

ARTICLE

Open Access

# Disentangling polygenic associations between attention-deficit/hyperactivity disorder, educational attainment, literacy and language

Ellen Verhoef<sup>1,2</sup>, Ditte Demontis<sup>3,4,5</sup>, Stephen Burgess<sup>6,7</sup>, Chin Yang Shapland<sup>1</sup>, Philip S. Dale<sup>8</sup>, Aysu Okbay<sup>9,10</sup>, Benjamin M. Neale<sup>11,12,13</sup>, Stephen V. Faraone<sup>14</sup>, iPSYCH-Broad-PGC ADHD Consortium, Evie Stergiakouli<sup>15,16</sup>, George Davey Smith<sup>15,17</sup>, Simon E. Fisher<sup>1,18</sup>, Anders D. Børglum<sup>3,4,5</sup> and Beate St Pourcain<sup>1,15,18</sup>

## Abstract

Interpreting polygenic overlap between ADHD and both literacy-related and language-related impairments is challenging as genetic associations might be influenced by indirectly shared genetic factors. Here, we investigate genetic overlap between polygenic ADHD risk and multiple literacy-related and/or language-related abilities (LRAs), as assessed in UK children ( $N \leq 5919$ ), accounting for genetically predictable educational attainment (EA). Genome-wide summary statistics on clinical ADHD and years of schooling were obtained from large consortia ( $N \leq 326,041$ ). Our findings show that ADHD-polygenic scores (ADHD-PGS) were inversely associated with LRAs in ALSPAC, most consistently with reading-related abilities, and explained  $\leq 1.6\%$  phenotypic variation. These polygenic links were then dissected into both ADHD effects shared with and independent of EA, using multivariable regressions (MVR). Conditional on EA, polygenic ADHD risk remained associated with multiple reading and/or spelling abilities, phonemic awareness and verbal intelligence, but not listening comprehension and non-word repetition. Using conservative ADHD-instruments ( $P$ -threshold  $< 5 \times 10^{-8}$ ), this corresponded, for example, to a 0.35 SD decrease in pooled reading performance per log-odds in ADHD-liability ( $P = 9.2 \times 10^{-5}$ ). Using subthreshold ADHD-instruments ( $P$ -threshold  $< 0.0015$ ), these effects became smaller, with a 0.03 SD decrease per log-odds in ADHD risk ( $P = 1.4 \times 10^{-6}$ ), although the predictive accuracy increased. However, polygenic ADHD-effects shared with EA were of equal strength and at least equal magnitude compared to those independent of EA, for all LRAs studied, and detectable using subthreshold instruments. Thus, ADHD-related polygenic links with LRAs are to a large extent due to shared genetic effects with EA, although there is evidence for an ADHD-specific association profile, independent of EA, that primarily involves literacy-related impairments.

Correspondence: Ellen Verhoef ([Ellen.Verhoef@mpi.nl](mailto:Ellen.Verhoef@mpi.nl)) or Beate St Pourcain ([Beate.StPourcain@mpi.nl](mailto:Beate.StPourcain@mpi.nl))

<sup>1</sup>Language and Genetics Department Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

<sup>2</sup>International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands

Full list of author information is available at the end of the article.

These authors contributed equally: Ellen Verhoef, Ditte Demontis

These authors jointly supervised this work: Anders D. Børglum, Beate St Pourcain

## Introduction

Children with Attention-Deficit/Hyperactivity Disorder (ADHD) often experience difficulties mastering literacy-related and/or language-related abilities (LRAs)<sup>1–3</sup>. It has been estimated that up to 40% of children diagnosed with clinical ADHD also suffer from reading disability (RD, also known as developmental dyslexia) and vice versa<sup>4</sup>. The spectrum of affected LRAs in ADHD may, however,

© The Author(s) 2019



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

also include writing<sup>5,6</sup>, spelling<sup>7,8</sup>, syntactic<sup>9,10</sup> and phonological<sup>9,10</sup> abilities. Both clinical ADHD and RD are complex childhood-onset neurodevelopmental conditions that affect about 5% and 7% of the general population, respectively<sup>11,12</sup>. ADHD is characterised by hyperactive, inattentive and impulsive symptoms<sup>13</sup>, whereas decoding and/or reading comprehension deficits are prominent in individuals with RD<sup>14</sup>.

To interpret the comorbidity of ADHD and RD, a multiple-deficit model including shared underlying aetiologies has been proposed, involving both genetic and environmental influences<sup>15</sup>. This model is supported by twin studies suggesting that the co-occurrence of ADHD symptoms and reading deficits is, to a large extent, attributable to shared genetic influences<sup>16–18</sup>. Further twin research suggests that the genetic covariance between reading difficulties and ADHD is largely independent of genetic factors shared with IQ<sup>19</sup>, although it is not known whether these findings extend to a wider spectrum of LRAs, beyond reading abilities. Furthermore, the interpretation of polygenic ADHD-LRA overlap using markers on genotyping arrays is more challenging. There is strong evidence that genetically predicted educational attainment (EA)<sup>20</sup> shares genetic variability with both ADHD<sup>21</sup> and reading abilities<sup>22,23</sup>. Genetically predicted EA is a genetic proxy of cognitive abilities, but also socioeconomic status<sup>20</sup> including, for example, associations with maternal smoking during pregnancy, parental smoking, household income or watching television<sup>24</sup>. Thus, observed genetic associations between ADHD and reading abilities may solely reflect shared genetic variation with EA, but not any other, more specific neuro-cognitive mechanisms. In other words, polygenic associations might be inflated or even induced<sup>25</sup> by genetically predictable traits that are related to both, ADHD and reading abilities (or other LRAs).

Here, we (a) study polygenic links between clinical ADHD and a wide range of population-ascertained literacy-related and language-related measures as captured by common variation, (b) evaluate to what extent such links reflect a shared genetic basis with EA and (c) assess whether there is support for shared genetic factors between clinical ADHD and LRAs conditional on genetically predicted EA.

Studied ADHD polygenic scores (ADHD-PGS) are based on ADHD genome-wide association study (GWAS) summary statistics from two large independent ADHD samples, the Psychiatric Genomics Consortium (PGC) and the Danish Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), and a combination thereof. Associations between ADHD-PGS and a wide spectrum of population-based literacy-related and language-related measures related to reading, spelling, phonemic awareness, listening comprehension, non-word

repetition and verbal intelligence skills, are examined in a sample of children from the UK Avon Longitudinal Study of Parents and Children (ALSPAC). Applying multi-variable regression (MVR) techniques, analogous to Mendelian Randomisation (MR) approaches<sup>26</sup>, we report here disentangled associations between polygenic ADHD risk and LRA measures and estimate effects independent of and shared with genetically predicted years of schooling, using summary statistics from the Social Science Genetic Association Consortium (SSGAC).

## Methods and materials

### Literacy-related and language-related abilities in the general population

LRAs were assessed in children and adolescents from ALSPAC, a UK population-based longitudinal pregnancy-ascertained birth cohort (estimated birth date: 1991–1992, Supplementary Information)<sup>27,28</sup>. Ethical approval was obtained from the ALSPAC Law-and-Ethics Committee (IRB00003312) and the Local Research-Ethics Committees. Written informed consent was obtained from a parent or individual with parental responsibility and assent (and for older children consent) was obtained from the child participants.

### Phenotype information

Thirteen measures capturing LRAs related to reading, spelling, phonemic awareness, listening comprehension, non-word repetition and verbal intelligence scores were assessed in 7 to 13 year-old ALSPAC participants ( $N \leq 5919$ , Table 1) using both standardised and ALSPAC-specific instruments. Detailed descriptions of all LRA measures are available in Table 1 and the Supplementary Information.

All LRA scores were rank-transformed to allow for comparisons of genetic effects across different psychological instruments with different distributions (Supplementary Information). Phenotypic correlations, using Pearson-correlation coefficients, were comparable for untransformed and rank-transformed scores (Table S1). To account for multiple testing, we estimated the effective number of phenotypes studied using Matrix Spectral Decomposition<sup>29</sup>(MatSpD), revealing seven independent measures (experiment-wide error rate of 0.007).

For sensitivity analysis, we excluded 188 children with an ADHD diagnosis at age 7, based on the Development and Wellbeing Assessment (DAWBA)<sup>30</sup> (Supplementary Information).

### Genetic analyses

ALSPAC participants were genotyped using the Illumina HumanHap550 quad chip genotyping platforms, and genotypes were called using the Illumina GenomeStudio software. Genotyping, imputation and genome-wide

**Table 1 Literacy-related and language-related abilities in the Avon Longitudinal Study of Parents and Children**

LRA (psychological instrument)	Mean Score (SE)	Mean Age (SE)	N (%males)	LRA combinations
Reading accuracy and comprehension (WORD <sup>69</sup> ), words	28.44 (9.24)	7.53 (0.31)	5891 (50.6)	-----Reading----- -----Spelling----- -----Global LRAs-----
Reading accuracy (ALSPAC specific: NBO <sup>70</sup> ), words	7.55 (2.44)	9.87 (0.32)	5738 (49.3)	
Reading speed <sup>a</sup> (NARA II <sup>71</sup> ), passages	105.50 (12.47)	9.88 (0.32)	5189 (49.1)	
Reading accuracy <sup>a</sup> (NARA II <sup>71</sup> ), passages	104.11 (13.62)	9.88 (0.32)	5201 (49.1)	
Reading speed (TOWRE <sup>72</sup> ), words	82.58 (10.28)	13.83 (0.20)	4247 (48.4)	
Non-word reading accuracy (ALSPAC specific: NBO <sup>70</sup> )	5.24 (2.48)	9.87 (0.32)	5731 (49.2)	
Non-word reading speed (TOWRE <sup>72</sup> )	50.82 (9.38)	13.83 (0.20)	4237 (48.3)	
Spelling accuracy (ALSPAC specific: NB)	7.89 (4.39)	7.53 (0.31)	5800 (50.2)	
Spelling accuracy (ALSPAC specific: NB)	10.27 (3.43)	9.87 (0.32)	5728 (49.2)	
Phonemic awareness (AAT <sup>73</sup> )	20.23 (9.51)	7.53 (0.31)	5919 (50.6)	
Listening comprehension (WOLD <sup>74</sup> )	7.50 (1.96)	8.63 (0.30)	5473 (49.9)	
Non-word repetition (CNRep <sup>75</sup> )	7.26 (2.51)	8.63 (0.30)	5464 (49.9)	
Verbal intelligence <sup>a</sup> (WISC-III <sup>76</sup> )	107.85 (16.74)	8.64 (0.31)	5456 (49.7)	

*Note:* Thirteen LRAs capturing aspects related to reading, spelling, phonemic awareness, listening comprehension, non-word repetition and verbal intelligence were assessed in 7 to 13 year-old ALSPAC participants using both standardised and ALSPAC-specific instruments (Supplementary Information)

LRAs literacy-related and language-related abilities, *WORD* Wechsler Objective Reading Dimension, *ALSPAC* Avon Longitudinal study of Parents and Children, *NBO* ALSPAC-specific assessment developed by Nunes, Bryant and Olson, *NARA II* The Neale Analysis of Reading Ability-Second Revised British Edition, *TOWRE* Test Of Word Reading Efficiency, *NB* ALSPAC-specific assessment developed by Nunes and Bryant, *AAT* Auditory Analysis Test, *WOLD* Wechsler Objective Language Dimensions, *CNRep* Children's Test of Nonword Repetition, *WISC-III* Wechsler Intelligence Scale for Children III

<sup>a</sup>Scores were derived using age norms and adjusted for sex and principal components only before transformation

**Table 2 Sample description**

Phenotype	Sample	Source	Ethnicity	Imputation reference panel	N
LRAs	ALSPAC	General population	White European	HRC r1.1	≤5891
ADHD	PGC	Clinical sample	Predominantly white European	HapMap phase 3	16,203 ( $N_{cases} = 4163$ )
ADHD	iPSYCH	Clinical sample	White European	1000 Genomes phase 3	37,076 ( $N_{cases} = 14,584$ )
ADHD	PGC + iPSYCH (EUR)	Clinical sample	White European	1000 Genomes phase 3	53,293 ( $N_{cases} = 19,099$ )
ADHD	PGC + iPSYCH	Clinical sample	Predominantly white European	1000 Genomes phase 3	55,374 ( $N_{cases} = 20,183$ )
EA	SSGAC	Predominantly general population	White European	1000 Genomes phase 3 <sup>a</sup>	326,041

<sup>a</sup>Predominantly 1000 Genomes phase 3 (20)

Abbreviations: LRAs literacy-related and language-related abilities, ADHD Attention-Deficit/Hyperactivity Disorder, EA educational attainment, ALSPAC Avon Longitudinal study of Parents and Children, PGC Psychiatric Genomics Consortium, iPSYCH The Lundbeck Foundation Initiative for Integrative Psychiatric Research, EUR European ancestry, SSGAC Social Science Genetic Consortium, HRC The Haplotype Reference Consortium

*Note:* There is no overlap between LRA, ADHD and EA samples

association analysis details are described in the Supplementary Information and Table 2.

### Clinical ADHD summary statistics

*Psychiatric Genomics Consortium (PGC).* GWAS summary statistics were obtained from a mega-analysis of clinical ADHD<sup>31</sup>, conducted by the PGC (4163 cases and 12,040 controls/pseudo-controls) (Table 2, Supplementary Information, [www.med.unc.edu/pgc/](http://www.med.unc.edu/pgc/)).

*The Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH).* An independent set of ADHD GWAS summary statistics were accessed through

the Danish iPSYCH project<sup>32</sup> (14,584 ADHD cases, 22,492 controls) (Table 2, Supplementary Information), using samples from the Danish Neonatal Screening Biobank hosted by Statens Serum Institute<sup>21,33</sup>.

*Combined PGC and iPSYCH ADHD sample (PGC + iPSYCH).* To maximise power, we also analysed meta-GWAS summary statistics from an ADHD sample containing both PGC and iPSYCH participants<sup>21</sup> (20,183 cases, 35,191 controls/pseudo-controls) (Table 2, [www.med.unc.edu/pgc/](http://www.med.unc.edu/pgc/)) and its European-only subset (*PGC + iPSYCH(EUR)*, 19,099 cases, 34,194 controls/pseudo-controls) (Table 2, [www.med.unc.edu/pgc/](http://www.med.unc.edu/pgc/)).

Detailed sample descriptions are available in Table 2 and the Supplementary Information.

#### Educational attainment summary statistics

GWAS summary statistics for EA<sup>20</sup> (discovery and replication sample combined, excluding ALSPAC and 23andMe samples,  $N = 326,041$ ) were obtained from the SSGAC consortium. EA was assessed as years of schooling<sup>20</sup>. A detailed sample description is available in Table 2 and the Supplementary Information.

#### Genome-wide complex trait analysis

SNP- $h^2$  and genetic correlations ( $r_g$ ) between LRAs were estimated using Restricted Maximum Likelihood (REML) analyses<sup>34,35</sup> as implemented in Genome-wide Complex Trait Analysis (GCTA) software<sup>36</sup>, including individuals with a genetic relationship  $< 0.05$ <sup>34</sup>. For this study, we selected only LRAs with evidence for SNP- $h^2$  and sample size  $N > 4000$  (Table S2).

#### Linkage disequilibrium score regression and correlation

Linkage Disequilibrium Score (LDSC) regression<sup>37</sup> was used to distinguish confounding biases from polygenic influences by examining the LDSC regression intercept. Unconstrained LD-score correlation<sup>38</sup> analysis was applied to estimate  $r_g$  (Supplementary Information).

#### Polygenic scoring analyses

ADHD-PGS<sup>39,40</sup> were created in ALSPAC using the independent PGC and iPSYCH GWAS summary statistics, and, to maximise power, also for GWAS summary statistics from the combined PGC + iPSYCH sample (Supplementary Information). ADHD-PGS have been previously linked to ADHD symptoms in ALSPAC participants<sup>41</sup>. Rank-transformed LRAs were regressed on Z-standardised ADHD-PGS (aligned to measure risk-increasing alleles) using ordinary least square (OLS) regression (R:stats library, Rv3.2.0). The proportion of phenotypic variance explained is reported as OLS-regression- $R^2$ . Beta-coefficients ( $\beta$ ) for ADHD-PGS quantify here the change in standard deviation (SD) units of LRA performance per one SD increase in ADHD-PGS.

#### Multivariable regression analysis

To study the genetic association between ADHD and LRAs conditional on genetic influences shared with EA, we applied MVR. This technique is analogous to MR methodologies<sup>26</sup> and controls for collider bias<sup>42</sup> through the use of GWAS summary statistics. Technically, it involves the regression of regression estimates from independent samples on each other<sup>26</sup> (Supplementary Information). Within this study we use MVR without inferring causality due to violations of classical MR assumptions<sup>26</sup> (see below).

**Genetic variant selection:** To disentangle ADHD-LRA associations, we selected two sets of instruments from the most powerful ADHD GWAS summary statistics (PGC + iPSYCH). The first set contained genome-wide significant variants ( $P < 5 \times 10^{-8}$ , conservative). The second set included variants passing a more lenient  $P$ -value threshold ( $P < 0.0015$ , subthreshold) to increase power, consistent with current guidelines for the selection of genetic instruments in MR (F-statistic  $< 10$ )<sup>43</sup>. All sets included independent (PLINK<sup>44</sup> clumping: LD- $r^2 < 0.25$ ,  $\pm 500$  kb), well imputed (INFO<sup>45</sup>  $> 0.8$ ) and common (EAF  $> 0.01$ ) variants. This resulted in 15 conservative and 2689  $< N_{\text{SNPs}} \leq 2692$  subthreshold ADHD-instruments (Table S8).

**Estimation of ADHD effects:** We extracted regression estimates for selected ADHD-instruments (conservative and subthreshold) from ADHD (PGC + iPSYCH), EA (SSGAC) and 13 LRA (ALSPAC) GWAS summary statistics. Analysing each set of variants independently, regression estimates for individual LRA measures ( $\beta$ ) were regressed on both ADHD ( $\beta$  as lnOR) and EA regression estimates ( $\beta$ ) using an OLS regression framework (R:stats library, Rv3.2.0). Outcomes were (1) a MVR regression estimate quantifying the change in SD units of LRA performance per log odds increase in ADHD risk conditional on years of schooling (ADHD effect independent of EA), and (2) a MVR regression estimate quantifying the change in SD units of LRA performance per year of schooling as captured by ADHD instruments (ADHD effect shared with EA). Latter MVR regression estimates capture here shared genetic effects between ADHD, EA and LRAs, including (1) genetic confounding (i.e., genetically predictable EA influences both ADHD and LRAs), (2) mediation (i.e., genetically predictable ADHD influences LRA indirectly through EA) and (3) biological pleiotropy (i.e., ADHD risk variants affect ADHD and EA through independent biological pathways). As ADHD risk and EA are inversely genetically related with each other<sup>21</sup>, they were reported to quantify change per missing year of schooling. To compare the magnitude of both MVR estimates, we also conducted analyses using fully standardised EA, ADHD and LRA regression estimates (Supplementary Information).

Finally, MVR regression estimates were meta-analysed and contrasted across reading-related, spelling-related and all LRA measures (excluding the composite measure verbal intelligence) (Table 1) using random-effects meta-regression, accounting for phenotypic correlations between LRAs (R:metafor library<sup>46</sup>, Rv3.2.0; Supplementary Information).

#### Sensitivity analyses

As the directionality of the relationship between ADHD, EA and LRAs cannot be inferred in this study, we also

examined the genetic association between EA and LRAs, conditional on ADHD, using MVR. Two sets of EA instruments (conservative and subthreshold, Table S8) were selected from EA (SSGAC) GWAS summary statistics, analogous to the selection of ADHD instruments, and MVR was conducted as described above. Note that we did not create LRA instrument sets, as GWAS summary statistics of LRAs were underpowered.

#### Attrition analysis

We carried out an *attrition* analysis in ALSPAC studying the genetic association between LRA-missingness and polygenic ADHD risk, using both polygenic scoring analyses and MVR (Supplementary Information).

## Results

### Genetic architecture of literacy-related and language-related abilities and clinical ADHD

Phenotypic variation in literacy-related and language-related measures (Table 1), including reading abilities (comprehension, accuracy and speed) assessed in words/passages and non-words, spelling abilities (accuracy), phonemic awareness, listening comprehension, non-word repetition and verbal intelligence scores, can be tagged by common variants, with SNP- $h^2$  estimates between 0.32 (SE = 0.07, non-word repetition age 8) and 0.54 (SE = 0.07, verbal intelligence age 8) (Table S2; GCTA-based and LDSC-based estimations). Importantly, all LRAs are phenotypically (Table S1) and genetically (Table S3) moderately to strongly interrelated. The observed LDSC-based evidence for genetic liability of clinical ADHD within the PGC (LDSC- $h^2$  = 0.08(SE = 0.03)), iPSYCH (LDSC- $h^2$  = 0.26(SE = 0.02)) and PGC + iPSYCH samples (Table S4) is consistent with previous reports<sup>21</sup>.

### Association between ADHD polygenic risk scores and literacy-related and language-related abilities

We observed robust evidence for an inverse genetic association between ADHD-PGS and reading accuracy/comprehension age 7 (PGC: OLS- $R^2$  = 0.1%,  $P$  =  $4.6 \times 10^{-3}$ ; iPSYCH: OLS- $R^2$  = 1.0%,  $P$  <  $1 \times 10^{-10}$ ), reading accuracy age 9 (PGC: OLS- $R^2$  = 0.1%,  $P$  =  $5.7 \times 10^{-3}$ ; iPSYCH: OLS- $R^2$  = 1.2%,  $P$  <  $1 \times 10^{-10}$ ), and spelling accuracy age 9 (PGC: OLS- $R^2$  = 0.2%,  $P$  =  $1.5 \times 10^{-3}$ ; iPSYCH: OLS- $R^2$  = 0.8%,  $P$  <  $1 \times 10^{-10}$ ) using independent ADHD discovery samples (Fig. 1, Table S5). The strongest evidence for association was observed when ADHD discovery samples were combined (PGC + iPSYCH; Fig. 1), including those of European ancestry only (PGC + iPSYCH (EUR)), with genetic trait-disorder overlap present for all LRAs studied (Table S5). For example, ADHD-PGS explain 1.49% phenotypic variation in reading accuracy age 9, translating into a genetic covariance of  $-0.11(95\%$

CI:  $-0.14$ ;  $-0.09$ ) (Supplementary Information). Polygenic scoring results are presented for a  $P$ -value threshold of 0.1, but other thresholds provided similar results (data not shown). Results were not affected by the exclusion of children with an ADHD diagnosis in ALSPAC (Table S6).

### Shared genetic liability between ADHD and LRA with EA

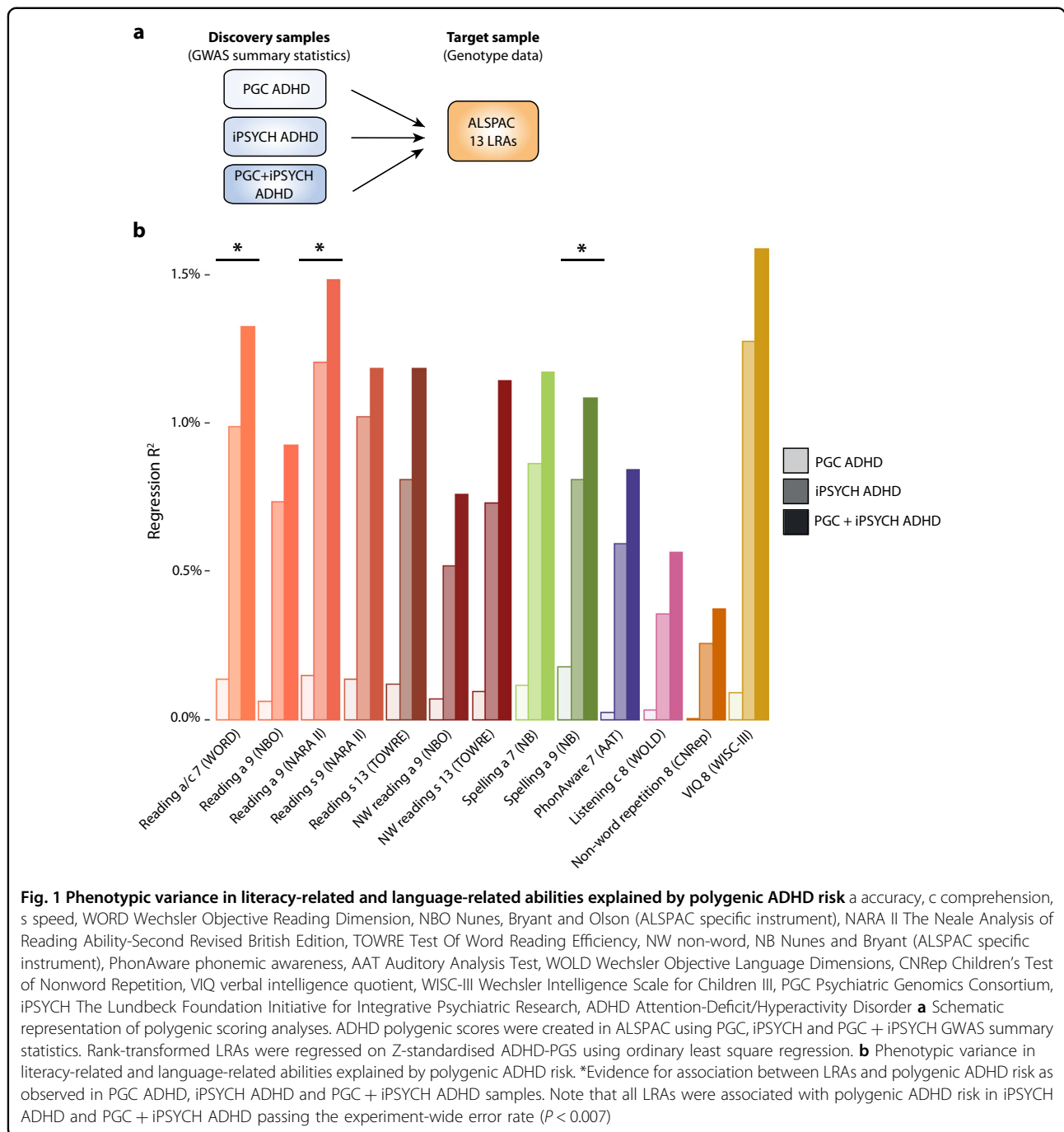
There was strong evidence for a moderate negative genetic correlation ( $r_g$  =  $-0.53$ (SE = 0.03),  $P$  <  $1 \times 10^{-10}$ ) between genetically predicted ADHD, as captured by the largest ADHD discovery sample (PGC + iPSYCH), and EA (LDSC- $h^2$  = 0.11(SE = 0.004)), consistent with previous findings<sup>21</sup>. Likewise, LRAs were moderately to highly positively correlated with EA (e.g., reading speed age 13  $r_g$  = 0.80(SE = 0.22),  $P$  =  $3.0 \times 10^{-4}$ ; Table S7), as previously reported<sup>22,23</sup>. Additionally, two independent variants reached genome-wide significance for both ADHD<sup>21</sup> and EA<sup>20</sup>, consistent with biological pleiotropy (i.e., single genetic loci influencing multiple traits)<sup>47</sup>. These findings indicate complex, potentially reciprocal cross-trait relationships (Fig. 2a) and violate MR causal modelling assumptions<sup>26</sup>. Consequently, ADHD instruments are not valid MR instruments as they are not independent of EA.

### Multivariable regression analyses

To disentangle the genetic overlap of polygenic ADHD risk with literacy-related and language-related measures into ADHD genetic effects independent of and shared with EA, we applied MVR<sup>26</sup> using ADHD instruments based on the most powerful ADHD discovery sample (PGC + iPSYCH) (Fig. 2b).

Using conservative ADHD instruments (Table S8), non-word reading accuracy at age 9 and pooled reading-related abilities were associated with polygenic ADHD risk, conditional on EA (Table 3). The latter translates into, for example, a decrease of 0.35 SD in pooled reading performance per log-odds increase in ADHD risk ( $\beta_{\text{ADHD}}$  =  $-0.35$ (SE = 0.09),  $P$  =  $9.2 \times 10^{-5}$ ,  $P_{\text{het}}$  = 0.19), an effect that was considerably stronger than for other LRAs ( $P_{\text{mod}}$  = 0.011, Table S10).

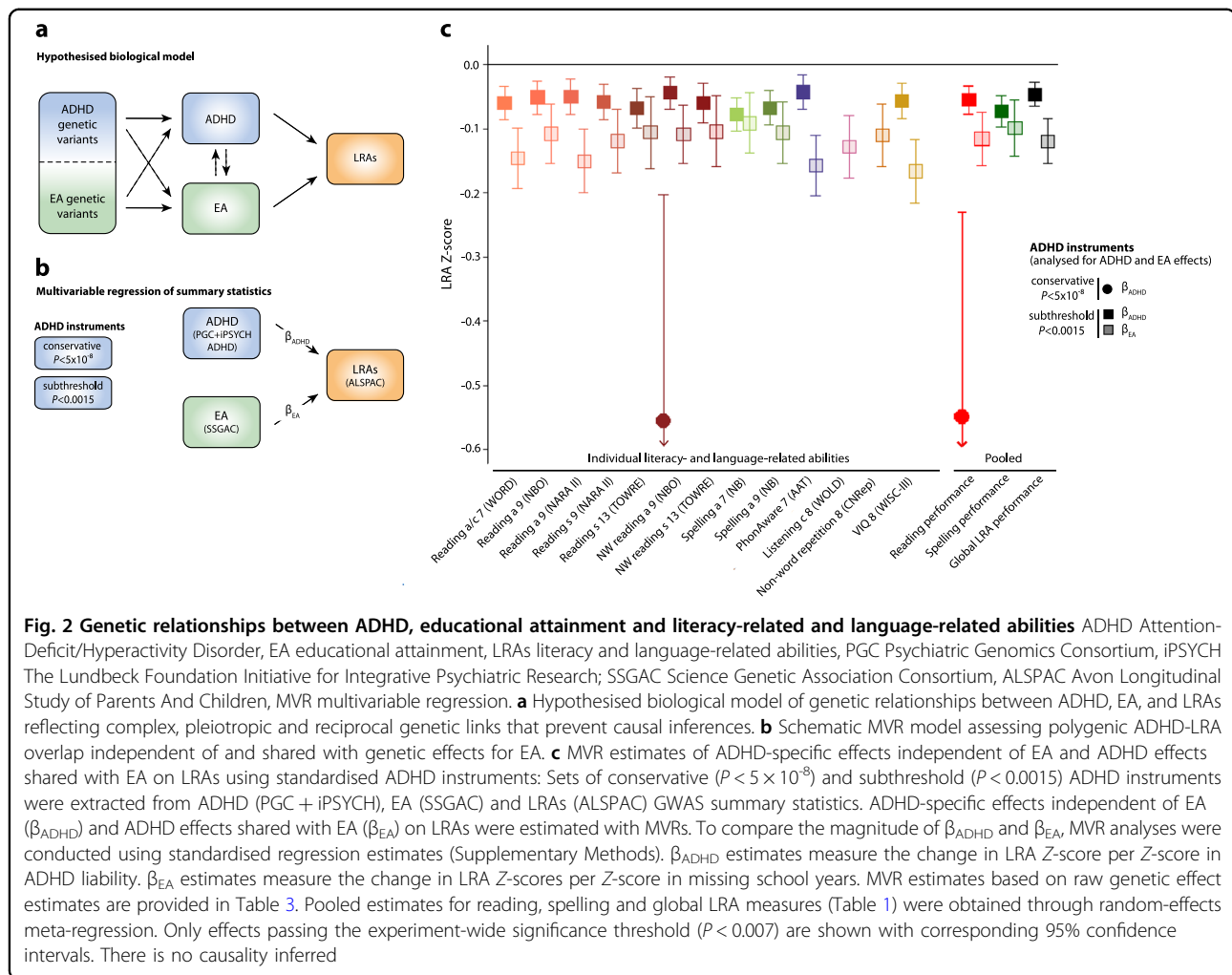
Using subthreshold ADHD instruments (Table S8), polygenic ADHD effects on LRA performance, conditional on EA, were detectable for all reading-related and spelling-related measures, phonemic awareness and verbal intelligence, but not other LRAs such as listening comprehension and non-word repetition (Table 3). Evidence was strongest for pooled reading and spelling abilities (Table 3, minimum  $P$  =  $1.1 \times 10^{-8}$ ). However, observable effects were smaller in magnitude compared to those captured by conservative ADHD instruments with, for example, a 0.03 SD decrease in pooled reading performance per log-odds increase in ADHD risk ( $\beta_{\text{ADHD}}$  =  $-0.03$ (SE = 0.01),  $P$  =  $1.4 \times 10^{-6}$ , Table 3). Comparing



ADHD-specific effects on both reading and spelling with ADHD-specific effects on all other LRAs provided evidence for effect differences ( $P_{mod} = 0.016$ ), with stronger ADHD effects on literacy-related abilities, in particular spelling (Table S10).

Polygenic ADHD effects that are shared with EA were identified for all LRAs studied using subthreshold, but not conservative ADHD instruments (Table 3). This translates into, for example, a further 0.50 SD units decrease in

pooled reading performance per missing school year ( $\beta_{EA} = -0.50$  (SE = 0.09),  $P = 4.9 \times 10^{-8}$ , Table 3). Thus, the observed association between polygenic ADHD risk and listening comprehension and non-word repetition is fully attributable to genetic effects shared with EA (Table 3). Contrary to ADHD-specific effects, ADHD effects shared with EA showed no evidence for effect differences between literacy-related versus other LRAs ( $P = 0.31$ ). Conducting MVR with fully standardised estimates



**Fig. 2 Genetic relationships between ADHD, educational attainment and literacy-related and language-related abilities** ADHD Attention-Deficit/Hyperactivity Disorder, EA educational attainment, LRAs literacy and language-related abilities, PGC Psychiatric Genomics Consortium, iPSYCH The Lundbeck Foundation Initiative for Integrative Psychiatric Research; SSGAC Science Genetic Association Consortium, ALSPAC Avon Longitudinal Study of Parents And Children, MVR multivariable regression. **a** Hypothesised biological model of genetic relationships between ADHD, EA, and LRAs reflecting complex, pleiotropic and reciprocal genetic links that prevent causal inferences. **b** Schematic MVR model assessing polygenic ADHD-LRA overlap independent of and shared with genetic effects for EA. **c** MVR estimates of ADHD-specific effects independent of EA and ADHD effects shared with EA on LRAs using standardised ADHD instruments: Sets of conservative ( $P < 5 \times 10^{-8}$ ) and subthreshold ( $P < 0.0015$ ) ADHD instruments were extracted from ADHD (PGC + iPSYCH), EA (SSGAC) and LRAs (ALSPAC) GWAS summary statistics. ADHD-specific effects independent of EA ( $\beta_{ADHD}$ ) and ADHD effects shared with EA ( $\beta_{EA}$ ) on LRAs were estimated with MVRs. To compare the magnitude of  $\beta_{ADHD}$  and  $\beta_{EA}$ , MVR analyses were conducted using standardised regression estimates (Supplementary Methods).  $\beta_{ADHD}$  estimates measure the change in LRA Z-score per Z-score in ADHD liability.  $\beta_{EA}$  estimates measure the change in LRA Z-scores per Z-score in missing school years. MVR estimates based on raw genetic effect estimates are provided in Table 3. Pooled estimates for reading, spelling and global LRA measures (Table 1) were obtained through random-effects meta-regression. Only effects passing the experiment-wide significance threshold ( $P < 0.007$ ) are shown with corresponding 95% confidence intervals. There is no causality inferred

showed that ADHD effects shared with EA were as large as or even larger compared to ADHD-specific effects (Fig. 2c, Table S9).

Using an analogous approach, we disentangled the genetic overlap between polygenic EA and LRAs into genetic EA effects independent of and shared with ADHD, based on EA instruments (Fig. S1). There was strong evidence for EA effects shared with ADHD using subthreshold, but not conservative EA instruments (Table S11). The magnitude of ADHD genetic effects shared with EA, captured by ADHD genetic instruments, compared to the magnitude of EA genetic effects shared with ADHD, captured by EA instruments, was largely consistent with each other in fully standardised analyses (Tables S9 and S11).

There was little evidence supporting the inclusion of regression intercepts in MVR that would imply additional genetic effect variation in LRAs estimates, not yet captured by either ADHD or EA effect estimates, based on the selected instruments. Therefore, all MVRs were performed using constrained intercepts<sup>26</sup>.

### Attrition in ALSPAC

Analyses of sample drop-out in ALSPAC, exemplified by missing reading accuracy and comprehension scores at age 7 (WORD), revealed a positive genetic association between missingness and polygenic ADHD risk (min  $P = 1.4 \times 10^{-8}$ , Supplementary Information, Table S12, Table S13).

### Discussion

This study identified strong and replicated evidence for an inverse association between polygenic ADHD risk and multiple population-based LRAs using a polygenic scoring approach. However, these associations involve shared genetic variation with genetically predictable EA. Accurate modelling of polygenic links using MVR techniques, conditional on EA, revealed an ADHD-specific association profile that primarily involves literacy-related impairments. Once shared genetic effects with EA were accounted for, polygenic ADHD risk was most strongly inversely associated with reading and/or spelling abilities,

**Table 3 Multivariable regression analysis of polygenic associations between ADHD and literacy-related and language-related abilities (raw estimates)**

LRAs	ADHD-specific effects independent of EA ( $\beta_{ADHD}$ )					EA genetic effects of ADHD-associated variants <sup>a</sup> ( $\beta_{EA}$ )				
	Conservative instruments ( $P_{thr} < 5 \times 10^{-8}$ )		Subthreshold instruments ( $P_{thr} < 0.0015$ )		$P_{het}$	Conservative instruments ( $P_{thr} < 5 \times 10^{-8}$ )		Subthreshold instruments ( $P_{thr} < 0.0015$ )		$P_{het}$
	$\beta$ (SE)	$P$	$\beta$ (SE)	$P$		$\beta$ (SE)	$P$	$\beta$ (SE)	$P$	
Reading a/c 7 (WORD)	-0.18 (0.13)	0.20	-0.03 (0.01)	$1.4 \times 10^{-5}$	—	-0.09 (1.19)	0.94	-0.63(0.10)	$1.2 \times 10^{-9}$	—
Reading a 9 (NBO)	-0.28 (0.09)	0.01	-0.03 (0.01)	$2.1 \times 10^{-4}$	—	1.33 (0.84)	0.14	-0.46 (0.10)	$6.3 \times 10^{-6}$	—
Reading s 9 (NARA II)	-0.29 (0.13)	0.04	-0.03 (0.01)	$3.0 \times 10^{-5}$	—	1.55 (1.16)	0.21	-0.51 (0.11)	$2.5 \times 10^{-6}$	—
Reading a 9 (NARA II)	-0.34 (0.12)	0.01	-0.03 (0.01)	$3.8 \times 10^{-4}$	—	1.57 (1.11)	0.18	-0.64 (0.11)	$4.1 \times 10^{-9}$	—
Reading s 13 (TOWRE)	-0.36 (0.15)	0.03	-0.04 (0.01)	$2.7 \times 10^{-5}$	—	2.06 (1.35)	0.15	-0.46 (0.12)	$1.8 \times 10^{-4}$	—
NW reading a 9 (NBO)	-0.37 (0.09)	0.002	-0.03 (0.01)	$2.4 \times 10^{-4}$	—	1.54 (0.85)	0.09	-0.46 (0.10)	$6.1 \times 10^{-6}$	—
NW reading s 13 (TOWRE)	-0.21 (0.17)	0.25	-0.03 (0.01)	$2.4 \times 10^{-4}$	—	1.43 (1.58)	0.38	-0.45 (0.12)	$1.9 \times 10^{-4}$	—
Spelling a 7 (NB)	-0.05 (0.12)	0.72	-0.05 (0.01)	$1.4 \times 10^{-5}$	—	-1.12 (1.13)	0.34	-0.63 (0.10)	$1.2 \times 10^{-9}$	—
Spelling a 9 (NB)	-0.23 (0.08)	0.01	-0.04 (0.01)	$8.7 \times 10^{-7}$	—	1.43 (0.72)	0.38	-0.45 (0.10)	$1.3 \times 10^{-5}$	—
PhonAware 7 (AAT)	-0.22 (0.15)	0.16	-0.02 (0.01)	0.002	—	-0.31 (1.33)	0.82	-0.67 (0.10)	$< 1 \times 10^{-10}$	—
Listening c 8 (WOLD)	-0.03 (0.08)	0.75	-0.02 (0.01)	0.02	—	-1.47 (0.70)	0.06	-0.55 (0.11)	$4.5 \times 10^{-7}$	—
Non-word repetition 8 (CNRep)	-0.14 (0.14)	0.34	-0.02 (0.01)	0.01	—	-0.16 (1.30)	0.90	-0.47 (0.11)	$1.5 \times 10^{-5}$	—
VIQ 8 (WISC-III)	-0.23 (0.15)	0.16	-0.03 (0.01)	$5.0 \times 10^{-5}$	—	-0.36 (1.37)	0.80	-0.71 (0.11)	$< 1 \times 10^{-10}$	—
Pooled reading	-0.35 (0.09)	$9.2 \times 10^{-5}$	-0.03 (0.01)	$1.4 \times 10^{-6}$	0.79	1.67 (0.78)	0.03	-0.50 (0.09)	$4.9 \times 10^{-8}$	0.09
Pooled spelling	-0.18 (0.11)	0.10	-0.04 (0.01)	$1.1 \times 10^{-8}$	0.28	0.40 (1.25)	0.75	-0.42 (0.10)	$1.3 \times 10^{-5}$	0.39
Pooled LRAs	-0.18 (0.07)	0.01	-0.03 (0.01)	$1.9 \times 10^{-6}$	0.05	0.29 (0.69)	0.67	-0.49 (0.08)	$< 1 \times 10^{-10}$	0.004

Note: Sets of conservative ( $P < 5 \times 10^{-8}$ ) and subthreshold ( $P < 0.0015$ ) ADHD instruments were extracted from ADHD (PGC + iPSYCH), EA (SSGAC) and LRAs (ALSPAC) GWAS summary statistics. ADHD-specific effects independent of EA ( $\beta_{ADHD}$ ) and ADHD effects shared with EA ( $\beta_{EA}$ ) on LRAs were estimated with MVRs (Fig. 2b). ADHD effects shared with EA genetic effect estimates of ADHD-associated variants and presented with respect to missing school years.  $\beta_{ADHD}$  quantifies the change in LRA Z-score per log odds increase in ADHD liability.  $\beta_{EA}$  quantifies the change in LRA Z-score per missing year of schooling. Pooled estimates for reading, spelling and global LRAs (Table 1) were obtained through random-effects meta-regression. Evidence for effect heterogeneity ( $P_{het}$ ) was monitored through Cochran's Q-test. To compare effect sizes of  $\beta_{ADHD}$  and  $\beta_{EA}$ , MVR was carried out using standardised genetic effect estimates for which results are provided in Fig. 2c.

LRAs literacy-related and language-related abilities; ADHD Attention-Deficit/Hyperactivity Disorder; EA educational attainment;  $P_{het}$ , heterogeneity P-value, a accuracy, c comprehension, s speed, WORD Wechsler Objective Reading Dimension, NBO Nunes, Bryant and Olson (ALSPAC specific instrument), NARA II The Neale Analysis of Reading Ability-Second Revised British Edition, TOWRE Test Of Word Reading Efficiency, NW non-word, NB Nunes and Bryant (ALSPAC specific instrument), PhonAware phonemic awareness, AAT Auditory Analysis Test, WOLD Wechsler Objective Language Dimensions, CNRep Children's Test of Nonword Repetition, VIQ verbal intelligence quotient, WISC-III Wechsler Intelligence Scale for Children III, MVR Multivariable regression

<sup>a</sup>ADHD genetic effects shared with EA as assessed through EA genetic effect estimates of ADHD-associated variants



in addition to phonemic awareness and verbal intelligence, but not listening comprehension and non-word repetition abilities. Importantly, genetic overlap between polygenic ADHD risk and all of the LRAs studied was inflated by genetic effects shared with EA.

Using independent ADHD discovery samples, these findings show that genetic overlap between ADHD and literacy-related impairments observed in twin and family studies<sup>16–18</sup> can be extended to genetic associations, as captured by common variation in general population samples. The identified association profile suggests that not only reading-related abilities (including both word and nonword reading skills), but also phonological and spelling-related abilities share genetic aetiology with ADHD. These interrelated LRAs may, as hypothesised for RD, arise from a phonological impairment<sup>48,49</sup>, which affects decoding and reading skills<sup>50</sup>, but also spelling abilities<sup>51</sup>. However, reading abilities can, once developed, also shape phonological skills<sup>52</sup>.

In addition, this study suggests that genetic associations between polygenic ADHD and LRAs reflect, at least partially, shared genetic influences with genetically predictable EA and that, equally likely, genetic associations between polygenic EA and LRAs share genetic influences with ADHD. The magnitude of these shared effects, modelled with different MVR approaches, was comparable with each other. This is consistent with reciprocal genetic influences between EA and ADHD (Fig. 2a) and supports an intergenerational multiple-deficit model proposed for reading disability<sup>15,53</sup>. Children growing up in disadvantaged environments, genetically predictable through polygenic EA scores<sup>54</sup>, might be more vulnerable to psychiatric illness including ADHD<sup>55</sup> that affects, in turn, their LRAs. In addition, adolescents with ADHD might be more likely to leave school at an earlier age, with lower LRA performance and EA, and pass on an increased genetic load to their own children<sup>56</sup>.

Here, we demonstrate that disentangling multivariate genetic interrelationships between ADHD, EA and LRAs using MVR can aid the interpretation of genetic overlap, while controlling for collider bias<sup>42</sup>. However, using MVR, the detection of these polygenic associations was strongly governed by the choice of genetic variants. Conservative ADHD instruments identified large ADHD-specific effects on reading as a domain and little evidence for genetic effects that are shared with EA, although they had limited power<sup>57</sup>. They comprised 15 independent SNPs only, including variation within *FOXP2*, a gene that has been implicated in childhood apraxia of speech and expressive and receptive language impairments (<http://omim.org/entry/602081>)<sup>58</sup>. On the other hand, subthreshold instruments, including thousands of variants, tagged ADHD-specific polygenic links with LRAs (conditional on EA) with smaller effects, but with higher predictive

accuracy. However, these instruments also captured shared genetic effects with EA, affecting polygenic links between ADHD and all of the LRAs studied. These shared genetic influences were of equal strength and at least equal magnitude compared to ADHD-LRA associations independent of EA. Contrary, a previous twin study showed that the genetic covariance between ADHD and reading difficulties was largely independent of genetic effects shared with IQ<sup>19</sup>, suggesting that our findings may also reflect socio-economic influences. Thus, in order to improve reading and, more generally, literacy-related deficits in children with ADHD, there is potentially a need for further intervention programmes targeting EA-independent underlying neurocognitive deficits, beyond general training programmes aiming at schooling outcomes<sup>59</sup>.

In general, our findings are consistent with an omnigenic<sup>60</sup> model of complex trait architectures, compatible with a general factor model of psychopathology<sup>61</sup>, including ADHD<sup>62</sup>. The omnigenic model construes that only the largest-effect variants will reflect ADHD specificity, and may thus tag the most trait-specific associations between ADHD and reading, independent of EA. The majority of variants, however, will capture pleiotropic (omnigenic) influences pointing to highly interconnected neural networks<sup>60</sup> that give rise to genetic confounding. Consequently, the majority of subthreshold variants, captured by both ADHD and EA subthreshold variants, are likely to represent highly powerful cross-trait genetic predictors that may enhance and induce genetic overlap.

Finally, the methodological framework within this work has not only relevance for studies investigating polygenic links between ADHD and LRAs, but for many studies examining multivariate trait interrelationships that involve shared genetic effects with a genetically predictable confounder. Specifically, our findings suggest that lower variant selection thresholds can introduce genetic variance sharing that is unspecific and needs to be accounted for before identified associations can be interpreted in terms of underlying mechanisms, including shared genetic aetiologies. This is especially important as current guidelines for studying polygenic links with allelic scores recommend aggregating genetic variants across less stringent significance thresholds to maximise genetic association between discovery and target samples<sup>63,64</sup>.

This study has several limitations. Firstly, ALSPAC, as other cohort studies, suffers from attrition<sup>65,66</sup>. Sensitivity analyses showed that this is unlikely to bias our findings based on conservative instruments. However, links identified using subthreshold ADHD variants, might have been underestimated given that individuals with a genetic predisposition to ADHD (but also smoking initiation, higher body mass index, neuroticism, schizophrenia and depression) are more likely to drop out<sup>66</sup>. Secondly, the

strength of the genetic overlap between polygenic ADHD risk and LRAs may vary according to ADHD symptom domain levels, implicating especially inattentiveness<sup>67</sup>, as well as the nature of the literacy-related or language-related ability involved (as we observed evidence for effect heterogeneity when combining all LRAs). It is conceivable that also other verbal abilities, not investigated in this study, such as grammar, expressive vocabulary or pragmatic skills, may genetically overlap with ADHD. Furthermore, we only studied the extent to which shared genetic variance with EA affects the genetic association between ADHD and LRAs. However, we found little evidence for the presence of additional unaccounted for genetic influences using these instruments, i.e., effects that are not yet captured by either genetically predicted ADHD or EA. Finally, the power of available LRA GWAS summary statistics is still too low to generate genetic instruments supporting reverse models. Larger and more detailed clinical and population-based samples, as well as extensive multivariate variance analyses of the spectrum of LRAs (that are currently computationally expensive<sup>68</sup>) will help to further characterise the overlap between ADHD and literacy-related and language-related cognitive processes.

## Conclusion

Polygenic associations of clinical ADHD and a range of LRAs are to a large extent attributable to genetic effects that are also shared with EA, especially when investigated with genetic variants typically selected for polygenic scoring approaches. Adjusting for these unspecific genetic effects reveals an ADHD-specific association profile that primarily involves literacy-related impairments.

## Acknowledgements

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. The ALSPAC GWAS data was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. This research was facilitated by the Social Science Genetic Association Consortium (SSGAC) and Psychiatric Genomics Consortium (PGC) by providing access to genome-wide summary statistics. This publication is the work of the authors and E.V. and B.S.T.P. will serve as guarantors for the contents of this paper. E.V., C.Y.S., B.S.T.P. and S.E.F. are supported by the Max Planck Society. The iPSYCH project is funded by the Lundbeck Foundation (grant no R102-A9118 and R155-014-1724) and the universities and university hospitals of Aarhus and Copenhagen. Genotyping of iPSYCH and PGC samples was supported by grants from the Lundbeck Foundation, the Stanley Foundation, the Simons Foundation (SFARI 311789 to M.J.D.), and NIMH (5U01MH094432-02 to M.J.D.). The Danish National Biobank resource was supported by the Novo Nordisk Foundation. Data handling and analysis was supported by NIMH (1U01MH109514-01 to Michael O'Donovan and Anders D Børglum). High-performance computer capacity for handling and statistical analysis of iPSYCH data on the GenomeDK HPC facility was provided by the Centre for Integrative Sequencing, iSEQ, Aarhus University,

Denmark (grant to Anders D Børglum). An earlier version of this manuscript is available as a preprint on bioRxiv.

## iPSYCH-Broad-PGC ADHD Consortium members:

Esben Agerbo<sup>3,19,20</sup>, Thomas Damm Als<sup>3,4,5</sup>, Marie Bækved-Hansen<sup>3,21</sup>, Rich Belliveau<sup>12</sup>, Anders D. Børglum<sup>3,4,5</sup>, Jonas Bybjerg-Grauholm<sup>3,21</sup>, Felicia Cerrato<sup>12</sup>, Kimberly Chambert<sup>12</sup>, Claire Churchhouse<sup>11,12,13</sup>, Søren Dalsgaard<sup>19</sup>, Mark J. Daly<sup>11,12,13</sup>, Ditte Demontis<sup>3,4,5</sup>, Ashley Dumont<sup>12</sup>, Jacqueline Goldstein<sup>11,12,13</sup>, Jakob Grove<sup>3,4,5,22</sup>, Christine S. Hansen<sup>3,21,23</sup>, Mads Engel Hauberg<sup>3,4,5</sup>, Mads V. Hollegaard<sup>3,21</sup>, David M. Hougaard<sup>3,21</sup>, Daniel P. Howrigan<sup>11,12</sup>, Hailiang Huang<sup>11,12</sup>, Julian Maller<sup>12,24</sup>, Alicia R. Martin<sup>11,12,13</sup>, Joanna Martin<sup>12,25,26</sup>, Manuel Mattheisen<sup>3,4,5,27,28</sup>, Jennifer Moran<sup>12</sup>, Ole Mors<sup>3,29</sup>, Preben Bo Mortensen<sup>3,4,19,20</sup>, Benjamin M. Neale<sup>11,12,13</sup>, Merete Nordentoft<sup>3,30</sup>, Jonatan Pallesen<sup>3,4,5</sup>, Duncan S. Palmer<sup>11,12</sup>, Carsten Bøcker Pedersen<sup>3,19,20</sup>, Marianne Giørtz Pedersen<sup>3,19,20</sup>, Timothy Poterba<sup>11,12,13</sup>, Jesper Buchhave Poulsen<sup>3,21</sup>, Stephan Ripke<sup>11,12,13,31</sup>, Elise B. Robinson<sup>11,32</sup>, F. Kyle Satterstrom<sup>11,12,13</sup>, Christine Stevens<sup>12</sup>, Patrick Turley<sup>11,12</sup>, Raymond K. Walters<sup>11,12</sup>, Thomas Werge<sup>3,23,33</sup>

<sup>19</sup>National Centre for Register-based Research, Aarhus University, Aarhus, Denmark

<sup>20</sup>Centre for Integrated Register-based Research, Aarhus University, Aarhus, Denmark

<sup>21</sup>Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark

<sup>22</sup>Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark

<sup>23</sup>Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, Roskilde, Denmark

<sup>24</sup>Genomics plc, Oxford, United Kingdom

<sup>25</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

<sup>26</sup>MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Cardiff, United Kingdom

<sup>27</sup>Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

<sup>28</sup>Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden

<sup>29</sup>Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark

<sup>30</sup>Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Copenhagen, Denmark

<sup>31</sup>Department of Psychiatry, Charité Universitätsmedizin Berlin Campus Benjamin Franklin, Berlin, Germany

<sup>32</sup>Department of Epidemiology, Harvard Chan School of Public Health, Boston, MA, USA

<sup>33</sup>Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

## Author details

<sup>1</sup>Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. <sup>2</sup>International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands. <sup>3</sup>The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus, Denmark. <sup>4</sup>Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark. <sup>5</sup>Department of Biomedicine-Human Genetics, Aarhus University, Aarhus, Denmark. <sup>6</sup>MRC Biostatistics Unit, University of Cambridge, Cambridge, UK. <sup>7</sup>Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. <sup>8</sup>Speech and Hearing Sciences, University of New Mexico, Albuquerque, USA. <sup>9</sup>Department of Complex Trait Genetics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. <sup>10</sup>Erasmus University Rotterdam Institute for Behavior and Biology, Rotterdam, The Netherlands. <sup>11</sup>Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. <sup>12</sup>Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA. <sup>13</sup>Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA. <sup>14</sup>Departments of Psychiatry and Neuroscience and Physiology, SUNY Upstate Medical University, New York, USA. <sup>15</sup>MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. <sup>16</sup>School of Oral and Dental Sciences, University of Bristol, Bristol, UK. <sup>17</sup>School of Social and Community Medicine, University of Bristol, Bristol, UK. <sup>18</sup>Donders Institute for

Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

#### Conflict of interest

The authors declare that they have no conflict of interest.

#### Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Supplementary Information** accompanies this paper at (<https://doi.org/10.1038/s41398-018-0324-2>).

Received: 17 October 2018 Accepted: 13 November 2018

Published online: 24 January 2019

#### References

- Geurts, H. M. & Embrechts, M. Language profiles in ASD, SLI, and ADHD. *J. Autism Dev. Disord.* **38**, 1931–1943 (2008).
- Helland, W. A., Posserud, M.-B., Helland, T., Heimann, M., Lundervold, A. J. Language impairments in children with ADHD and in children with reading disorder. *J. Atten. Disord.* 1087054712461530 (2012).
- Germanò, E., Gagliano, A. & Curatolo, P. Comorbidity of ADHD and dyslexia. *Dev. Neuropsychol.* **35**, 475–493 (2010).
- Maughan, B. & Carroll, J. Literacy and mental disorders. *Curr. Opin. Psychiatry* **19**, 350–354 (2006).
- Mayes, S. D. & Calhoun, S. L. Learning, attention, writing, and processing speed in typical children and children with ADHD, autism, anxiety, depression, and oppositional-defiant disorder. *Child Neuropsychol. J. Norm. Abnorm. Dev. Child Adolesc.* **13**, 469–493 (2007).
- Mayes, S. D. & Calhoun, S. L. Frequency of reading, math, and writing disabilities in children with clinical disorders. *Learn Individ Differ.* **16**, 145–157 (2006).
- Czamara, D. et al. Children with ADHD symptoms have a higher risk for reading, spelling and math difficulties in the GINplus and LISplus Cohort Studies. *PLoS ONE* **8**, e63859 (2013).
- Åsberg, J., Kopp, S., Berg-Kelly, K. & Gillberg, C. Reading comprehension, word decoding and spelling in girls with autism spectrum disorders (ASD) or attention-deficit/hyperactivity disorder (AD/HD): performance and predictors. *Int. J. Lang. Commun. Disord.* **45**, 61–71 (2010).
- Green, B. C., Johnson, K. A. & Bretherton, L. Pragmatic language difficulties in children with hyperactivity and attention problems: an integrated review. *Int. J. Lang. Commun. Disord.* **49**, 15–29 (2014).
- Hawkins, E., Gathercole, S., Astle, D. & Holmes, J. The Calm Team. Language problems and ADHD symptoms: how specific are the links?. *Brain Sci* **6**, 50 (2016).
- Polanczyk, G., de Lima, M. S., Horta, B. L., Biederman, J. & Rohde, L. A. The worldwide prevalence of ADHD: a systematic review and meta-regression analysis. *Am. J. Psychiatry* **164**, 942–948 (2007).
- Peterson, R. L. & Pennington, B. F. Developmental dyslexia. *Lancet Lond. Engl.* **379**, 1997–2007 (2012).
- Thapar, A. & Cooper, M. Attention deficit hyperactivity disorder. *Lancet* **387**, 1240–1250 (2016).
- Gough, P. B. & Tunmer, W. E. Decoding, reading, and reading disability. *Remedial Spec. Educ.* **7**, 6–10 (1986).
- Pennington, B. F. From single to multiple deficit models of developmental disorders. *Cognition* **101**, 385–413 (2006).
- Martin, N. C., Levy, F., Pieka, J. & Hay, D. A. A genetic study of attention deficit hyperactivity disorder, conduct disorder, oppositional defiant disorder and reading disability: aetiological overlaps and implications. *Int. J. Disabil. Dev. Educ.* **53**, 21–34 (2006).
- Willcutt, E. G., Pennington, B. F. & DeFries, J. C. Twin study of the etiology of comorbidity between reading disability and attention-deficit/hyperactivity disorder. *Am. J. Med. Genet.* **96**, 293–301 (2000).
- Willcutt, E. G., Pennington, B. F., Olson, R. K. & DeFries, J. C. Understanding comorbidity: a twin study of reading disability and attention-deficit/hyperactivity disorder. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **144B**, 709–714 (2007).
- Paloyelis, Y., Rijdsdijk, F., Wood, A. C., Asherson, P. & Kuntsi, J. The genetic association between ADHD symptoms and reading difficulties: the role of inattentiveness and IQ. *J. Abnorm. Child Psychol.* **38**, 1083–1095 (2010).
- Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
- Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* **51**, 63 (2019).
- Selzam, S. et al. Genome-wide polygenic scores predict reading performance throughout the school years. *Sci. Stud. Read* **21**, 334–349 (2017).
- Luciano, M. et al. Single nucleotide polymorphisms associated with reading ability show connection to socio-economic outcomes. *Behav. Genet.* **47**, 469–479 (2017).
- Krapohl, E. et al. Widespread covariation of early environmental exposures and trait-associated polygenic variation. *Proc. Natl Acad. Sci* **114**, 11727–11732 (2017).
- Krapohl, E. et al. Multi-polygenic score approach to trait prediction. *Mol. Psychiatry* **23**, 1368–1374 (2017).
- Burgess, S. et al. Dissecting causal pathways using mendelian randomization with summarized genetic data: application to age at menarche and risk of breast cancer. *Genetics* **207**, 481–487 (2017).
- Fraser, A. et al. Cohort profile: the avon longitudinal study of parents and children: ALSPAC mothers cohort. *Int. J. Epidemiol.* **42**, 97–110 (2013).
- Boyd, A. et al. Cohort Profile: the children of the 90s—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* **42**, 111–127 (2013).
- Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* **95**, 221–227 (2005).
- Goodman, R., Ford, T., Richards, H., Gatward, R. & Meltzer, H. The development and well-being assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J. Child Psychol. Psychiatry* **41**, 645–655 (2000).
- Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
- Mors, O., Perto, G. P. & Mortensen, P. B. The Danish psychiatric central research register. *Scand. J. Public Health* **39**(7 Suppl), 54–57 (2011).
- Pedersen, C. B. et al. The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol. Psychiatry* **23**, 6–14 (2017).
- Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
- Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M. & Wray, N. R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinforma. Oxf. Engl.* **28**, 2540–2542 (2012).
- Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
- Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S. & Yang, J. Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
- International Schizophrenia Consortium, Purcell, S. M. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
- Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS. Genet.* **9**, e1003348 (2013).
- Stergiakouli, E. et al. Shared genetic influences between dimensional ASD and ADHD symptoms during child and adolescent development. *Mol. Autism* **8**, 18 (2017).
- Aschard, H., Vilhjálmsson, B. J., Joshi, A. D., Price, A. L. & Kraft, P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am. J. Hum. Genet.* **96**, 329–339 (2015).
- Pierce, B. L., Ahsan, H. & Vanderweele, T. J. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int. J. Epidemiol.* **40**, 740–752 (2011).
- Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

45. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS. Genet.* **5**, e1000529 (2009).
46. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).
47. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**(R1), R89–R98 (2014).
48. Carroll, J. M. & Snowling, M. J. Language and phonological skills in children at high risk of reading difficulties. *J. Child Psychol. Psychiatry* **45**, 631–640 (2004).
49. Hulme, C., Snowling, M. J. *Developmental Disorders of Language Learning and Cognition*. 1 edn. (Wiley-Blackwell, Chichester, Malden, 2009). 446 p.
50. Rack, J., Snowling, M. & Olson, R. The nonword reading deficit in developmental. *Dyslexia a Rev. Read. Res Q.* **27**, 28–53 (1992).
51. Treiman, R. *Beginning to Spell: A Study of First-grade Children*. (Oxford University Press, New York, 1992). 380 p.
52. Kamhi, A. G. & Catts, H. W. *Language and Reading Disabilities*. 3 edn. (Pearson, Boston, 2011). p. 320.
53. van Bergen, E., van der Leij, A. & de Jong, P. F. The intergenerational multiple deficit model and the case of dyslexia. *Front. Hum. Neurosci.* **8**, 346 (2014).
54. Ayorech, Z., Krapohl, E., Plomin, R. & Stumm, S. Genetic influence on intergenerational educational attainment. *Psychol. Sci.* **28**, 1302–1310 (2017).
55. Reiss, F. Socioeconomic inequalities and mental health problems in children and adolescents: a systematic review. *Soc. Sci. Med.* **90**, 24–31 (1982).
56. Russell, G., Ford, T., Rosenberg, R. & Kelly, S. The association of attention deficit hyperactivity disorder with socioeconomic disadvantage: alternative explanations and evidence. *J. Child Psychol. Psychiatry* **55**, 436–445 (2014).
57. Schatzkin, A. et al. Mendelian randomization: how it can—and cannot—help confirm causal relations between nutrition and cancer. *Cancer Prev. Res.* **2**, 104–113 (2009).
58. Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F. & Monaco, A. P. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519–523 (2001).
59. Daley, D. & Birchwood, J. ADHD and academic performance: why does ADHD impact on academic performance and what can be done to support ADHD children in the classroom? *Child Care Health Dev.* **36**, 455–464 (2010).
60. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
61. Lahey, B. B. et al. Is there a general factor of prevalent psychopathology during adulthood? *J. Abnorm. Psychol.* **121**, 971–977 (2012).
62. Franke, B. What's in a name: the "omnigenic" model as a point of departure for polygenic psychiatric disorders. *JPBS* **2**, S7 (2017).
63. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: polygenic risk score software. *Bioinformatics* **31**, 1466–1468 (2015).
64. Wray, N. R. et al. Research review: polygenic methods and their application to psychiatric traits. *J. Child Psychol. Psychiatry* **55**, 1068–1087 (2014).
65. Martin, J. et al. Association of genetic risk for schizophrenia with non-participation over time in a population-based cohort study. *Am. J. Epidemiol.* **183**, 1149–1158 (2016).
66. Taylor, A. et al. Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* **47**, 1207–1216 (2018).
67. Greven, C. U., Harlaar, N., Dale, P. S. & Plomin, R. Genetic overlap between ADHD symptoms and reading is largely driven by inattentiveness rather than hyperactivity-impulsivity. *J. Can. Acad. Child Adolesc. Psychiatry* **20**, 6–14 (2011).
68. St Pourcain, B. et al. Developmental changes within the genetic architecture of social communication behaviour: a multivariate study of genetic variance in unrelated individuals. *Biol. Psychiatry* **83**, 598–606 (2017).
69. WORD. *Wechsler Objective Reading Dimensions Manual*. (Psychological Corporation, London, 1993) 146 p.
70. Nunes, T., Bryant, P. & Olsson, J. Learning morphological and phonological spelling rules: an intervention study. *Sci. Stud. Read.* **7**, 289–307 (2003).
71. Neale, M. D. *Neale analysis of reading ability: second revised British Edition*. (NFER-Nelson, London, 1997).
72. TOWRE. *Test of word reading efficiency: examiner's manual. PRO-ED* (1999) 108 p.
73. Rosner, J. & Simon, D. P. The auditory analysis test: an initial report. *J. Learn. Disabil.* **4**, 384–392 (1971).
74. Rust, J., Wechsler, D. WOLD: *Wechsler Objective Language Dimensions*. (Psychological Corp., London, Artsberg Enterprises Ltd., Hong Kong, 1996).
75. Gathercole, S. E., Willis, C. S., Baddeley, A. D. & Emslie, H. The children's test of nonword repetition: a test of phonological working memory. *Mem. Hove Engl.* **2**, 103–127 (1994).
76. Wechsler, D., Golombok, S., Rust, J. WISC-III U. K. *Wechsler Intelligence Scale for Children—Third Edition UK Manual*. (The Psychological Corporation, Sidcup, 1992).