Title
Novel Loci Associated With Attention-Deficit/Hyperactivity Disorder Are Revealed by Leveraging Polygenic Overlap With Educational Attainment

Permalink
https://escholarship.org/uc/item/7153w9nx

Journal
JOURNAL OF THE AMERICAN ACADEMY OF CHILD AND ADOLESCENT PSYCHIATRY, 57(2)

ISSN
0890-8567

Authors
Shadrin, AA
Smeland, OB
Zayats, T
et al.

Publication Date
2018-02-01

DOI
10.1016/j.jaac.2017.11.013

Peer reviewed
Objective: Attention-deficit/hyperactivity disorder (ADHD) is a common and highly heritable psychiatric condition. By exploiting the reported relationship between ADHD and educational attainment (EA), we aimed to improve discovery of ADHD-associated genetic variants and to investigate genetic overlap between these phenotypes.

Method: A conditional/conjunctional false discovery rate (condFDR/conjFDR) method was applied to genome-wide association study (GWAS) data on ADHD (2,064 trios, 896 cases, and 2,455 controls) and EA (n = 328,917) to identify ADHD-associated loci and loci overlapping between ADHD and EA. Identified single nucleotide polymorphisms (SNPs) were tested for association in an independent population-based study of ADHD symptoms (n = 17,666). Genetic correlation between ADHD and EA was estimated using LD score regression and Pearson correlation.

Results: At levels of condFDR < 0.01 and conjFDR < 0.05, we identified 5 ADHD-associated loci, 3 of these being shared between ADHD and EA. None of these loci had been identified in the primary ADHD GWAS, demonstrating the increased power provided by the condFDR/conjFDR analysis. Leading SNPs for 4 of 5 identified regions are in introns of protein coding genes (KDM4A, MEF2C, PINK1, RUNX1T1), whereas the remaining one is an intergenic SNP on chromosome 2 at 2p24. Consistent direction of effects in the independent study of ADHD symptoms was shown for 4 of 5 identified loci. A polygenic overlap between ADHD and EA was supported by significant genetic correlation (r_g = −0.403, p = 7.90 × 10^−8) and >10-fold mutual enrichment of SNPs associated with both traits.

Conclusion: We identified 5 novel loci associated with ADHD and provided evidence for a shared genetic basis between ADHD and EA. These findings could aid understanding of the genetic risk architecture of ADHD and its relation to EA.

Key words: attention-deficit/hyperactivity disorder, educational attainment, conditional/conjunctional false discovery rate, genetic overlap

ADHD is consistently associated with lower levels of EA\(^1\)\(^{-14}\): the percentage of US adolescents not completing high school is 5%, whereas it is approximately 35% for adolescents diagnosed with ADHD.\(^{15}\) There are several ways in which ADHD may relate to lower EA, which are not mutually exclusive. First, the clinical and cognitive symptoms of ADHD (e.g., attention deficits) may directly perturb EA. Second, ADHD has a number of common comorbidities, including learning disabilities,\(^16\) mood disorders,\(^16\) and disruptive behavior,\(^16\) which are associated with lower EA. Another possibility is that ADHD and EA share causative factors. Recent findings demonstrate negative genetic correlation between ADHD and EA (\(r_g = -0.305, \text{SE} = 0.141, \rho = 3.00 \times 10^{-2}\))\(^17\), suggesting that genetic variants conferring risk for ADHD may contribute to lower EA in the general population. Thus, we can hypothesize that ADHD and EA may have a shared genetic basis and may amplify association signal by combining these phenotypes in the condFDR/conjFDR method.

In contrast to ADHD, where the currently published largest GWASs contain fewer than 4,000 cases,\(^18,19\) the latest GWAS on EA contains more than 300,000 individuals, uncovering multiple genome-wide significantly associated loci.\(^20\) Combining this EA GWAS with moderately powered GWAS of ADHD\(^18\) in the condFDR/conjFDR approach, we aimed here to identify novel loci associated with ADHD as well as loci shared between ADHD and EA. The latter may provide insights into the molecular genetic mechanisms jointly influencing ADHD and EA and inform their biological underpinnings. Applying novel statistical methods, we also tested whether the observed phenotypic correlation between ADHD and EA implies a genetic correlation between these traits. In addition, for the identified ADHD-associated variants, we assessed consistency of effect directions in an independent population-based study of ADHD symptoms and performed in silico analyses of their functional effects (eQTL, expression quantitative trait loci).

**METHOD**

**Participant Samples**

We used ADHD data from the Psychiatric Genomics Consortium (PGC).\(^{18}\) The data set contains information from 2,064 trios, 896 individuals with ADHD, and 2,455 controls. EA data were obtained from the Social Science Genetic Association Consortium (SSGAC),\(^{20}\) where EA was measured as the number of years of schooling completed that was harmonized between different educational systems. For our analyses, we used summary statistics generated by the meta-analysis of all discovery and replication cohorts, except the 23andMe sample (64 datasets with total \(n = 328,917\)).

Top association signals identified in our analyses were examined in the summary statistics from an independent GWAS of ADHD symptoms performed by EArly Genetics and LifeCourse Epidemiology (EAGLE) consortium.\(^21\) Unlike the PGC case-control ADHD GWAS, EAGLE GWAS represents a meta-analysis of 9 population-based pediatric cohorts containing information on 17,666 children under the age of 13 years with measures of ADHD symptom scores.

Detailed description of data used for analysis and data preprocessing steps is given in Supplement 1, available online.

**Statistical Analyses**

To assess genetic overlap between ADHD and EA and thus warrant subsequent condFDR/conjFDR analysis, we generated conditional QQ plots and fold-enrichment plots in both directions: conditioning ADHD on EA and vice versa.\(^9\) To explore the nature of the polygenic overlap and to test the hypothesis that the investigated phenotypes correlate genetically, we calculated Pearson correlations between association \(z\) scores of ADHD and EA SNPs within nested subset (strata) of SNPs with increasing significance of \(p\) values in either ADHD or EA (formal definition of SNP stratum is given in Supplement 1, available online). To further support this hypothesis, we estimated genetic correlation between ADHD and EA using LD score regression.\(^8\) Details of these analyses are described in Supplement 1, available online.

To identify specific loci associated with ADHD, we applied the condFDR method described previously.\(^9\) The condFDR method takes summary statistics that reflect genetic association of a phenotype of interest (primary) together with those of an auxiliary (conditional) phenotype and estimates a posterior probability that an SNP is null (has no association) in the primary phenotype, given that \(p\) values of the SNP in both the primary and conditional phenotypes are lower than observed \(p\) values. Thus, the condFDR method increases the power to discover loci associated with a primary phenotype by leveraging associations with a secondary phenotype. It does so by re-ranking SNPs compared to nominal \(p\) value-based ranking.\(^9\) In contrast, ranking SNPs based on unconditional FDR (e.g., using Benjamini–Hochberg or Benjamini–Yekutieli procedure) does not change their order (compared to nominal \(p\) values).

Although both conditional QQ plots and genetic correlation based on the LD score regression can be useful to get a general idea of whether

---

**TABLE 1** Glossary of Core Terms Used in the Study

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWAS (genome-wide association study)</td>
<td>A study that performs genome-wide scans of common genetic variants aiming to identify variants associated with the trait.</td>
</tr>
<tr>
<td>LD (linkage disequilibrium)</td>
<td>The statistical correlation between alleles at 2 loci.</td>
</tr>
<tr>
<td>FDR (false discovery rate)</td>
<td>A posterior probability that an identified association is false.</td>
</tr>
<tr>
<td>condFDR (conditional FDR)</td>
<td>The method that uses association (p) values of 2 traits (primary and conditional) to estimate a posterior probability that a variant has no association in a primary trait, given that (p) values of the variant in both the primary and conditional traits are lower than their observed (p) values.</td>
</tr>
<tr>
<td>conjFDR (conjunctional FDR)</td>
<td>An extension of condFDR method that estimates a posterior probability that a variant has no association for either phenotype or both at the same time, given that the (p) values for both phenotypes are lower than the observed (p) values.</td>
</tr>
<tr>
<td>QQ (Quantile–Quantile)</td>
<td>A visual tool used to compare 2 probability distributions (e.g., observed vs. expected) by plotting their quantiles against each other.</td>
</tr>
<tr>
<td>Conditional QQ plot</td>
<td>A QQ plot comparing probability distributions of association (p) values in primary trait for different strata of variants (e.g., variants with different levels of significance in conditional trait).</td>
</tr>
<tr>
<td>eQTL (expression quantitative trait loci)</td>
<td>Genetic variants that affect gene expression levels.</td>
</tr>
</tbody>
</table>
2 traits have a significant genetic overlap, they are unable to find specific susceptibility loci shared by the traits. The conjFDR approach is an extension of condFDR allowing the identification of specific loci associated with both traits.\textsuperscript{9} The conjFDR is defined as the maximum of the 2 condFDR values (taking one phenotype as primary and another as conditional and vice versa) for a specific SNP. Thus, the conjFDR approach estimates a posterior probability that an SNP is null for either phenotype or both at the same time, given that the $p$ values for both phenotypes are lower than the observed $p$ values. The method, therefore, uncovers loci associated with both phenotypes simultaneously.

To avoid inflation of the results due to LD dependency in fold-enrichment and QQ plots as well as in condFDR/conjFDR analyses, we randomly pruned all SNPs across 500 iterations. For each iteration, all but one random SNP in each LD-independent region (clump of SNPs in strong LD, $r^2 > 0.2$) were removed, and finally the results were averaged across all iterations. LD ($r^2$ values) was estimated based on the 1000 Genomes Project phase 3 European subpopulation data using PLINK.\textsuperscript{22}

As for meta-analyses based on multiple data sources, the quality of our condFDR/conjFDR analysis will depend on the robustness of the primary data. More details about condFDR and conjFDR methods can be found in Supplement 1, available online, and in the original publication.\textsuperscript{9}

**RESULTS**

**Evaluation of the Detected ADHD Loci in an Independent Study of ADHD Symptoms**

We used genetic data on association of ADHD symptoms obtained from the EAGLE consortium to test whether our results could be supported by data from the independent sample. For this purpose, we examined whether effects of the most significant SNPs in the loci identified by condFDR/conjFDR analyses were consistent between PGC ADHD and EAGLE data sets.

**In Silico Identification of Allele-Specific Effects of Significant SNPs on Transcription**

Identifying and investigating genetic variants that might affect gene expression (expression quantitative trait loci or eQTLs) may shed light on how associated variants may contribute to biological mechanisms underlying a phenotype. eQTLs vary significantly both between different tissues and over time.\textsuperscript{23} Existing GWASs on ADHD and EA clearly demonstrate remarkable enrichment of association signals in genomic regions implicated in the regulation of gene expression in the brain.\textsuperscript{18,20} Hence, we focused on eQTL analysis of genes expressed in brain tissues. Significant associations identified with condFDR and conjFDR analyses were queried for known eQTLs using the GTEx portal (http://gtexportal.org) and the Braineac database (http://www.braineac.org). The latter database contains information on cis-eQTLs for 10 brain regions: cerebellar cortex, frontal cortex, hippocampus, medulla (specifically inferior olivary nucleus), occipital cortex (specifically primary visual cortex), putamen, substantia nigra, thalamus, temporal cortex, and intralobular white matter. In addition, we checked age-dependent variations of expression in genes containing identified significant SNPs using the Human Brain Transcriptome database (http://hbatlas.org).\textsuperscript{24}

**Evaluation of Genetic Overlap and Correlation**

In the absence of genetic overlap between 2 traits, it is expected that $p$ values for association with 1 trait are independent of the $p$ values for association with the other. However, conditional QQ plots in Figure 1 clearly demonstrate an increasing degree of leftward deflection for strata of more significant SNPs. This is observed both when conditioning ADHD on EA (Figure 1A) and vice versa (Figure 1B), suggesting substantial cross-trait polygenic enrichment. Enrichment of association signals for 1 trait among those of another is also clearly visible in the fold-enrichment plots, with more than 10-fold enrichment of SNPs from the strictest stratum ($q_{\text{conditional trait}} < 1.00 \times 10^{-7}$) for both traits (Figure S1, available online).

In addition, association $z$ scores of ADHD and EA demonstrate increasing negative correlation in more strictly defined strata of SNPs, both when strata are defined based on ADHD $p$ values (Figure 1C) and on EA $p$ values (Figure 1D). Moreover, LD score regression analysis also showed significant negative genetic correlation ($r_g = -0.403$, $SE = 0.075$, $p = 7.90 \times 10^{-8}$) between these phenotypes.

**Identification of ADHD-Associated Loci and Loci Shared Between ADHD and EA**

Using the condFDR/conjFDR method, we identified 5 LD-independent regions significantly associated with ADHD (condFDR $< 0.01$, conjFDR $< 0.05$), 3 of which were also identified as shared between ADHD and EA. From each of these regions, a single SNP with the lowest condFDR/conjFDR value (strongest association signal) was selected to represent their loci. These SNPs are presented in Table 2. Manhattan plots resulting from condFDR and conjFDR analyses are presented in Figures 2 and 3, respectively. Four of 5 identified most significant SNPs revealed the opposite directions of effect in ADHD and EA.

**Identified Loci and Related Genes**

Two loci (represented in Table 2 by variants rs618678 and rs412458) were identified both in condFDR and conjFDR analyses. Rs618678 was identified as the strongest signal in the conjFDR analysis (condFDR $= 3.82 \times 10^{-3}$ and the second strongest in the condFDR analysis (condFDR $= 3.77 \times 10^{-3}$). This SNP is an intronic variant within $KDM4A$ on chromosome 1p34.2 (Figure 4B).\textsuperscript{25} Figure 4B and Figure S2B (available online) show the genetic context of rs618678, indicating, respectively, the conjFDR and condFDR values of adjacent regions significantly enriched. Enrichment of association signals for 1 trait (Figure S2A, D, available online) demonstrated remarkable enrichment of association signals in genomic regions. More details about condFDR and conjFDR methods can be found in Supplement 1, available online, and in the original publication.\textsuperscript{9}

None of the SNPs identified either in condFDR or conjFDR reached genome-wide significance in previously published GWAS of ADHD.\textsuperscript{18} Rs618678 reached genome-wide significance in EA ($p = 1.05 \times 10^{-10}$),\textsuperscript{20} Rs412458, which was identified by both condFDR and conjFDR, was not reported as genome-wide significant (Figure 1A) and vice versa (Figure 1B), suggesting substantial cross-trait polygenic enrichment. Enrichment of association signals for 1 trait among those of another is also clearly visible in the fold-enrichment plots, with more than 10-fold enrichment of SNPs from the strictest stratum ($q_{\text{conditional trait}} < 1.00 \times 10^{-7}$) for both traits (Figure S1, available online).

In addition, association $z$ scores of ADHD and EA demonstrate increasing negative correlation in more strictly defined strata of SNPs, both when strata are defined based on ADHD $p$ values (Figure 1C) and on EA $p$ values (Figure 1D). Moreover, LD score regression analysis also showed significant negative genetic correlation ($r_g = -0.403$, $SE = 0.075$, $p = 7.90 \times 10^{-8}$) between these phenotypes.
by the published EA GWAS ($p = 3.73 \times 10^{-6}$), but it is in LD ($r^2 = 0.35$) with rs588282, which did reach genome-wide significance in that study (previously reported $p = 1.69 \times 10^{-10}$). Other loci identified in our analyses were below the genome-wide significance threshold in EA. It is also worth noting that the unconditional FDR values for all identified SNPs were above 0.01 and 0.05 in condFDR and conjFDR analysis, respectively.

Evaluation of the Detected ADHD Loci in an Independent Study of ADHD Symptoms

To assess the robustness of our results, we examined the loci identified in either the condFDR or conjFDR analyses (Table 2) in the association summary statistics from the independent GWAS of ADHD symptoms conducted by the EAGLE consortium.\(^2\) Four of 5 loci (represented by SNPs rs17441302, rs412458, rs618678, rs4324303) have the same direction of effect in the PGC and EAGLE GWASs, whereas the last locus (represented by rs4477079 SNP) has an opposite direction of effect in these GWASs. These results are presented in Table S1 (available online).

In Silico Identification of Allele-Specific Effects on Transcription

According to Human Brain Transcriptome data,\(^2\) all 6 implicated genes (Table 2, genes in the region) have a pronounced expression in different brain regions during the whole life cycle (Figure S3, available online). Therefore, alterations in the expression level of these genes (where the detected SNPs are located) may affect a broad variety of processes over an extended period. We scanned the Braineac database to check whether SNPs identified in either the condFDR or conjFDR analyses are...
associated with gene expression in brain tissues. We found that 4 of 5 SNPs from Table 2 may operate as eQTLs, significantly \( p < 0.001 \) associated with the expression of 13 different genes in several brain regions (Table S2, available online). Among those 13 genes, the most significant eQTL was observed between rs618678 and ST3GAL3. Furthermore, significant eQTL effects of rs618678 on ST3GAL3 were identified in muscle–skeletal tissue \( (p = 3.40 \times 10^{-5}) \) in the GTEx database (https://gtexproject.org/), but not in brain tissue.

**DISCUSSION**

The present study sought to investigate the genetic overlap between ADHD and EA, to leverage their potentially common genetics to improve the discovery of ADHD-associated loci and to help our understanding of the correlation between EA and ADHD observed in clinical studies demonstrating poor academic performance and decreased rates of high school graduation and postsecondary education in individuals with diagnosed ADHD.\(^{14}\) Altogether, these findings provide new insights into the genetic architecture of ADHD, suggesting shared molecular genetic mechanisms with EA. Furthermore, the findings may suggest that individuals with a high load of ADHD genetic risk factors, but not necessarily with the disorder itself, may be at higher risk for lower EA.

The most significant locus shared between ADHD and EA \( (rs618678) \) is located on chromosome 1 and represents a broad region of association spanning over more than 200,000 bp in 1p34.2 and 1p34.1 (Figure 4B; Figure S2B, available online). This region contains 3 protein

**TABLE 2 The Most Significant Single Nucleotide Polymorphisms (SNPs) for Each Linkage Disequilibrium (LD)–Independent Region Identified Either With Conditional False Discovery Rate (condFDR; condFDR \( < 0.01 \)) or With Conjunctive False Discovery Rate (conjFDR; conjFDR \( < 0.05 \)) Analysis**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr Region</th>
<th>Position</th>
<th>Value (p)</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs17414302</td>
<td>1p36.12</td>
<td>20976535</td>
<td>8.17 \times 10^{-7}</td>
<td>-0.090</td>
</tr>
<tr>
<td>rs618678</td>
<td>1p34.2</td>
<td>44133299</td>
<td>1.05 \times 10^{-10}</td>
<td>-0.053</td>
</tr>
<tr>
<td>rs4324303</td>
<td>2p24</td>
<td>13817678</td>
<td>6.69 \times 10^{-3}</td>
<td>0.009</td>
</tr>
<tr>
<td>rs412458</td>
<td>5q14.3</td>
<td>88029627</td>
<td>3.73 \times 10^{-6}</td>
<td>0.014</td>
</tr>
<tr>
<td>rs4477079</td>
<td>8q12.1</td>
<td>93059038</td>
<td>1.62 \times 10^{-3}</td>
<td>-0.009</td>
</tr>
</tbody>
</table>

**Note:** CondFDR/conjFDR values that are below the predefined significance threshold of 0.01/0.05 are shown in boldface type. Chromosome (Chr) and position are indicated according to GRCh37. For both attention-deficit/hyperactivity disorder (ADHD) and educational attainment (EA), \( p \) values without genomic inflation correction are shown. The effect size is given as \( \log_{10}(\text{OR}) \) for ADHD and as \( \beta \) regression coefficient for EA. Genes in the region are defined as genes containing SNPs at either condFDR \( < 0.01 \) or conjFDR \( < 0.05 \) and in LD \( (r^2 > 0.20) \) with the most significant SNP of the locus. Genes containing the leading SNP are marked in boldface type. Annotation was generated with Biomart Variant Effect Predictor (http://www.ensembl.org/Homo_sapiens/Tools/VEP).
coding genes: PTPRF, KDM4A, and ST3GAL3. rs618678 is an intronic variant within KDM4A, a member of the Jumonji domain 2 family, which encodes a protein that demethylates histone residues, and acts as an epigenetic transcriptional regulator.27 Genome-wide significant variants within KDM4A were reported in a recent GWAS of schizophrenia,9 a disorder that may share a genetic background with ADHD. The protein encoded by PTPRF is a member of the protein tyrosine phosphatase (PTP) family, which regulates a variety of cellular processes, including cell growth, differentiation, mitotic cycle, and oncogenic transformation. Murine studies showed that PTPRF promotes neurogenesis in the hippocampus,28 a brain region linked to memory. ST3GAL3 encodes a sialyltransferase responsible for the terminal sialylation of brain gangliosides and glycoproteins, which constitute a major part of the surface glycan coat of neurons and glia and acts as an interface for cellular interactions.29 Interestingly, mutations of ST3GAL3 may impair the development of higher cognitive functions30 and are associated with severe infantile epilepsy.31 Our eQTL analysis with the Brainex database revealed strong associations of rs618678 with altered expression of ST3GAL3 (Table S2, available online), suggesting that this may be a potential mechanism whereby this locus affects ADHD and EA. However, this association was not detected using the GTEx database. The discrepancy between the results from the different eQTL datasets could be attributed to differences in methodological techniques or sample configuration between the eQTL databases, or could reflect the relatively small sample sizes. The eQTL results should be reassessed when larger brain-eQTL databases are available.

The second locus shared between ADHD and EA (rs412458) is an intronic variant within MEF2C (Figure S2A, D, available online), which has multiple LD-linked variants with low condFDR/conjFDR values. MEF2C encodes 1 of 4 transcription factors constituting the myocyte enhancer factor 2 (MEF2) family.32 MEF2 is involved in neuronal survival and may regulate the growth and pruning of neurons as well as the number of synapses in the hippocampus, with potential relevance for memory and learning.33 Mutations of MEF2C cause severe mental retardation with stereotypic movements, seizures, and/or cerebral malformations.34 Furthermore, genome-wide significant SNPs within MEF2C have been reported to be associated with schizophrenia,35 which shares polygenic risk with ADHD.35 In addition, mutations in MEF2 genes have been found in patients with different neurological disorders, including Rett-like disorder and Parkinson disease.35 MEF2C expression is particularly enriched in the cerebral cortex36 (Figure S3, available online).

The third locus identified as susceptible for both ADHD and EA by conjFDR is an intronic variant within PINK1 on chromosome 1 (rs17414302). PINK1 encodes a serine/threonine protein kinase that primarily localizes to mitochondria and protects against progressive mitochondrial damage and dysfunction.37 This protein is thought to be involved in regulating neurite morphogenesis, enhancing anterograde mitochondrial transport and density of mitochondria in dendrites, and upregulating expression of neuronal differentiation proteins.38 PINK1 is important for the maintenance of mitochondria in part by selective degradation of compromised mitochondria (mitophagy).39 Mutations in this gene are a common cause of autosomal recessive Parkinson disease.40 However, rs17414302 represents an isolated signal with rather poor LD support (Figure S2E, available online), and it should thus be examined in more detail.

The strongest SNP association with ADHD revealed by the condFDR analysis was rs4324303. This SNP was not significant in the conjFDR analysis, but showed consistent direction of effect with ADHD symptoms in the EAGLE sample, possibly suggesting a putative role...
specific to ADHD, rs4324303 is an intergenic variant located approximately 1 mega base upstream of the nearest protein coding gene (TRIB2). It is therefore difficult to speculate about the potential role of this variant in different cellular processes.

Another variant identified by the condFDR analysis is rs4477079, an intronic variant within RUNX1T1 on chromosome 2. RUNX1T1 acts as a co-repressor of Notch41 and Wnt42 pathways. RUNX1T1 was reported to have high expression levels in adult and fetal brains43 and may influence the axon guidance process.44 RUNX1T1 was previously identified among the top associations (although not reaching genome-wide significance) in the context of oppositional defiant disorder (ODD), which is a frequent psychiatric disorder seen in individuals with ADHD.45 Notably, unlike the other loci identified in our analyses, this locus shows an inconsistent direction of effect between PGC ADHD risk and quantitative measures of ADHD symptoms in pediatric populations (Table S1, available online) and a co-directional effect between PGC ADHD risk and EA (Table 2). The latter is contrary to expectations (Table S1, available online) and a co-directional effect between PGC ADHD risk and EA (Table S1, available online).

To further evaluate the ADHD-associated variants identified in this study using the data from PGC ADHD case-control and EA GWAs, we examined our top hits in light of the ADHD symptoms’ GWAS. Four of 5 loci identified here revealed consistent direction of effect in the independent GWAS of ADHD symptoms (Table S1, available online). Of note, twin studies provide strong evidence that the diagnosis of ADHD can be considered the extreme of a continuous trait,46 and several studies show that the polygenic risk score computed from an association study of ADHD diagnosis predicts the variability of ADHD symptoms in population samples.21,47 In addition, it has been shown that the continuous measure of ADHD (such as symptom score) and the ADHD diagnosis share more than 90% of their genetic background.48 Thus, the results of the performed exploration may be viewed as confirmatory of our findings.

It is also worth mentioning that 2 loci identified in our analyses (corresponding to rs618678 and rs412458 in Table 2) were reported to reach genome-wide significance in the largest GWAS on ADHD performed to date, with a total number of 20,183 individuals with ADHD and 35,191 controls. In this GWAS, ADHD diagnosis was based on either the International Classification of Diseases—10th Revision (ICD-10) or DSM-IV. At the time of writing the study is yet unpublished, but a preprint is available in bioRxiv.48

As children with ADHD have been reported to be at high risk for academic failure, school dropout, grade repetition, and placement in special education,49,50 it is likely that the prevalence of ADHD cases among individuals with lower EA would be increased compared to the prevalence among individuals with higher EA. Moreover, ADHD is known to have a complex pattern of comorbid conditions1 (including dyslexia,52 ODD,53 and others), many of them also associated with lower EA. This potential overlap of phenotypes prevents us from translating the genetic correlation into actual pleiotropy, which is defined as the same gene variant affecting independent diseases or traits. Furthermore, it is challenging to evaluate small effect sizes and to speculate about molecular mechanisms behind the effective variants when examining such potentially overlapping phenotypes. Another general problem is that the effects of the associated variants are small, and their functional roles have not been directly investigated. Associated genetic loci contain several genes, and it is difficult to establish an arrow of causality when studying association between traits. Thus, the question of whether ADHD is diagnosed because of observed educational problems or because ADHD is the

---

**FIGURE 3** Manhattan plot of conjunctional –log10(false discovery rate [FDR]) for attention-deficit/hyperactivity disorder (ADHD) and educational attainment (EA).

Note: Data are unpruned. Small points represent nonsignificant single nucleotide polymorphisms (SNPs), bold points represent significant SNPs (conjunctional false discovery rate [condFDR] < 0.05). Points corresponding to significant SNPs with lowest conjunctional FDR in each linkage disequilibrium (LD)-independent region (r2 > 0.20) have a black border and the name of the corresponding gene written above it. Horizontal gray dotted line shows the significance threshold of condFDR (0.05).
cause of subsequent educational problems—or there is another common underlying factor—needs further exploration.

Also of possible relevance is the sample overlap between PGC ADHD and EA datasets (both GWASs include the WTCCC58C cohort12), which may inflate the results of our FDR analyses. However, the results of LD score regression, which are in line with those of our FDR analyses, are not affected by the sample overlap.8

In conclusion, we have identified 5 loci associated with ADHD and have provided evidence for a shared genetic basis between ADHD and EA, implicating 3 genetic loci in this overlap. Four of 5 identified loci showed consistent effects in the independent data set of ADHD symptoms, and inverse correlation with EA, in line with prior epidemiological and genetic studies. Altogether, the findings provide new insights into the relationship between ADHD and EA, suggesting shared molecular genetic mechanisms. On a cautious note, the identified risk variants are not informative clinically due to their small effect sizes. Further research is required to clarify the biological effects of the identified genetic variants and how these may influence EA and ADHD pathogenesis.

Accepted November 21, 2017.

Drs. Shadrin, Smeland, Frei, Bettella, Witoelar, Li, Eriksen, Krull, Wang and Andreassen are with NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo and Innovation Programme, and the National Institute of Mental Health.

The authors thank the Psychiatric Genetics Consortium (PGC) and the Social Science Genetic Association Consortium (SSGAC) for access to GWAS data. The authors thank Thomas Bjella, PhD, of the Oslo University Hospital & University of Copenhagen, Copenhagen, Denmark. Dr. Dyrvold is with Oslo University Hospital, Oslo, and NORMENT, KG Jebsen Centre for Psychosis Research, University of Bergen. Dr. Faraone is with KG Jebsen Centre for Neuropsychiatric Disorders, University of Bergen, SUNY Upstate Medical University, Syracuse, New York. Dr. Reichborn-Kjennerud is with Division of Mental Health, Norwegian Institute of Public Health, Oslo, and Institute of Clinical Medicine, University of Oslo. Dr. Thompson is with University of California, San Diego. Dr. Dale is with NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, and University of California, San Diego.

This work was supported by the Research Council of Norway (248778, 223273, 213694, 248980), the KG Jebsen Stiftelsen (SKGJ-MED-008), the National Institutes of Health (ROI GM0410400), and the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 667302. Dr. Wang was also supported by The Research Council of Norway through a FRIPRO Mobility Grant (contract no 251134). The FRIPRO Mobility grant scheme (FRICON) is co-funded by the European Union’s Seventh Framework Programme for research, technological development and demonstration under Marie Curie grant agreement no 608695. Dr. Faraone has received grant and research support from the K.G. Jebsen Centre for Research on Neuropsychiatric Disorders, the University of Bergen, Bergen, Norway, the European Union’s Seventh Framework Programme for research, technological development and demonstration, the European Union’s Horizon 2020 research and innovation programme, and the National Institute of Mental Health.

The authors thank the Psychiatric Genetics Consortium (PGC) and the Social Science Genetic Association Consortium (SSGAC) for access to GWAS data. The authors thank Thomas Bjella, PhD, of the Oslo University Hospital & Institute of Clinical Medicine, for support with the database.

Disclosure: Dr. Faraone has received income, potential income, travel expenses, continuing education support, research support from, and/or has served on the advisory boards of/as a consultant to Lundbeck, Rhodes, Arbor, KenPharm, Ironshore, Neurovance, Impact, Takeda, Shire, Akili Interactive Labs, CogCubed, Alcobra, VAYA Pharma, Sunovion, Genomind, and...
REFERENCES


18. Bernabeu R, Yang T, Xie Y, Mehta B, Ma SY, Longo FM. Downregulation of the LAR protein tyrosine phosphatase receptor is associated with increased dentate gyrus neurogenesis and an increased number of granule cell layer neurons. Mol Cell Neurosci. 2006;31:723-738.


29. Valente EM, Abou-Sleiman PM, Caputo V, et al. MEF2C haploinsufficiency caused by either microdeletion of the 1q43.4 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. J Med Genet. 2010;47:22-29.

30. Valente EM, Abou-Sleiman PM, Caputo V, et al. MEF2C haploinsufficiency caused by either microdeletion of the 1q43.4 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. J Med Genet. 2010;47:22-29.


32. Valente EM, Abou-Sleiman PM, Caputo V, et al. MEF2C haploinsufficiency caused by either microdeletion of the 1q43.4 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. J Med Genet. 2010;47:22-29.

33. Valente EM, Abou-Sleiman PM, Caputo V, et al. MEF2C haploinsufficiency caused by either microdeletion of the 1q43.4 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. J Med Genet. 2010;47:22-29.

34. Valente EM, Abou-Sleiman PM, Caputo V, et al. MEF2C haploinsufficiency caused by either microdeletion of the 1q43.4 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. J Med Genet. 2010;47:22-29.


