementary table 5**).

To determine if this variant is a novel ASD SNP or represents evidence of genetic heterogeneity within ADHD, we employed multinomial logistic regression (MLR) (**Methods**). This test extends the two-group logistic regression comparing ADHD with or without an adjacent ASD diagnosis to additionally compare individuals with an ASD, but not ADHD, diagnosis and controls with neither diagnosis. We observed that the frequency of the minor (T) allele of rs8178395 was significantly increased in the group with ADHD-adjacent ASD relative to all other groups (vs. ADHD only,  $p_{\text{MLR}} = 2.4 \times 10^{-8}$ ; vs. ASD only,  $p_{\text{MLR}} = 1.0 \times 10^{-6}$ ; vs. controls,  $p_{\text{MLR}} = 2.9 \times 10^{-7}$ ; **Figure 2b**, **Supplementary table 6**). rs8178395 indexes the first locus reported to have a specific association to individuals diagnosed with both ADHD and ASD.

The locus indexed by rs8178395 contains 38 genes (**Figure 2a**). rs8178395 falls within an intron of *LPO* and has been associated with expression of *DYNLL2*, *BZRAP1*, and *SKA2* in brain and *MPO*, *RAD51C*, *RNF43*, *SEPT4*, *SMG8*, *SUPT4H1*, *TEX14*, and *TRIM37* in other tissues (**Methods, Supplementary table 7**). Partitioned LD-score regression did not identify significant enrichment in any selected adult or fetal brain derived genome annotations (**Methods, Supplementary figures 8-10, Supplementary table 8**). So, we performed a regional transcription association study (TWAS) (**Methods**) using expression levels imputed via multiple adult<sup>40</sup> and fetal brain<sup>41</sup> expression quantitative trait loci (eQTLs) (**Methods; Supplementary table 9**). This did not identify significant candidates after Bonferroni correction ( $p < 0.006$ ); imputed adult-brain *SKA2* expression was the most significant ( $p = 0.008$ ). The proxy SNP for rs8178395 (rs8178289,  $r^2$  LD=0.9) is also associated with *BZRAP1* expression in brain tissue (**Figure 2c**) and, additionally, is reported by GTEx as a member of a haplotype containing splice QTLs (sQTL) for an associated antisense RNA, *BZRAP-AS1*, active in brain (**Figure 2d**) and other tissues<sup>42</sup>. Regulation of *BZRAP1* may be an especially plausible mechanism for this association. *BZRAP1* is a brain expressed gene<sup>43</sup> encoding a binding protein that couples voltage gated calcium channels to neurotransmitter vesicles in the presynaptic active zone<sup>44</sup>, has a well-characterized role in neurotransmitter release and synaptic transmission<sup>45</sup>, and exonic CNVs have been associated with ASD in multiplex families<sup>46</sup>.

#### *ADHD-adjacent traits share polygenes with psychiatric, cognitive, and socio-behavioral traits*

Next, we pursued a novel approach for mapping *polygenetic* contributions to clinical heterogeneity, extending concepts used in studies of disorder onset to increase power beyond single locus tests. We first estimated the genetic correlations ( $\rho_{G,SNP}$ ; **Methods**) between ADHD-adjacent traits and 41 psychiatric, cognitive, and socio-behavioral traits using GWAS summary statistics (**Figure 3, Supplementary table 10**). Trends were broadly similar across reference traits for a first ADHD diagnosis as an adult and ADHD-adjacent SUD, and generally opposite to those estimated for ADHD-adjacent ASD, consistent with phenotypic correlations (**Supplementary figure 1, Supplementary table 3**). Individually, an adult first diagnosis was genetically correlated ( $\rho_{G,SNP}$ , FDR<0.05) with psychiatric traits, reproductive behaviors, education, and cognitive performance (**Figure 3a**), and, of note, showed a large negative trend with clinically ascertained and defined childhood ADHD<sup>47</sup> ( $\rho_{G,SNP} = -0.5$ , s.e.=0.27,  $p = 0.07$ ). ADHD-adjacent SUD showed similar estimates of  $\rho_{G,SNP}$  with psychiatric outcomes, reproductive behaviors, education, and cognitive performance, but also smoking behavior (**Figure 3b**). ADHD-adjacent ASD, however, had different  $\rho_{G,SNP}$  for education and cognitive performance (**Figure 3c**). In summary, ADHD-adjacent traits likely share polygenes with psychiatric, cognitive, and socio-behavioral traits and may define genetic heterogeneity in ADHD.



# *Polygenic scores for psychiatric, cognitive, and socio-behavioral traits define aspects of heterogeneity in ADHD*

We extend these more qualitative descriptions with a novel polygenic profiling approach providing robust, quantitative tests that ADHD-adjacent traits manifest as a result of genetic heterogeneity indexed by profiles of PGS (**Methods; Figure 4, Supplementary tables 11-13**). The multi-PGS profiles were different among individuals receiving a first diagnosis for ADHD as an adult (1<sup>st</sup> Dx adult; n=3,323), a first diagnosis for ADHD before adulthood (1<sup>st</sup> Dx child; n=10,761), and controls (ADHD-; n=21,409). The mean PGS for ADHD, depressive symptoms, childhood IQ, years of schooling, and age at first birth varied significantly among the three groups ( $p < 4 \times 10^{-4}$ ; **Figure 4a**, stars; **Supplementary table 11**) and these three-group-wise differences were driven by specific, highly significant pairwise contrasts (**Figure 4a**, black and red brackets). The two ADHD case groups generally deviated significantly from controls and in the same direction. Individuals first diagnosed as an adult had, on average, significantly less positive PGS (i.e., closer to the population average) for ADHD and more negative PGS (i.e., further from the population average) for years of schooling than ADHD cases diagnosed before adulthood. Additional trends suggested more negative cognitive, education, and reproductive PGS and more positive psychiatric PGS for adult diagnosed ADHD cases. Simulations suggest the overall trends here are not consistent with models where an adult diagnosis is simply due to ADHD liability associating with onset, a misdiagnosis of MDD, or education as an age dependent, heritable exposure (**Supplementary Note, Supplementary Figures 11-15**).

The multi-PGS profiles for individuals with an ADHD-adjacent SUD diagnosis (ADHD+, SUD+; n=2 627), without an ADHD-adjacent SUD diagnosis (ADHD+, SUD-; n=11 457), with an SUD but no ADHD diagnosis (ADHD-, SUD+; n=5 943), and with neither ADHD nor SUD diagnoses (ADHD-, SUD-, controls; n=20 509) were also different and with a similar pattern to the previous result (**Figure 4**). The mean PGS for ADHD, depressive symptoms, schizophrenia, education as in years of schooling, age at first birth, and smoking initiation varied significantly among the four groups ( $p < 4 \times 10^{-4}$ ; **Figure 4b**, stars; **Supplementary table 12**). As above, all case groups generally deviated from controls significantly and in the same direction. Notable pairwise differences (**Figure 4b**, black and red brackets) included individuals with ADHD-adjacent SUD diagnoses showing a trend towards less positive ADHD PGS, significantly more positive PGS for depressive symptoms, schizophrenia, and smoking initiation, and significantly more negative PGS for years of schooling, that ADHD without an adjacent SUD diagnosis. Individuals with ADHD-adjacent SUD diagnoses showed, relative to those with SUD diagnoses only (i.e., no ADHD diagnosis), significantly more negative PGS for education and more positive PGS for smoking initiation, but no significant differences in ADHD, depressive symptoms, or schizophrenia PGS.

The multi-PGS profiles for individuals with ADHD-adjacent ASD diagnoses (ADHD+, ASD+; n=2 284), with ADHD but no adjacent ASD diagnosis (ADHD+, ASD-; n=11 800), with ASD but not ADHD diagnoses (ADHD-, ASD+; n=9 804), and with neither (ADHD-, ASD- controls; n=21 197) were also

different, but trends were broadly different from the previous two analyses. The mean PGS for ADHD, ASD, intelligence, and years of schooling varied significantly among the four groups ( $p < 4 \times 10^{-4}$ ; **Figure 4c**, stars, **Supplementary table 13**). The individuals with ADHD-adjacent ASD diagnoses had profiles that appeared to reflect both single-diagnosed groups, simultaneously. The ADHD-adjacent ASD group had significantly more positive ADHD and ASD PGS than controls, significantly more positive ASD PGS than, but similar ADHD PGS as, the ADHD only group, and significantly more positive ADHD PGS than, but similar ASD PGS as, the ASD only (**Figure 4b**, black and red brackets). Interestingly, the single diagnosed ASD and ADHD groups had opposite trends (i.e., above vs. below the population average) for intelligence and education PGS, and the ADHD-adjacent ASD group fell nearly perfectly in between the two. Simulations suggest the overall trends here are not consistent with misdiagnosis and most consistent with models where diagnosed individuals have diagnosis that relate to an excess of liability for both disorders, simultaneously (**Supplementary Note, Supplementary Figures 16-24**).

Large differences among groups that are not described as statistically significant (e.g., **Figure 4b**: college completion) are due to collinearity among constituent profile PGS that are fitted jointly (**Methods**) and are significant when fitted separately (**Supplementary figure 25, 26, Supplementary tables 14-16**). Trends depicted as unadjusted levels of PGS in **Figure 4** are directionally consistent with trends when depicted as partial PGS effects (i.e., regression coefficients) estimated in the joint models (**Supplementary figure 27, Supplementary tables 11-13**).

*A polygenic score for ADHD-adjacent ASD is associated with cognitive performance in an independent, typically developing cohort*

Finally, we sought external support for our finding that polygenic contributions to clinical heterogeneity in ADHD, as indexed by ADHD-adjacent traits, are shared with psychiatric, cognitive, socio-behavioral traits. To pursue this, we used the results of our three ADHD-adjacent trait GWAS to construct PGS in the Adolescent Brain and Cognitive Development (ABCD) study<sup>48</sup>, an independent, typically developing child cohort. Each PGS was tested for an association with each of 51 behavioral and cognitive assessments, adjusting for ADHD PGS and other potential confounders (**Methods, Figure 5, Supplementary table 17**). These analyses confirmed an association between the polygenic basis of an ADHD-adjacent ASD diagnoses and cognitive performance. Specifically, a higher PGS was significantly ( $p < 0.001$ ) associated with increased performance on an assessment of overall cognitive function that appears driven by higher scores in verbal cognition (e.g., crystalized composite, and picture vocabulary scores; **Figure 5c**). No other tests were strictly significant, although a few relevant trends are noted: the signs of the relationships of the different PGS and cognitive or psychiatric traits were in line with our previous analyses and multiple individual PGS trends were consistent for specific psychiatric or behavioral symptoms. We find consistent, independent support for a polygenic relationship between an ADHD-adjacent ASD



diagnosis and cognitive performance, but the small discovery GWAS and indirect, modest replication sample limit our power for broader replication.

## Discussion

In this study we used a unique data resource and a novel analysis strategy to identify new evidence for genetic heterogeneity underlying clinical heterogeneity in a widely studied complex disorder. We showed that multiple, clinically relevant ADHD-adjacent traits were related to genetic variability among diagnosed individuals and that these genetic contributions were complex (i.e., marked by few large effects despite significant  $h^2_{SNP}$ ) but characterizable. Our multivariate analysis of polygenic profiles demonstrated that variability in polygenes associated with psychiatry, cognitive, and socio-behavioral traits is likely to underlie differences in clinical presentations of individuals diagnosed with ADHD. Our findings regarding age of first ADHD diagnosis and ADHD-adjacent ASD diagnoses, in particular, make timely contributions to the understanding of the genetic etiology of ADHD, where other results reflect important themes that are relevant for the study of complex disorders more broadly. Finally, our approach may enable better characterizations of heterogeneity in other complex disorders.

Despite evidence of clinical co-occurrence in individuals and families, shared genetic risk factors<sup>18,49,50</sup>, and biological overlap<sup>51</sup>, no previous GWAS has focused specifically on individuals with diagnoses for both ADHD and ASD<sup>50,52,53</sup>. We identified a locus that appears novel for both disorders, and specific to this doubly diagnosed group. eQTL and previous trait associations suggest the locus is functional, has a plausible connection to regulating neurobiology<sup>43-45</sup>, and the association is supported in a complementary design using multiplex ASD families<sup>46</sup>. While this association needs further replication, it suggests that studying specific comorbidity patterns or presentations may help identify pleiotropic SNPs and disentangle antagonistic or heterogeneous variant effects.

Until the introduction of DSM-5 in 2013<sup>1</sup>, formal guidelines did not allow ASD and ADHD to be diagnosed as comorbid<sup>54</sup>. The recognition of overlap in cognition<sup>55</sup>, genetics<sup>55</sup>, symptoms scores<sup>56</sup>, and neuroanatomic characteristics<sup>57</sup> challenge this and the possibility that the core etiologies of both disorders could exist in an individual, simultaneously, are active areas of research and debate<sup>54</sup>. Our polygenic profile approach suggests that individuals with an ADHD-adjacent ASD diagnosis carry, at minimum, the *polygenic* etiologies of both individual disorders. This idea is supported by a recent register study in the Swedish population<sup>50</sup> that reported comorbid ADHD and ASD diagnosis were most often made on the same day, by the same clinician, observing concurrent symptoms. This trend stands in contrast to similar studies of bipolar disorder and schizophrenia, that suggest an appreciable level of lifetime comorbidity may relate to evolving symptoms and difficulties in diagnosing first episode psychosis<sup>58</sup>. Our polygenic profiling approach supports the notion that ASD and ADHD can be experienced simultaneously, consistent with emerging trends in clinical epidemiology and evolving psychiatry nosology.

The nature of adult-onset ADHD, its validity, existence, and whether it represents a distinct clinical or biological construct, is a particularly active and contentious research topic<sup>3,59,60</sup>. The crux of this debate rests on the timing of symptom onset. As above, it was DSM-5<sup>1</sup> that introduced criteria for an adult ADHD diagnosis. Here, adult ADHD requires retrospective self or informant reported childhood symptoms for a diagnosis, but these reports may suffer recall bias, and objective, representative longitudinal studies of premorbid symptoms are sparse. One perspective (e.g.<sup>29</sup>), then, has been that a first diagnosis as an adult should be thought of as persistent ADHD, perhaps missed during childhood, and, as symptoms may remit with age, be considered a more severe or biological form of ADHD. Here, we observed that, contrary to this perspective, individuals with a first registered diagnosis as an adult had *less* ADHD PGS suggesting they may have a more environmental, less reliable, or less severe form of the disorder<sup>61</sup>. Our polygenic profile approach offers varied support for each of these hypotheses in that adult diagnosed individuals had lower education PGS, higher non-ADHD psychiatric PGS, and similar profiles to individuals with ADHD-adjacent SUD diagnoses. Our data and approach implicate genetic heterogeneity in important debates surrounding adult ADHD where other studies<sup>29</sup> with similar aims have been less successful.

Beyond ADHD etiology, our study highlights a few themes that are broadly relevant for the study of complex disorders and associated heterogeneity. First, as demonstrated for MDD<sup>23</sup>, male pattern baldness<sup>62</sup>, and in simulations<sup>23,25,63</sup>, the polygenic background of case groups may become skewed if evolving nosology, case definitions, or censoring are not considered carefully. Here, we expect that evolving guidelines around comorbid ADHD and ASD may change the genetic landscape of typical ADHD and ASD case groups, implicitly, and more active ascertainment or censoring mechanisms could have similar consequences. Second, PGS have received a lot of attention as potential clinical instruments for various applications<sup>64,65</sup>, but the focus has often been on primary disorder scores (i.e., here, ADHD) and distinguishing cases from controls. Here we saw using profiles of multiple polygenic scores is a powerful approach for distinguishing among ADHD cases, and that the strongest predictors were not primary disorder PGS. The idea that the polygenic background of a patient can alter a clinical trajectory is not new, but as PGS are aimed at more diverse clinical decisions (i.e., beyond case-control discrimination), more comprehensive summaries of polygenic background should be considered. Finally, methods for detecting heterogeneity might benefit from relaxing strong *a priori* restrictions to onset associated variants while also focusing more broadly on polygenes. Our approach has broad implications for the study of complex disorders beyond ADHD.

Our results must be considered in light of a few important limitations. We use register diagnoses, which are given with high reliability by trained psychiatrists, but are assigned per hospital contact, and may shift over time as patient symptoms evolve or clarify<sup>58</sup>. This could lead to some misdiagnosis. This issue is not unique to register studies as it reflects clinical practice but may offer a different perspective from studies of research diagnoses that may include retrospective censoring or integration according to a hierarchy. Some individuals may seek treatment exclusively through primary care and those contacts are not available in

iPSYCH. This may be especially relevant when considering symptom onset and first diagnosis as earlier contacts could be missed. A registered diagnosis may still represent an increase in severity, if not onset, as a hospital referral could represent a change in need of care. The iPSYCH cohort is relatively young and may underestimate the prevalence of adult and later onset outcomes. iPSYCH includes only a combined ADHD subtype (ICD-10: F90.0), and important genetic differences may emerge if individuals diagnosed with inattentive (ICD-10: F98.8) and combined subtypes (ICD-10: F90.1, F90.8, or F90.9) were included. Methodologically, we compared groups of ADHD cases to estimate SNP-heritability and performed GWAS, using these to implicate genetic contributions to clinical heterogeneity of ADHD. However, just as with case-control analysis, these approaches are sensitive to spurious differences. This could occur due to population stratification or if heritable, disorder irrelevant features (e.g., hair color) were used to define patient groups<sup>12,30</sup>. In our work, we mitigate this by adjusting for ancestry related principal components, carefully selecting ADHD adjacent-traits with plausible clinical relevance or specificity to ADHD and including controls in our polygenic profile approach.

This work is part of an emerging body of evidence in psychiatric genetics that suggests we are now, after decades of data aggregation, at a point where we can begin to study not just what makes diagnosed individuals different from healthy controls, but what may differentiate diagnosed individuals from each other with respect to outcomes. Next, we must consider the implications these differences have across multiple areas of research and clinical care.

## Online Methods

### *iPSYCH2012 case-cohort study*

The Lundbeck Foundation initiative for Integrative Psychiatric Research (iPSYCH)<sup>33</sup> is a case-cohort study of individuals born in Denmark between 1981 and 2005 (n=1 472 762). 87 764 individuals were sampled, including a random sample of 30 000 and 59 996 ascertained for diagnoses of attention-deficit hyperactivity disorder (ADHD), autism spectrum disorders (ASD), anorexia (ANO), affective disorder (AFF), bipolar disorder (BP), or schizophrenia (SCZ). DNA was extracted from dried bloodspots in the Danish Neonatal Screening Biobank<sup>66</sup>. Diagnoses were obtained from the Danish Psychiatric Central Research Register (PCR)<sup>36</sup> and for anorexia also from the Danish National Patient Register (DNPR)<sup>35</sup> from 1 year birthday or 10 year birthday of study individuals to December 31, 2012. Linkage across registers uses the Danish Civil Registration System<sup>37</sup>. Diagnoses given by psychiatrists in primary care are not recorded in these registers. This study focused specifically on 14 084 of 18 726 individuals ascertained for ADHD (ICD-10: F90.0), 12 088 of 16 146 ascertained for ASD (ICD-10: F84.0, F84.1, F84.5, F84.8 or F84.9), and 21 197 of 30 000 controls diagnosed with neither ADHD nor ASD. Furthermore, 8 498 individuals with substance use disorders (SUD) (ICD-10: F1) were selected from among the 30 000 population controls and 57 764 psychiatric cases and finally, 20 509 of 30 000 controls with neither ADHD nor SUD. These subsets passed quality control (see below). The use of this data follows standards of the Danish Scientific Ethics Committee, the Danish Health Data Authority, the Danish Data Protection Agency, and the Danish Neonatal Screening Biobank Steering Committee. Data access was via secure portals in accordance with Danish data protection guidelines set by the Danish Data Protection Agency, the Danish Health Data Authority, and Statistics Denmark.

Genotyping was performed on the Infinium PsychChip v1.0 array with amplified DNA extracted from dried bloodspots. Data quality control is described in detail elsewhere<sup>20,33</sup>. Briefly, 246 369 of the ~550 000 genotyped SNPs were deemed good quality, phased using SHAPEIT3<sup>67</sup>, and imputed using the 1 000 genomes project phase3<sup>68</sup> as a reference with Impute2<sup>69</sup>. Imputed additive genotype dosages and best-guess genotypes were checked for imputation quality (INFO>0.2), Hardy–Weinberg equilibrium (HWE;  $p < 1 \times 10^{-6}$ ), association with genotyping wave ( $p < 5 \times 10^{-8}$ ), association with imputation batch ( $p < 5 \times 10^{-8}$ ), differing imputation quality between cases and controls ( $p < 1 \times 10^{-6}$ ), and minor allele frequency (MAF>0.01). 8 019 760 dosages and best-guess genotypes remained. Subjects of homogeneous genetic ancestry were selected after principal components analysis using EIGENSOFT v6.0.1<sup>70</sup>. One from each pair of individuals with closer than third degree kinship as estimated with KING v1.9<sup>71</sup> was excluded, and no samples had abnormal heterozygosity, high levels of missing genotypes (>1%), nor genotype/recorded sex discordance.

## Selecting ADHD-adjacent traits

DNPR<sup>35</sup> and PCR<sup>36</sup> contain information (e.g., date of admission, ICD diagnostic code, etc.) on all inpatient hospital contacts since 1977 and 1969, respectively, and outpatient and emergency room contacts since 1995. From these registers we defined two sets of ADHD-adjacent traits to capture psychiatric comorbidity. First, we coded the presence of specific psychiatric disorders as defined by the iPSYCH2012 case-cohort ascertainment criteria<sup>33</sup>, including ASD, AFF, ANO, and combined psychotic disorders (SCZ and BP). As a second set, we summarized psychiatric diagnoses more broadly, recording each ICD-10 Behavioral and Mental Disorders subchapter<sup>2</sup> recorded in either PCR or DNPR until 2016 (F1, F2, F3, F4, F5, F6, F7, F8, F9 excluding ADHD (F90.0-9 and F98.8)). As before, we created variables representing a diagnosis from each subchapter. We also counted the number of hospitalizations with ADHD (ICD-10 F90.0) recorded as the main diagnosis of action in the PCR up until 2016, also dichotomized as a split by the median. A variable splitting cases on the first recorded ADHD diagnosis (ICD-10 F90.0) occurring before or after 18 years of age. Finally, a variable was created recording sex as reported at birth.

The Danish National Prescription Register (NPR)<sup>34</sup> holds information on prescriptions redeemed from pharmacies in Denmark since 1995. The NPR does not cover drugs used during hospital admissions, by certain institutionalized individuals (e.g. psychiatric), or drugs supplied directly by hospitals or treatment centers. We defined three traits from the NPR noting cases with a record of at least two prescriptions after the age of 3 for drugs with the Anatomical Therapeutic Chemical (ATC) codes for stimulants (N06BA01, N06BA12, N06BA02, N06BA04), or atomoxetine (N06BA09), or both (i.e., any ADHD medication). A fourth trait counted the number of prescriptions an individual had filled and dichotomized on the median number (in this cohort) of total prescriptions. Further details are provided in **Supplementary table 1**.

We used tetrachoric correlations to describe the co-occurrence of ADHD-adjacent traits, estimating them with the R package *polycorr*<sup>72</sup> and adding 0.5 to zero cells<sup>73</sup>. Hierarchical clustering was performed using the *heatmaply*<sup>74</sup> package.

## SNP-based heritability

SNP heritability ( $h^2_{SNP}$ ) was estimated on the observed scale using GREML in GCTA v1.92.1 beta6<sup>75</sup> with 25 ancestry principal components as fixed effect covariates. The genetic relationship matrix (GRM) was constructed using GCTA and from best-guess genotypes with a MAF greater than 0.01. Significance was adjusted by Bonferroni correction ( $p < 0.05/22 = 0.002$ ), and nominally significant tests ( $p < 0.05$ ) were deemed suggestive.

## Genome-wide Association Studies

Case-case genome-wide association studies (GWAS) were performed within the ADHD case group (n=14 084). Logistic regression in plink v1.90b3.34<sup>76</sup> was used to test the association between imputed additive allele counts and case subgroup membership, with 25 ancestry principal components as covariates. Genomic inflation was estimated as the unconstrained LD-score regression<sup>77</sup> intercept following: <https://github.com/bulik/ldsc/wiki>. Individual SNPs with  $p < 5 \times 10^{-8}$  were declared genome-wide significant. Lead SNPs were defined as the most significant SNP within a 2 mega base (mb) locus.

For genome-wide significant loci, multinomial logistic regression (MLR)<sup>78</sup> was performed using the R package *nnet*<sup>79</sup>. The MLR tests the logistic regression comparison of allele counts at a lead SNP between ADHD subgroups (i.e., individuals with comorbid ADHD and ASD, n=2 284, and individuals with ADHD but not ASD, n=11 800) to simultaneously include comparisons with individuals diagnosed with relevant complementary disorder (i.e., individuals with ASD but not ADHD, n=9 804) and undiagnosed controls (i.e., individuals with neither ADHD nor ASD, n=21 197).

For genome wide significant loci, candidate genes were selected according to three criteria: 1) positional candidate genes were selected by ANNOVAR<sup>80</sup> in FUMA<sup>81</sup> as overlapping a lead SNP's genomic position, 2) eQTL candidate genes with previous expression association to a lead SNP in one of multiple sources aggregated by FUMA, 3) transcriptional association candidate genes with predicted expression in adult and/or fetal brain. Here a transcription association (TWAS) analysis was used integrating published per-SNP effects on expression for fetal<sup>41</sup> and adult<sup>82</sup> brain, and FUSION<sup>83</sup> to generate aggregate predicted expression in each individual. TWAS p-values were considered significant after Bonferroni correction ( $p < 0.05/9 = 0.006$ ). For lead SNPs from genome-wide significant loci, the GWASatlas v20191115<sup>84</sup> (<https://atlas.ctglab.nl/>) was used to pursue a pheWAS across 4 756 studies grouped according to the provided ontology. The proxy SNP for rs8178395 was identified using LDlink (<https://ldlink.nci.nih.gov/?tab=help#LDproxy>)<sup>85</sup>.

We used LD score regression (LDSC)<sup>77</sup> to partition  $h^2_{SNP}$  among selected genome annotations, accounting for LD and baseline annotations. Annotations of interest included those defined by eQTLs from fetal brain<sup>41</sup>, eQTLs from adult<sup>40</sup> brain, and regions of open chromatin measured by ATAC sequencing in fetal<sup>86</sup> and adult<sup>87</sup> brain. A “full baseline model” of 53 functional categories was employed following Finucane et al<sup>88</sup>. We followed the author protocols for analysis (<https://github.com/bulik/ldsc/wiki/Partitioned-Heritability>). eQTL and ATAC LD-scores were created for the subset of human SNPs genotyped in HapMap v3 SNPs (HM3) using the LD-score regression software by identifying HM3 SNPs within a 500bp window ( $\pm 250$ ) around each eQTL SNP or within an ATAC open chromatin window. Category-specific LD-scores were the sum of LD ( $r^2$ ) for each HM3 SNP with SNPs meeting the previous functional criteria. The European subset of the 1000 Genomes Project Phase3 was used as an LD reference. Enrichment was estimated for each annotation as the proportion of heritability explained by each annotation divided by the proportion of SNPs in the genome falling in that category. Enrichment p-values were declared significant after Bonferroni correction ( $p < 0.05/28 = 0.0018$ ) or suggestive when  $p < 0.05$ .



## SNP-based genetic correlations

SNP-based genetic correlations ( $\rho_{G,SNP}$ ) were estimated using LD-Score regression in LD-hub (<http://ldsc.broadinstitute.org/>)<sup>89</sup>, an online-tool for estimating  $\rho_{G,SNP}$  against a catalog of published GWAS studies. Summary statistics were uploaded for the three ADHD variable case-case GWAS performed for this study and were estimated with each of 41 cataloged traits from 10 psychiatric, cognitive, and socio-behavioral categories: smoking, neurology, personality, sleeping, cognition, reproduction, education, neuroimaging, psychiatry, and ageing. Prioritized traits had FDR<0.05.

## Polygenic Scores in iPSYCH

GWAS summary statistics (**Supplementary table 18**) were downloaded from public repositories and PGS for each were calculated using LDpred v01 for individuals in iPSYCH<sup>90</sup>. Reference GWAS were selected to ensure no subject overlap with iPSYCH. SNP inclusion criteria were: MAF>0.05, INFO>0.98, and HWE p-value>1x10<sup>-5</sup>. Palindromic (A/T, C/G) SNPs, SNPs not uniquely mapped to hg19 positions, and SNPs not having unique IDs in dbSNP v151 were excluded. Only one SNP from a group of SNPs in LD (pairwise  $r^2$  LD>0.99) was retained. We used an LDpred p-parameter of 1, corresponding to an infinitesimal model as a prior assumption for the per-SNP effects.

We used multivariate, multinomial logistic regression (MLR)<sup>78</sup> implemented in the R package *nnet*<sup>79</sup> to test for heterogeneity in PGS profiles among ADHD case subgroups, non-diagnosed controls, and individuals diagnosed with ASD or SUD but not ADHD. For each of the three ADHD variables, we jointly fit primary disorder PGS (i.e., ADHD, ASD), the set of PGS for traits that showed significant  $\rho_{G,SNP}$  with the ADHD variable in the LDSC analysis described above, and 25 ancestry principal components. MLR models fitting each PGS individually, along with ancestry covariates, are presented in the **Supplementary tables 14-16**. Three MLR were of primary interest. First, for ADHD-adjacent ASD, we compared individuals diagnosed with neither ADHD nor ASD (n=21 197), both ADHD and ASD (n=2 284), ADHD but not ASD (n=11 800), and ASD but not ADHD (n=9 804) on PGS profiles containing scores for ADHD<sup>47</sup>, ASD<sup>91</sup>, intelligence<sup>92</sup> and years of schooling<sup>93</sup>. For ADHD-adjacent SUD, we compared individuals with neither ADHD nor SUD (n=20 509), both ADHD and SUD (n=2 627), ADHD but not SUD (n=11 457), and SUD but not ADHD (n=5 943) on PGS profiles containing scores for ADHD<sup>47</sup>, depressive symptoms<sup>94</sup>, schizophrenia<sup>95</sup>, intelligence<sup>92</sup>, college completion<sup>96</sup>, years of schooling<sup>93</sup>, age at first birth<sup>97</sup>, number of children ever born<sup>97</sup>, and smoking initiation<sup>98</sup>. For age of first ADHD diagnosis we compared individuals without a diagnosis for ADHD (n=21 409), with a first diagnosis of ADHD as an adult (>18 years of age) n=3 323, and with a first diagnosis of ADHD before age 18 (n=10 761) on PGS profiles containing scores for ADHD<sup>47</sup>, depressive symptoms<sup>94</sup>, major depressive disorder<sup>99</sup>, mother's age at death<sup>100</sup>, childhood IQ<sup>101</sup>,

college completion<sup>96</sup>, years of schooling<sup>93</sup>, age at menopause<sup>102</sup>, and age at first birth<sup>97</sup>. A likelihood ratio test implemented in R with the *anova* function was used to compare the goodness of fit of the joint MLR models with and without each individual PGS and p-values were used to determine significance of the variability in the mean PGS across groups ( $p_{\text{global}}$ ). Significance was declared after Bonferroni correction for 127 p-values ( $p < 0.05/127 = 0.0004$ ).

### *Extension cohort: The adolescent brain cognitive development (ABCD) Study*

The Adolescent Brain and Cognitive Development (ABCD) study<sup>48,103</sup> (<http://abcdstudy.org>) is an observational study following 11 875 children from across the US, starting at age nine or ten with limited exclusion criteria to create a socio-economically and demographically representative cohort. ABCD administers biennial assessments of physical health, mental health, neuro cognition, family, cultural and environmental variables, SUD, genetic and other biomarkers, and multi-modal neuroimaging. Genotyping used the Smokescreen array, imputations used the Michigan Imputation Server, and quality control is described in Loughnan et al.<sup>104</sup>. We used measures of 51 assessments (**Supplementary table 19**) covering categorical and dimensional psychopathology, mania symptoms, prodromal psychosis, impulsivity, and behavioral activation and inhibition collected at enrollment<sup>48</sup>. To protect against confounding by population structure, we focus on the homogenous, European ancestry sub-cohort of 5 455 children selected following genetic ancestry estimation using fastStructure<sup>105</sup>.

### *Polygenic Scores in ABCD*

PGS in ABCD were calculated with PRSice<sup>106</sup> and summary statistics from each of the three ADHD case variable GWAS described above. SNPs were pruned and clumped (clumping  $r^2 \leq 0.1$ , distance  $\leq 250$  kb), but no threshold on p-value was set (i.e., all independent SNPs,  $p < 1$ , were included). SNPs within the major histocompatibility complex were removed. Generalized linear models (GLMs) were used to test the association of each of the three PGS across each of 51 assessments, separately. GLMs included fixed effects of ADHD case-control PGS, age, sex at birth, data collection site, and 10 ancestry principal components. To control for family effects (twins and siblings within the sample), we iteratively fit 100 models for each behavior taking a random selection on singletons (down sampling to one individual from each family,  $n = 4\,622$ ) and report the median across iterations. The distribution of each of the baseline assessment (response variable) was analyzed to ensure normality or to handle zero inflation and right skewness. Non-zero inflated distributions were rank normalized (as was performed for PGS) and the GLM was fitted using the family=gaussian option. For the right-skewed, zero-inflated distributions, data were shifted to ensure non-negativity and GLM were fitted with a gamma distribution and a log link function. Summary measures making up the Kiddie Schedule Disorders and for Schizophrenia (KSADS), except for

the two KSADS Total Symptoms measures, were binarized using a median split and fit using logistic regression. This decision to binarize was due to model convergence issues resulting from very heavy distribution skews when attempting to fit using a gamma distribution. Bonferroni adjustment was used to declare significance ( $p < 0.05/51 = 0.001$ ) and nominally significant tests were noted suggestive ( $p < 0.05$ ).

## Acknowledgments

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (<https://abcdstudy.org>), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, and U24DA041147. A full list of supporters is available at <https://abcdstudy.org/nih-collaborators>. A listing of participating sites and a complete listing of the study investigators can be found at <https://abcdstudy.org/principal-investigators.html>. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The iPSYCH Initiative is funded by the Lundbeck Foundation (grant numbers R102-A9118 and R155-2014-1724), the Mental Health Services Capital Region of Denmark, University of Copenhagen, Aarhus University and the university hospital in Aarhus. Genotyping of iPSYCH samples was supported by grants from the Lundbeck Foundation, the Stanley Foundation, the Simons Foundation (SFARI 311789), and NIMH (5U01MH094432-02). The IPSYCH Initiative utilize the Danish National Biobank resource that is supported by the Novo Nordisk Foundation. IPSYCH data was stored and analyzed at the Computerome HPC facility ([http:// www.computerome.dtu.dk/](http://www.computerome.dtu.dk/)) and authors are grateful for continuous support from the HPC team led by A. Syed of DTU Bioinformatics, Technical University of Denmark. AJS acknowledges support from Lundbeckfonden under the Fellowship R335-2019-2318 and the National Institute for Aging of the National Institutes of Health under awards U19AG023122, U24AG051129S1, UH2AG064706, and UH2AG064706S1. SLB acknowledges support from the Research Fund of the Mental Health Services – Capital Region of Denmark R4A92, The Lundbeck Foundation R208-2015-3951 and Fonden for Faglig Udvikling af Speciallægepraksis 38850/16. SD acknowledges research support from the European Commission (Horizon 2020, grant no 667302), Helsefonden (grant no 19-8-0260) and the European Union’s Horizon 2020 research and innovation programme under grant agreement No 847879.

## Author Contributions

The aims of this paper were conceived of jointly by S.L., I.B., S.D., T.M.W., and A.J.S and supervised broadly by S.D., T.M.W., A.J.S. The overall study design was developed by S.L., I.B., and A.J.S, with various components of the paper conducted with input and guidance from collaborators. Variable extraction from and definition in registers was conducted by S.L., I.B., and D.H., with assistance and guidance from E.A., M.G.P, and S.D. SNP heritability, genetic correlations, GWAS, and PGS generation were conducted by S.L. and A.J.S, with assistance from J.G, V.A., M.V., A.I., supervised by A.J.S. Functional annotations were conducted by S.L., with assistance, design, and supervision from R.W., D.G., and A.J.S. Single locus and polygenic multinomial tests were conducted by S.L., J.M., and A.J.S., using a statistical implementation from A.D. and N.Z with supervision from A.D., N.Z., and A.J.S. Analysis of ABCD was designed and conducted by R.L. and C.P., supervised by T.J. Simulations were conducted by A.J.S., with support from S.L., M.K., and K.K. A.D.B., D.M.H., O.M., M.N., P.B.M, T.M.W. contributed iPSYCH data. T.J. contributed ABCD data. The initial draft was written by S.L., with subsequent versions written by S.L., I.B., and A.J.S. All authors discussed results, commented on drafts, and provided critical feedback throughout.

## References

- 1 American Psychiatric Association. & American Psychiatric Association. DSM-5 Task Force. 1  
online resource (xlv, 947 pages).
- 2 World Health Organization. *ICD-10, the ICD-10 classification of mental and behavioural disorders :  
clinical descriptions and diagnostic guidelines*. (World Health Organization, 1992).
- 3 Franke, B. *et al*. Live fast, die young? A review on the developmental trajectories of ADHD across  
the lifespan. *Eur Neuropsychopharmacol* **28**, 1059-1088, doi:10.1016/j.euroneuro.2018.08.001  
(2018).
- 4 Luo, Y., Weibman, D., Halperin, J. M. & Li, X. A Review of Heterogeneity in Attention  
Deficit/Hyperactivity Disorder (ADHD). *Front Hum Neurosci* **13**, 42,  
doi:10.3389/fnhum.2019.00042 (2019).
- 5 Thapar, A., Cooper, M. & Rutter, M. Neurodevelopmental disorders. *Lancet Psychiatry* **4**, 339-346,  
doi:10.1016/S2215-0366(16)30376-5 (2017).
- 6 Dalsgaard, S., Mortensen, P. B., Frydenberg, M. & Thomsen, P. H. Long-term criminal outcome of  
children with attention deficit hyperactivity disorder. *Crim Behav Ment Health* **23**, 86-98,  
doi:10.1002/cbm.1860 (2013).
- 7 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).
- 8 Dalsgaard, S. *et al*. Association of Mental Disorder in Childhood and Adolescence With Subsequent  
Educational Achievement. *JAMA Psychiatry* **77**, 797-805, doi:10.1001/jamapsychiatry.2020.0217  
(2020).
- 9 Daley, D. *et al*. *Costing adult attention deficit hyperactivity disorder : impact on the individual and  
society*. First edition. edn, (Oxford University Press, 2015).
- 10 Plana-Ripoll, O. *et al*. Exploring Comorbidity Within Mental Disorders Among a Danish National  
Population. *JAMA Psychiatry* **76**, 259-270, doi:10.1001/jamapsychiatry.2018.3658 (2019).
- 11 McClellan, J. & King, M. C. Genetic heterogeneity in human disease. *Cell* **141**, 210-217,  
doi:10.1016/j.cell.2010.03.032 (2010).
- 12 Dahl, A. & Zaitlen, N. Genetic Influences on Disease Subtypes. *Annu Rev Genomics Hum Genet* **21**,  
413-435, doi:10.1146/annurev-genom-120319-095026 (2020).
- 13 Charney, A. W. *et al*. Evidence for genetic heterogeneity between clinical subtypes of bipolar  
disorder. *Transl Psychiatry* **7**, e993, doi:10.1038/tp.2016.242 (2017).
- 14 Bipolar, D., Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic  
address, d. r. v. e., Bipolar, D. & Schizophrenia Working Group of the Psychiatric Genomics, C.  
Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell* **173**,  
1705-1715 e1716, doi:10.1016/j.cell.2018.05.046 (2018).
- 15 Faraone, S. V. & Larsson, H. Genetics of attention deficit hyperactivity disorder. *Mol Psychiatry* **24**,  
562-575, doi:10.1038/s41380-018-0070-0 (2019).
- 16 Williams, N. M. *et al*. Rare chromosomal deletions and duplications in attention-deficit hyperactivity  
disorder: a genome-wide analysis. *Lancet* **376**, 1401-1408, doi:10.1016/S0140-6736(10)61109-9  
(2010).
- 17 Olsen, L. *et al*. Prevalence of rearrangements in the 22q11.2 region and population-based risk of  
neuropsychiatric and developmental disorders in a Danish population: a case-cohort study. *Lancet  
Psychiatry* **5**, 573-580, doi:10.1016/S2215-0366(18)30168-8 (2018).
- 18 Satterstrom, F. K. *et al*. Autism spectrum disorder and attention deficit hyperactivity disorder have a  
similar burden of rare protein-truncating variants. *Nat Neurosci* **22**, 1961-1965, doi:10.1038/s41593-  
019-0527-8 (2019).
- 19 Demontis, D. *et al*. Discovery of the first genome-wide significant risk loci for attention  
deficit/hyperactivity disorder. *Nat Genet* **51**, 63-75, doi:10.1038/s41588-018-0269-7 (2019).



- 20 Schork, A. J. *et al.* A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment. *Nat Neurosci* **22**, 353-361, doi:10.1038/s41593-018-0320-0 (2019).
- 21 Sullivan, P. F. *et al.* Psychiatric Genomics: An Update and an Agenda. *Am J Psychiatry* **175**, 15-27, doi:10.1176/appi.ajp.2017.17030283 (2018).
- 22 Sullivan, P. F. & Geschwind, D. H. Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell* **177**, 162-183, doi:10.1016/j.cell.2019.01.015 (2019).
- 23 Cai, N. *et al.* Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nat Genet* **52**, 437-447, doi:10.1038/s41588-020-0594-5 (2020).
- 24 Wray, N. R., Lee, S. H. & Kendler, K. S. Impact of diagnostic misclassification on estimation of genetic correlations using genome-wide genotypes. *Eur J Hum Genet* **20**, 668-674, doi:10.1038/ejhg.2011.257 (2012).
- 25 Kendler, K. S., Chatzinakos, C. & Bacanu, S. A. The impact on estimations of genetic correlations by the use of super-normal, unscreened, and family-history screened controls in genome wide case-control studies. *Genet Epidemiol* **44**, 283-289, doi:10.1002/gepi.22281 (2020).
- 26 Wimberley, T. *et al.* Genetic liability to ADHD and substance use disorders in individuals with ADHD. *Addiction* **115**, 1368-1377, doi:10.1111/add.14910 (2020).
- 27 Jansen, A. G. *et al.* Psychiatric Polygenic Risk Scores as Predictor for Attention Deficit/Hyperactivity Disorder and Autism Spectrum Disorder in a Clinical Child and Adolescent Sample. *Behav Genet* **50**, 203-212, doi:10.1007/s10519-019-09965-8 (2020).
- 28 Martin, J. *et al.* A Genetic Investigation of Sex Bias in the Prevalence of Attention-Deficit/Hyperactivity Disorder. *Biol Psychiatry* **83**, 1044-1053, doi:10.1016/j.biopsych.2017.11.026 (2018).
- 29 Rovira, P. *et al.* Shared genetic background between children and adults with attention deficit/hyperactivity disorder. *Neuropsychopharmacology* **45**, 1617-1626, doi:10.1038/s41386-020-0664-5 (2020).
- 30 Liley, J., Todd, J. A. & Wallace, C. A method for identifying genetic heterogeneity within phenotypically defined disease subgroups. *Nat Genet* **49**, 310-316, doi:10.1038/ng.3751 (2017).
- 31 Nadeau, J. H. Modifier genes in mice and humans. *Nat Rev Genet* **2**, 165-174, doi:10.1038/35056009 (2001).
- 32 Fanous, A. H. & Kendler, K. S. Genetic heterogeneity, modifier genes, and quantitative phenotypes in psychiatric illness: searching for a framework. *Mol Psychiatry* **10**, 6-13, doi:10.1038/sj.mp.4001571 (2005).
- 33 Pedersen, C. B. *et al.* The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol Psychiatry*, doi:10.1038/mp.2017.196 (2017).
- 34 Kildemoes, H. W., Sorensen, H. T. & Hallas, J. The Danish National Prescription Registry. *Scand J Public Health* **39**, 38-41, doi:10.1177/1403494810394717 (2011).
- 35 Lynge, E., Sandegaard, J. L. & Rebolj, M. The Danish National Patient Register. *Scand J Public Health* **39**, 30-33, doi:10.1177/1403494811401482 (2011).
- 36 Mors, O., Perto, G. P. & Mortensen, P. B. The Danish Psychiatric Central Research Register. *Scand J Public Health* **39**, 54-57, doi:10.1177/1403494810395825 (2011).
- 37 Pedersen, C. B. The Danish Civil Registration System. *Scand J Public Health* **39**, 22-25, doi:10.1177/1403494810387965 (2011).
- 38 Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* **51**, 431-444, doi:10.1038/s41588-019-0344-8 (2019).
- 39 Cross-Disorder Group of the Psychiatric Genomics Consortium. Electronic address, p. m. h. e. & Cross-Disorder Group of the Psychiatric Genomics, C. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469-1482 e1411, doi:10.1016/j.cell.2019.11.020 (2019).
- 40 Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, doi:10.1126/science.aat8464 (2018).

- 41 Walker, R. L. *et al.* Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* **179**, 750-771 e722, doi:10.1016/j.cell.2019.09.021 (2019).
- 42 Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585, doi:10.1038/ng.2653 (2013).
- 43 Mittelstaedt, T. & Schoch, S. Structure and evolution of RIM-BP genes: identification of a novel family member. *Gene* **403**, 70-79, doi:10.1016/j.gene.2007.08.004 (2007).
- 44 Hibino, H. *et al.* RIM binding proteins (RBPs) couple Rab3-interacting molecules (RIMs) to voltage-gated Ca(2+) channels. *Neuron* **34**, 411-423, doi:10.1016/s0896-6273(02)00667-0 (2002).
- 45 Acuna, C., Liu, X., Gonzalez, A. & Sudhof, T. C. RIM-BPs Mediate Tight Coupling of Action Potentials to Ca(2+)-Triggered Neurotransmitter Release. *Neuron* **87**, 1234-1247, doi:10.1016/j.neuron.2015.08.027 (2015).
- 46 Bucan, M. *et al.* Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. *PLoS Genet* **5**, e1000536, doi:10.1371/journal.pgen.1000536 (2009).
- 47 Neale, B. M. *et al.* Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* **49**, 884-897, doi:10.1016/j.jaac.2010.06.008 (2010).
- 48 Barch, D. M. *et al.* Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. *Dev Cogn Neurosci* **32**, 55-66, doi:10.1016/j.dcn.2017.10.010 (2018).
- 49 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. *Eur Child Adolesc Psychiatry* **19**, 281-295, doi:10.1007/s00787-010-0092-x (2010).
- 50 Ghirardi, L. *et al.* The familial co-aggregation of ASD and ADHD: a register-based cohort study. *Mol Psychiatry* **23**, 257-262, doi:10.1038/mp.2017.17 (2018).
- 51 Martin, J. *et al.* Biological overlap of attention-deficit/hyperactivity disorder and autism spectrum disorder: evidence from copy number variants. *J Am Acad Child Adolesc Psychiatry* **53**, 761-770 e726, doi:10.1016/j.jaac.2014.03.004 (2014).
- 52 LaBianca, S. *et al.* Brief Report: Clusters and Trajectories Across the Autism and/or ADHD Spectrum. *J Autism Dev Disord* **48**, 3629-3636, doi:10.1007/s10803-018-3618-6 (2018).
- 53 LaBianca, S. *et al.* Copy Number Variants and Polygenic Risk Scores Predict Need of Care in Autism and/or ADHD Families. *J Autism Dev Disord* **51**, 276-285, doi:10.1007/s10803-020-04552-x (2021).
- 54 Young, S. *et al.* Guidance for identification and treatment of individuals with attention deficit/hyperactivity disorder and autism spectrum disorder based upon expert consensus. *BMC Med* **18**, 146, doi:10.1186/s12916-020-01585-y (2020).
- 55 Pinto, R., Rijdsdijk, F., Ronald, A., Asherson, P. & Kuntsi, J. The Genetic Overlap of Attention-Deficit/Hyperactivity Disorder and Autistic-like Traits: an Investigation of Individual Symptom Scales and Cognitive markers. *J Abnorm Child Psychol* **44**, 335-345, doi:10.1007/s10802-015-0037-4 (2016).
- 56 Panagiotidi, M., Overton, P. G. & Stafford, T. Co-Occurrence of ASD and ADHD Traits in an Adult Population. *J Atten Disord* **23**, 1407-1415, doi:10.1177/1087054717720720 (2019).
- 57 Aoki, Y. *et al.* Association of White Matter Structure With Autism Spectrum Disorder and Attention-Deficit/Hyperactivity Disorder. *JAMA Psychiatry* **74**, 1120-1128, doi:10.1001/jamapsychiatry.2017.2573 (2017).
- 58 Laursen, T. M., Agerbo, E. & Pedersen, C. B. Bipolar disorder, schizoaffective disorder, and schizophrenia overlap: a new comorbidity index. *J Clin Psychiatry* **70**, 1432-1438, doi:10.4088/JCP.08m04807 (2009).
- 59 Asherson, P. & Agnew-Blais, J. Annual Research Review: Does late-onset attention-deficit/hyperactivity disorder exist? *J Child Psychol Psychiatry* **60**, 333-352, doi:10.1111/jcpp.13020 (2019).
- 60 Faraone, S. V. & Biederman, J. Can Attention-Deficit/Hyperactivity Disorder Onset Occur in Adulthood? *JAMA Psychiatry* **73**, 655-656, doi:10.1001/jamapsychiatry.2016.0400 (2016).

- 61 Zaitlen, N. *et al.* Informed conditioning on clinical covariates increases power in case-control association studies. *PLoS Genet* **8**, e1003032, doi:10.1371/journal.pgen.1003032 (2012).
- 62 Yap, C. X. *et al.* Misestimation of heritability and prediction accuracy of male-pattern baldness. *Nat Commun* **9**, 2537, doi:10.1038/s41467-018-04807-3 (2018).
- 63 van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H. & Wray, N. R. Genetic correlations of polygenic disease traits: from theory to practice. *Nat Rev Genet* **20**, 567-581, doi:10.1038/s41576-019-0137-z (2019).
- 64 Wray, N. R. *et al.* From Basic Science to Clinical Application of Polygenic Risk Scores: A Primer. *JAMA Psychiatry* **78**, 101-109, doi:10.1001/jamapsychiatry.2020.3049 (2021).
- 65 Lewis, C. M. & Vassos, E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med* **12**, 44, doi:10.1186/s13073-020-00742-5 (2020).
- 66 Norgaard-Pedersen, B. & Hougaard, D. M. Storage policies and use of the Danish Newborn Screening Biobank. *J Inherit Metab Dis* **30**, 530-536, doi:10.1007/s10545-007-0631-x (2007).
- 67 O'Connell, J. *et al.* Haplotype estimation for biobank-scale data sets. *Nat Genet* **48**, 817-820, doi:10.1038/ng.3583 (2016).
- 68 The 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).
- 69 Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529, doi:10.1371/journal.pgen.1000529 (2009).
- 70 Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet* **2**, e190, doi:10.1371/journal.pgen.0020190 (2006).
- 71 Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-2873, doi:10.1093/bioinformatics/btq559 (2010).
- 72 polycor: Polychoric and Polyserial Correlations v. 0.7-10 (The Comprehensive R Archive Network (CRAN), 2019).
- 73 Savalei, V. What to Do About Zero Frequency Cells When Estimating Polychoric Correlations. *Structural Equation Modeling: A Multidisciplinary Journal* **18**, 253-273, doi:<https://doi.org/10.1080/10705511.2011.557339> (2011).
- 74 Galili, T., O'Callaghan, A., Sidi, J. & Sievert, C. heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics* **34**, 1600-1602, doi:10.1093/bioinformatics/btx657 (2018).
- 75 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82, doi:10.1016/j.ajhg.2010.11.011 (2011).
- 76 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575, doi:10.1086/519795 (2007).
- 77 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295, doi:10.1038/ng.3211 (2015).
- 78 Morris, A. P. *et al.* A powerful approach to sub-phenotype analysis in population-based genetic association studies. *Genet Epidemiol* **34**, 335-343, doi:10.1002/gepi.20486 (2010).
- 79 nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models v. 7.3-16 (The Comprehensive R Archive Network (CRAN), 2021).
- 80 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164, doi:10.1093/nar/gkq603 (2010).
- 81 Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826, doi:10.1038/s41467-017-01261-5 (2017).
- 82 Gusev, A. *et al.* Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nat Genet* **50**, 538-548, doi:10.1038/s41588-018-0092-1 (2018).
- 83 Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-252, doi:10.1038/ng.3506 (2016).
- 84 Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* **51**, 1339-1348, doi:10.1038/s41588-019-0481-0 (2019).

85 Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific  
haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**,  
3555-3557, doi:10.1093/bioinformatics/btv402 (2015).

86 de la Torre-Ubieta, L. *et al.* The Dynamic Landscape of Open Chromatin during Human Cortical  
Neurogenesis. *Cell* **172**, 289-304 e218, doi:10.1016/j.cell.2017.12.014 (2018).

87 Bryois, J. *et al.* Evaluation of chromatin accessibility in prefrontal cortex of individuals with  
schizophrenia. *Nat Commun* **9**, 3121, doi:10.1038/s41467-018-05379-y (2018).

88 Finucane, H. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant  
tissues and cell types. *bioRxiv* (2017).

89 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).

90 Vilhjalmsón, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk  
Scores. *Am J Hum Genet* **97**, 576-592, doi:10.1016/j.ajhg.2015.09.001 (2015).

91 Autism Spectrum Disorders Working Group of The Psychiatric Genomics, C. Meta-analysis of  
GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at  
10q24.32 and a significant overlap with schizophrenia. *Mol Autism* **8**, 21, doi:10.1186/s13229-017-  
0137-9 (2017).

92 Sniekers, S. *et al.* Genome-wide association meta-analysis of 78,308 individuals identifies new loci  
and genes influencing human intelligence. *Nat Genet* **49**, 1107-1112, doi:10.1038/ng.3869 (2017).

93 Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational  
attainment. *Nature* **533**, 539-542, doi:10.1038/nature17671 (2016).

94 Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and  
neuroticism identified through genome-wide analyses. *Nat Genet* **48**, 624-633, doi:10.1038/ng.3552  
(2016).

95 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).

96 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).

97 Barban, N. *et al.* Genome-wide analysis identifies 12 loci influencing human reproductive behavior.  
*Nat Genet* **48**, 1462-1472, doi:10.1038/ng.3698 (2016).

98 Erzurumluoglu, A. M. *et al.* Meta-analysis of up to 622,409 individuals identifies 40 novel smoking  
behaviour associated genetic loci. *Mol Psychiatry* **25**, 2392-2409, doi:10.1038/s41380-018-0313-0  
(2020).

99 Major Depressive Disorder Working Group of the Psychiatric, G. C. *et al.* A mega-analysis of  
genome-wide association studies for major depressive disorder. *Mol Psychiatry* **18**, 497-511,  
doi:10.1038/mp.2012.21 (2013).

100 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).

101 Benyamin, B. *et al.* Childhood intelligence is heritable, highly polygenic and associated with  
FBNP1L. *Mol Psychiatry* **19**, 253-258, doi:10.1038/mp.2012.184 (2014).

102 Day, F. R. *et al.* Large-scale genomic analyses link reproductive aging to hypothalamic signaling,  
breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat Genet* **47**, 1294-1303,  
doi:10.1038/ng.3412 (2015).

103 Volkow, N. D. *et al.* The conception of the ABCD study: From substance use to a broad NIH  
collaboration. *Dev Cogn Neurosci* **32**, 4-7, doi:10.1016/j.dcn.2017.10.002 (2018).

104 Loughnan, R. J. *et al.* Gene-expression correlation during cognitive development: Evidence from the  
Adolescent Brain Cognitive Development (ABCD) Study. *bioRxiv*, doi:10.1101/637512v3 (2021).

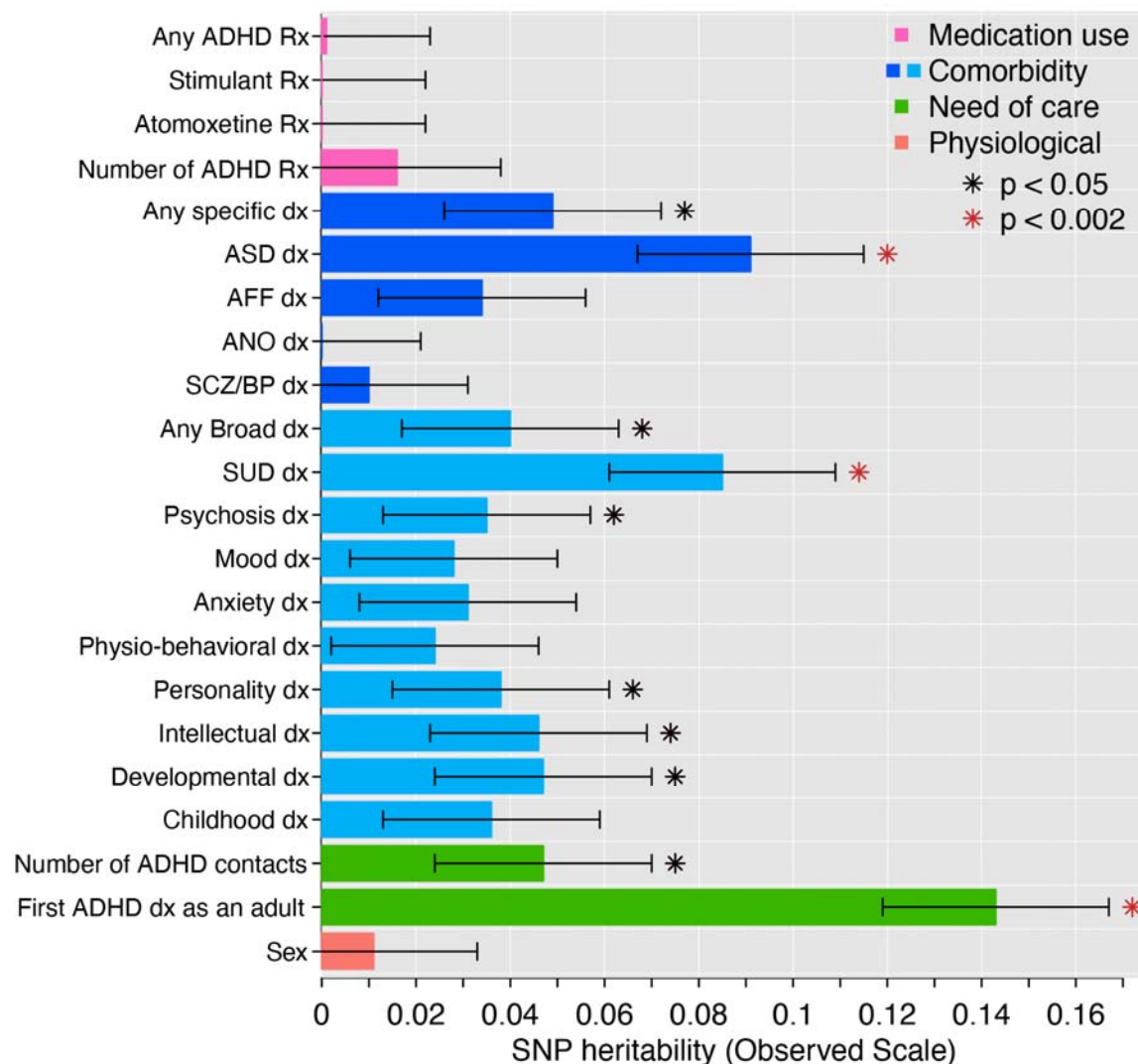
105 Raj, A., Stephens, M. & Pritchard, J. K. fastSTRUCTURE: variational inference of population  
structure in large SNP data sets. *Genetics* **197**, 573-589, doi:10.1534/genetics.114.164350 (2014).

106 Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics*  
**31**, 1466-1468, doi:10.1093/bioinformatics/btu848 (2015).



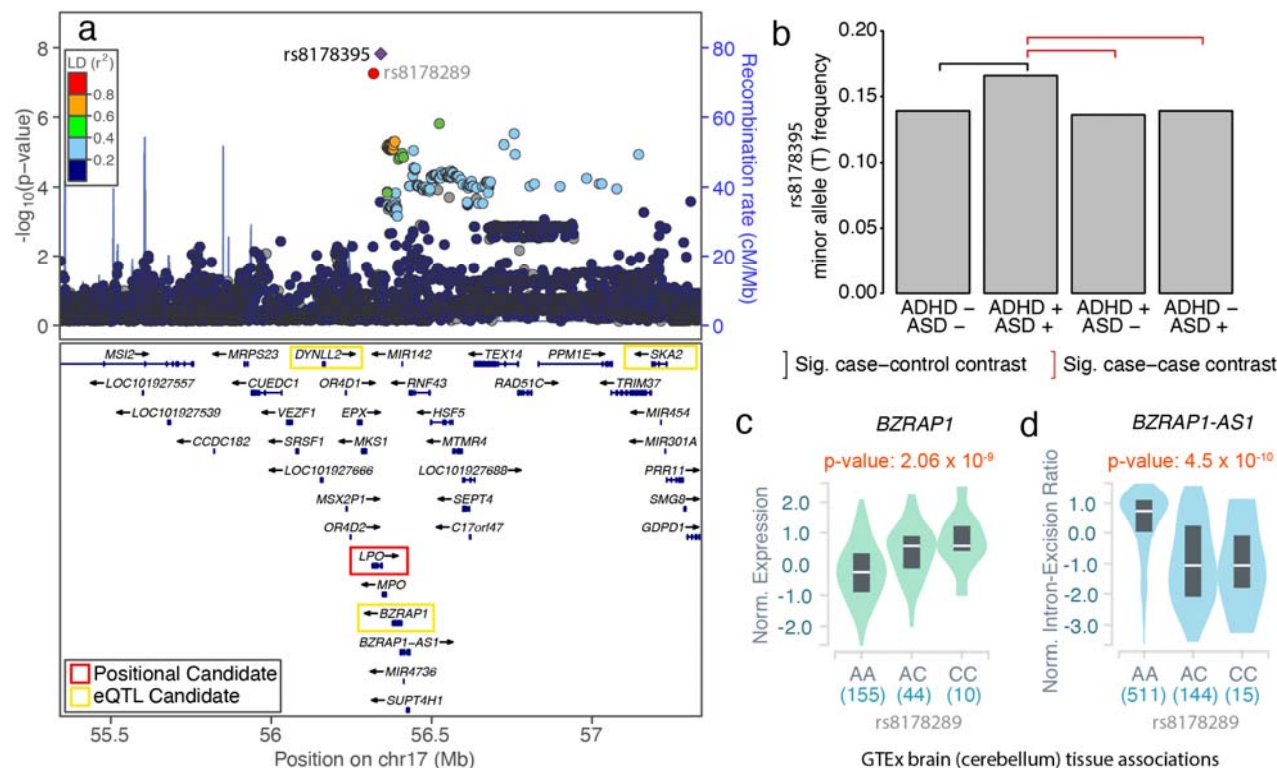
Category	Trait	Description	Reference Group (0)	N <sub>0</sub>	Outcome Group (1)	N <sub>1</sub>
Medication	Any ADHD Rx	Recorded prescription for any ADHD medication	never	1756	ever	12328
	Stimulant Rx	Recorded prescription for stimulant medication	never	2167	ever	11917
	Atomoxetine Rx	Recorded prescription for atomoxetine	never	9948	ever	4136
	Number of ADHD Rx	Number of recorded prescriptions for any ADHD medication	≤ 29	7052	> 29	7032
Adjacent diagnosis (Specific)	Any specific dx	Any recorded diagnosis for an iPSYCH ascertained, specific disorder	never	10478	ever	3606
	ASD dx	Autism spectrum disorder (ICD-10: F84.0, F84.1, F84.5, F84.8 F84.9)	never	11800	ever	2284
	AFF dx	Affective disorder (ICD-10: F30-39)	never	12796	ever	1288
	ANO dx	Anorexia (ICD-10: F50.0)	never	14020	ever	64
Adjacent diagnosis (Broad)	SCZ/BP dx	Schizophrenia or bipolar disorder (ICD-10: F20, F30-31)	never	13769	ever	315
	Any broad dx	Any recorded diagnosis for a mental, behavioral, and/or neurodevelopmental disorders according to ICD-10, chapter F	never	3171	ever	10913
	Substance use disorder (SUD) dx	Mental and behavioral disorders due to psychoactive substance use (ICD-10: F10-F19)	never	11457	ever	2627
	Psychosis dx	Schizophrenia, schizotypal, delusional, and other non-mood psychotic disorders (ICD-10: F20-F29)	never	13133	ever	951
	Mood dx	Mood/affective disorders (ICD-10: F30-F39)	never	12250	ever	1843
	Anxiety dx	Anxiety, dissociative, stress-related, somatoform and other nonpsychotic mental disorders (ICD-10: F40-F49)	never	10373	ever	3711
	Physio-behavioral dx	Behavioral syndromes from physiological disturbances and physical factors (ICD-10: F50-F59)	never	13578	ever	506
	Personality dx	Disorders of adult personality and behavior (ICD-10: F60-F69)	never	12391	ever	1693
	Intellectual dx	Intellectual disabilities (ICD-10: F70-F79)	never	12592	ever	1492
	Developmental dx	Pervasive and specific developmental disorders (ICD-10: F80-F89)	never	8713	ever	5371
	Childhood dx	Behavioral and emotional disorders with childhood and adolescent onset (ICD-10: F90-F99). Excludes F90.0.	never	10264	ever	3820
	Need of Care	Number of recorded hospital contacts with an ADHD diagnosis (ICD-10: F90.0)	≤ 2	9732	> 2	4352
Sex	First ADHD dx as an adult	Age at first recorded ADHD diagnosis (ICD-10: F90.0)	< 18 years	10761	≥ 18 years	3323
	Sex	Biological sex recorded at birth	female	3668	male	10416

**Table 1. 14,084 individuals diagnosed with ADHD vary on a number of clinically relevant adjacent traits.** Dx, diagnosis; Rx, prescription; ICD-10, international classification of disease, 10<sup>th</sup> revision; ADHD, attention-deficit hyperactivity disorder.

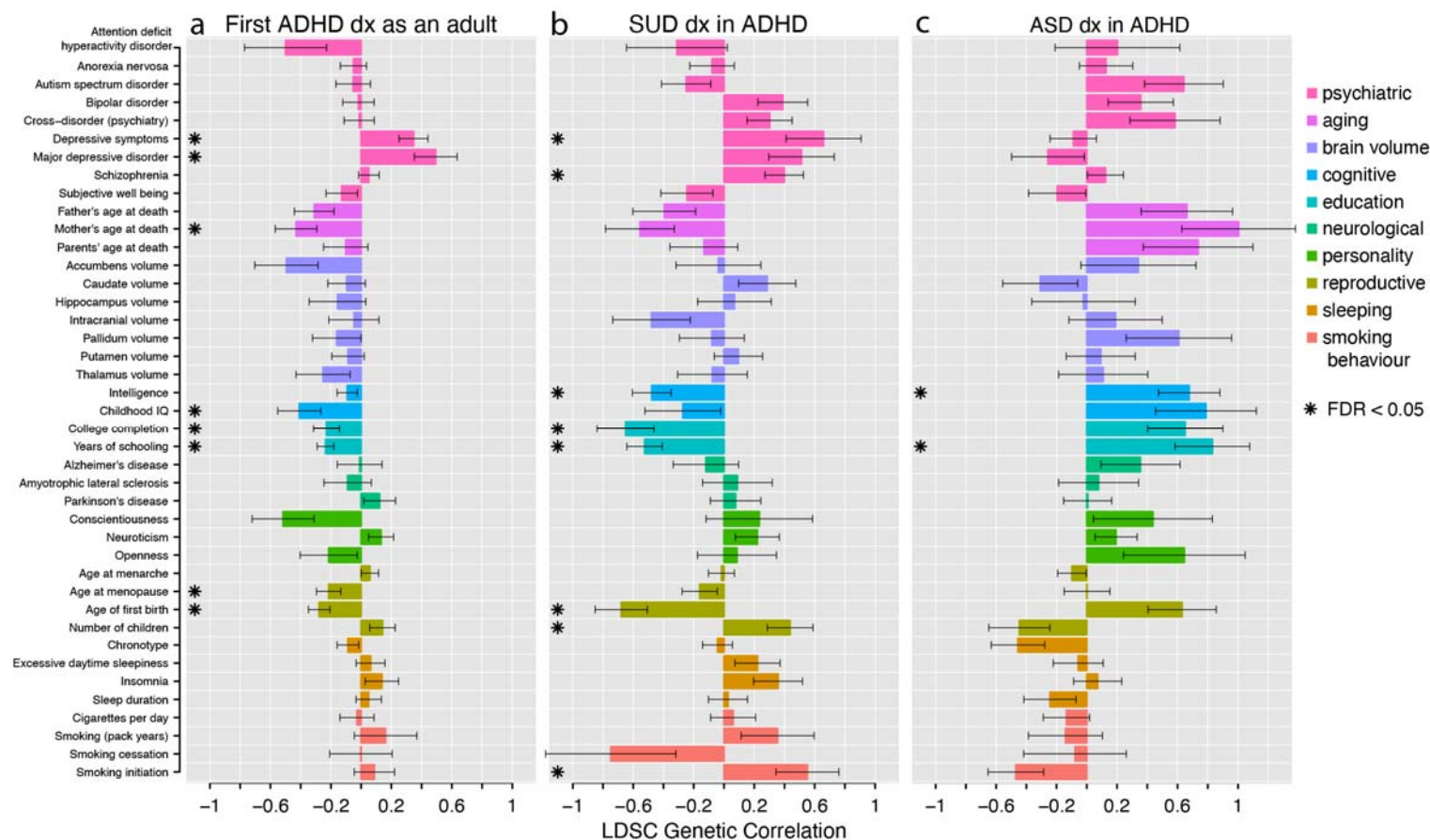


**Fig. 1 | ADHD-adjacent traits associate with genetic variability among diagnosed individuals.** The same 14,084 individuals diagnosed with ADHD were repeatedly partitioned into 22 groups on the basis of ADHD-adjacent traits and the SNP-heritability of each trait was estimated with GCTA. Full statistical results are available in Supplementary table 3. Significance (red star) is after Bonferroni correction for 22 tests. Rx, prescription; dx, diagnosis; ADHD, attention-deficit hyperactivity disorder; ASD, autism spectrum disorders; AFF, affective disorders; ANO, anorexia; SCZ/BP, schizophrenia or bipolar disorder; SUD, substance use disorder.

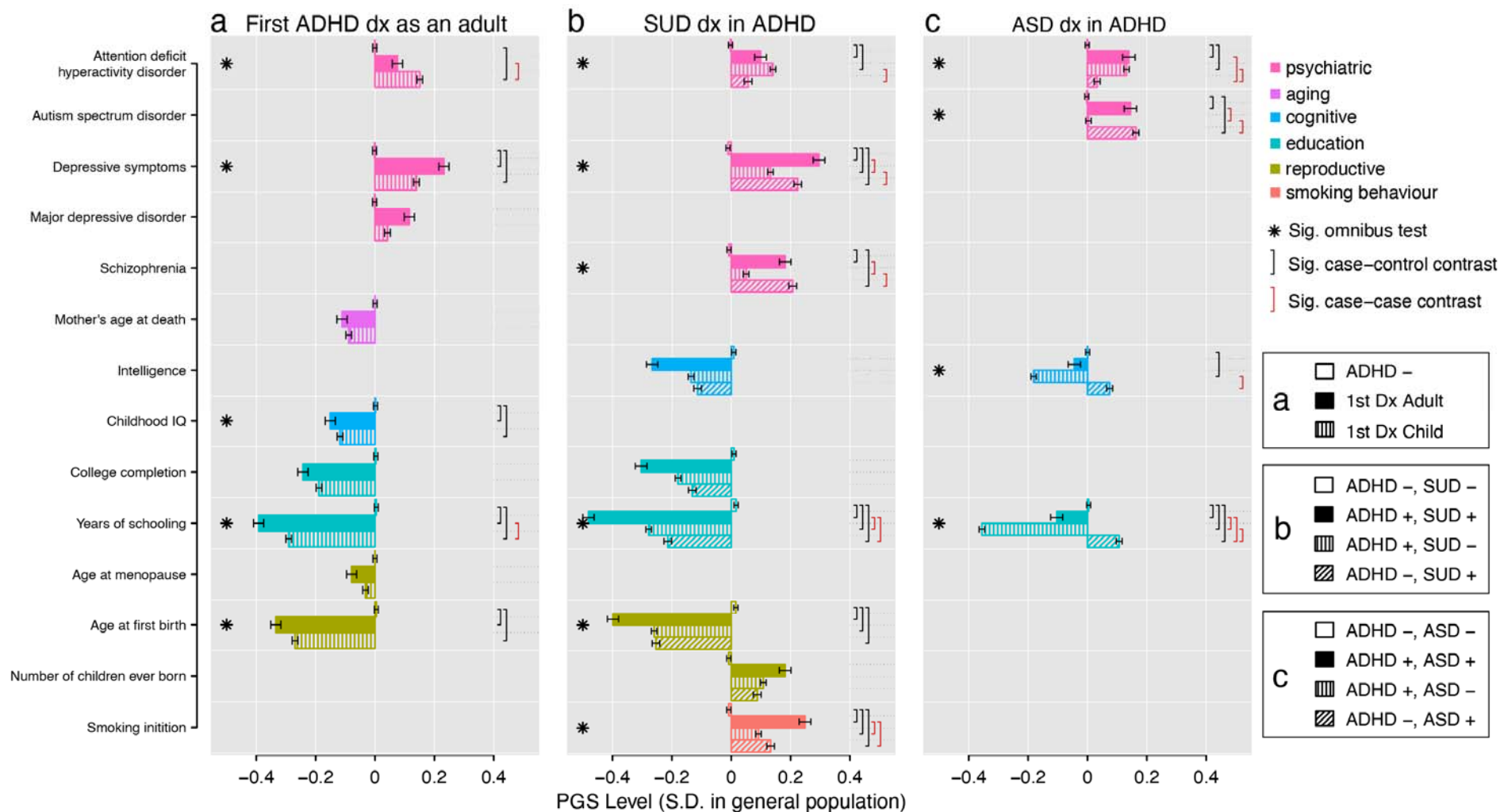




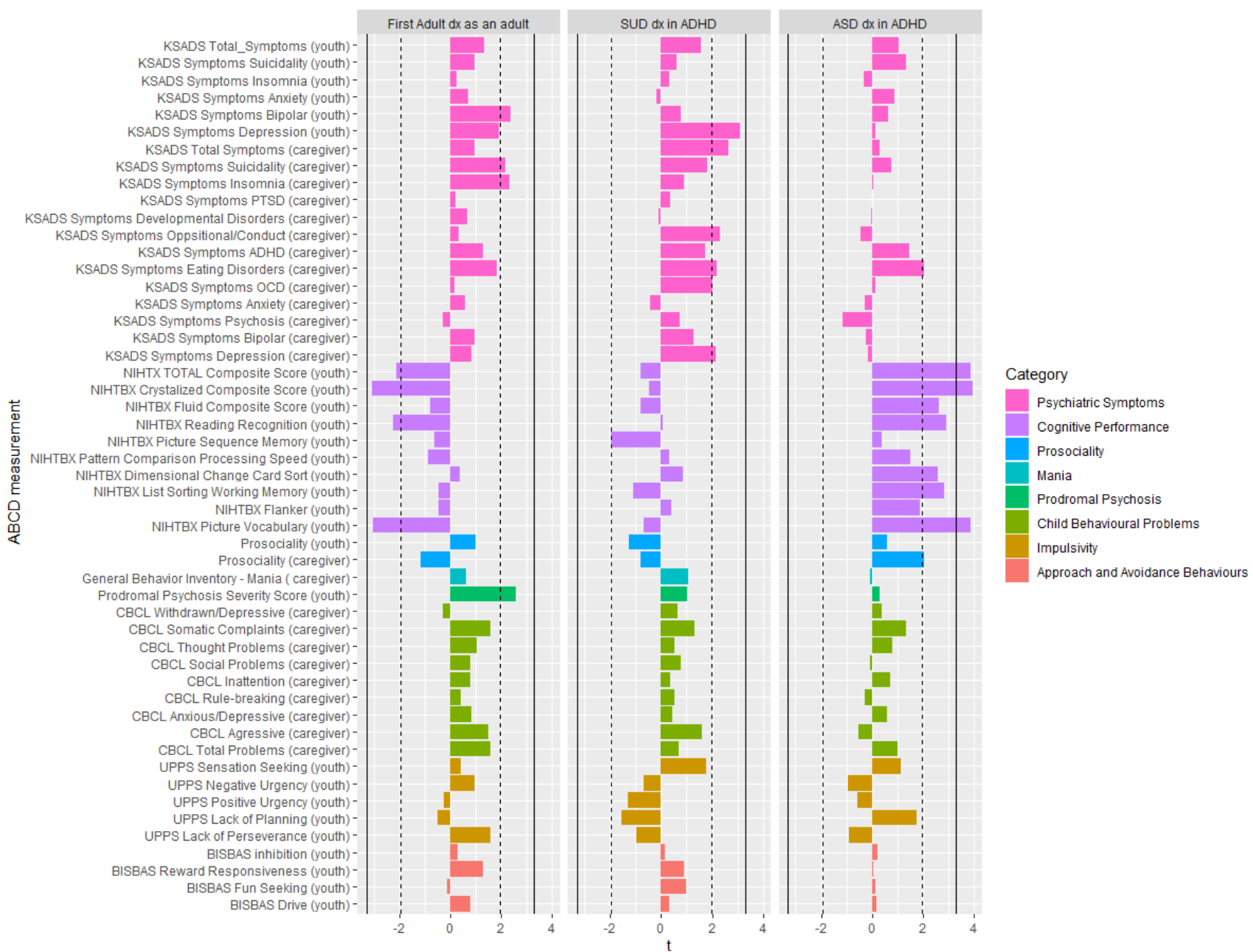
**Fig. 2 | rs8178395 is specifically associated with an ADHD-adjacent ASD diagnosis.** **a**, Locus zoom of lead SNP (rs8178395) identified in GWAS comparing individuals diagnosed with both ADHD and ASD to those only diagnosed with ADHD (iPSYCH2012\_ASD). *LPO*, *DYNLL2*, *BZRAP1*, and *SKA2* are prioritized as candidate genes according to different criteria. LD, linkage disequilibrium; cM, centimorgans; Mb, megabases; TAS, transcription association study; **b**, The minor allele (T) of rs8178395 is significantly increased in frequency in the group of individuals diagnosed with both ADHD and ASD (ADHD +, ASD +) relative to those diagnosed with neither ADHD, nor ASD (black bracket; ADHD -, ASD -:  $p=2.9 \times 10^{-7}$ ) and either, exclusively (red brackets; ADHD +, ASD -:  $p=2.4 \times 10^{-8}$ ; ADHD -, ASD +:  $p=1.0 \times 10^{-6}$ ). Significance is based on multinomial logistic regression. See Supplementary table 4 for additional details. Sig, Significant. **c**, The proxy SNP for rs8178395 (rs8178289,  $r^2$  LD: 0.9) is associated with *BZRAP1* expression in brain (and other) tissue(s) in GTEx. **d**, rs8178289 is also reported in GTEx as a member of a haplotype containing a splice QTL for a *BZRAP1* associated antisense RNA, *BZRAP1-AS1* in the same (and other) tissue(s).



**Fig. 3 | ADHD-adjacent traits share polygenes with psychiatric, cognitive, and socio-behavioral traits.** LD score regression was used to estimate the genetic correlations between ADHD-adjacent traits and a reference set of psychiatric, cognitive, and socio-behavioral traits. **a**, A first diagnosis of ADHD as an adult, **b**, an ADHD-adjacent SUD diagnosis, and **c**, an ADHD-adjacent ASD diagnosis show different patterns of genetic correlation with 41 reference traits. Prioritization determined by FDR < 0.05. Error bars denote standard error of estimates. See Supplementary table 10 for full statistical results. LDSC, LD score regression; FDR, false discovery rate; dx, diagnosis; ADHD, attention-deficit hyperactivity disorder; ASD, autism spectrum disorders; SUD, substance use disorder.



**Fig. 4 | Profiles of polygenic scores for psychiatric, cognitive, and socio-behavioral traits define aspects of heterogeneity in ADHD.** The mean level of polygenic scores (PGS) are displayed for ADHD subgroups, controls, and complementary disorder groups after centering and standardizing based on a population random sample. **a.** Individuals not diagnosed with ADHD (ADHD -), first diagnosed as an adult (1<sup>st</sup> Dx Adult), and first diagnosed as a child (1<sup>st</sup> Dx Child), **b.** individuals diagnosed with neither ADHD, nor SUD (ADHD-, SUD-), both ADHD and SUD (ADHD+, SUD+), ADHD but not SUD (ADHD+, SUD-), and SUD but not ADHD (ADHD-, SUD+), and **c.** individuals diagnosed with neither ADHD, nor ASD (ADHD-, ASD-), both ADHD and ASD (ADHD+, ASD+), ADHD but not ASD (ADHD+, ASD-), and ASD but not ADHD (ADHD-, ASD+), vary significantly with respect to multivariate profiles of PGS. Stars denote a significant global test for differences in PGS level among all groups, while brackets denote significant pairwise group contrasts, accounting for all other PGS and 25 ancestry principle components in a joint multinomial logistic regression. Bonferroni adjustment for 127 contrasts. Errors bars denote standard error of mean. See Supplementary tables 11-13 for complete statistical results. ADHD, attention deficit hyperactivity disorder; ASD, autism spectrum disorder; SUD, substance use disorder; Dx, diagnosis; Sig., Significant.



**Fig. 5 | Polygenes for ADHD-adjacent ASD are associated with cognitive performance in an independent, typically developing cohort.** Polygenic scores (PGS) constructed using summary statistics from ADHD-adjacent trait GWAS were tested for association with cognitive, behavioral, and psychiatric traits in 5449 nine or ten year-old children from the Adolescent brain cognitive development (ABCD) Study. No significant associations were found for **a** PGS for adult diagnosed ADHD or **b** ADHD-adjacent SUD, although consistent trends are prevalent. **c** PGS for an ADHD-adjacent ASD diagnosis associated significantly with cognitive performance, replicating our previous finding that polygenes are shared among these two domains. Significance was after Bonferroni correction for 51 tests (solid line) and suggestive trends are noted for  $p < 0.05$  (dotted line). Caregiver or youth in parenthesis denotes the informant for the assessment of the child. t, t-statistic. See Supplementary table 17 for complete statistical results and supplementary table 19 for a description of the questionnaire used for each ABCD assessment.