Differences in the genetic architecture of common and rare variants in childhood, persistent and late-diagnosed attention-deficit hyperactivity disorder

Rajagopal, Veera M.

2022-08

Rajagopal , V M , Duan , J , Vilar-Ribo , L , Grove , J , Zayats , T , Antoni Ramos-Quiroga , J , Satterstrom , F K , Soler Artigas , M , Bybjerg-Grauholm , J , Baekvad-Hansen , M , Als , T D , Rosengren , A , Daly , M J , Neale , B M , Nordentoft , M , Werge , T , Mors , O , Hougaard , D M , Mortensen , P B , Ribases , M , Borglum , A D & Demontis , D 2022 , ' Differences in the genetic architecture of common and rare variants in childhood, persistent and late-diagnosed attention-deficit hyperactivity disorder ' , Nature Genetics , vol. 54 , no. 8 , pp. 1117-+ . https://doi.org/10.1038/s41588-022-01143-7

http://hdl.handle.net/10138/354126
https://doi.org/10.1038/s41588-022-01143-7

other
acceptedVersion

Differences in the genetic architecture of common and rare variants in childhood, persistent and late-diagnosed attention deficit hyperactivity disorder

Veera M. Rajagopal[1,2,3], Jinjie Duan[1,2,3], Laura Vilar-Ribó[4,5,6], Jakob Grove[1,2,3,7], Tetyana Zayats[8,9,10], J. Antoni Ramos-Quiroga[4,5,6,11], F. Kyle Satterstrom[8,9], María Soler Artigas[4,5,6,11], Jonas Bybjerg-Grauholm[2,12], Marie Bækvad-Hansen[2,13], Thomas D. Als[1,2,3], Anders Rosengren[2,14], Mark J. Daly[8,9,15,16], Benjamin M. Neale[8,9], Merete Nordentoft[2,17], Thomas Werge[2,14], Ole Mors[2,18], David M. Hougaard[2,13], Preben B. Mortensen[2,19,20], Marta Ribasés[4,5,6,21], Anders D. Børglum[1,2,3], Ditte Demontis[1,2,3]

[1]Department of Biomedicine (Human Genetics), Aarhus University, Aarhus, Denmark
[2]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark
[3]Center for Genomics and Personalized Medicine, Aarhus, Denmark
[4]Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain
[5]Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain
[6]Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain
[7]BiRC, Bioinformatics Research Centre, Aarhus University, Denmark
[8]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, 02114, USA
[9]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA
[10]PROMENTA, Department of Psychology, University of Oslo, Oslo, Norway
[11]Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain
[12]Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark
[13]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark
[14]Mental Health Centre Sct. Hans, Capital Region of Denmark, Institute of Biological Psychiatry, Copenhagen University Hospital, Copenhagen, Denmark
[15]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA
[16]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
[17]Mental Health Centre Copenhagen, Capital Region of Denmark, Copenhagen University Hospital, Copenhagen, Denmark
[18]Psychosis Research Unit, Aarhus University Hospital-Psychiatry, Denmark
[19]NCRR – National Centre for Register-Based Research, Business and Social Sciences, Aarhus University, Aarhus, Denmark
[20]Centre for Integrated Register-based Research, CIRRAU, Aarhus University, Aarhus, Denmark
[21]Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain

Corresponding author: Ditte Demontis, Department of Biomedicine, Aarhus University, Denmark, email: ditte@biomed.au.dk

Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder, with onset in childhood ('childhood ADHD'), two thirds continue to have ADHD in adulthood ('persistent ADHD'), and sometimes ADHD is diagnosed in adulthood ('late-diagnosed ADHD'). We evaluated genetic differences between childhood (N=14,878), persistent (N=1,473) and late-diagnosed ADHD (N=6,961), alongside 38,303 controls and rare variant differences in 7,650 ADHD cases and 8,649 controls. We identified four genome-wide significant loci for childhood ADHD and one for late-diagnosed ADHD. We show an increased polygenic score (PGS) for ADHD in persistent ADHD compared to the other two groups. Childhood ADHD had higher genetic overlap with hyperactivity and autism compared to late-diagnosed ADHD and the highest burden of rare protein truncating variants (rPTVs) in evolutionarily constrained genes. While, late-diagnosed ADHD had larger genetic overlap with depression than childhood ADHD and no increased burden in rPTVs. Overall, this suggest that genetics influence age at first ADHD diagnosis, persistence of ADHD and different comorbidity patterns in the groups.

INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is a childhood neurodevelopmental disorder characterised by age-inappropriate levels of hyperactivity, impulsivity and inattention. The disorder affects around 5%-6% of school-age children, and around 3% of adults[1,2]. It is a complex disorder with both environmental and genetic factors contributing to the risk. Genetics explains a large part of the aetiology, with an estimated twin heritability of 0.74[3], and the single nucleotide polymorphism (SNP) heritability (i.e. the contribution from common genetic variants) is substantial, explaining 22% of the phenotypic variance[4].

Around two thirds of children diagnosed with ADHD will continue to have symptoms in adulthood[5], which is referred to as 'persistent ADHD'. Persistent ADHD is associated with more

severe outcomes compared with the one third of individuals who do not have ADHD as adults (remitters) – for example, increased risk of substance use disorders[6,7], nicotine dependence[8] and comorbidity with other psychiatric disorders[9-11]. Several studies have reported a lower heritability for persistent ADHD than for childhood ADHD[12,13]; however, these findings have been questioned due to methodological differences of assessment between children and adults[12,14]. It has also been suggested that persistence of symptoms has a genetic risk component specific to persistence rather than baseline symptoms[15] and that a trajectory of persistent symptoms is associated with a high load of common ADHD risk variants[16].

According to the ICD10 diagnostic criteria, ADHD is a childhood-onset disorder, and the behavioural symptoms should be present prior to 7 years of age (prior to 12 years of age according to DSM-5 diagnostic criteria) with a duration of at least 6 months. However, the disorder is often diagnosed in adolescence and can also be diagnosed in adult life, which according to the current diagnostic criteria is termed 'late-diagnosed ADHD'. Currently, there are insufficient data to clarify if ADHD diagnosed in adulthood has the same underlying causes as childhood ADHD or if late-diagnosed ADHD is a disorder having a somehow different aetiology than childhood ADHD resulting in a later diagnosis or even adult onset-ADHD[17].

ADHD symptoms such as hyperactivity and inattention are believed to have continuous distributions in the population, with diagnosed ADHD representing the extreme end, which is supported by genetic findings[4]. The symptoms have an impairing impact on an individual's life when the accumulation of environmental and genetic risk factors exceeds a threshold. Age at first diagnosis might therefore differ depending on when this threshold is passed. Environmental factors influencing this could be age-related, such as increased educational demands in college or university, resulting in a delayed 'late-diagnosed ADHD'. Age at first diagnosis might also be influenced by genetic factors affecting symptom heterogeneity and/or severity, and a recent study found that individuals with late-

diagnosed ADHD have a burden of common ADHD risk alleles at a level comparable to individuals without ADHD[18] when analysing polygenic scores (PGSs) based on variant weights from the latest ADHD genome-wide association study (GWAS) meta-analysis[4]. The sample size was very small (N=98 for late-diagnosed individuals), and further investigation is needed to elucidate the impact of genetics on age at first diagnosis and to determine whether individuals diagnosed with ADHD in adulthood differ genetically from individuals diagnosed as children.

We have previously performed a large GWAS to evaluate the genetic architecture of childhood and persistent ADHD including in total 17,149 ADHD cases and 32,411 controls[19]. The genetic correlation (r_g) between the two groups was high (r_g=0.81), suggesting that childhood and persistent ADHD to a large extent have the same underlying genetic architecture. However, we noticed that the genetic correlation was significantly different from 1 (P=0.02), suggesting that further dissection of the genetic architecture might reveal genetic differences. Moreover, in that study all adult individuals with ADHD were grouped together, meaning that the persistent group consisted of individuals diagnosed in childhood with symptoms persisting into adulthood and individuals diagnosed as adults (i.e., late-diagnosed ADHD). Further subgrouping of individuals with ADHD depending on age at first diagnosis could therefore reveal unknown information about the genetic architecture underlying the disorder and comorbidities.

Here we perform in-depth characterisation of the polygenic architecture of childhood, persistent and late-diagnosed ADHD in a large Danish population-based case-cohort of ADHD cases and controls generated by iPSYCH[20]. We identify interesting differences among the groups with respect to common ADHD risk variants and rare protein-truncating variants (rPTVs). We also report several significant differences in genetic overlap of ADHD subgroups with other phenotypes, including an increased load of autism risk variants in individuals with childhood compared with late-

diagnosed ADHD and a larger genetic overlap of persistent and late-diagnosed ADHD with depression compared with childhood ADHD.

RESULTS

Sample characteristics

Individuals diagnosed with ADHD were identified in the large nation-wide population-based case-cohort established by iPSYCH[20] consisting of 133,296 genotyped individuals (iPSYCH1+2, see online methods). ADHD cases were divided into three groups depending on age at first diagnosis (see Online Methods for detailed definition of the groups): (1) childhood ADHD (N=14,878), defined as individuals diagnosed with ADHD in childhood; (2) persistent ADHD (N=1,473), defined as individuals that received an ADHD diagnosis as a child and again as adults; and (3) late-diagnosed ADHD (N=6,961), defined as individuals diagnosed with their first ADHD diagnosis as adults. Controls were randomly selected from the same nationwide birth cohort and not diagnosed with ADHD (N=38,303).

The sex distribution was different in the three groups. Females composed 23% of childhood ADHD cases, 36% of persistent ADHD cases and 41% of late-diagnosed cases, and the male/female ratio was significantly different among all three groups (Supplementary Table 1). Moreover, comorbidity patterns were different in the three groups. Autism spectrum disorder was very frequent in childhood (23% comorbid) and persistent (18% comorbid) ADHD compared with late-diagnosed ADHD (6.2% comorbid) (Supplementary Table 2). The adolescence/adulthood onset disorders schizophrenia, bipolar disorder and major depressive disorder were more frequent among individuals with persistent and late-diagnosed ADHD. As much as 27% of individuals with late-diagnosed ADHD had comorbid major depressive disorder (Supplementary Table 2).

Genome-wide association analyses of ADHD subgroups

We conducted a GWAS for each of the three ADHD subgroups. The GWAS of childhood ADHD revealed four genome-wide significant loci on chromosomes 1, 5, 18 and 20 (Table 1 and Supplementary Figures 1.a and 2.a-d). Two were new ADHD risk loci (on chromosomes 18 and 20) and two known risk loci (on chromosomes 1 and 5), identified in our previous GWAS meta-analysis of ADHD[4], which included an earlier and smaller iPSYCH sample than analysed here. One genome-wide significant locus was identified for late-diagnosed ADHD on chromosome 7, located in *FOXP2* (Table 1 and Supplementary Figures 1.b and 2.e). The effect size was significantly higher in late-diagnosed ADHD (odds ratio [OR]=1.11, standard error [SE]=0.01) compared with childhood ADHD (OR=1.05, SE=0.02) (P=0.012). No genome-wide significant loci were found for persistent ADHD, which was expected due the low number of cases.

SNP heritability and genetic correlations

We estimated the SNP heritability (h^2_{SNP}) by using best guess genotypes and GCTA[21] assuming a prevalence of 5% for childhood ADHD and 3% for persistent and late-diagnosed ADHD. We found the highest h^2_{SNP} in the persistent group (h^2_{SNP}=0.29), followed by late-diagnosed ADHD (h^2_{SNP} =0.27) and childhood ADHD (h^2_{SNP}=0.24) (Supplementary Table 3.A). None of the estimates were significantly different (Supplementary Table 3). Because the population prevalence of ADHD sub-types is not known precisely, we also estimated the heritability over a range of prevalence values. At all points persistent ADHD demonstrated the highest h^2_{SNP}, followed by late-diagnosed ADHD (Supplementary Figure 3 and Supplementary Table 3.B).

Pair-wise genetic correlations between ADHD subgroups revealed a high genetic correlation between childhood and persistent ADHD (r_g=0.82, SE=0.08), and between persistent and late-

diagnosed ADHD (r_g=0.77, SE=0.08), while the genetic correlation between childhood ADHD and late-diagnosed ADHD was moderate (r_g=0.65, SE=0.04) (Supplementary Table 4).

ADHD risk polygenic load in ADHD subgroups

The polygenic risk load of variants associated with the general liability to ADHD risk in the three ADHD subgroups was evaluated by PGS analyses. All groups demonstrated a highly significantly increased ADHD-PGS load compared with controls (Supplementary Table 5). The highest mean ADHD-PGS was found for persistent ADHD (mean=0.41, SD=0.95), followed by late-diagnosed ADHD (mean=0.27, SD=0.98) and then childhood ADHD (mean=0.26, SD=0.96) (Supplementary Table 5). The ADHD-PGS load in persistent ADHD was significantly higher than childhood ADHD ($P=3.0\times10^{-4}$) and nominally significantly higher than late-diagnosed ADHD (P=0.02). The results did not change in the sensitivity analysis when splitting the childhood group into those younger than 18 years and those older than 18 years of age by the end of follow-up (Supplementary Figure 4).

To replicate the findings, we performed PGS analysis in a Spanish sample consisting of 453 individuals with childhood ADHD, 270 with persistent ADHD, 889 with late-diagnosed ADHD and 3,440 controls. We did not replicate the findings, with trends in the opposite direction when comparing with controls (ADHD-PGS childhood ADHD: beta=0.27, SE=0.05; persistent ADHD: beta=0.21, SE=0.06; late-diagnosed ADHD: beta=0.19, SE=0.04). However, the differences were not significant, and we cannot make any strong conclusions based on this small replication sample.

Genetic overlap with ADHD symptoms in the general population

Genetic overlap with ADHD symptoms in the general population was estimated, using results from GWAS of ADHD subgroups and the results from GWAS meta-analyses of measures of inattention and hyperactivity (N=43,117) in the general population[22]. Inattention and hyperactivity were

highly correlated with both childhood ADHD ($r_{g\_inattention}$=0.86, SE=0.08; $r_{g\_hyperactivity}$=0.95, SE=0.08) and persistent ADHD ($r_{g\_inattention}$=0.87, SE=0.14; $r_{g\_hyperactivity}$=~1, SE=0.15) (Supplementary Table 6), but showed a considerably lower correlation with late-diagnosed ADHD ($r_{g\_inattention}$=0.57, SE=0.08; $r_{g\_hyperactivity}$=0.59, SE=0.07). The genetic correlation of hyperactivity with late-diagnosed ADHD was significantly lower than observed for childhood ADHD (P=0.004). In addition, the genetic correlations of ADHD symptoms with late-diagnosed ADHD were significantly less than 1 ($P_{diff\_1\_inattention}=7.66\times10^{-8}$; $P_{diff\_1\_hyperactivity}=4.71\times10^{-9}$; Supplementary Table 6).

PGS analyses to test for enrichment in the three ADHD subgroups of variants associated with inattention and hyperactivity identified a nominally significantly lower PGS for hyperactivity in late-diagnosed ADHD compared with childhood ADHD (P=0.04; Supplementary Table 7).

Genetic overlap with psychiatric disorders and other traits

The observed differences in comorbidity patterns with other psychiatric disorders could reflect age differences among the groups, but could also be influenced by differences in the genetic architecture. To evaluate this, we performed genetic correlation and PGS analyses for major psychiatric disorders (schizophrenia[23], bipolar disorder[24], major depressive disorder[25], autism spectrum disorder[26], anorexia[27], obsessive compulsive disorder [OCD][28], cannabis use disorder[29] and alcohol use disorder[30]). We found positive genetic correlations of ADHD subgroups with autism, schizophrenia, bipolar disorder, major depressive disorder, alcohol use disorder and cannabis use disorder and negative genetic correlations with OCD and anorexia (Figure 1 and Supplementary Table 8), in line with previous findings[4]. We identified a significantly higher genetic correlation of childhood ADHD with autism (r_g=0.48, SE=0.05) compared with late-diagnosed ADHD (r_g=0.27, SE=0.06), and a significantly higher genetic correlation of depression and alcohol use disorder with late-diagnosed ADHD ($r_{g\_depression}$=0.69, SE=0.04; $r_{g\_alcohol\_use\_disorder}$=0.82, SE=0.2) compared with childhood ADHD

($r_{g\_depression}$=0.45, SE=0.04; $r_{g\_alcohol\_use\_disorder}$=0.39, SE=0.09) ($P_{diff\_depression}=8.7\times10^{-7}$; $P_{diff\_alcohol\_use\_disorder}=3.8\times10^{-5}$) (Figure 1 and Supplementary Table 8). The PGS results demonstrated the same pattern, with a significantly increased autism-PGS in childhood ADHD compared with late-diagnosed ADHD, a significantly higher PGS in persistent and late-diagnosed ADHD compared with childhood ADHD for depression and cannabis use disorder and a significantly increased PGS in late-diagnosed ADHD compared with childhood ADHD for schizophrenia and bipolar disorder (Figure 2 and Supplementary Table 9).

We also performed genetic correlation and PGS analyses for phenotypes representing domains that had previously[4] demonstrated high genetic correlations with ADHD: cognition (educational years[31]), overweight (body mass index [BMI][32]), reproduction (age at first birth[33]), mortality (maternal age of death[34]) and sleep (insomnia[35]). We identified stronger negative genetic correlations of late-diagnosed ADHD compared with childhood ADHD for educational years ($P_{difference}=1.7\times10^{-5}$; $r_{g\_late\text{-}diagnosed}$=-0.61, SE=0.03; $r_{g\_childhood}$=-0.46, SE=0.03), increased age at first birth ($P_{difference}=8.9\times10^{-5}$; $r_{g\_late\text{-}diagnosed}$=-0.73, SE=0.04; $r_{g\_childhood}$=-0.54, SE=0.04) and increased mother´s age at death ($P_{difference}=2.6\times10^{-4}$; $r_{g\_late\text{-}diagnosed}$=-0.79, SE=0.10; $r_{g\_childhood}$=-0.48, SE=0.08) (Figure 1 and Supplementary Table 8). Furthermore, we identified a significantly less negative PGS in childhood ADHD compared with persistent and late-diagnosed ADHD for number of educational years ($P_{childhood\_vs\_persistent}=8.02\times10^{-8}$; $P_{childhood\_vs\_late\text{-}diagnosed}=4.35\times10^{-14}$) and a less negative PGS for age at first birth for childhood ADHD compared with late-diagnosed ADHD (Figure 2 and Supplementary Table 9).

Except for autism and OCD, the highest PGS was observed for persistent ADHD (Supplementary Table 10.A); however, due to the small sample size of this group, we had limited power to detect pairwise PGS differences for this group compared with the other two groups.

Finally, we performed two PGS sensitivity analyses. First, we evaluated the PGS for autism, schizophrenia, bipolar disorder and depression in the three ADHD groups in which we excluded individuals with the disorder of the PGS being analysed. The results revealed the same patterns seen in the full sample, but the PGS-autism was only nominal significantly higher in childhood ADHD compared to late-diagnosed ADHD (P=0.02) (Supplementary Figure 5 and Supplementary Tables 10.B and 10.C).

Second, we redid the PGS analyses but this time with the childhood group split in two, namely younger than 18 years and older than 18 years of age by the end of follow-up. The PGS load in the two childhood ADHD groups were generally at the same level except for schizophrenia and bipolar disorder, for which the load was higher in the individuals older than 18 years of age (Supplementary Figure 6 and Supplementary Table 11).

Burden of rare variants ADHD subgroups

We have previously demonstrated an enrichment of rPTVs in highly constrained genes in ADHD cases[36]. We wanted to explore this further by evaluating the load of rPTVs in the three ADHD subgroups. For this, we used whole exome-sequencing (WES) data available for a subset of the iPSYCH cohort (childhood ADHD, N=4,987; persistent ADHD, N=748; late-diagnosed ADHD, N=1,915; controls, N=8,649). The burden of rPTVs and rare synonymous variants (rSYNs) in the three ADHD subgroups was tested in three gene sets: (1) 'highly constrained genes', genes being evolutionary intolerant to loss-of-function mutations with a pLI score >0.9[37] (3,488 genes); (2) '*de novo* constrained genes', the subset of highly constrained genes that overlap another gene-set of 285 genes found to be enriched with *de novo* mutations in individuals with neurodevelopmental disorders[38] (241 genes); and (3) 'low constrained genes', genes being less constrained i.e. genes relatively tolerant to loss-of-function mutations with a pLI score <0.1 (9,662 genes). When compared

with controls, the load of rPTVs in highly constrained genes was comparable and significantly increased in childhood and persistent ADHD (childhood ADHD beta=0.13, SE=0.02, $P=2.41\times10^{-11}$; persistent ADHD beta=0.12, SE=0.04, $P=1.90\times10^{-3}$) but lower and not significantly enriched in late-diagnosed ADHD (beta=0.06, SE=0.03, P=0.02). The same pattern was observed for *de novo* highly constrained genes (Figure 3 and Supplementary Table 12). No pair-wise comparisons among ADHD sub-groups were significant, but there was a tendency towards a higher burden of rPTVs in childhood compared with late-diagnosed ADHD in *de novo* highly constrained genes (P=0.096). For comparison, we did not find enrichment of any rSYNs in in the gene sets or an enrichment of rPTVs in low constrained genes (Figure 3 and Supplementary Table 12).

DISCUSSION

We identified differences in the genetic architecture of childhood, persistent and late-diagnosed ADHD based on unique data from a large population-based Danish case-cohort. We identified the first four genome-wide significant loci associated with childhood ADHD, two of them novel ADHD risk loci located on chromosomes 18 and 20. The chromosome 18 index variant is located in *DCC*, a gene recently linked to the general liability to psychiatric disorders[39]; thus, it does not seem specific to ADHD. The chromosome 20 locus is intergenic, and the index variant has previously demonstrated genome-wide significant association with weight-related phenotypes[34]. We also identified a genome-wide significant locus associated with late-diagnosed ADHD in *FOXP2*. When this locus was first reported as a risk locus for ADHD[4], it received a lot of attention due to the role of *FOXP2* in cognition, language and speech development[40-42], and recently we also found *FOXP2* to be a risk gene for cannabis use disorder[29]. The effect size was significantly higher in late-diagnosed ADHD compared with childhood ADHD, which suggests that the association of *FOXP2* with ADHD is driven to a greater extent by late-diagnosed ADHD than childhood ADHD.

When assessing the polygenic architecture, we identified the highest SNP heritability for persistent ADHD. In concordance with this, we observed the highest polygenic risk load for the general liability to ADHD in individuals with persistent ADHD. This observation is consistent with the hypothesis that individuals with a higher genetic risk load for ADHD are those who will continue to have ADHD symptoms as adults. This finding is in line with a previous study reporting an association of ADHD-PGS with persistence of ADHD symptoms in the general population[16], and a recent smaller study reporting higher ADHD-PGS in persistent ADHD compared with late-diagnosed ADHD (but not significantly different)[18]. The ADHD-PGS analyzed represents the general liability to diagnosed ADHD because the scores are derived from data representing all individuals with ADHD in the Danish population born between 1981 and 2008 (see online methods). In relation to this, it should be noted that the training data included a higher proportion of childhood ADHD cases than persistent and late-diagnosed ADHD, which potentially could result in a better prediction of childhood ADHD. Despite this we observe the opposite, a higher ADHD-PGS in the adult groups compared to childhood ADHD (with significantly increased ADHD-PGS in persistent ADHD and a slight increase in late-diagnosed ADHD), which reinforces the validity of the results.

The findings did not replicate in the Spanish cohort. This could be due to the relatively small replication cohort or differences in ascertainment. The iPSYCH cohort reflects the genetic architecture across all ADHD cases in the Danish population, whereas the Spanish cohort is a smaller clinical dataset that might be influenced by unknown ascertainment biases.

The genetic correlation of childhood ADHD with persistent ADHD was high (r_g=0.82) and at the same level as reported previously (r_g=0.81)[19], while the genetic correlation of childhood ADHD with late-diagnosed ADHD was lower (r_g=0.65). This suggests some differences in the polygenic architecture of childhood and late-diagnosed ADHD. A part of this could be due to a lower load of variants associated with hyperactivity and inattention in individuals with late-diagnosed ADHD,

because we observed a higher genetic correlation of ADHD symptoms with childhood and persistent ADHD compared with late-diagnosed ADHD. Likewise, PGS analyses also suggested a lower burden of ADHD-symptom-associated variants in late-diagnosed ADHD than in the other groups. We cannot exclude that different age distributions in the GWAS meta-analyses of ADHD symptoms influenced the results. However, the age distribution is similar in the persistent and late-diagnosed groups (all are above 18 years of age), indicating that the decreased genetic overlap with late-diagnosed ADHD is not caused by age differences. A part of the explanation behind a later diagnosis of ADHD could therefore be genetic, with late-diagnosed individuals being less genetically predisposed to be inattentive and hyperactive, leaving their ADHD unnoticed until later in life.

The comorbidity pattern in the three groups differed, with a higher comorbidity of autism spectrum disorder among childhood (23%) and persistent ADHD (18%) compared with late-diagnosed ADHD (6.2%), in line with previous reports concerning comorbid autism among children with ADHD[43]. The observed comorbidity patterns were reflected in the genetic analyses where we found a significantly higher genetic correlation of autism with childhood ADHD compared with late-diagnosed ADHD, and higher PGS-autism in childhood ADHD compared with late-diagnosed ADHD. Therefore, childhood ADHD seems genetically more related to autism than late-diagnosed ADHD.

The comorbidities of psychiatric disorders with onset in adolescence/adulthood were higher among persistent and late-diagnosed ADHD (Supplementary Table 2). Part of this is likely due to the age difference, because many individuals in the childhood group are too young to develop these disorders. However, our results suggest that age alone cannot explain the comorbidity patterns. Genetics might play a role, as we in general observed a higher genetic correlation or PGS for several of the disorders (schizophrenia, bipolar disorder, alcohol use disorder, cannabis use disorder and depression) in persistent and late-diagnosed ADHD compared with childhood ADHD (Figures 1 and

2 and Supplementary Tables 8 and 9). Depression was particularly striking with a significantly higher PGS in individuals with persistent and late-diagnosed ADHD compared with childhood ADHD. The high comorbidity of depression among adults with ADHD is well known, but the causes are not. ADHD in itself could be a risk factor[44,45], but genetics is also considered a risk factor due to the high genetic correlation of ADHD with depression[4]. Our results suggest genetic heterogeneity among ADHD cases: individuals with persistent and late-diagnosed ADHD are at higher risk for comorbid depression due to the underlying genetic architecture of the disorder in these groups.

In analyses of five selected phenotypes (educational years, insomnia, mother's age at death, age at first birth and BMI) representing domains highly genetically correlated with ADHD[4], we observed the highest genetic correlations and the highest PGS load in persistent ADHD, followed closely by late-diagnosed ADHD and the lowest in childhood ADHD (except for BMI), suggesting a similar polygenic architecture of persistent and late-diagnosed ADHD for these phenotypes (Figures 1 and 2, see Supplementary information, note 1, regarding mother´s age at death). These results also support the idea that the negative outcomes associated with persistent ADHD, such as decreased school performance[46] and sleep problems[47], are influenced by genetics to a greater extent than in childhood ADHD.

We found an increased burden of rPTVs in persistent and childhood ADHD compared with controls in highly constrained genes, but not in late-diagnosed ADHD. There was also a tendency towards a significantly higher burden of rPTVs in *de novo* highly constrained genes in childhood ADHD compared with late-diagnosed ADHD. These findings suggest that when considering rPTVs, which are variants expected to have greater impact on the disorder than common variants, the genome of individuals with late-diagnosed ADHD is less burdened. When considering both common and rare variants, the emerging picture suggests that childhood ADHD is genetically more similar to autism (high genetic correlation with autism and increased rPTV burden), whereas late-diagnosed ADHD

genetically is more similar to depression (high genetic correlation with depression and no significant increase in rPTVs in highly constrained genes compared with controls).

We cannot rule out ascertainment differences among children and adults. However, the genetic correlations of late-diagnosed and persistent ADHD (which requires an ADHD diagnosis in childhood) with depression and alcohol use disorder are very similar, suggesting that such ascertainment bias between the two groups is limited. Likewise, in the PGS sensitivity analyses, where individuals with the disorder of the PGS being analysed were excluded, the genetic differences among the groups demonstrated the same patterns as observed in the full sample. This further reinforces the conclusion that comorbid psychiatric disorders do not have a strong influence on the observed genetic differences among the ADHD sub-groups.

In summary, our results are population-based and thus reflect the genetic architecture of ADHD and comorbidity patterns across ADHD subgroups in the Danish population. Persistent ADHD demonstrated the highest load of ADHD risk variants, while late-diagnosed ADHD was less enriched for variants associated with hyperactivity and inattention and did not, unlike childhood and late-diagnosed ADHD, demonstrate an increased burden of rPTVs compared with controls. This suggests that genetics in part might explain why some individuals are diagnosed late as adults. The comorbidity of depression and alcohol use disorder was highest in the late-diagnosed group. If this observation was only due to age differences among groups, we would not expect the genetic correlations to differ, but we found a higher genetic overlap of these disorders with late-diagnosed ADHD compared with childhood ADHD. This suggests that the higher comorbidity among individuals with late-diagnosed ADHD is not only due to those individuals being older, but also due to a higher genetic risk. Conversely, childhood ADHD demonstrated a higher genetic overlap with autism and a higher burden of rPTVs in highly constrained genes than the other two groups. Overall, we have identified genetic heterogeneity among ADHD subgroups, and our findings suggest that

genetic factors influence time of first ADHD diagnosis, persistence of ADHD into adulthood and comorbidity patterns.

Acknowledgements
D.D was supported by the Novo Nordisk Foundation (NNF20OC0065561) and the Lundbeck Foundation (R344-2020-1060). The iPSYCH team was supported by grants from the Lundbeck Foundation (R102-A9118, R155-2014-1724 and R248-2017-2003), the EU FP7 Program (Grant No. 602805, 'Aggressotype') and the H2020 Program (Grant No. 667302, 'CoCA'), NIMH (1U01MH109514-01 to A.D.B.) and the universities and university hospitals of Aarhus and Copenhagen. The Danish National Biobank resource was supported by the Novo Nordisk Foundation. High-performance computer capacity for handling and statistical analysis of iPSYCH data at the GenomeDK HPC facility was provided by the Center for Genomics and Personalized Medicine and the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark (grant to A.D.B.). M.S.A. is a recipient of a Juan de la Cierva Incorporación contract from the Ministry of Science, Innovation and Universities, Spain (IJC2018-035346-I). The research leading to these results has received funding from the Instituto de Salud Carlos III (PI17/00289, PI18/01788, P19/01224 and PI20/00041) and from the Agència de Gestió d'Ajuts Universitaris i de Recerca-AGAUR, Generalitat de Catalunya (2017SGR1461) and co-financed by the European Regional Development Fund (ERDF). The Norwegian Mother, Father and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. We are grateful to all the participating families in Norway who have taken part in this on-going cohort study. L.V.R. is a recipient of a pre-doctoral fellowship from the Instituto de Salud Carlos III, Spain (FI18/00285). M.R. was a recipient of a Miguel de Servet contract from the Instituto de Salud Carlos III, Spain (CP09/00119 and CPII15/00023). TZ is funded by R37MH107649-07S1 and by the Research Council of Norway, NRC, Grant No. 288083.

Author contributions
Analysis: V.M.R, J.D., L. V-R. Sample and/or data provider and processing: J.G., T. Z., J.A.R-Q., F.K.S., M.S.A., J.B.-G., M.B.-H., T.D.A., A.R., M.J.D., B.M.N., M.N., T.W., O.M., D.M.H., P.B.M. Writing: D.D., V.M.R. Manuscript revision, core group: D.D., V.M.R., F.K.S., A.D.B., M.R. Study Design: D.D., V.M.R. All authors contributed with critical revision of the manuscript.

Competing interest
D.D. has received a speaker fee from Takeda. J.A.R.Q has been on the speaker's bureau and/or has acted as a consultant for Janssen-Cilag, Novartis, Shire, Takeda, Bial, Shionogi, Sincrolab, Novartis, BMS, Medice and Rubiand Raffo in the last 3 years. He has also received travel awards (air tickets and hotel) for taking part in psychiatric meetings from Janssen-Cilag, RubiShire, Takeda, Shionogi, Bial and Medice. The Department of Psychiatry chaired by him has received unrestricted educational and research support from the following companies in the last 3 years: Janssen-Cilag, Shire, Oryzon, Roche, Psious and Rubió. The remaining authors declare no competing interests.

Table 1. Index variants for the genome-wide significant loci identified in the genome-wide association studies (GWAS) of childhood and late-diagnosed attention-deficit hyperactivity disorder (ADHD). The variant ID (SNP), chromosome position (CHR), base position in hg19 (BP), effect allele (A1), other allele (A2), minor allele frequency of A1 (MAF), odds ratio (OR) with respect to A1, standard error (SE), association P-value from logistic regression (P) and nearest gene are given.

SNP	CHR	BP	A1	A2	MAF	OR	SE	P	Nearest gene
Childhood ADHD									
rs7511800	1	44214269	T	A	0.31	0.91	0.01	7.4×10^{-11}	*ST3GAL3*
rs12653396	5	87847273	T	A	0.42	0.91	0.01	2×10^{-11}	*MEF2C*
rs28718037	18	50572697	A	G	0.33	0.92	0.01	8.7×10^{-9}	*DCC*
rs6035830	20	21265728	C	T	0.28	1.10	0.01	1.5×10^{-9}	*XRN2*
Late-diagnosed ADHD									
rs1229758	7	114229139	G	A	0.43	0.90	0.01	2.1×10^{-8}	*FOXP2*

Figure legends

Figure 1. Genetic correlations of ADHD subgroups with major psychiatric disorders and phenotypes representing domains with high genetic correlation with ADHD, estimated by linkage disequilibrium (LD) score regression. Results from genome-wide association analyses of ADHD subgroups was used including childhood (N=14,878 individuals), late-diagnosed (N=6,961 individuals) and persistent (N=1,473 individuals) attention-deficit hyperactivity disorder (ADHD) against 38,303 control individuals. Error bars (horizontal lines) represent standard errors. *Indicates a significant difference (after Bonferroni correction) with a two-sided P-value lower than P=0.0013 in the genetic correlation observed for childhood ADHD compared with late-diagnosed ADHD.

Figure 2. Results from polygenic score (PGS) analysis demonstrating the association of PGS with childhood (N=14,878 individuals), persistent (N=1,473 individuals) and late-diagnosed attention-deficit hyperactivity disorder (ADHD) (N=6,961 individuals). PGSs for psychiatric disorders: autism spectrum disorder (ASD), depression (MDD), schizophrenia (SZ), bipolar disorder (BD), anorexia, obsessive compulsive disorder (OCD), alcohol use disorder (AUD) and cannabis use disorder (CUD). PGSs for five phenotypes representing domains highly correlated with ADHD: educational attainment (EA), insomnia, mother's age at death and age of first birth (AOB). On the y-axis is the beta from multi-nominal regression against controls (N=38,303 individuals); error bars (vertical lines) represent standard errors (see also Supplementary Table 10). Significant pair-wise comparisons (after Bonferroni correction) with a two-sided P-value lower than P=0.0013 are given in the right corner of the header: 'C' indicates childhood ADHD, 'L' indicates late-diagnosed ADHD, 'P' indicates persistent ADHD and the direction of the difference in the betas is given by '>' (see also Supplementary Table 9).

Figure 3. The load of rare protein truncating variants (rPTVs) and rare synonymous (rSyn) in three gene-sets (1) 'Highly constrained', genes intolerant to loss-of-function mutations (pLI > 0.9); (2) 'Denovo', highly constrained genes which in another study has been found to be enriched with *de novo* mutations in individuals with neurodevelopmental disorders; and (3) 'Low constrained', genes tolerant to loss-of-function mutations (pLI score <0.1). The y-axis represents the beta from multiple logistic regression of childhood (N=4,987 individuals), persistent (N=748 individuals) and late-diagnosed (N=1,915 individuals) ADHD with comparison to controls (N=8,649 individuals), error bars (vertical lines) represent the standard error. For pair-wise comparisons of attention-deficit hyperactivity disorder (ADHD) subgroups, see Supplementary Table 12.

References

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

2 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).

3 Faraone, S. V. & Larsson, H. Genetics of attention deficit hyperactivity disorder. *Molecular psychiatry*, doi:10.1038/s41380-018-0070-0 (2018).

4 Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature genetics* **51**, 63-75, doi:10.1038/s41588-018-0269-7 (2019).

5 Faraone, S. V., Biederman, J. & Mick, E. The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies. *Psychological medicine* **36**, 159-165, doi:10.1017/S003329170500471X (2006).

6 Fatseas, M. *et al.* Addiction severity pattern associated with adult and childhood Attention Deficit Hyperactivity Disorder (ADHD) in patients with addictions. *Psychiatry Res* **246**, 656-662, doi:10.1016/j.psychres.2016.10.071 (2016).

7 Agnew-Blais, J. C. *et al.* Young adult mental health and functional outcomes among individuals with remitted, persistent and late-onset ADHD. *The British journal of psychiatry : the journal of mental science* **213**, 526-534, doi:10.1192/bjp.2018.97 (2018).

8 Ilbegi, S. *et al.* Substance use and nicotine dependence in persistent, remittent, and late-onset ADHD: a 10-year longitudinal study from childhood to young adulthood. *J Neurodev Disord* **10**, 42, doi:10.1186/s11689-018-9260-y (2018).

9 Yoshimasu, K. *et al.* Adults With Persistent ADHD: Gender and Psychiatric Comorbidities-A Population-Based Longitudinal Study. *J Atten Disord* **22**, 535-546, doi:10.1177/1087054716676342 (2018).

10 Kim, K. M. *et al.* Psychopathological, temperamental, and characteristic factors in adults with remaining childhood attention-deficit hyperactivity symptoms. *Int J Psychiatry Clin Pract* **21**, 236-241, doi:10.1080/13651501.2017.1297835 (2017).

11 Tandon, M., Tillman, R., Agrawal, A. & Luby, J. Trajectories of ADHD severity over 10 years from childhood into adulthood. *Atten Defic Hyperact Disord* **8**, 121-130, doi:10.1007/s12402-016-0191-8 (2016).

12 Kan, K. J. *et al.* Genetic and environmental stability in attention problems across the lifespan: evidence from the Netherlands twin register. *Journal of the American Academy of Child and Adolescent Psychiatry* **52**, 12-25, doi:10.1016/j.jaac.2012.10.009 (2013).

13 Larsson, H. *et al.* Genetic and environmental influences on adult attention deficit hyperactivity disorder symptoms: a large Swedish population-based study of twins. *Psychological medicine* **43**, 197-207, doi:10.1017/S0033291712001067 (2013).

14 Brikell, I., Kuja-Halkola, R. & Larsson, H. Heritability of attention-deficit hyperactivity disorder in adults. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **168**, 406-413, doi:10.1002/ajmg.b.32335 (2015).

15 Pingault, J. B. *et al.* Genetic and Environmental Influences on the Developmental Course of Attention-Deficit/Hyperactivity Disorder Symptoms From Childhood to Adolescence. *JAMA Psychiatry* **72**, 651-658, doi:10.1001/jamapsychiatry.2015.0469 (2015).

16 Riglin, L. *et al.* Association of Genetic Risk Variants With Attention-Deficit/Hyperactivity Disorder Trajectories in the General Population. *JAMA Psychiatry* **73**, 1285-1292, doi:10.1001/jamapsychiatry.2016.2817 (2016).

17 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).
18 Agnew-Blais, J. C. *et al.* Polygenic Risk and the Course of Attention-Deficit/Hyperactivity Disorder From Childhood to Young Adulthood: Findings From a Nationally-Representative Cohort. *Journal of the American Academy of Child and Adolescent Psychiatry*, doi:10.1016/j.jaac.2020.12.033 (2021).
19 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).
20 Pedersen, C. B. *et al.* The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Molecular psychiatry*, doi:10.1038/mp.2017.196 (2017).
21 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).
22 Zayats, T., QIMR Group, SickKids Group, TEDS Group, CATSS Group, MoBa Group, NTR Group, ALSPAC Group, ATGU Group. Genome-wide association meta-analysis of ADHD symptomatology: inattention and hyperactivity/impulsivity. *European Neuropsychopharmacology* **29**, S1190, doi:https://doi.org/10.1016/j.euroneuro.2018.08.226 (2019).
23 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).
24 Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature genetics* **51**, 793-803, doi:10.1038/s41588-019-0397-8 (2019).
25 Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature neuroscience* **22**, 343-352, doi:10.1038/s41593-018-0326-7 (2019).
26 Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics* **51**, 431-444, doi:10.1038/s41588-019-0344-8 (2019).
27 Watson, H. J. *et al.* Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nature genetics* **51**, 1207-1214, doi:10.1038/s41588-019-0439-2 (2019).
28 International Obsessive Compulsive Disorder Foundation Genetics, C. & Studies, O. C. D. C. G. A. Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Molecular psychiatry* **23**, 1181-1188, doi:10.1038/mp.2017.154 (2018).
29 Johnson, E. C. *et al.* A large-scale genome-wide association study meta-analysis of cannabis use disorder. *Lancet Psychiatry*, doi:10.1016/S2215-0366(20)30339-4 (2020).
30 Walters, R. K. *et al.* Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nature neuroscience* **21**, 1656-1669, doi:10.1038/s41593-018-0275-1 (2018).
31 Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics* **50**, 1112-1121, doi:10.1038/s41588-018-0147-3 (2018).
32 Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet* **27**, 3641-3649, doi:10.1093/hmg/ddy271 (2018).

33 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).
34 Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nature genetics* **51**, 1339-1348, doi:10.1038/s41588-019-0481-0 (2019).
35 Jansen, P. R. *et al.* Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nature genetics* **51**, 394-403, doi:10.1038/s41588-018-0333-3 (2019).
36 Satterstrom, F. K. *et al.* Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. *Nature neuroscience* **22**, 1961-1965, doi:10.1038/s41593-019-0527-8 (2019).
37 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).
38 Kaplanis, J. *et al.* Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* **586**, 757-762, doi:10.1038/s41586-020-2832-5 (2020).
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 Enard, W. *et al.* Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* **418**, 869-872, doi:10.1038/nature01025 (2002).
41 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).
42 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).
43 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).
44 Skirrow, C. & Asherson, P. Emotional lability, comorbidity and impairment in adults with attention-deficit hyperactivity disorder. *Journal of affective disorders* **147**, 80-86, doi:10.1016/j.jad.2012.10.011 (2013).
45 Chang, Z., D'Onofrio, B. M., Quinn, P. D., Lichtenstein, P. & Larsson, H. Medication for Attention-Deficit/Hyperactivity Disorder and Risk for Depression: A Nationwide Longitudinal Cohort Study. *Biological psychiatry* **80**, 916-922, doi:10.1016/j.biopsych.2016.02.018 (2016).
46 Biederman, J., Petty, C. R., Evans, M., Small, J. & Faraone, S. V. How persistent is ADHD? A controlled 10-year follow-up study of boys with ADHD. *Psychiatry Res* **177**, 299-304, doi:10.1016/j.psychres.2009.12.010 (2010).
47 Van Veen, M. M., Kooij, J. J., Boonstra, A. M., Gordijn, M. C. & Van Someren, E. J. Delayed circadian rhythm in adults with attention-deficit/hyperactivity disorder and chronic sleep-onset insomnia. *Biological psychiatry* **67**, 1091-1096, doi:10.1016/j.biopsych.2009.12.032 (2010).

Methods

Sample characteristics

Individuals included in the study were identified in a nationwide population-based case-cohort established by iPSYCH[20] comprising 133,296 genotyped individuals, among which 91,378 have been diagnosed with at least one of six mental disorders (i.e. schizophrenia, bipolar disorder, major depressive disorder, autism spectrum disorder, ADHD and anorexia) and the remaining are population-based controls. Samples were selected from a baseline birth cohort comprising all singletons born in Denmark between 1st of May 1981 and 31 of December 2008, who have a known mother (99.9% of all individuals born in Denmark since 1970 have a known mother[48]) and were residents in Denmark on their first birthday (N=1,472,762). We included all individuals in the cohort diagnosed with ADHD by psychiatrists according to the ICD10 criteria (F90.0 diagnosis code) identified in the Danish Psychiatric Central Research Register[49]. See Supplementary Information, note 2, for information on cases potentially missed by the diagnostic system.

The ICD10 diagnosis code F90.0 describes a disorder characterised by early onset, usually in the first five years of life, with hyperactivity and decreased attention. According to the current diagnostic criteria, individuals diagnosed with ADHD as adults should be able to describe ADHD symptoms in childhood retrospectively. Diagnoses were given in 2016 or earlier for individuals at least 1 year old. Cases were divided into three groups depending on age at first diagnosis: (1) childhood ADHD, defined as cases diagnosed with ADHD and less than 18 years of age in 2016 or cases older than 18 years by the end of follow-up (2016) who did not receive another ADHD diagnosis when older than 18 years (N=15,338 before QC); (2) persistent ADHD, defined as cases diagnosed with ADHD as a child (before 18 years of age) and again as adults (after 18 years of age) (N=1,709 before QC); and (3) late-diagnosed ADHD, defined as individuals diagnosed with ADHD as adults

(after 18 years of age) (N=7,815 before QC). Controls were randomly selected from the same nationwide birth cohort and not diagnosed with ADHD (N=45,398 before QC).
The study was approved by the Danish Data Protection Agency and the Scientific Ethics Committee in Denmark.

Testing for difference in the female/male ratio between ADHD subgroups was done by using a chi-square test. Information about comorbidity for other major psychiatric disorders was obtained from the Danish Psychiatric Central Research Register[49]: autism spectrum disorder (ICD10 diagnosis code F84), schizophrenia (ICD10 diagnosis code F20), bipolar disorder (ICD10 diagnosis codes F30-F31), major depressive disorder (ICD10 diagnosis codes F32-F33), OCD (ICD10 diagnosis code F42), anorexia (ICD10 diagnosis codes F50), alcohol use disorder (ICD10 diagnosis code F10.1-F10.9) and cannabis use disorder (ICD10 diagnosis code F12.1-F12.9).

Genotyping and QC

The study subjects were linked to their biological sample (dried blood spots) stored in the Danish Newborn Screening Biobank[50], through the personal identification number[51] assigned to all individuals with residence in Denmark. DNA was extracted from the dried blood spots and whole genome amplified in triplicates[52,53]. Genotyping was done in two rounds. In round one (iPSYCH1), 79,492 individuals were genotyped by using Illumina's Beadarrays (PsychChip; Illumina, San Diego, CA, USA). In round two (iPSYCYH2), 53,804 individuals were genotyped by using Illumina's Global Screening array. iPSYCH1 genotypes were a result of merging call sets from two different calling algorithms, GenCall(1.6.2.2)[54] and Birdseed(1.6)[55], which were used to call genotypes with a minor allele frequency (MAF) >0.01. iPSYCH2 genotypes were called by using GenTrain V3.

All downstream analyses were performed at our secure server (GenomeDK [http://genome.au.dk]). Stringent QC was applied to the full iPSYCH sample. Only individuals with

a high call rate (>0.95) were included, and only genotypes with a high call rate (>0.98), no strong deviation from Hardy-Weinberg equilibrium ($P>1\times10^{-6}$ in controls or $P>1\times10^{-10}$ in cases) and low heterozygosity rates ($|F_{het}|<0.2$) were included. Genotypes were phased and imputed using the Haplotype Reference Consortium[56] data as reference panel, while pre-phasing was done by using Eagle v2.3.5[57] and imputation with Minimac3[58].

Relatedness and population stratification were evaluated for ADHD cases and controls by using merged data from iPSYCH1 and iPSYCH2 and a set of high-quality markers (best guess genotypes with MAF >0.05, HWE $P>1\times10^{-9}$, SNP call rate >0.99, imputation info score [INFO] >0.9), which were pruned for linkage disequilibrium (LD) ($r^2<0.1$), resulting in a set of 37,986 pruned variants (variants located in long-range LD regions defined by Price et al.[59] were excluded). Genetic relatedness was estimated by using PLINK v1.9[60,61] to identify first- and second-degree relatives ($\hat{\pi}$ >0.2), and one individual was excluded from each related pair (cases kept preferably over controls). Genetic outliers were excluded based on principal component analyses (PCA) using EIGENSOFT 6.1.3[62,63]. After the first PCA, the principal components (PCs) from a set of individuals born in Denmark for three generations were used as a reference to generate an ellipsoid based on information from the first 6 PCs and their standard deviations (8 SDs were used). Those who fell outside this ellipsoid were removed. The PCA was repeated, and the new PCs were used as covariates to correct for any remaining population substructure in subsequent analyses. After QC, the number of included individuals was: (1) childhood ADHD, N=14,878; (2) persistent ADHD, N=1,473; and (3) late-diagnosed ADHD, N=7,188. The control group contained 38,3030 individuals.

GWAS

A flow chart of the genetic analyses performed in this study can be found in Supplementary Figure 7. We performed GWAS for each ADHD subgroup against a common set of controls (N=38,3030).

We used merged iPSYCH1 and iPSYCH2 high-quality best guess genotypes (MAF >0.01, INFO >0.80, missing rate <1%; N=5,826,893 variants) and tested for association using logistic regression (in PLINK1.9[60]) and the following covariates: 10 PCs from PCA, sex and a covariate for genotyping round (iPSYCH1 or iPSYCH2).

We tested whether the effect size of the genome-wide significant locus in late-diagnosed ADHD was significantly higher than the effect size observed for the other groups using a z-test and effect sizes from association analyses based on non-overlapping samples. The numbers of non-overlapping controls were: childhood controls, 24,443; persistent controls, 2,289; and late-diagnosed controls, 11,571.

SNP heritability and genetic correlations of ADHD subgroups

h^2_{snp} was estimated for each group against the same controls (N=38,3030) by using univariate GREML implemented in GCTA[21] (and the same covariates as used in the GWAS) and a population prevalence of childhood ADHD = 0.05[1,64], persistent ADHD = 0.03[3] and late-diagnosed ADHD = 0.03. To test for differences in h^2_{snp} estimates among groups, we also derived estimates by using non-shared controls (control numbers: childhood controls, 24,443; persistent controls, 2,289; and late-diagnosed controls, 11,571). Testing for difference in $h^2_{snp}$was done by using a z-test. Additionally, h^2_{snp} in the sub-groups was estimated over a range of population prevalence values spanning from 1% to 15%.

Genetic correlations between ADHD subgroups were calculated by using bivariate GREML analysis in GCTA and non-shared controls.

PGS analyses of ADHD and other phenotypes

The PGSs for ADHD were generated by using a 5-fold cross-validation approach, similarly to what we did previously[4]. In short, the sample was split into five groups, aiming for equal numbers of ADHD cases and controls within each group. We then conducted a GWAS using four out of five groups to derive effect sizes with respect to ADHD risk. These effect sizes were then used to estimate the PGS for the remaining target group. Thus, the training data were independent of the target data. This procedure was repeated five times until PGSs were estimated in all target groups. Indels and variants in the extended major histocompatibility complex region (chromosome 6: 25-34 Mb) were also removed. Only independent variants were used to generate the score and clumping of the training data was done on the summary statistics by employing PLINK and the flags -clump-p1 1, -clump-p2 1, -clump-r2 0.1 and -clump-kb 500. PGS was estimated for each individual in the target sample by using a range of P-value thresholds in the training data (5×10^{-8}, 1×10^{-6}, 1×10^{-4}, 1×10^{-3}, 0.01, 0.05, 0.1, 0.2, 0.5 and 1.0), multiplying the natural log of the OR of each variant by the allele-count of each variant. The whole-genome PGS was obtained by summing values over variants for each individual. The PGSs were standardised for each of the five target sample groups (subtracting the mean and dividing by the standard deviation). The scores from the five target groups were then pooled at each threshold and tested for association with general ADHD (i.e., all cases vs controls) and the P-value threshold with the scores explaining the maximum variance (estimated by Nagelkerke's R^2) in the target data of general ADHD (i.e., all ADHD cases vs controls) was used to test for differences in the ADHD-PGS load across ADHD sub-groups ($P_{threshold}<0.1$). Because our ADHD cohort is population-based, including all individuals with ADHD born in Denmark between 1981 and 2008 and diagnosed before or in 2016, the generated PGS represents the general liability to diagnosed ADHD (because the cross-validation approach is based on all population-based cases). We would like to stress that the PGS only reflects the general liability with respect to diagnosed ADHD as some cases potentially might be missed by the diagnostic system (see Supplementary Information, note 2).

PGSs for ADHD symptoms and 13 other phenotypes (schizophrenia, autism, bipolar disorder, alcohol use disorder, cannabis use disorder, OCD, anorexia, depression, educational years, mother's age at death, BMI, age at first birth and insomnia) were generated by using summary statistics from large GWAS of the phenotypes (see Supplementary Table 9 for references) and the approach described above (without 5-fold cross validation). The data on ADHD symptoms (inattention and hyperactivity/impulsivity) come from a genome-wide association meta-analysis on up to 43,117 children and adolescents[22]. In the PGS analyses, P-value thresholds in the training GWAS that captured most variance (estimated by Nagelkerke's R^2) in the target data were used as thresholds for analyses of the PGS load in the subgroups (threshold information in Supplementary Table 9).

We tested for differences in the PGS load among ADHD subgroups by using multi-nominal regression in R v. 3.6.0 and the packages 'multcomp' and ´fmsb´, with ADHD coded as four factors: controls, childhood, adulthood and persistent ADHD, and we included covariates to correct for genotyping round (iPSYCH1 or iPSYCH2), sex and 10 ancestry PCs. Correction for multiple testing was done separately for the following three analyses: (1) the PGS-ADHD load among subgroups correcting for three pair-wise comparisons, (2) the PGS load for ADHD symptoms (inattention and hyperactivity) correcting for six pair-wise comparisons and (3) the PGS load for 13 other phenotypes correcting for 39 pair-wise comparisons.

We also performed two sensitivity PGS analyses. (1) We evaluated whether the differences in the PGS load of four major psychiatric disorders (depression, schizophrenia, bipolar disorder and autism) could be caused primarily by individuals with comorbidity. We excluded all individuals in the target sample with the diagnosis being analysed – that is, all depression cases were excluded in the analysis of depression-PGS (sample sizes are given in Supplementary Table 10.B). (2) We evaluated the potential genetic heterogeneity in the childhood group caused by age. This was done by splitting the childhood group into two groups, those younger than 18 years of age (N=8,664) and

those older than 18 years of age (N=6,214) by the end of follow-up. We then redid the PGS analyses including the two childhood ADHD groups, persistent and late-diagnosed ADHD.

PGS analysis in the Spanish cohort[19] consisting of 453 individuals with childhood ADHD, 270 with persistent ADHD, 889 with late-diagnosed ADHD and 3,440 controls was done by following the same approach described above (See Supplementary Information, note 3, for patient exclusion criteria).

Genetic correlations with ADHD symptoms and other phenotypes

Genetic correlations (r_g) of the three ADHD subgroups with ADHD symptoms[22] and the 13 phenotypes listed above were calculated by using summary statistics from GWAS and LD score regression[65]. No sample overlap and no population stratification were assumed when calculating r_g with ADHD symptoms and therefore the intercept was restricted by setting the single-trait intercepts to 1 and cross-trait intercepts to 0.

Statistical difference between two r_g estimates was calculated by using the block jackknife method implemented in the LD score regression software[65,66]. The genome was divided into 200 blocks and jackknife deleted values were calculated by excluding one block at a time. The jackknife deleted values were then used to calculate corresponding jackknife pseudovalues. Based on the mean and variance of the jackknife pseudovalues, Z-score and corresponding P-values were computed, testing the null hypothesis that the difference between the genetic correlations is equal to zero. A z-test was used to test whether the genetic correlation differed from 1.

Correction for multiple pair-wise comparisons was done separately for the following two evaluations: (1) difference in r_g of ADHD subgroups with ADHD symptoms correcting for six pair-wise comparisons and (2) r_g difference of ADHD subgroups with 13 other phenotypes correcting for 39 pair-wise comparisons.

Burden of rare variants in ADHD subgroups

WES data were available for a subset of the iPSYCH samples. It has previously been shown that WES of DNA from dried blood spots results in high-quality data[67]. DNA was extracted from dried blood spot samples of the study subjects and whole genome amplified in triplicates[52,53], the coding regions of the genome were extracted by using the Illumina Nextera capture kit and sequencing was performed in multiple waves (Pilot 1, Wave 1, Wave 2 and Wave 3) by using the Illumina HiSeq platform at the Genomics Platform of the Broad Institute.

Part of the data (Pilot 1, Wave 1, Wave 2) were also included in the recent study by Satterstrom et al.[36], in which the authors examined the overall burden of rPTVs in ADHD, and the same QC procedure was used in this study, including all data (Pilot 1, Wave 1, Wave 2 and Wave 3). In short, the raw sequencing data were aligned to the reference genome Hg19 using BWA[68]. Calling of genotypes was done by using the best practice recommended by the Genome Analysis Toolkit[69] (GATK) v.3.4. Most QC steps were performed with Hail 0.1 (Hail Team, Hail 0.2, https://github.com/hail-is/hail). All variants annotated to American College of Medical Genetics (ACMG)[70] genes were removed due to Danish legislation. Samples were removed if they lacked complete phenotype information, sex inconsistencies of the imputed sex with the reported sex, if they were duplicates or genetic outliers identified by PCA, if they had an estimated level of contamination >5% or if they had an estimated level of chimeric reads >5%.

Only autosomal genotypes were included in our analyses. Genotypes were removed if they did not pass GATK variant quality score recalibration (VQSR) or had a read depth <10 or >1,000. Homozygous alleles were removed if they had reference calls with genotype quality <25, homozygous alternate alleles with PL(HomRef) (i.e., the phred-scaled likelihood of being homozygous reference) <25 or <90% reads supporting alternate allele. Heterozygous alleles were

removed if they had PL(HomRef) <25 or <25% reads supporting the alternate allele, <90% informative reads (i.e. number of reads supporting the reference allele plus number of reads supporting the alternate allele <90% of the read depth) or a probability of the allele balance calculated from a binomial distribution centred on 0.5 less than $1x10^{-9}$. After the application of these basic genotype filters, variants with a call rate <90% were removed, then samples with a call rate <95% and then variants with a call rate <95% were removed. Between the sample call rate filter and the final variant call rate filter, one of each pair of related samples was removed, defining relatedness as individuals with a pair-wise pi-hat value ≥0.2. After QC, the number of individuals were: childhood ADHD, N=4,987; persistent ADHD, N=748; late-diagnosed ADHD, N=1,915; and controls, N=8,649.

Following QC, variants were annotated by using SnpEff [71] version 4.3t. The variants were also annotated with the gnomAD[72] exomes r2.1.1 database by using SnpSift[71] version 4.3t. Variants were only included if they were located in consensus high-confident coding regions with a high read depth in both the iPSYCH data and the gnomAD dataset (80% of the samples in both datasets had at least 10× sequencing coverage in the region). Variants were defined as rPTVs if they were annotated as having large effects on gene function (nonsense variant, frameshift, splice site) and were rare in the sample, defined as having an allele count ≤5 across the combined counts in iPSYCH (N=16,299) and non-Finnish Europeans in the nonpsychiatric gnomAD exome database (N=44,779).

The burden of rPTVs and rSYNs in ADHD subgroups and controls was tested in (1) highly constrained genes (N=3,488), defined as genes being highly intolerant to loss-of-function mutations having a pLI score >0.9[37], and (2) *de novo* highly constrained genes (N=241), defined as highly constrained genes that overlap with another set of genes (N=285) previously found to be enriched with *de novo* mutations in individuals with neurodevelopmental disorders[38] (3) for comparison, we also tested a set of 9,662 evolutionary less constrained genes with a pLI score <0.1. Testing for

enrichment in rPTVs and rSYNs variants was done by using multiple logistic regression with the three ADHD groups and controls included in the same regression model (using R v. 3.6.0 and the R packages foreign, nnet, ggplot2, reshape2). The outcome (or dependent) variable was rPTV count and the predictor (or independent) variables were ADHD given as a categorical variable with multiple factors –: childhood, persistent, late-diagnosed and controls (with controls as the reference factor) – and relevant covariates – rPTV counts ~ ADHD (controls | childhood | persistent | late-diagnosed) + covariate 1 + covariate 2 + … + covariate N). ADHD was coded as the independent variable rather than the dependent variable to make pair-wise comparisons between ADHD subtypes in the same analysis. Covariates were: birth year, sex, the first 10 PCs from ancestry PCA, the number of rSYNs, the percentage of target with coverage >20x, the mean read depth at sites within the exome target passing VQSR, the total number of variants and the sequencing wave. Testing for enrichment of rPTVs in ADHD subgroups compared with controls was corrected for nine tests (three groups × three gene sets, i.e. new P-value threshold=0.006), and testing for differences between groups was corrected for nine pair-wise comparisons (three gene-sets × three pair-wise comparisons for each set).

Statistics and reproducibility

GWAS of ADHD subgroups were performed by logistic regression with an additive model of imputed dosage to estimate the association of the effect allele with childhood, persistent and late-diagnosed ADHD. Differences in the PGS load among ADHD subgroups was tested by using multi-nominal regression. Genetic correlations were calculated using LD score regression and statistical difference between two r_g estimates was calculated by using the block jackknife method. Analysis of the burden of rare variants in the three ADHD subgroups was done using multiple logistic regression with the three ADHD groups and controls included in the same regression model. All analyses were corrected using relevant covariates and Bonferroni correction was applied when appropriate (see Methods

section for details). No statistical method was used to determine sample size. The sample size was fixed, since we initially (i.e., before QC) included all individuals born in Denmark between 1981 and 2008 and diagnosed with ADHD in 2016 or before.

DATA AVAILABILITY

Summary statistics from GWAS of childhood, persistent and late-diagnosed ADHD are available at the iPSYCH website (https://ipsych.dk/en/research/downloads/). All relevant iPSYCH data are available from the authors after approval by the iPSYCH Data Access Committee and can only be accessed on the secured Danish server (GenomeDK [https://genome.au.dk]) as the data are protected by Danish legislation. For data access please contact: Ditte Demontis or Anders D. Børglum (anders@biomed.au.dk). Correspondence and requests for materials should be addressed to**:** Ditte Demontis, ditte@biomed.au.dk.

CODE AVAILABILITY

No previously unreported custom computer code or algorithm were used to generate results, all software used in the study are publicly available from the Internet as described in Methods and Reporting Summary.

Methods only references

48 Pedersen, C. B., Gotzsche, H., Moller, J. O. & Mortensen, P. B. The Danish Civil Registration System. A cohort of eight million persons. *Danish Medical Bulletin* **53**, 441-449 (2006).

49 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).

50 Nørgaard-Pedersen, B. & Hougaard, D. M. Storage policies and use of the Danish Newborn Screening Biobank. *Journal of Inherited Metabolic Disease* **30**, 530-536 (2007).

51 Pedersen, C. B. The Danish Civil Registration System. *Scand J Public Health* **39**, 22-25, doi:10.1177/1403494810387965 (2011).

52 Borglum, A. D. *et al.* Genome-wide study of association and interaction with maternal cytomegalovirus infection suggests new schizophrenia loci. *Molecular psychiatry* **19**, 325-333, doi:10.1038/mp.2013.2 (2014).

53 Hollegaard, M. V. *et al.* Robustness of genome-wide scanning using archived dried blood spot samples as a DNA source. *BMC Genet* **12**, 58, doi:10.1186/1471-2156-12-58 (2011).

54 Illumina. illumina GenCall Data Analysis Software. *Illumina Tech Note* (2005).

55 Korn, J. M. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nature genetics* **40**, 1253-1260, doi:10.1038/ng.237 (2008).

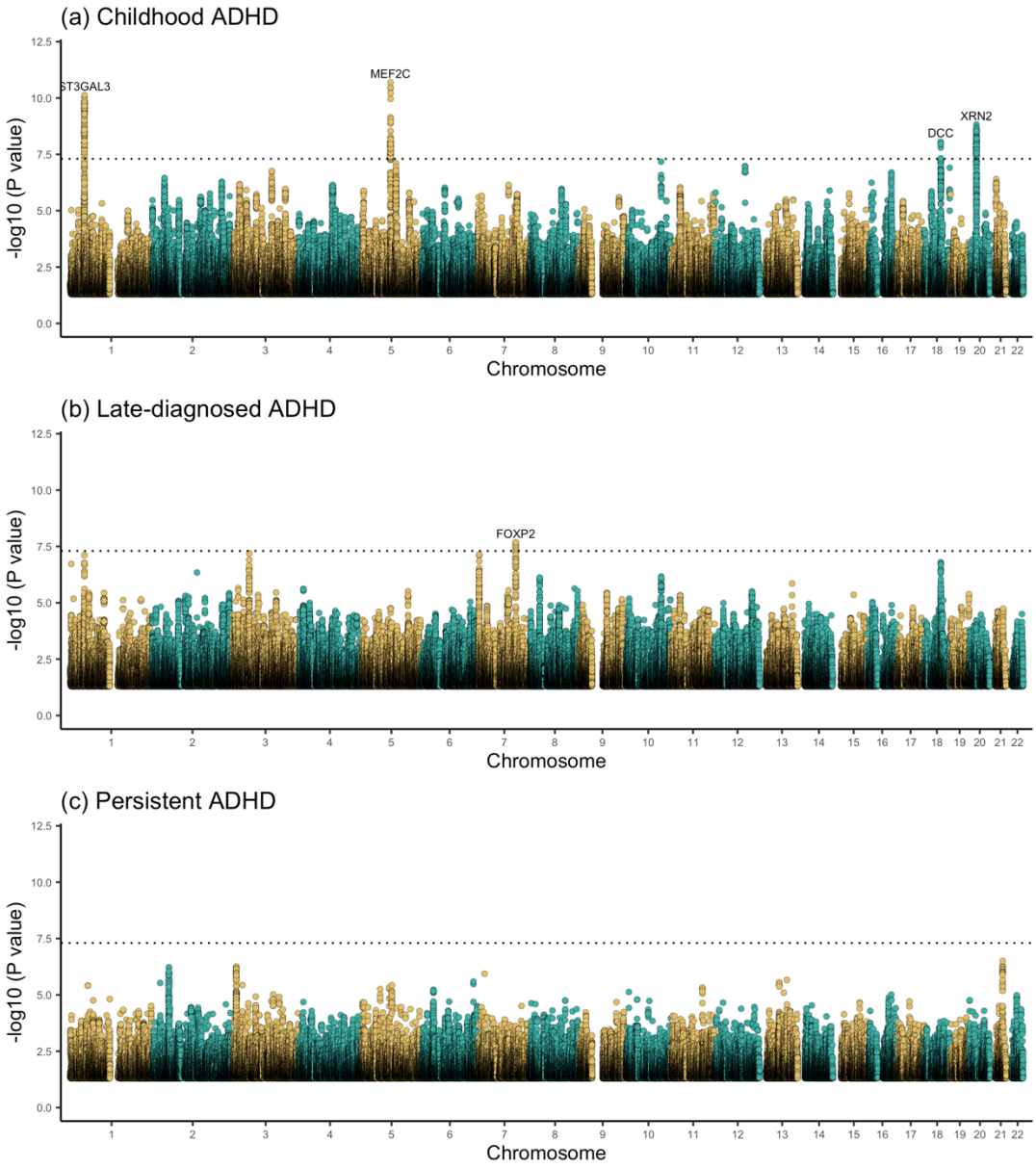
56 Iglesias, A. I. *et al.* Haplotype reference consortium panel: Practical implications of imputations with large reference panels. *Hum Mutat* **38**, 1025-1032, doi:10.1002/humu.23247 (2017).
57 Loh, P. R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nature genetics* **48**, 1443-1448, doi:10.1038/ng.3679 (2016).
58 Das, S. *et al.* Next-generation genotype imputation service and methods. *Nature genetics* **48**, 1284-1287, doi:10.1038/ng.3656 (2016).
59 Price, A. L. *et al.* Long-range LD can confound genome scans in admixed populations. *American journal of human genetics* **83**, 132-135; author reply 135-139, doi:10.1016/j.ajhg.2008.06.005 (2008).
60 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).
61 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).
62 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).
63 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).
64 Dalsgaard, S. *et al.* Incidence Rates and Cumulative Incidences of the Full Spectrum of Diagnosed Mental Disorders in Childhood and Adolescence. *JAMA Psychiatry* **77**, 155-164, doi:10.1001/jamapsychiatry.2019.3523 (2020).
65 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).
66 Hubel, C. *et al.* Genomics of body fat percentage may contribute to sex bias in anorexia nervosa. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **180**, 428-438, doi:10.1002/ajmg.b.32709 (2019).
67 Poulsen, J. B. *et al.* High-Quality Exome Sequencing of Whole-Genome Amplified Neonatal Dried Blood Spot DNA. *PloS one* **11**, e0153253, doi:10.1371/journal.pone.0153253 (2016).
68 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:btp324/10.1093/bioinformatics/btp324 (2009).
69 Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **11**, 11 10 11-11 10 33, doi:10.1002/0471250953.bi1110s43 (2013).
70 Kalia, S. S. *et al.* Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med* **19**, 249-255, doi:10.1038/gim.2016.190 (2017).
71 Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80-92, doi:10.4161/fly.19695 (2012).
72 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443, doi:10.1038/s41586-020-2308-7 (2020).

Supplementary Information

Differences in the genetic architecture of common and rare variants in childhood, persistent and late-diagnosed attention deficit hyperactivity disorder

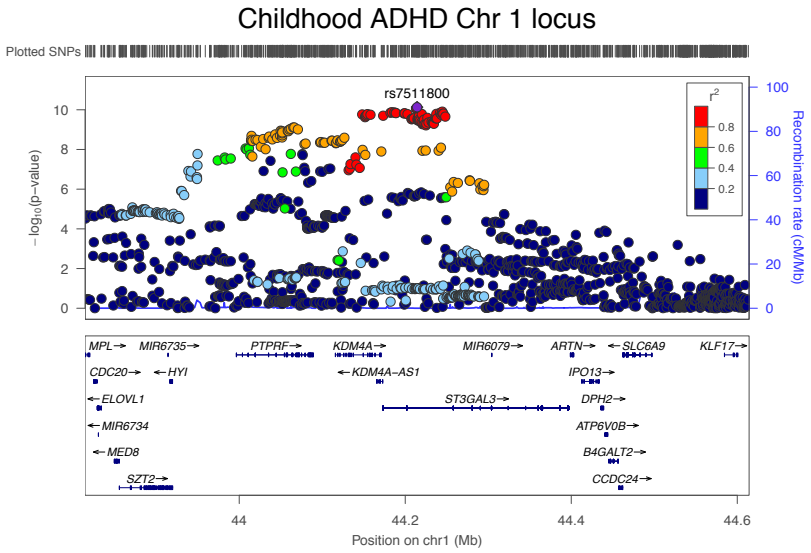
Supplementary figures 1-7 and supplementary notes 1-3

Supplementary Figure 1. Manhattan plots from GWAS of ADHD subgroups
Results from logistic regression corrected for sex and ancestry principal components 1-10. Y-axes represent two-sided -log(P-values) and x-axes represent location on autosomal chromosomes. The red horizontal line represents the threshold for genome-wide significant association ($P = 5x10^{-8}$) **(a)** GWAS of childhood ADHD (14,878 cases; 38,303 controls) **(b)** GWAS of late-diagnosed ADHD (6,961 cases; 38,303 controls) **(c)** GWAS of persistent ADHD (1,473 cases; 38,303 controls).

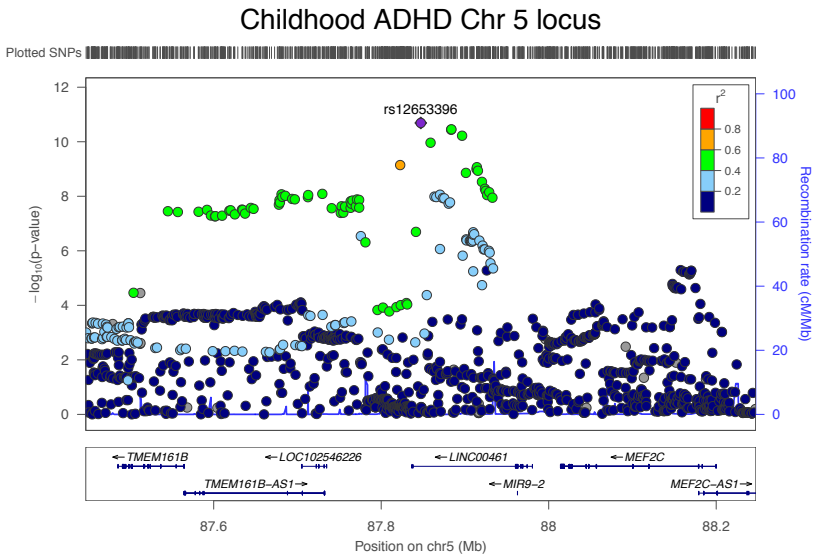


Supplementary Figure 2. Regional association plots of genome-wide significant loci
Regional association plots of the local association results from the GWAS of ADHD sub-groups **(a-d)** the four genome-wide significant loci identified in the GWAS of childhood ADHD (14,878 cases; 38,303 controls) **(e)** the genome-wide significant locus identified in the GWAS of late-diagnosed ADHD. The y-axis represents –log(P-values) of variant association; the P-values are two-sided from logistic regression corrected using relevant covariates. Location and orientation of the genes in the regions are indicated on the X-axis, LD estimates of surrounding SNPs with the index SNP (r^2 values estimated based on 1KGP3) is indicated by colour (colour bar in upper left corner indicates r^2 values). Additionally, the local estimation of recombination rate is indicated in blue (legend on vertical axis at right).

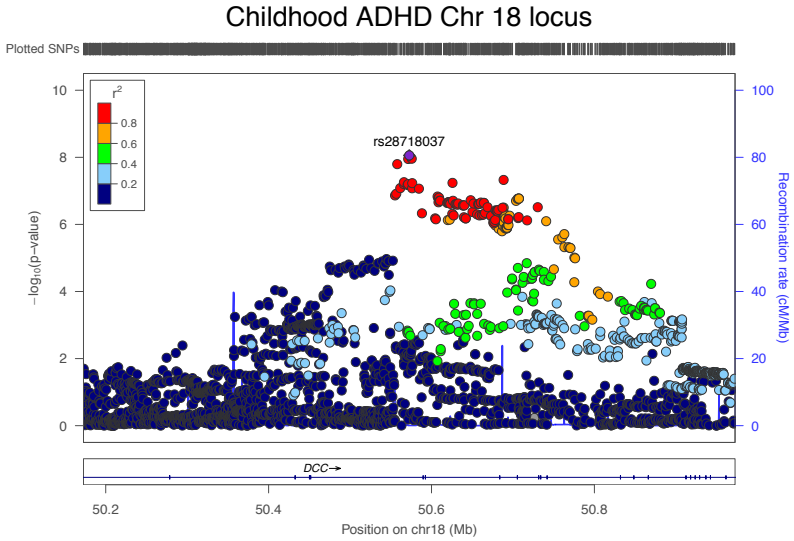
a.



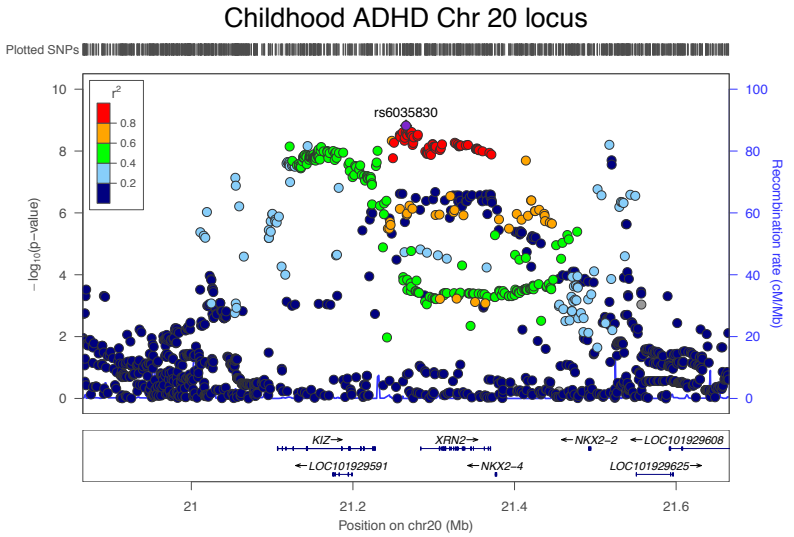
b.



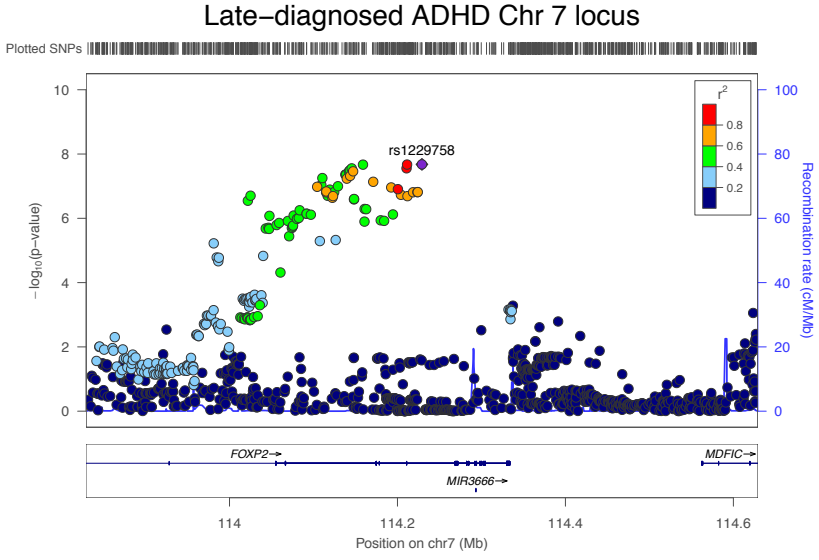
c.


Childhood ADHD Chr 18 locus
Plotted SNPs
rs28718037
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
DCC
50.2
50.4
50.6
50.8
Position on chr18 (Mb)

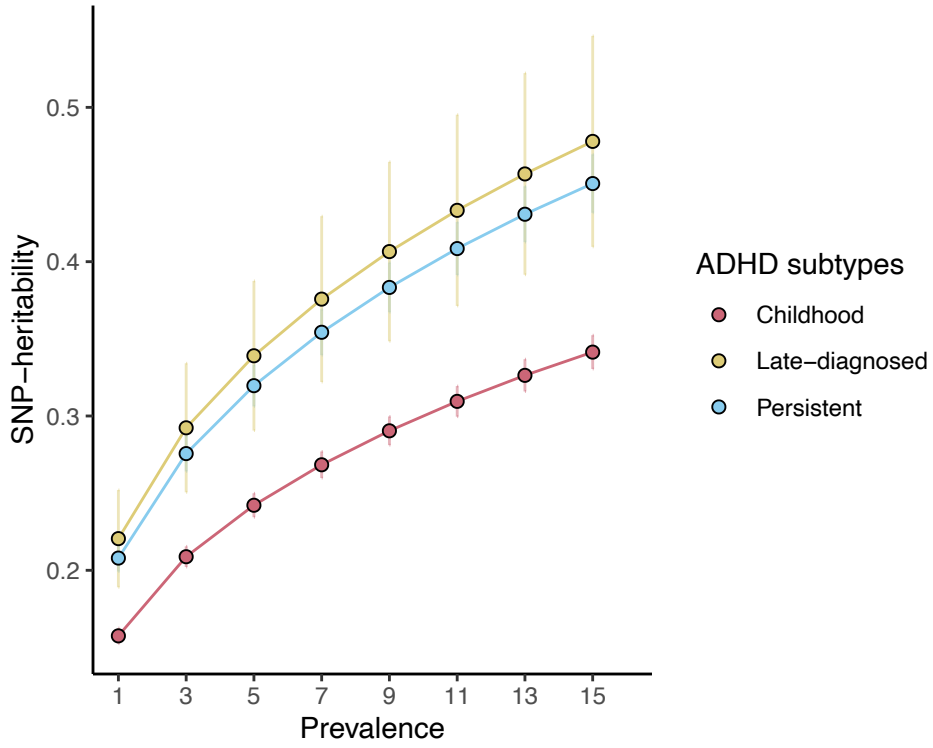
d.


Childhood ADHD Chr 20 locus
Plotted SNPs
rs6035830
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
KIZ
XRN2
NKX2−2
LOC101929608
LOC101929591
NKX2−4
LOC101929625
21
21.2
21.4
21.6
Position on chr20 (Mb)

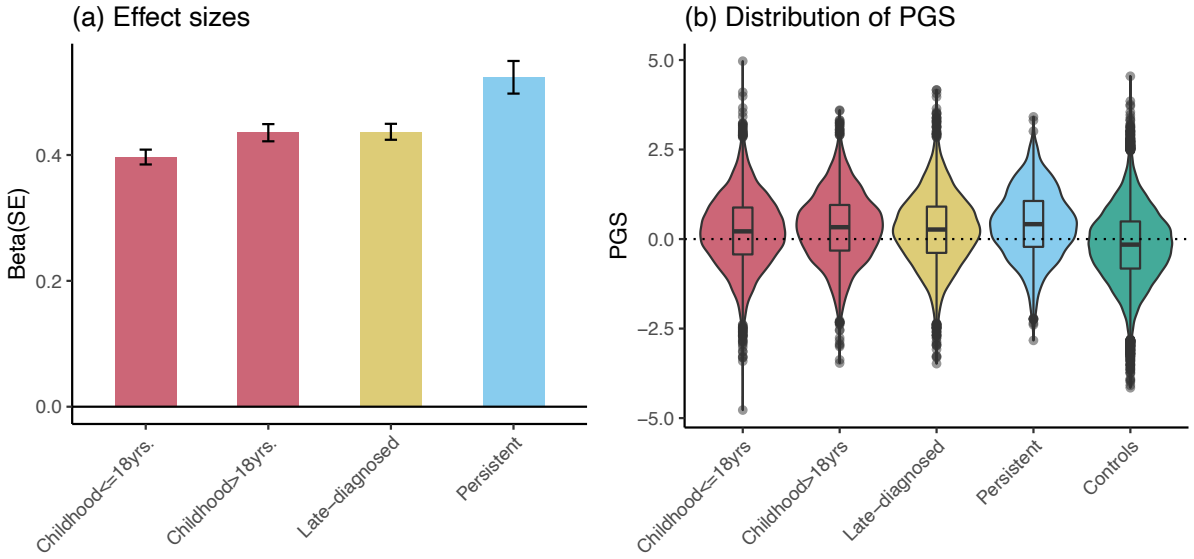
e.


Late−diagnosed ADHD Chr 7 locus
Plotted SNPs
rs1229758
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
FOXP2
MDFIC
MIR3666
114
114.2
114.4
114.6
Position on chr7 (Mb)

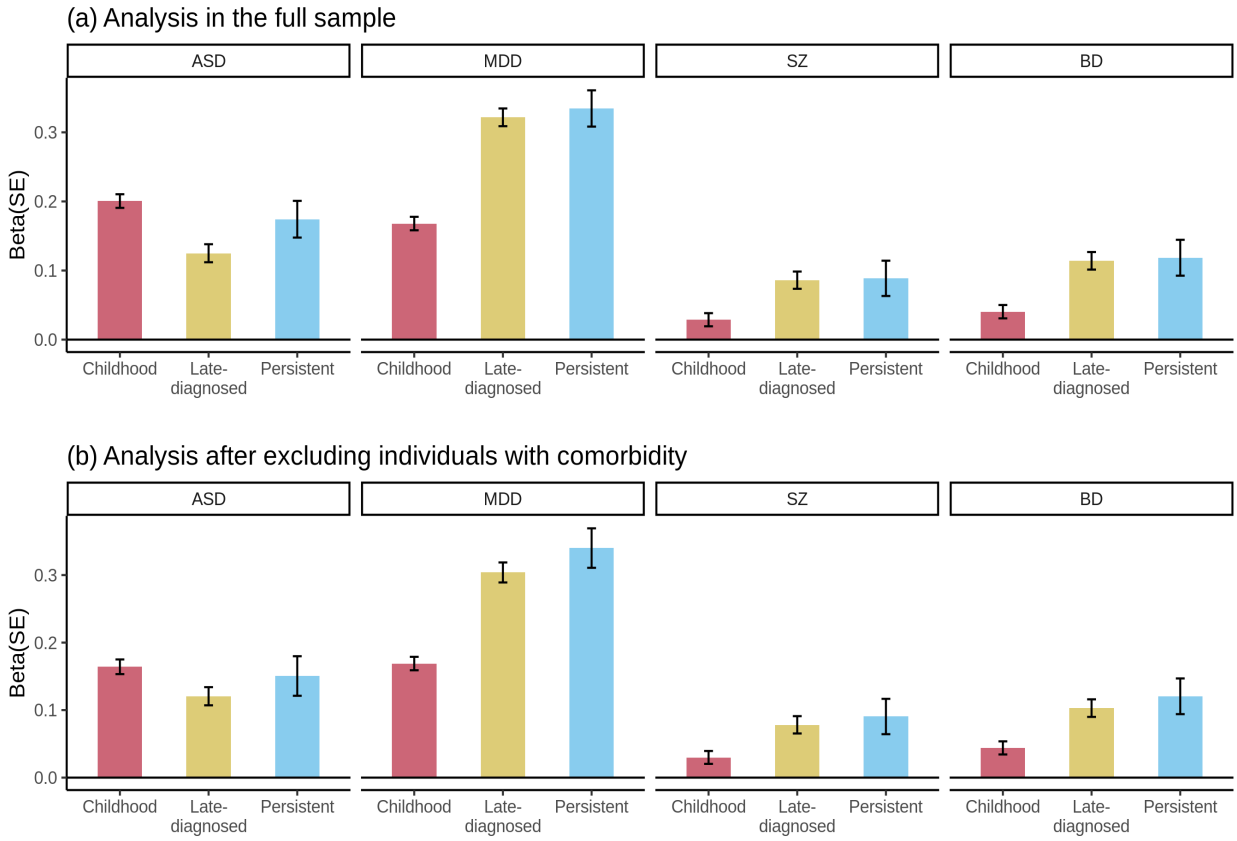
Supplementary Figure 3. SNP-heritability (h2) estimates in childhood (N=14,878), persistent (N=1,473) and late-diagnosed ADHD (N=6,961) compared to controls (N=38,303). The h2 estimates are indicated with dots and are calculated using GCTA. Vertical bares represent standard errors.



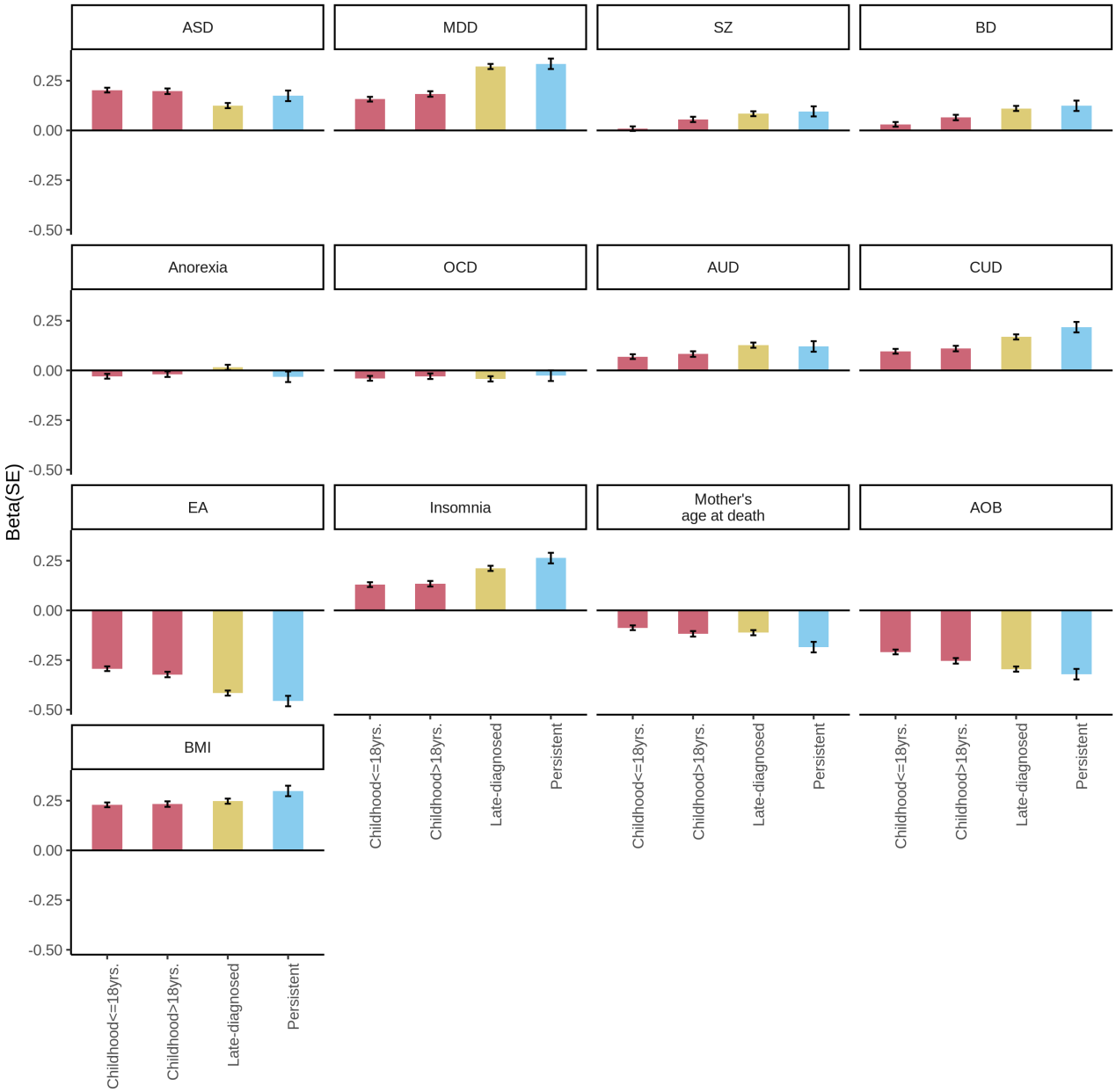
Supplementary Figure 4. ADHD-PGS load in childhood ADHD individuals younger than 18 years of age (N=8,664), childhood ADHD individuals older than 18 years of age (N= 6,214), Persistent ADHD (N=1,473), late diagnosed ADHD (N=6,961) compared to controls (N=38,303) **(a)** The beta from multiple-nominal regression is given on the y-axis. Vertical bares represent standard errors **(b)** Distribution of the PGS in the four groups are displayed in the violin plots. The vertical line in the box represents the mean PGS, the box represents the interquartile range, and the thin lines represent the rest of the distribution except for dots determined to be outliers.



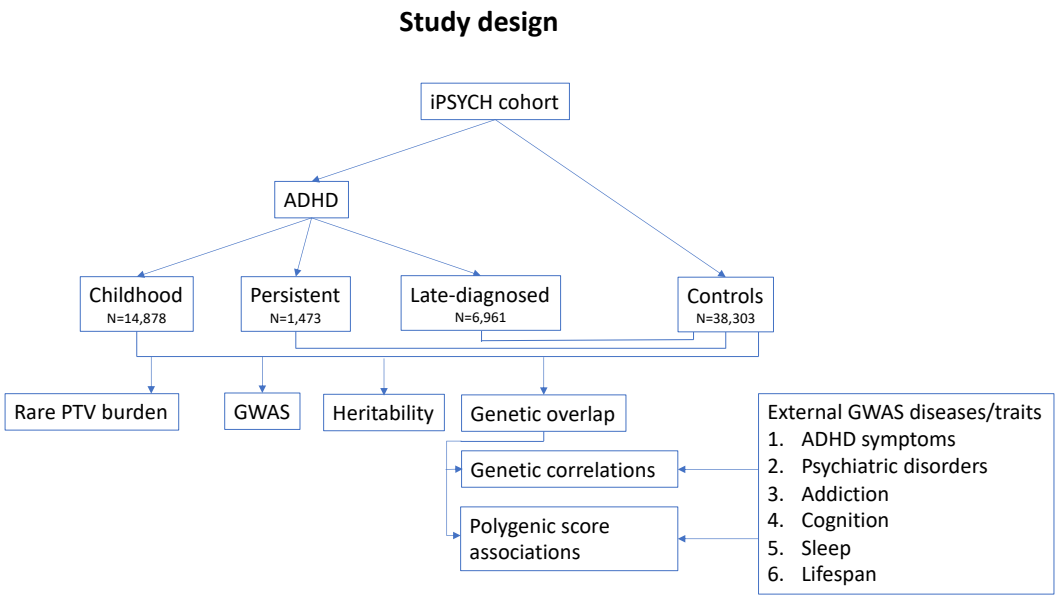
Supplementary Figure 5. Results from PGS analyses, demonstrating the association of PGS for autism (ASD), depression (MDD), schizophrenia (SZ), and bipolar disorder (BD) with childhood, late-diagnosed and persistent ADHD compared to controls in **(a)** the full sample (childhood ADHD N=14,878, late-diagnosed ADHD N=6,961, persistent ADHD N=1,473) and **(b)** in samples where individuals are excluded if they are diagnosed with the disorder of the analyzed PGS (sample sizes can be found in Supplementary Table 10.B).The beta from multiple-nominal regression is given on the y-axis. Vertical bares represent standard errors.



Supplementary Figure 6. Results from PGS analysis demonstrating the association of PGS with childhood ADHD individuals younger than 18 years of age (N=8,664), childhood ADHD individuals older than 18 years of age (N= 6,214), persistent ADHD (N=1,473) and late-diagnosed ADHD. (N=6,961). PGS for psychiatric disorders: autism spectrum disorder (ASD), major depressive disorder (MDD), schizophrenia (SZ), bipolar disorder (BD), anorexia, obsessive compulsive disorder (OCD), alcohol use disorder (AUD), cannabis use disorder (CUD). PGS for five phenotypes representing domains highly correlated with ADHD: educational attainment (EA), insomnia, mother´s age at death and age of first birth (AOB). On the y-axis is the beta from multi-nominal regression against controls (N=38,303), vertical bares represent standard errors (see also Supplementary Table 11).



Supplementary Figure 7. Flow chart demonstrating the genetic analyses that was performed in the study.



Supplementary note 1. Genetic correlation with mother´s age at death

We found a stronger negative genetic correlation of mother's age at death with persistent and late-diagnosed ADHD compared with childhood ADHD. The phenotype mother's age at death captures all variants that affect longevity, for example, variants associated with cardiovascular diseases[1], and our results suggest that ADHD in adulthood is more enriched in variants that decrease longevity than childhood ADHD.

Supplementary note 2. Cases potentially missed by the diagnostic system

We acknowledge that we might have missed cases that never came into contact with the health care system or cases that only briefly interacted with the system and never got diagnosed. This could be due to various reasons such as lack of resources (mentally or physically) to engage with the system, lack of persistency if the diagnostic investigation took too long or unequal access to mental health care. Regarding the latter, we think Denmark, world-wide is one of the countries with least bias caused by unequal access to health care. Denmark is a small welfare

state with health care facilities distributed across the country and citizens have equal access to the health care system that is free of charge.

Additionally, ADHD is considered a childhood-onset disorder and therefore, and according to the current diagnostic criteria, individuals diagnosed with ADHD as adults should be able to describe ADHD symptoms in childhood retrospectively. We would therefore like to note that some adults with ADHD might be missed if they were not able to recall having ADHD symptoms in childhood. In line with this, we also acknowledge that late-diagnosed ADHD probably is more likely to be missed than childhood and persistent ADHD, because the disorder seems to be underdiagnosed among adults[65].

Supplementary note 3. Patient exclusion criteria in the Spanish cohort

Exclusion criteria for patients in the Spanish ADHD cohort were the following: intelligence quotient (IQ) <70, having pervasive developmental disorders, schizophrenia or other psychotic disorders, adoption, sexual or physical abuse, birth weight <1.5 kg, and any significant neurological or systemic disease that might explain ADHD symptoms. Comorbid oppositional defiant disorder, conduct disorder, depression and anxiety disorders were allowed unless determined to be the primary cause of ADHD symptomatology.

1 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).