Title

Brain structure, phenotypic and genetic correlates of reading performance

Author list

Amaia Carrión-Castillo[1*], Pedro M. Paz-Alonso[1,2], Manuel Carreiras[1,2,3,*]

* Corresponding author

Emails: a.carrion@bcbl.eu; m.carreiras@bcbl.eu

Affiliations

[1] Basque Center on Cognition, Brain and Language (BCBL), Donostia-San Sebastián, Spain

[2] Ikerbasque, Basque Foundation for Science, Bilbao, Spain

[3] University of the Basque Country, Bilbao, Spain

Abstract

Reading is an evolutionarily recent development that recruits and tunes brain circuitry connecting primary- and language-processing regions. We investigated whether metrics of the brain's physical structure correlate with reading performance and whether genetic variants affect this relationship. To this aim, we used the ABCD dataset (N=9,013) of 9-to-10-year-olds and focused on 150 measures of cortical regional surface area (CSA) and thickness. Our results reveal that reading performance is associated with nine measures of brain structure including relevant regions of the reading network. Furthermore, we show that this relationship is partially mediated by genetic factors for two of these measures: the cortical surface area of the entire left hemisphere and, specifically, of the left superior temporal gyrus CSA. These effects emphasise the complex and subtle interplay between genes, brain and reading, which is a partly heritable polygenic skill that relies on a distributed network.

Main text

Introduction

Reading requires a brain system capable of integrating orthographic, phonological, and lexico-semantic features of written words [1]. The invention of reading approximately 5000 years ago does not provide enough time on an evolutionary scale to develop such a specific circuitry. Thus, reading recruits already available networks in the brain implicated in language and visual processing [2], including the inferior frontal gyrus (IFG), the superior temporal gyrus (STG), the inferior parietal lobe, and occipito-temporal regions (fusiform gyrus, inferior temporal gyrus; e.g. [3,4]). The development of this reading network, which includes a dorsal (phonological) and a ventral (lexico-semantic) processing stream [5,6], is shaped by the literacy environment and genetic constraints.

The most convincing evidence for the importance of environmental factors in developing this network is that learning to read requires instruction and that illiterate individuals do not show the landmark of the literate brain, a functionally defined region in the occipito-temporal gyrus that is specialized for word recognition [7,8]. Furthermore, the development of reading skills is influenced by socioeconomic factors (caregiver education and home-literacy environment), reading instruction methodology/practice and the orthography of the language through which such skills are being learned [9,10].

The reading network has also been investigated in individuals with developmental dyslexia (DD), which is defined as a reading disability despite normal intelligence, adequate education and lack of obvious sensory or neurological damage [11,12]. Although many different deficits have been characterized in DD, there is a general agreement that DD involves phonologically-related reading processing deficits (e.g., [13,14]. In spite of mixed literature on neuroanatomical markers of DD [15], several of the key regions that are important for reading show reduced volume or surface area in individuals with DD [16]. This suggests that differences in the development of the reading network might contribute towards the reading deficits that characterise DD.

Genetic variation explains a substantial component of reading abilities, with twin-based heritability (twin-h^2) estimates of 0.66 for general reading performance (reading speed and accuracy) [17] and population-based heritability (SNP-h^2) estimates of up to 0.50 for reading accuracy [18]. The largest genome-wide association study (GWAS) of language and reading-related traits to date (N~34,000) has confirmed the heritability estimates for these traits, identifying a single genome-wide significant locus for word reading accuracy in chromosome 1 [19]. This study also highlighted a shared genetic component of reading-related measures with other cognitive components and the cortical surface area (CSA) of the banks of the left superior temporal sulcus [19]. DD also has a complex genetic and environmental aetiology, with twin-h^2 estimates of 0.40-0.60 [20] and SNP-h^2 estimates of 0.15-0.19 [21,22]. Another recent GWAS study with an unprecedented sample size (over 51,000 cases) identified 42 loci associated with self-reported dyslexia status at the genome-wide significant level, consistent with high polygenicity of the trait [21]. Recently, the brain imaging genetics field has also been revolutionized by meta-analytic efforts [23] and large-scale datasets such as the UK Biobank [24,25], providing insights into the role genetics play in shaping brain structure. However, a mechanistic account of how these genetic effects contribute to the neurobiology of cognitive functions and human behaviour is still lacking.

Over the last decade, there has been an increase in studies on the brain imaging genetics of reading performance, mostly examining the association between a wide range of neuroimaging phenotypes and candidate genes for reading (see [26] for a review), as well as a few genome-wide studies [27]. Functional studies so far have relied on small samples (range: 33-427 participants [26]) and have produced mixed findings, reflecting the challenging task of characterizing the reading phenotype and combining it with informative task designs, as well as the difficulty in transferring those functional

designs into larger datasets. Structural imaging studies have had larger sample sizes (range: 56-1,717 [26]), but these are still too small to robustly identify the expected small genetic effects, given the polygenicity of both reading [19] and brain phenotypes [25]. Moreover, reading itself is a complex ability that can be measured through different constructs. For instance, word reading fluency measures both reading accuracy and speed and is used as a proxy for skilled reading, whereas other tasks such as nonword reading tap into the phonological decoding component [28]. These differences in the behavioural measures may also have contributed to the mixed results in the imaging literature [29,30]. Hence, these studies provide the groundwork to understand the link from genetic variation to brain phenotypes and reading, but systematicity in the analytical approaches is lacking as, until now, only a handful of phenotypes and genetic loci have been considered.

It is critical to use large datasets to identify robust and scalable brain correlates of reading to perform genetic analyses and to seek replicable results. The goal of the present study was twofold: (A) to identify brain correlates of reading in the developing brain; and (B) to examine whether genetics influence brain-reading associations (see Figure 1). To address these goals we used three complementary analytical approaches, namely (1) conducting a regression analysis in the Adolescent Brain Cognitive Development (ABCD) dataset of 9-to-10-year-old US children (N=9,013) to define structural cortical measures consistently associated with reading performance (goal A; Figure 1). (2) examining the genetic architecture of reading-related cognitive and brain measures by estimating the heritability of these traits using the ABCD and other publicly available datasets (goal B); and (3) exploring the shared genetic influences on brain and reading through genetic correlation and polygenic scoring combined with mediation analyses (goal B).

In sum, we expect to reliably identify brain-behaviour associations within the known reading network, and to take advantage of the large ABCD dataset to unravel other more subtle but reliable associations that have not been detected in smaller datasets [31]. As the reading network is mostly left-lateralized [32], we first consider regional measures of the left-hemisphere cortical thickness (CT) and CSA. However, in a second stage we look at the homotopic right-hemisphere regions of interest to check for hemispheric specificity. As speech comprehension ontogenetically precedes literacy [33], we hypothesize that hub regions of the speech processing network such as the IFG or STG will be associated with reading performance. Moreover, if genetic effects mediate this effect, they are likely to act upon those regions, more than on other regions such as the ventral occipito-temporal cortex, which is a highly plastic and environmentally malleable area [8].

Results

Left hemisphere structural correlates of reading performance

To identify reliable cortical structural correlates of reading in the developing brain (Figure 1A) we performed a regression analysis using the ABCD study dataset (N=9,013, Supplementary Fig.1). Since the ABCD dataset is a population-based dataset, we have focused on reading performance, treating it as a quantitative trait (see Supplementary Fig.1 for trait distribution). Variable definitions and descriptive statistics are shown in Supplementary Tables 1 and 2.

In a first baseline model we assessed the effect of covariates (Equation 1), in which fixed effects explained 18.26% of the variance in reading (marginal R^2). Age was a strong predictor for reading performance (std beta=0.2, se=0.01, F(1,8449.37)=537.7, p=2.27e-115), despite the relatively narrow age range in the sample (see Supplementary Table 2), whereas sex was not a significant predictor (see Supplementary Fig.2 and Supplementary Table 3). Socioeconomic factors were also strong predictors, with higher caregiver education and higher household income being associated with better reading performance. The first four genetic components were also associated with reading (Supplementary Table 3, Supplementary Fig.2).

We next tested the effect of 150 left-hemisphere measures on reading performance: the global measures of total CSA and mean CT, and 74 regional measures for CSA and CT using the Destrieux

parcellation [34]. For each regional measure, we included an absolute model (Model 1, Equation 2) without adjustment for global brain measures, and a second model (Model 2, Equation 3) that assessed the relative regional expansion/thickening after accounting for global measures (Supplementary Table 4). The global measures included in Model 2 were mean left-hemisphere CT for regional CT measures and total left-hemisphere CSA for regional CSA measures. We focused our analysis on regions that were consistently associated with reading in both models. Considering both models allowed us to assess global and local brain effects of each imaging measure [35,31].

The strongest association was observed with total CSA (Likelihood Ratio Test between nested models with and without variable of interest: $\chi^2(1)=115.11$; q-value=1.27e-24). This was reflected in 69 out of 74 of the regional CSA measures being significantly associated with reading in Model 1 (Figure 2A, Supplementary Table 4), although the majority of these (67/69) were no longer significant after adjusting for total CSA (Model 2), suggesting that in those regions the association with reading performance was driven by total CSA. Nevertheless, two regional CSA measures were significantly associated with reading in both models: the CSA of the lateral STG, which was consistently positively associated with reading (Model 1: $\chi^2(1)=96.14$;q-value=8.99e-21 and Model 2: $\chi^2(1)=16.65$, q-value=0.002; Figure 2A,C), and the CSA of the superior parietal gyrus (SPG), which showed a positive relationship to reading that was shifted to negative after adjustment (Model 1: $\chi^2(1)=9.48$; tq-value=0.005 and Model 2: $\chi^2(1)=9.63$; q-value=0.036). This reversal in the direction of effect reflects that although a larger overall CSA of the SPG is related to better reading performance, the relative size of CSA in this region is negatively associated with it.

CT of six regions was consistently associated with reading performance (Figure 2B, Supplementary Table 4). The postcentral gyrus and three occipital regions (occipital inferior gyrus and sulcus, occipito-temporal medial lingual gyrus, occipito-temporal lateral fusiform gyrus) were consistently associated with reading. Two superior frontal regions (middle frontal gyrus, superior frontal gyrus) showed a negative association, and the extent of the relation increased when adjusting for mean CT (Model 2), indicating that relatively thinner CT in these regions is associated with better reading performance.

Effect sizes were comparable across the three analysed subsets (Supplementary Fig.3). There was no significant correlation between CT and CSA beta estimates for reading in any of the analyses we ran (see Supplementary Fig.4). This suggests that the effects CSA and CT measures may have on reading are independent. Furthermore, there was no interaction with age or sex for either of the brain measures associated with reading (neither in Model 1 or 2, see Supplementary Table 5), which suggests that the effect we observed is stable across sexes at this relatively narrow developmental timepoint.

The specificity of the observed associations was assessed through sensitivity analysis by adjusting for additional cognitive variables that were correlated with reading performance (Supplementary Fig.5, Supplementary Table 6), namely fluid intelligence, matrix reasoning and/or picture vocabulary. Total CSA, CSA of the lateral STG, CT of the occipital inferior gyrus and sulcus and CT of the occipito-temporal lateral fusiform gyrus were significantly associated with reading performance across all of these analyses, while the associations were weaker and no longer significant for the other five measures after adjusting for some of the additional cognitive measures, suggesting that the association we observed with reading may be part of a more general cognitive effect in those specific regions (see Supplementary Table 6, Supplementary Fig.6).

To establish whether the observed associations were left-hemisphere specific, we also explored the effect of homotopic right-hemisphere measures on reading, which were imperfectly correlated with the left-hemisphere measures (Pearson's r ranging from 0.38 to 0.8 for the regional measures, see Supplementary Fig.7). The right total CSA was associated with reading ($\chi^2(1)=109.924$; p=1.16e-25) to a similar extent as the left CSA. At the regional level, the association was robust to adjustment for global measures (Model 2) for three measures (see Supplementary Fig.8). The CT of the right occipito-temporal lateral fusiform gyrus was consistently associated with reading performance (Model 1: $\chi^2(1)=8.93$; p=0.003 and Model 2: $\chi^2(1)=8.10$; p=0.005), while for the other two measures the

direction of the associations reversed when adjusting for global measures: the CSA of the right SPG reversed the direction of effect after adjustment for total right CSA (Model 1: $\chi^2(1)=8.47$; p=0.004 and Model 2: $\chi^2(1)=5.05$; p=0.024) and the CT of the right superior frontal gyrus was negatively associated in both models and the association became stronger after adjustment for mean right CT (Model 1: $\chi^2(1)=10.01$; p=0.002 and Model 2: $\chi^2(1)=35.57$; p=2e-9).

Correlations among these nine reading-associated measures showed two independent clusters reflecting the type of measurement: one for the CSA measures and the other for CT measures (Supplementary Fig.7a). After correction for global measures, the correlations showed an overall decrease in the strength of the effect, indicating that they are mostly independent of each other and therefore not likely to reflect the same relationship with reading performance (see Supplementary Fig.7b). After this adjustment, the direction of some correlations was reversed, which may reflect the complex relationship between the regional and global measures: e.g. a weak negative correlation between the CSA of the lateral STG and the CSA of the SPG, and weak negative correlations between the CT of the superior and frontal gyri with the rest of the CT measures.

In sum, a set of nine cortical structural measures are associated with reading performance in this sample consistently before and after correcting for global measures. Sensitivity analyses support that several of these effects are specific to reading when controlling for other cognitive measures, and that the associations are mostly left-hemisphere specific.

Genetic architecture of reading-related cognitive and brain measures

As a first step to investigate whether genetics affect the identified brain-behavioural associations, we established whether our traits of interest were heritable (Figure 1B.2). Heritability was estimated for reading, related cognitive measures and reading-associated brain measures (as defined in Supplementary Tables 2 and 3) using different methods and datasets: in the ABCD dataset SNP-h^2 was estimated using genome-based restricted maximum likelihood (GREML [36]) in a set of unrelated individuals of European ancestry (N=4,716) whereas summary statistics of GWASes from publicly available datasets were used to compute SNP-h^2 through linkage-disequilibrium score regression (LDSC [37]) (see Supplementary Table 7). The heritability measures across all traits and methods are reported in Supplementary Table 8 and Supplementary Fig.9.

Reading had a SNP-h^2 of 0.14 in the ABCD dataset (see Supplementary Table 8), which was nominally significantly different from 0 , while the remaining cognitive measures were low to moderately heritable, ranging from 0.09 for WISC-V to 0.25 for vocabulary (95%CI [0.1;0.4]; corrected p=0.01). All nine brain structural measures associated with reading were also heritable (Supplementary Fig.9b), with the CT of the postcentral gyrus having the highest estimate (SNP-h^2=0.32, 95%CI [017;0.46], corrected p=2.26e-5), and the inferior occipital gyrus and sulcus having the lowest (SNP-h^2=0.11). All were nominally significantly different from zero, although four (two regional CSA and two occipital CT measures) did not survive the correction for the 21 tested measures (Supplementary Table 8). After the adjustment for global brain measures, most regional estimates were lower (Supplementary Table 8; Supplementary Fig.9b), and only two CT measures were nominally (CT of middle frontal gyrus) or significantly (CT of the postcentral gyrus) different from zero.

SNP-h^2 estimates from summary statistics from GWASes were moderate for word reading, dyslexia, educational attainment and cognitive performance, consistent with original GWAS reports (Supplementary Tables 7 and 8; [19,22,38]). All of the SNP-h^2 (LDSC) estimates for brain measures, derived from the UK Biobank, were significantly different from zero (all corrected p-values<0.05; see Supplementary Table 8), ranging from 0.31 (total CSA) to 0.12 (CT of the occipital inferior gyrus and sulcus).

In sum, these analyses support that reading and reading-associated cognitive and brain measures have a modest genetic component that partly explains part of the variation in these traits.

Shared genetic influences on brain and reading

Having established that genetic variation explains part of the variance of the reading-associated brain measures, we explored the extent to which genetics contribute to the brain-behaviour associations (Figure 1B.3). To this end, we first computed genetic correlation estimates. Then, we followed up the most promising brain measures with polygenic score (PGS) analyses. Finally, we conducted a mediation analysis to examine whether CSA measures mediate PGS effects of cognitive performance on reading.

Genetic correlation

Genetic correlation (r_g) is the proportion of variance that two traits share due to genetic variance. We estimated genetic correlations across cognitive traits and brain measures using LDSC. Bivariate GREML analysis was not considered as it was not well-powered to detect genetic correlations between brain measures and reading in the unrelated European ancestry ABCD subset (see Supplementary Fig.10).

There was no evidence of genetic correlation that survived multiple testing comparison (Bonferroni corrected for 36 measures) between any of the brain measures and educational attainment, cognitive performance, developmental dyslexia or word reading (Supplementary Fig.11a, Supplementary Table 9). We further analyzed the two measures showing the strongest association to reading using PGS: total CSA and the CSA of the lateral STG.

GWAS signals are normally interpreted as the effect a genetic variant has on a given phenotype. However, recent evidence points out that they can also measure confounds such as population stratification or indirect genetic effects, i.e. such as effects of parental genotypes that may influence the individual's environment and are also correlated with their genotype [39]. We thus compared the r_g estimates between brain imaging measures (population-based GWAS) and cognitive measures (cognitive function and years of schooling) from sibling-based and population-based GWASes [39]. Total CSA and sibling-based cognitive function had a positive genetic correlation; while several CT measures had nominally significant negative genetic correlations with sibling-based and/or population-based educational attainment and cognitive function (see Supplementary Table 10, Supplementary Fig.11b).

Polygenic scores

PGS are individual level predictors derived from the sum of effect alleles at a SNP, weighted by the regression coefficient describing each SNP's level of association with a trait. PGSes can also be used to study the genetic relationship between two traits by making predictions across traits [40].

We first used GWASes for word reading, educational attainment and cognitive performance to define the PGS that best predicted reading in the ABCD dataset (unrelated European ancestry subset). The PGS for cognitive performance (PGS_{CP}) at the p-value threshold of 0.05 explained the largest amount of variance in reading performance (ΔR^2=3.6%, estimate=0.1909, se=0.015; corrected p=8.73e-37; Supplementary Figs. 12-15). Therefore, we used this best PGS to perform cross-trait prediction in brain measures.

This PGS_{CP} was a significant predictor of the left CSA brain measures associated with reading, explaining up to 0.75% of the variance of the total CSA (estimate=0.089, se=0.013, corrected p=4e-11), and up to 0.56% of the CSA of the lateral STG (estimate=0.076, se=0.014, corrected p=3e-07). These results indicate that genetic effects that affect cognitive performance also influence CSA measures associated with reading. The regional prediction was significantly diminished and no longer significant when including the total CSA as a covariate in the regression analysis (ΔR^2=8.35e-05, estimate=0.015, se=0.112, corrected p=0.5300.707), suggesting that the genetic effect is shared between the total CSA and the CSA of the lateral STG (Figure 3).

Mediation analysis

The relation between reading performance, total CSA, the CSA of the lateral STG and PGS of cognitive performance was further explored through a set of mediation analyses that are summarized in Table 1. The effect of PGS_{CP} on reading was partially mediated by the CSA measure (total or STG), the indirect effect of total CSA explaining 4.9% of the total PGS-reading effect, and the CSA of the lateral STG explaining 4.2% of it (Figure 4).

When adjusting the CSA measures with each other, the indirect effect was no longer significant, which again supports that the genetic effect is common to both CSA measures (Table 1). For sensitivity, we also repeated the mediation analyses adjusting reading for general intelligence measures and picture vocabulary, after which the mediated effect was diminished but remained significant (total CSA: 3.3%; CSA of STG=3.3%; see Table 1).

In sum, through complementary approaches, we find evidence for a shared genetic component between reading performance and CSA measures associated with reading, namely the total left-hemisphere CSA and the CSA of the lateral STG.

Discussion

In the current study, we have established a set of morphometric measures associated with reading performance in the ABCD dataset, constituted by 3 CSA and 6 CT measures, including relevant regions of the reading network. Several of the effects were robust when controlling for other cognitive measures, highlighting the specificity of these effects for reading. Next, we explored whether genes played a role in the brain-behaviour relationship and found evidence for genetic overlap with two CSA measures (total CSA and CSA of the lateral STG) but not for any of the CT measures. Finally, through mediation analysis, we showed that the gene-behaviour association is partially mediated through the CSA measures. These results are discussed next.

A set of nine cortical structural measures were consistently associated with reading in the ABCD study of 9-to-10-year-old US children. The scale of this dataset (N>9,000) allowed us to perform a whole-brain search for morphometric (CSA and CT) correlates of reading in the left hemisphere, and to uncover small but consistent effects. For each regional CT or CSA measure, we assessed two effects: a global effect which is a measure of variability within a given region, and a second relative effect that reflects variability of the regional expansion (i.e. correcting for global measures). The relative regional CSA and CT measures are also of interest. For instance, a gradient of relatively greater CSA and relatively thinner CT has been positively associated with general cognitive ability; this pattern seems to be driven by genetic associations, reflecting cortical expansion during development and evolution [41].

Total CSA was associated with reading performance, and this effect was reflected by a global association of most regional CSA measures. Smaller brain volume [15] and lower CSA [42] has been reported in dyslexic individuals. A previous study that used the ABCD dataset also demonstrated that performance on crystallised cognitive measures, including reading, was more strongly associated with total CSA compared to the regional CSA measures [43].

We also observed a regional effect of CSA beyond the generalized effect, as the relative CSA of the lateral part of the STG was associated with reading skills after adjusting for total left CSA. Sensitivity analyses showed that the association was not driven by general cognitive processes, as it was robust after adjustment for other cognitive variables (i.e. fluid intelligence, matrix reasoning and picture vocabulary), which in is line with previous findings of distinct regionalization patterns of CSA across cognitive tasks in the ABCD [43]. The STG is a known hub within the speech and reading networks [44,33]: Functional MRI studies have shown that the posterior part of the STG is a key multimodal area for audiovisual integration of speech and print, involved in the grapheme-to-phoneme correspondence mapping within the dorsal reading pathway [45,4,46], while the anterior STG seems to be more related to the ventral pathway of speech processing [33,3]. The present study included the CSA of the lateral STG as a single unit, and we cannot, therefore, relate the observed effects to the different specific

processes within it. Nevertheless, it is noteworthy that this specific region, the key hub for the ventral and dorsal reading networks, was the area most strongly associated with reading performance. The lateral temporal cortex is one of the areas of greatest local cortical expansion in humans compared to macaques [47]. Recent studies looking at grey matter measures and reading have also highlighted the relevance of the STG. Plonski et al. [48] performed multivariate classification of children with dyslexia based on morphometric features and identified the mean curvature of the lateral STG as a feature discriminating dyslexics from controls, and Perdue et al. [49] found that the CT of the left STG was positively correlated with word and nonword reading performance. Moreover, reduced grey matter volume in the bilateral superior temporal cortex was found in dyslexic readers [50], and an altered pattern of asymmetry of the planum temporale, within the superior surface of the posterior part of the superior temporal gyrus, has also been reported in dyslexic young males [51]. Nevertheless, the literature on neuroanatomical differences in DD is inconsistent, possibly due to small sample sizes and methodological heterogeneity [15].

Six regional CT measures were associated with reading performance, although mean CT was not. The postcentral gyrus and three occipital CT measures (inferior occipital gyrus and sulcus, lingual gyrus, fusiform gyrus) were positively associated with reading performance after adjusting for other cognitive covariates. These occipital regions are part of the visual system and are involved in the left ventral occipito-temporal network of reading that performs visual-orthographic processing [52], with the left fusiform gyrus including the so-called visual word form area [53]. Bilateral fusiform gyri have been reported to have reduced surface area in pre-readers who develop DD, although no differences in CT were seen [54]. The three other regional measures associated with reading performance were the CT of the superior and middle frontal gyri and the CSA of the SPG. The left parietal cortex has been implicated in letter position coding [55], while activation in the middle frontal gyrus is activated by reading in Chinese and French children [56]. The association of these measures with reading performance shifted in the direction of effect (for the SPG) and the strength of the association became bigger (for frontal CT regions) after controlling for the global measures, implying that the relative regional change of these measures is in the opposite direction to the global effect.

In line with previous findings, the effects of CSA and CT measures on reading were independent of each other. These two measures reflect different features of cerebral cortical structure that are relatively independent both phenotypically and genetically [57,58], although relative greater area and thinner CT have also been linked to cortical stretching that takes place over development [23,59] and may share genetic influences without a clear pattern of sign concordance [60]. Previous studies have shown that CSA measures are overall more heritable than CT measures [61,23,25] and total CSA has been linked to several cognitive traits [23,19].

Despite the apparent specificity of the association of the identified cortical measures to reading, this specificity was not tested beyond the cognitive variables described above, so it is possible that these effects are shared with other cognitive and psychological traits. For instance, children with attention deficit/hyperactivity disorder (ADHD) have reduced total CSA [62] and some of the right-hemisphere homologue regions that were reading-associated in the present study are also among the several sMRI measures that have been found to predict ADHD symptomatology (in the ABCD dataset): namely, lower CSA of the right lateral temporal cortex, lower CT of the right postcentral cortex and greater CT of the right lateral occipital cortex [63].

In sum, we established a set of structural correlates of reading performance, which included key regions such as the CSA of the STG, which is involved in the dorsal and ventral reading networks, and the CT of occipital regions implicated in the ventral reading network. These associations were subtle but consistent across sensitivity analyses, and were not modulated by age or sex. No association was found between structural measures of other key regions of the reading network, such as the inferior frontal gyrus or inferior parietal regions, which may indicate the multifaceted nature of these regions and that their implication in reading is not reflected in morphological changes, at least at the developmental stage (late childhood, 9-10yo) at which we tested this association.

Reading was heritable in the ABCD dataset, although with a lower estimate relative to previously reported results for reading in a large GWAS meta-analysis [19,18]. In this dataset, morphometric brain measures associated with reading performance showed low to modest heritabilities, which were significantly different from zero in most cases, as has been previously reported for these measures [23]. Total left CSA was the measure with the highest estimates across methods, while regional measures were more variable. Relative regional measures (adjusted for global measures) had lower heritability estimates in the ABCD dataset.

On the other hand, estimates based on summary statistics from the UK Biobank dataset for the same imaging measures ranged between 0.12-0.23, as previously reported [25]. Note that these brain GWAS summary statistics from the UK Biobank had been adjusted for multiple imaging confounds, including global measures (i.e. head size [25,64]). The ABCD dataset consists of children (9-10 year-olds), whereas the UK Biobank contains older adults (mean age at recruitment: 55.2 years). As heritability is a relative measure that depends on the amount of total variance, it is possible that the differences in heritability of brain structures reflect either distinct genetic influences and/or differential environmental influences on the developing and the ageing brain. Twin-h^2 of brain imaging measures slightly increases from childhood to adulthood and then decreases in older ages [65,66]. A longitudinal twin-study identified distinct genetic factors that influence cortical thickness of regions across space and time throughout adolescence [67]. Nevertheless, there are multiple other aspects that differ between the two datasets that we analysed (i.e. sample size, the use of individual-level data or summary statistics data, specific global brain covariates used: adjusting for totalCSA or meanCT in the ABCD vs adjusting for head size in the UK Biobank). We cannot therefore establish whether age or some of these other factors are driving the difference we observe.

Overall, we confirmed that reading and reading-related cognitive and brain traits were moderately heritable across methods. Next, three complementary analyses were performed to further elucidate possible shared genetic effects for reading and reading-associated brain measures. First, we performed genetic correlation analyses, we then examined the most promising signals through polygenic scores, and finally assessed these relations through mediation analyses.

There was a consistent trend of a small positive genetic correlation between the total CSA and the CSA of the lateral STG with reading-related cognitive measures (educational attainment and cognitive performance). Genetic correlation between total CSA and educational attainment has previously been reported in the ENIGMA dataset, albeit with a considerably stronger genetic correlation of 0.22 (p=1.9e-13) [23], using the same GWAS summary statistics for educational attainment [38] and the ENIGMA GWAS meta-analysis data for total CSA. The difference between the previous and our genetic correlation estimates for CSA and educational attainment may be partially explained by the fact that the UK Biobank GWAS summary statistics we used here was run adjusting for head size [25], and it was therefore a more relative measure of cortical expansion, while the ENIGMA meta-analysis did not correct for this [23]. Lateral STG also showed genetic correlation with word reading and a negative genetic correlation with dyslexia, although these associations did not survive multiple comparisons correction for the 36 tests. The recent GWAS meta-analysis of reading- and language-related traits [19] found a significant genetic correlation between reading and the CSA of the banks of the superior temporal sulcus, which partially overlaps the lateral STG we report here. They used the same UK Biobank GWAS resource used in the present study [25] but a different brain parcellation to compute the genetic correlations between the reading and language GWAS meta-analysis and 58 structural neuroimaging traits with known links to reading and language. The word reading GWAS we included in this analysis was from Eising et al. [19], so the concordance of these results is reassuring but also expected. Overall, our results are in line with the literature, supporting a possible genetic overlap between reading and related cognitive traits with total CSA and the CSA around the superior temporal regions.

We did not find evidence for a genetic overlap between reading and any of the other reading associated CT and CSA measures. The power to detect genetic correlations depends on the strength of the phenotypic association, the strength of the genetic effects on each trait (i.e. heritability) and the

actual genetic overlap that exists [68]. In our results, the heritability estimates were of a similar order of magnitude for all cortical measures (Supplementary Fig.9), but the strength of the association with reading was lower for these other measures, which may have resulted in diminished power to detect low to moderate genetic correlations. Note that a lack of a genetic correlation between two traits may also occur despite there being a genetic overlap, if there are genetic variants that contribute to both traits with mixed directions of effect [69].

In an attempt to tease apart direct and indirect genetic effects, we conducted a comparison of genetic correlations between sibling-based GWAS and population-based GWASes for years of education and cognitive function [39] and reading-related brain phenotypes (UK Biobank GWAS from population-based estimates). These analyses revealed few nominally significant genetic correlations with the sibling-based, but in most cases not the population-based cognitive GWASes, none of which survived multiple comparison corrections. Although this could potentially indicate that the genetic correlation is robust to detect at least some indirect genetic effects, it should be noted that the GWAS for the brain measures was population-based, and hence these r_g estimates could still be inflated due to demographic and indirect effects on the brain measures. As new and larger family-based GWASes become available, it will be important to further study the direct and indirect genetic influences on the relationship between cognitive traits and reading-related brain measures.

To further examine the shared genetic correlates to reading and CSA measures we used polygenic scoring. A polygenic score of cognitive performance (PGS_{CP}) predicted 3.6% of the variance in reading performance and was also a significant predictor of total CSA and the CSA of the STG region. These CSA measures partially mediate the effect of PGS_{CP} on reading, explaining up to 4.8% of the effect of total CSA, and 4.2% of the effect of the CSA of the STG. The effect seems to be shared between the two CSA measures, as the adjustment of the CSA measures with each other diminished the mediation or made it disappear. Of note, these effects are small: less than 0.2% of the total variation in reading performance is explained through this route (i.e. the 4.8% of the 3.6% that the PGS_{CP} explains in the variation of reading performance). This relationship is similar to the one reported for intelligence measures in young-adult samples: CSA measures mediated up to 3.4% of the effect of the PGS of educational attainment on intelligence in two twin-based datasets [70] and vertex-wise surface area measures were shown to partly mediate the effect of PGS of educational attainment on the ‘g’ general intelligence factor [71], which is also in line with previously reported positive genetic correlations between CSA and educational attainment in genomic [23] and twin-studies [72]. Based on these results, we suggest that a small part of the genetic effects captured by the PGS act through the CSA measures to affect reading performance. Hence, the observed effects highlight that the biology underlying reading ability is partly heritable and polygenic, and that it relies on a distributed network throughout the brain.

Although the current study showed that it is possible to triangulate reading with the brain and its genetic effects, it also has some limitations. First, all the effect sizes we report here are small. The sum of demographic variables in the baseline model explains up to 18.3% of the variability in reading, while the maximum brain-reading association (i.e. CSA total) only explains an additional 1%. This is expected for univariate brain wide association studies that attempt to establish brain-behaviour correlates based on inter-individual variability [73]. Hence, the current results do not support the existence of morphometric brain features with large effect sizes in reading performance. Second, we focused on the relationship between reading performance and morphological cortical measures (CSA and CT). Although cortical morphometry can be seen as an indirect measure of function [73], other imaging modalities, such as functional task-related measures, may be more closely associated with reading performance. Large datasets such as the ABCD and the UK Biobank do not contain relevant reading-related task measures [74,75], while smaller datasets do not provide enough power to perform the type of analyses carried out in the current study. However, functional connectivity from resting-state data, available in the ABCD and UK Biobank, could be used as markers of interest in follow-up studies. Furthermore, the present study only considered the neocortex, and did not include subcortical

structures, despite regional differences in volume of the thalamus or cerebellum having been related to DD or reading performance in typically developing children [76].

Moreover, we used a variety of analyses and datasets with different demographic characteristics. The goal was to find converging evidence across them, but this approach also adds heterogeneity that may hinder the interpretation of our results. We provided plausible explanations of why some results may not be congruent in terms of this heterogeneity, and, instead, focused on the strongest signals that replicated throughout the analyses to overcome this caveat. Lastly, some crucial aspects to understand how reading affects the brain and its genetic underpinnings, such as development, were overlooked to some extent in the present study. Our modelling does not capture the full complexity of brain-genetic-behaviour relations. The effects could be inflated by gene-environment interactions. For instance, recreational reading has been shown to mediate the relationship between polygenic scores of intelligence and crystallised measures of intelligence (including reading) in the ABCD dataset [74], while sociodemographic factors such as caregiver education and income (modelled in the present study as covariates) can also influence brain structure [75]. As the ABCD is a multimodal and phenotypically-dense longitudinal study that has been following children since 2018 and will continue to do so for the next 7 years, follow-up studies will have the opportunity to assess the stability of the effects reported here.

The present study illustrates the type of analytical approach that could be used to understand the biological bases of reading performance, utilizing openly available datasets and tools. It should be noted that the goal of the current study was to dissect the complex relationship between traits. Thus, this work does not have any direct implication for efforts to support reading skills, as the effect sizes we have uncovered are very small and do not have any predictive power at the individual-subject level. Nevertheless, this approach allows us to better understand the possible key brain features and regions (e.g. total CSA, STG) implicated in a complex behaviour such as reading. In the future, this work could be extended to other relevant phenotypes, such as functional or structural connectivity measures. In addition, possible factors that are likely to shape the effects we describe (such as age or sex) should also be taken into consideration to provide a more comprehensive account.

In summary, we identified the cortical correlates of reading performance, including total CSA and key reading-network measures such as the CSA of the STG and the CT of a cluster of occipital regions that are involved in the dorsal and ventral reading pathways. The effects reported in these analyses were consistent and predominantly left-hemisphere specific. Nevertheless, these were also modest effects, which argue against large effects of cortical features in reading performance, and highlight the need for large datasets to be able to address these types of questions in an unbiased manner. Further, while there was no indication of a genetic overlap with any of the CT measures, we found evidence that suggests that genetic effects contribute to the association with CSA. These findings revealed insights into the structural brain correlates of reading, and indicate that the total CSA and the superior temporal gyrus CSA partially mediate the association between genetic factors and reading performance.

Methods

ABCD study data

The Adolescent Brain Cognitive Development (ABCD) study is a longitudinal study across 21 data acquisition sites following ~11,878 children starting at 9 and 10 years old [76]. Parents/caregivers provided written informed consent, and children verbal assent, to a research protocol approved by a central Institutional Review Board (cIRB) at the University of California, San Diego (UCSD) for most sites or by a local IRB for a few sites (https://abcdstudy.org/study-sites/) [77]. The study used specific recruitment strategies to create a population-based, demographically diverse sample [78]. However, it is not necessarily representative of the U.S. national population [79]. The current study analysed the full baseline sample (N~11,878) from the ABCD data release 3.0 RDS (DOI: 10.15154/1520591) and the Genotyping Data from the ABCD Curated Annual Release 3.0 (NDA Study 901; DOI: 10.15154/1519007). All variables included in the current study are listed and described in Supplementary Table 1, with Supplementary Table 2 providing their descriptive statistics.

Behavioural data

The dependent variable for our primary analysis was the Toolbox Oral Reading Recognition Task from the NIH Toolbox® Cognition Battery, which is a reading test that asks individuals to pronounce single words [80]. Three additional cognitive variables were included for the sensitivity analyses: Toolbox Picture Vocabulary Task (NIH Toolbox® Cognition Battery), which measures language skills and verbal intellect; the "fluid composite" cognitive score, which is a composite score derived as the average normalized scores from several fluid ability measures from the NIH Toolbox® Cognition Battery, namely: Flanker, Dimensional Change Card Sort, Picture Sequence Memory, List Sorting and Pattern Comparison; and the Matrix Reasoning task from the Wechsler Intelligence Scale for Children-V (WISC-V), which measures fluid intelligence and visuospatial reasoning.

Structural magnetic resonance imaging data

The ABCD MRI data were collected using harmonized scanning protocols across 21 research sites, using Siemens Prisma, GE 750 and Philips 3T scanners. Full details of the imaging acquisition protocols and preprocessing can be found in Casey et al. [81] and Hagler et al. [82]. Cortical surfaces were constructed from T1 weighted MRI volumes and segmented to calculate measures of apparent CT and CSA using Freesurfer v5.3.0. We used ABCD tabulated imaging data that provided CT and CSA measures per region. We restricted the analysis to subjects who passed quality control for the cortical surface reconstruction based on manual inspection of the data (i.e. had “1” on the fsqc_qc variable; see Hagler et al., 2019).

Freesurfer atlas parcellation measures (Destrieux et al., [34]) for left-hemisphere CT and CSA were included in the analysis. Seventy-four regional and one global measure (total CSA or mean CT) were included for CT and CSA. Thus, 150 measures were analyzed in total. The main analyses were focused on the left hemisphere, although some of the homotopic right-hemisphere measures were also included for completeness and sensitivity.

Additional data

The study comprises different data acquisition sites and associated MRI scanners, as well as related individuals (twins and siblings). Therefore, the scanner ID was included as a random factor to account for the dependence of samples, and family relatedness (family ID) was included as a random factor nested within scanner in all analyses that included related individuals. The ABCD is a diverse dataset with a wide range of individuals across different ethnicities and socioeconomic backgrounds (Garavan

et al., 2018). We included the following demographic variables in all analyses, unless otherwise specified: sex, age, genetic principal components (PC1-PC10; see "Genetic analyses" section below), household income and caregiver education. Note that the genetic PCs were computed for the full sample (including diverse ancestries) and for the European-PCA cluster for some genetic analyses, which aims to minimize variation in non-genetic and genetic factors (see below). The PCs corresponding to the specific dataset used in each analysis were included as covariates. Furthermore, as some genetic analyses require the use of unrelated individuals, we defined a subset of unrelated individuals for those analyses (see "Genetic quality control").

Brain-behaviour association analyses

Quality control

Individuals that passed quality control for both neuroimaging (N=11,265) and genetic data (N=11,092) were selected. Next, we filtered out individuals that had missing data for covariates, or variables of interest (1,868 individuals excluded), or with extreme outlier values (i.e. more than 7 standard deviations away from the mean for that variable: 47 individuals). Finally, we also trimmed extreme values of the dependent variable (i.e. reading; see above) by excluding the 0.01 quantiles at each end of the distribution (N=164 individuals excluded) to avoid having long tails that would affect the regression models (kurtosis was reduced from 4.56 to 3.43; see Supplementary Fig.1). This final dataset of 9,013 individuals was used for the main brain-behaviour association analyses. Descriptive statistics per covariate before and after the final trimming are presented in Supplementary Table 2.

Regression analyses

All numeric variables were z-transformed (centered and scaled to have a mean of 0 and variance of 1). Linear mixed-effect regressions were run using R (v 4.0.3) package *lme4* v 1.1-25; [83] to identify structural ROIs associated with reading.

$$reading \sim (1|scanner) + (1|scanner{:}family) + sex + age + high.educ + household.income + PC1{:}PC10 + \varepsilon$$

(Eq 1)

$$reading \sim (1|scanner) + (1|scanner{:}family) + sex + age + high.educ + household.income + PC1{:}PC10 + imaging\ phenotype + \varepsilon$$

(Eq 2)

$$reading \sim (1|scanner) + (1|scanner{:}family) + sex + age + high.educ + household.income + PC1{:}PC10 + global\ imaging\ phenotype + imaging\ phenotype + \varepsilon$$

(Eq 3)

First, we included all random variables and covariates of interest in a baseline model (Equation 1; see Supplementary Table 1 for the definition of all covariates and random factors). Each brain measure was then added as an independent variable in separate regressions to assess their effect on reading (Equation 2; Model 1). Next, we checked that the association of regional measures was robust after controlling for the corresponding global measures (Equation 3; Model 2), i.e. the left-hemisphere total CSA for CSA measures or the left-hemisphere mean CT for CT measures. A likelihood ratio test (LRT) between each pair of nested models (Baseline *vs* Model 1; Model 1 *vs* Model 2) was used to assess the significance of the term of interest. We defined brain measures consistently associated

with reading as those that were significant after multiple testing corrections in both models 1 and 2 (False Discovery Rate q-value<0.05).

We next assessed whether any of the identified brain measures had a significant interaction with demographic variables such as age and sex, by separately adding these interaction terms to Models 1 and 2 for each brain measure, and assessing the significance of each interaction through a LRT between tested models.

To assess whether the observed associations were specific to reading, we also explored whether the reading-associated measures were robust to adjustment for other cognitive covariates related to reading: fluid intelligence score, matrix reasoning, and picture vocabulary (see Supplementary Table 1 for the definition of variables included in this sensitivity analysis).

Although the main analysis focused on left-hemisphere measures, the reading-associated measures were followed up by performing identical analyses with their homotopic right-hemisphere measures. We also computed the correlations between these homotopic regions and across all reading-associated brain measures, using the residuals after adjusting for the covariates in Models 1 and 2 (R package *psych* v2.0.12 [84]).

We re-ran these analyses again restricting them to subsets of the data (see section below for exact subset definitions) to confirm that these results were not driven by the family component or were ancestry-specific: unrelated individuals only (after QC: N=7,502), and unrelated individuals of European ancestry (after QC: N=4,080).

In order to clarify the inter-dependence of the CSA and CT effects, we computed Pearson's correlation of the beta estimates that we obtained for the CSA and CT measures for reading performance in the ABCD dataset. We performed this comparison for the whole dataset, as well as for the two other smaller subsets from this dataset.

All effects were visualized in brainplots using the R packages *ggseg* v 1.6.4 [85] and *ggsegDesterieux* v 1.0.1.002 [86].

Genetic analyses

Genetic quality control

Plink [87] (v1.90b6.15) was used to perform SNP and sample quality control (QC) following Coleman et al. [88]. 11,099 individuals had available genotype data for 516,598 SNPs. SNPs were excluded if they had genotyping rate <0.99, Hardy Weinberg Equilibrium (HWE) p-value <1e-6, or minor allele frequency (MAF) < 0.01. After this QC, 340,003 SNPs were kept. Samples were excluded if they had a missing genotyping rate greater than 95% (N=7).

A set of 291,223 common (MAF>0.05), autosomal, independent SNPs were selected by pruning LD using a window of 1500 variants and a shift of 150 variants between windows, with an r^2 cut-off of 0.2, and excluding high-LD and non-autosomal regions [88]. These SNPs were used to identify sex mismatches, assess relatedness and flag outliers for IBD or heterozygosity, and to perform ancestry analyses. To identify a set of unrelated individuals, we randomly excluded one subject from each pair with a pi-hat > 0.1875. This resulted in 11,047 subjects passing genetic QC, of which 9,061 were unrelated (i.e. pi-hat>0.1875).

Genetic principal components (PCs) were calculated using SmartPCA (EIGENSOFT v 6.1.4 [89]) for the total sample. These PCs were then used as covariates in the brain-behaviour regression analyses (see above). PCs computed with reference populations of the 1000 genomes reference dataset (v37; [90]) were plotted to visualize the genetic ancestry of the total sample.

In order to define a subset of homogeneous ancestry, we then selected individuals with a > 90% of European ancestry (as defined by the variable "genetic_ancestry_factor_european" from the ABCD data release 3.0; N=6,103) and removed outlier individuals exceeding 6 standard deviations along

one of the top 10 PCs using SmartPCA [89]. The genetic PCs for this subset were used in downstream genetic analyses that are sensitive to population stratification effects. This subset consisted of 5,740 individuals of European ancestry, of which 4,716 were unrelated.

Heritability and genetic correlation analyses

A genetic relationship matrix (GRM) of 4,716 unrelated individuals of European ancestry from the ABCD dataset was built in GCTA (v1.93.2beta [36]) using 331,460 autosomal, directly genotyped SNPs. We further excluded one random participant from each pair having a kinship coefficient higher than 0.05 based on the calculated GRM (as this analysis is especially sensitive to higher levels of relatedness), resulting in 4,633 participants. Genome-based restricted maximum likelihood (GREML; [36]) analyses were performed to estimate the SNP-based heritabilities (h^2-SNP), using residuals after controlling for the sex, age and the first 10 genetic PCs for this subset as covariates and scanner (for brain measures) or site (for non-brain measures) as a random factor. As power was limited for this analysis, we maximized the sample size by not excluding individuals who had missing variables that were not used in each specific analysis (as for the brain-behaviour analysis), and therefore the sample sizes were slightly larger, ranging from 4,285 to 4,363 depending on the specific measure. The significance of the heritability estimates was Bonferroni corrected for multiple comparisons for the number of tested measures, i.e. 21 measures (9 brain measures, 8 brain measures adjusted for global measures and 4 cognitive measures). Bivariate GREML analyses were run to compute the genetic correlation (r_g) between regional and global brain measures, and we then assessed whether the r_g was significantly different to 1 through a one-tailed z-test (where $z = (1-r_g)/se$). In order to determine whether genetic correlations between reading performance and brain imaging phenotypes could be detected, given the sample size and heritability estimates for the different traits in the ABCD dataset, we ran power analyses using the GCTA-GREML power calculator [68]. This analysis showed that the current dataset is not well powered to detect these genetic correlations (power < 80% for all measures, see Supplementary Fig.10).

Heritability and genetic correlations based on summary statistics

The linkage disequilibrium score regression (LDSC) allows the computation of genetic correlation from GWAS summary statistics, without relying on individual level data [37]. LDSC (v1.0.1) was used to calculate the heritability and genetic correlations of the identified nine brain measures and four reading-related cognitive measures (as defined in Supplementary Table 7). These included GWAS summary statistics from much larger datasets than the ABCD dataset used for the genetic analysis, and therefore we expected that they would be better powered to detect potentially subtle genetic correlations. The following traits were included: word reading (reaction time) [19], developmental dyslexia (defined based on word reading and spelling tests) [22], educational attainment and cognitive performance (for fluid intelligence) [38], and CT and CSA brain measures [25]. Multiple comparisons corrections were applied for the 13 heritability estimates (9 brain imaging and 4 cognitive; α=0.05/9=0.0056) and 36 genetic correlations (9 imaging x 4 cognitive; α=0.05/36=0.0014) tested.

In order to disentangle whether potential genetic correlations were driven by direct or indirect genetic effects, we performed additional genetic correlation analyses using LDSC between the same brain measures (from the UK Biobank) and GWAS summary statistics based on the population and sibling estimates for “years of schooling” and “cognitive function” from Howe et al. [39]. Multiple comparisons corrections were applied for 36 tests (9 imaging measures x 4 cognitive measures: 2 sib- + 2-population based; α=0.0014).

Additional genetic correlations were computed to assess differences of CSA measures across studies (with a partially overlapping sample): mean total CSA (ENIGMA, including 10,083 individuals from the UK Biobank [23]) and total left CSA in the UK Biobank [25].

Polygenic Scoring

PRSice (v2.2.12) [91] was used to compute polygenic scores (PGS) for each of the traits of interest (i.e. using the published cognitive GWAS summary statistics listed in Supplementary Table 7 as base datasets).

The target dataset was the ABCD unrelated European ancestry subset (N=4,080). The ABCD imputed genotype data was used, which had been imputed using the TOPMed imputation server with Eagle v2.4 phasing and TOPMed mixed ancestry reference (http://dx.doi.org/10.15154/1519007). The imputed data was filtered using plink to keep only HRC biallelic genotype calls with a minimum genotype probability of 0.9, and SNPs with imputation quality scores above 0.7, and lifted to the hg19 reference panel (27,817,000 SNPs). Next, SNPs were filtered on HWE (p-value >1e-6), MAF>0.05 and missing genotype rate < 0.1 (4,034,417 SNPs kept).

The PGSs were then computed for each trait through clumping (clump-kb 250kb, clump-p: 1, clump-r2=0.1) and thresholding (8 p-value thresholds: 5e-08, 1e-5, 0.0001, 0.001, 0.05, 0.1, 0.5, 1). Next, linear mixed-effects models were run in R (v 4.0.3; package *lme4* v 1.1-25). A baseline model included the dependent variable and all covariates specified above (see section "Regression analyses") and separate PGS models were run including each PGS (for Educational Attainment, Cognitive Performance or Reading) as an independent variable. The reading GWASMA used as the base dataset was from Eising et al. [19] excluding the ABCD dataset. The significance of each PGS was assessed by a likelihood ratio test between the baseline and PGS model. The proportion of variance explained by the PGS (ΔR^2) was computed as the difference of the R^2 between the models divided by one minus R^2 of the baseline model ($\Delta R^2 = \frac{R^2 PGS \ - R^2 baseline}{(1 - R^2 baseline)}$).

The most predictive PGS for reading was first identified (largest ΔR^2) to define the most predictive base dataset and p-value threshold for reading in the ABCD dataset (see Supplementary Figs. 9 and 10). Next, the selected base dataset and p-value threshold were used to perform the cross-trait regression (i.e. with the two CSA measures as dependent variables) in order to maximize power and limit the number of performed comparisons. An additional regression was run with STG as dependent variable and total left CSA included as a covariate. The significance of the PGS was Bonferroni adjusted for the three tested brain measures (α=0.05/3=0.0167).

Mediation analysis

For each model of interest (see Table 1), two regression models were computed using the R package *lme4* [86]: an initial model that computed the effect of the independent variable PGS_{CP} (IV) on the CSA measure (mediator), and another that computed the indirect effect of the IV on reading (dependent variable), after accounting for the mediator. All models included the covariates and random structure indicated in Equation 1, and additional cognitive covariates were also included for sensitivity analyses (as specified in Table 1). The *mediation* (v 4.5.0) package [92] was then used to estimate causal mediation effects [93]. The significance of the direct (ADE) and indirect (ACME) paths was assessed by a quasi-Bayesian approximation of the confidence intervals (10,000 simulations). Sensitivity analysis included additional covariates to assess if the mediation effects were robust to those adjustments (see Table 1). The significance of the mediation indirect path was Bonferroni corrected for the 8 mediation models (α=0.05/8=0.0063).

Data availability

ABCD data are publicly available through the National Institute of Mental Health (NIHM) Data Archive (https://data-archive.nimh.nih.gov/abcd). The current study analysed the full baseline sample (N~11,878) from the ABCD data release 3.0 RDS (DOI: 10.15154/1520591) and the Genotyping Data from the ABCD Curated Annual Release 3.0 (NDA Study 901; DOI: 10.15154/1519007). All variables included in the current study are listed and described in Supplementary Table 1. GWAS summary statistics used in this study are available from the NHGRI-EBI GWAS Catalog

https://www.ebi.ac.uk/gwas/downloads/summary-statistics and the Oxford Brain Imaging Genetics Server – BIG40 (https://open.win.ox.ac.uk/ukbiobank/big40/). Supplementary Table 7 contains the references and field ID's for all analysed traits.

Code availability

The custom code associated with this study is publicly available at https://github.com/amaiacc/MS-brain-reading-genetics/.

Acknowledgements

We thank James Barry and Caroline Handley for helping with manuscript proofreading. This research is supported by the Basque Government through the BERC 2022-2025 program and by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010-S. A. C-C. received funding from the Spanish Ministry of Science and Innovation and the Agencia Estatal de Investigación through Ayudas Juan de la Cierva-Incorporación (ref. IJC2018-036023-I), the Programa Fellows Gipuzkoa de atracción y retención de talento from the Diputación Foral de Gipuzkoa. P.M.P-A. is supported by grants from the Spanish Ministry of Science and Innovation (PID2021-123574NB-I00), from the Basque Government (PIBA-2021-1-0003), and from the Red guipuzcoana de Ciencia, Tecnología e Innovación of the Diputación Foral de Gipuzkoa (FA/OF 422/2022). M. C. is supported by the Spanish Ministry of Science and Innovation grant (PID2021-122918OB-I00) and "la Caixa" Foundation (ID 100010434), under the agreement HR18-00178-DYSTHAL. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Data used in the preparation of this article were obtained from the ABCD Study (https://abcdstudy.org/) and are held in the NIMH Data Archive. This is a multisite, longitudinal study designed to recruit more than 10,000 children aged 9–10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National Institutes of Health (NIH) and additional federal partners under award numbers U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, and U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners/. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/principal-investigators/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from https://doi.org/10.15154/1520591 and https://doi.org/10.15154/1519007 (genotyping data).

Author contributions

All authors conceived and designed the study. A.C-C. analysed the data, made figures and wrote the first draft of the paper. All authors discussed the results and contributed towards the writing of the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Tables

Table 1: Mediation model testing the significance of CSA measures as mediators of the PGS_{CP} (independent variable) and reading (dependent variable) association (N=4,080). All models included MRI scanner as a random factor, and the following covariates: sex, age, high.educ, household.income, PC1:PC10 (genetic principal components). IV: independent variable. Adjustment: additional covariates included in mediation models. Estimate: standardized beta for the parameter. 95% CI: confidence intervals . A: Effect of the independent variable on the mediator. B: Effect of the mediator on the dependent variable. ACME= Average Causal Mediation Effect; ADE= Average Direct Effect. The significance of the direct (ADE) and indirect (ACME) paths was assessed by quasi-Bayesian Monte Carlo simulation of the confidence intervals (two-sided, 10,000 simulations). P (unadj)= unadjusted p-value. Note that the *mediation* R package [92] used for this analysis does not provide exact p-values for values < 2.2e-16. P (adj) = the significance of the mediation indirect path was Bonferroni corrected for the 8 mediation models (α=0.05/8=0.0063). PGS_{CP}= polygenic score for cognitive performance at the p-value threshold of 0.05; CSA= Cortical surface area; STG lat= lateral part of the superior temporal gyrus. Lh= left-hemisphere. WISCV= WISC-V Matrix Reasoning Total Scaled Score.

Mediator	Adjustment	Parameter	Estimate	95% CI	P (unadj)	P (adj)	% Mediation
Total CSA (lh)		A	0.089	0.06-0.11			
		B	0.106	0.07-0.14			
		ACME	0.009	0.01-0.01	<2e16	<0.0063	4.9
		ADE	0.182	0.15-0.21	<2e16		
		Total	0.191	0.16-0.22	<2e16		
Total CSA (lh)	G-TEMP-SUP-LATERAL (CSA lh)	A	0.045	0.02-0.06			
		B	0.054	0.01-0.1			
		ACME	0.002	0-0	0.014	0.208	1.3
		ADE	0.181	0.15-0.21	<2e16		
		Total	0.183	0.16-0.21	<2e16		
Total CSA (lh)	Fluid Component, WISCV	A	0.081	0.05-0.11			
		B	0.083	0.05-0.12			
		ACME	0.007	0-0.01	<2e16	<0.0063	4.4
		ADE	0.145	0.12-0.17	<2e16		
		Total	0.151	0.12-0.18	<2e16		
Total CSA (lh)	Fluid Component, WISCV, Vocabulary	A	0.072	0.05-0.1			
		B	0.052	0.02-0.08			
		ACME	0.004	0-0.01	0.0004	0.0032	3.3
		ADE	0.105	0.08-0.13	<2e16		
		Total	0.109	0.08-0.13	<2e16		
G-TEMP-SUP-LATERAL (CSA lh)		A	0.076	0.05-0.1			
		B	0.108	0.08-0.14			
		ACME	0.008	0-0.01	<2e16	<0.0063	4.2
		ADE	0.183	0.15-0.21	<2e16		
		Total	0.191	0.16-0.22	<2e16		
G-TEMP-SUP-LATERAL (CSA lh)	Total CSA (lh)	A	0.015	-0.01-0.04			
		B	0.076	0.04-0.12			
		ACME	0.001	0-0	0.176	1	0.6
		ADE	0.181	0.15-0.21	<2e16		
		Total	0.182	0.15-0.21	<2e16		
G-TEMP-SUP-LATERAL (CSA lh)	FluidComponent, WISCV	A	0.069	0.04-0.1			
		B	0.091	0.06-0.12			
		ACME	0.006	0-0.01	<2e16	<0.0063	4
		ADE	0.145	0.12-0.17	<2e16		
		Total	0.152	0.12-0.18	<2e16		
G-TEMP-SUP-LATERAL (CSA lh)	FluidComponent, WISCV, Vocabulary	A	0.059	0.03-0.09			
		B	0.061	0.03-0.09			
		ACME	0.004	0-0.01	<2e16	<0.0063	3.3
		ADE	0.105	0.08-0.13	<2e16		
		Total	0.109	0.08-0.14	<2e16		

Figure legends

Figure 1. Overview of study goals and analytical approaches. A and B state the main goals of the study, and the analytical steps taken to address them are depicted in the central panels (1, 2, 3). The measures used in each analysis are specified in these panels. The data used for each analysis type are indicated by the colour: green (individual level data from the ABCD study) and blue (GWAS summary statistics). sMRI= structural MRI measures of cortical thickness (CT) and cortical surface area (CSA); LMM= linear mixed effect model; h^2_{SNP}(GREML)= SNP-heritability estimated from unrelated samples using the genome-based restricted maximum likelihood (GREML) estimation approach; h^2_{SNP} (LDSC)= SNP-heritability estimated from GWAS summary statistics using linkage disequilibrium score regression (LDSC); PGS= polygenic scoring analysis. Cognitive: cognitive measures (i.e. Reading performance, fluid intelligence component, NIHTBX picture vocabulary, WISCV: WISC-V Matrix Reasoning Total Scaled Score).

Figure 2. Effect of left-hemisphere cortical measures on reading performance (N=9,013). Beta estimates for **A** cortical surface area measures and **B** cortical thickness measures. Points indicate beta estimates for the independent brain variables and error bars represent the 95% confidence intervals. Covariates: $sex + age + high.educ + household.income + PC1{:}PC10$; Model 1: test model assessing the effect of each brain measure by comparing it to a nested model including only the covariates, Model 2: test model after adjusting for global measure (total CSA or mean CT). The regions that survived multiple comparisons correction (FDR) are shaded in grey. **C**, **D**: brainplots of T-values associated for each brain region in Model 2. The upper panels show T-values for all regions, and the lower panels show only T-values for regions that survive multiple comparisons correction (FDR). CSA= cortical surface area; CT= cortical thickness; LH= left-hemisphere; G = gyrus; S = sulcus.

Figure 3. Decile plots for PGS of cognitive performance on left-hemisphere CSA measures (N=4,080). The points indicate the mean CSA for each decile; the error bars show the 95% confidence intervals. PGS_{CP}= polygenic score for cognitive performance at the p-value threshold of 0.05; CSA= Cortical surface area; STG: superior temporal gyrus; ΔR^2: variance explained by the PGS, assessed by a likelihood ratio test between the baseline and PGS model. The proportion of variance explained by the PGS (ΔR^2) was computed as the difference of the R^2 between the models divided by one minus R^2 of the baseline model ($\Delta R^2 = \frac{R^2 PGS \ - R^2 baseline}{(1 - R^2 baseline)}$).

Figure 4. Mediation of the CSA on the association between the PGSCP and reading. (A) total CSA as mediator and (B) CSA of the lateral STG as mediator. The standardized estimates of the paths are provided, with the 95% confidence intervals in brackets. ACME= Average Causal Mediation Effect; ADE= Average Direct Effect. A: Effect of the independent variable on the mediator. B: Effect of the mediator on the dependent variable. The significance of the direct (ADE) and indirect (ACME) paths was assessed by quasi-Bayesian Monte Carlo simulation of the confidence intervals (10,000 simulations). *** indicates p-values <0.001. Bonferroni corrected p-value threshold for the 8 mediation models is α=0.05/8=0.0063. PGSCP= polygenic score for cognitive performance; CSA= Cortical surface area; STG lat= lateral part of the superior temporal gyrus.

825 # References

1. Sandak, R., Mencl, W. E., Frost, S. J. & Pugh, K. R., The Neurobiological Basis of Skilled and Impaired Reading: Recent Findings and New Directions. *Scientific Studies of Reading* **8**, 273-292 (2004).

2. Dehaene, S. & Cohen, L., Cultural Recycling of Cortical Maps. *Neuron* **56**, 384-398 (2007).

3. Friederici, A. D., The cortical language circuit: from auditory perception to sentence comprehension. *Trends in Cognitive Sciences* **16**, 262-268 (2012).

4. Lau, E. F., Phillips, C. & Poeppel, D., A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience* **9**, 920-933 (2008).

5. Martin, A., Schurz, M., Kronbichler, M. & Richlan, F., Reading in the brain of children and adults: A meta-analysis of 40 functional magnetic resonance imaging studies. *Human Brain Mapping* **36**, 1963-1981 (2015).

6. Pugh, K. R. *et al.*, Neurobiological studies of reading and reading disability. *Journal of Communication Disorders* **34**, 479-492 (2001).

7. Carreiras, M. *et al.*, An anatomical signature for literacy. *Nature* **461**, 983-986 (2009).

8. Dehaene, S., Cohen, L., Morais, J. & Kolinsky, R., Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience* **16**, 234-244 (2015).

9. Jednoróg, K. *et al.*, The Influence of Socioeconomic Status on Children's Brain Structure. *PLoS ONE* **7**, e42486 (2012).

10. Romeo, R. R., Uchida, L. & Christodoulou, J. A., Socioeconomic status and reading outcomes: Neurobiological and behavioral correlates. *New Directions for Child and Adolescent Development* (2022).

11. Association, A. P., *Diagnostic and statistical manual of mental disorders. (5th. Ed.)* (American Psychiatric Association Publishing, Washington D.C., 2013).

12. Organization, W. H., *International statistical classification of diseases and related health problems, 10th revision, Fifth edition* (World Health Organization, 2016).

13. Goswami, U., Phonology, reading development, and dyslexia: A cross-linguistic perspective. *Annals of Dyslexia* **52**, 139-163 (2002).

14. Ramus, F., Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology* **13**, 212-218 (2003).

15. Ramus, F., Altarelli, I., Jednoróg, K., Zhao, J. & Covella, L. S., Neuroanatomy of developmental dyslexia: Pitfalls and promise. *Neuroscience & Biobehavioral Reviews* **84**, 434-452 (2018).

16. Richlan, F., Developmental dyslexia: dysfunction of a left hemisphere reading network. *Frontiers in Human Neuroscience* **6** (2012).

17. Andreola, C. *et al.*, The heritability of reading and reading-related neurocognitive components: A multi-level meta-analysis. *Neuroscience & Biobehavioral Reviews* **121**, 175-200 (2021).

18. Verhoef, E., Shapland, C. Y., Fisher, S. E., Dale, P. S. & Pourcain, B. S., The developmental origins of genetic factors influencing language and literacy: Associations with early-childhood vocabulary. *Journal of Child Psychology and Psychiatry* (2020).

19. Eising, E. *et al.*, Genome-wide association analyses of individual differences in quantitatively assessed reading- and language-related skills in up to 34,000 people. *Proceedings of the National Academy of Science* **119** (35), e2202764119 (2022).

20. Fisher, S. E. & DeFries, J. C., Developmental dyslexia: genetic dissection of a complex cognitive trait. *Nature Reviews Neuroscience* **3**, 767-780 (2002).

21. Doust, C. *et al.*, Discovery of 42 Genome-Wide Significant Loci Associated with Dyslexia. *medRxiv* (2021).

22. Gialluisi, A. *et al.*, Genome-wide association study reveals new insights into the heritability and genetic correlates of developmental dyslexia. *Molecular Psychiatry* (2020).

23. Grasby, K. L. *et al.*, The genetic architecture of the human cerebral cortex. *Science* **367** (2020).

24. Elliott, L. T. *et al.*, Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210-216 (2018).

25. Smith, S. M. *et al.*, An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature Neuroscience* (2021).

26. Landi, N. & Perdue, M. V., Neuroimaging genetics studies of specific reading disability and developmental language disorder: A review. *Language and Linguistics Compass* **13** (2019).

27. Roeske, D. *et al.*, First genome-wide association scan on neurophysiological endophenotypes points to trans-regulation effects on SLC2A3 in dyslexic children. *Molecular Psychiatry* **16**, 97-107 (2009).

28. Scarborough, H. S. & Brady, S. A., Toward a Common Terminology for Talking about Speech and Reading: A Glossary of the Phon Words and Some Related Terms. *Journal of Literacy Research* **34**, 299-336 (2002).

29. Taylor, J. S. H., Rastle, K. & Davis, M. H., Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. *Psychological Bulletin* **139**, 766-791 (2013).

30. Jobard, G., Crivello, F. & Tzourio-Mazoyer, N., Evaluation of the dual route theory of reading: a metanalysis of 35 neuroimaging studies. *NeuroImage* **20**, 693-712 (2003).

31. Dick, A. S. *et al.*, Meaningful associations in the adolescent brain cognitive development study. *NeuroImage* **239**, 118262 (2021).

32. Vigneau, M. *et al.*, Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage* **30**, 1414-1432 (2006).

33. Rueckl, J. G. *et al.*, Universal brain signature of proficient reading: Evidence from four contrasting languages. *Proceedings of the National Academy of Sciences* **112**, 15510-15515 (2015).

34. Destrieux, C., Fischl, B., Dale, A. M. & Halgren, E., A sulcal depth-based anatomical parcellation of the cerebral cortex. *NeuroImage* **47**, S151 (2009).

35. Hyatt, C. S. *et al.*, The quandary of covarying: A brief review and empirical examination of covariate use in structural neuroimaging studies on psychological variables. *NeuroImage* **205**, 116225 (2020).

36. Yang, J. *et al.*, Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565-569 (2010).

37. Bulik-Sullivan, B. *et al.*, An atlas of genetic correlations across human diseases and traits. *Nature Genetics* **47**, 1236-1241 (2015).

38. Lee, J. J. *et al.*, Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics* **50**, 1112-1121 (2018).

39. Howe, L. J. *et al.*, Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nature Genetics* **54**, 581-592 (2022).

40. Maier, R. M., Visscher, P. M., Robinson, M. R. & Wray, N. R., Embracing polygenicity: a review of methods and tools for psychiatric genetics research. *Psychological Medicine* **48**, 1055-1067 (2018).

41. Vuoksimaa, E. *et al.*, Is bigger always better? The importance of cortical configuration with respect to cognitive ability. *NeuroImage* **129**, 356-366 (2016).

42. Dynak, A. *et al.*, Separating the influences of late talking and dyslexia on brain structure. *Journal of Abnormal Psychology* **130**, 286-296 (2021).

43. Palmer, C. E. *et al.*, Distinct Regionalization Patterns of Cortical Morphology are Associated with Cognitive Performance Across Different Domains. *Cerebral Cortex* **31**, 3856-3871 (2021).

44. Hickok, G. & Poeppel, D., The cortical organization of speech processing. *Nature Reviews Neuroscience* **8**, 393-402 (2007).

45. Atteveldt, N., Formisano, E., Goebel, R. & Blomert, L., Integration of Letters and Speech Sounds in the Human Brain. *Neuron* **43**, 271-282 (2004).

46. Boets, B. *et al.*, Intact But Less Accessible Phonetic Representations in Adults with Dyslexia. *Science* **342**, 1251-1254 (2013).

47. Essen, D. C. V. & Dierker, D. L., Surface-Based and Probabilistic Atlases of Primate Cerebral Cortex. *Neuron* **56**, 209-225 (2007).

48. Płoński, P. *et al.*, Multi-parameter machine learning approach to the neuroanatomical basis of developmental dyslexia. *Human Brain Mapping* **38**, 900-908 (2016).

49. Perdue, M. V., Mednick, J., Pugh, K. R. & Landi, N., Gray Matter Structure Is Associated with Reading Skill in Typically Developing Young Readers. *Cerebral Cortex* **30**, 5449-5459 (2020).

50. Richlan, F., Kronbichler, M. & Wimmer, H., Structural abnormalities in the dyslexic brain: A meta-analysis of voxel-based morphometry studies. *Human Brain Mapping* **34**, 3055-3065 (2012).

51. Altarelli, I. *et al.*, Planum temporale asymmetry in developmental dyslexia: Revisiting an old question. *Human Brain Mapping* **35**, 5717-5735 (2014).

52. Ben-Shachar, M., Dougherty, R. F., Deutsch, G. K. & Wandell, B. A., Differential Sensitivity to

Words and Shapes in Ventral Occipito-Temporal Cortex. *Cerebral Cortex* **17**, 1604-1611 (2007).

53. Cohen, L. *et al.*, The visual word form area. *Brain* **123**, 291-307 (2000).

54. Beelen, C., Vanderauwera, J., Wouters, J., Vandermosten, M. & Ghesquière, P., Atypical gray matter in children with dyslexia before the onset of reading instruction. *Cortex* **121**, 399-413 (2019).

55. Carreiras, M., Quiñones, I., Hernández-Cabrera, J. A. & Duñabeitia, J. A., Orthographic Coding: Brain Activation for Letters, Symbols, and Digits. *Cerebral Cortex* **25**, 4748-4760 (2015).

56. Feng, X. *et al.*, A universal reading network and its modulation by writing system and reading ability in French and Chinese children. *eLife* **9** (2020).

57. Lyall, A. E. *et al.*, Dynamic Development of Regional Cortical Thickness and Surface Area in Early Childhood. *Cerebral Cortex* **25**, 2204-2212 (2014).

58. Panizzon, M. S. *et al.*, Distinct Genetic Influences on Cortical Surface Area and Cortical Thickness. *Cerebral Cortex* **19**, 2728-2735 (2009).

59. Rakic, P., Specification of cerebral cortical areas. *Science* **241**, 170-176 (1988).

60. Meer, D. *et al.*, Quantifying the Polygenic Architecture of the Human Cerebral Cortex: Extensive Genetic Overlap between Cortical Thickness and Surface Area. *Cerebral Cortex* (2020).

61. Eyler, L. T. *et al.*, A Comparison of Heritability Maps of Cortical Surface Area and Thickness and the Influence of Adjustment for Whole Brain Measures: A Magnetic Resonance Imaging Twin Study. *Twin Research and Human Genetics* **15**, 304-314 (2012).

62. Hoogman, M. *et al.*, Brain Imaging of the Cortex in ADHD: A Coordinated Analysis of Large-Scale Clinical and Population-Based Samples. *American Journal of Psychiatry* **176**, 531-542 (2019).

63. Owens, M. M. *et al.*, Multimethod investigation of the neurobiological basis of ADHD symptomatology in children aged 9-10: baseline data from the ABCD study. *Translational Psychiatry* **11** (2021).

64. Alfaro-Almagro, F. *et al.*, Confound modelling in UK Biobank brain imaging. *NeuroImage* **224**, 117002 (2021).

65. Batouli, S. A. H., Trollor, J. N., Wen, W. & Sachdev, P. S., The heritability of volumes of brain structures and its relationship to age: A review of twin and family studies. *Ageing Research Reviews* **13**, 1-9 (2014).

66. Jansen, A. G., Mous, S. E., White, T., Posthuma, D. & Polderman, T. J. C., What Twin Studies Tell Us About the Heritability of Brain Development, Morphology, and Function: A Review. *Neuropsychology Review* **25**, 27-46 (2015).

67. Teeuw, J. *et al.*, Genetic Influences on the Development of Cerebral Cortical Thickness During Childhood and Adolescence in a Dutch Longitudinal Twin Sample: The Brainscale Study. *Cerebral Cortex* **29**, 978-993 (2018).

68. Visscher, P. M. *et al.*, Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Samples. *PLoS Genetics* **10**, e1004269 (2014).

69. Frei, O. *et al.*, Bivariate causal mixture model quantifies polygenic overlap between complex traits

beyond genetic correlation. *Nature Communications* **10** (2019).

70. Mitchell, B. L. *et al.*, Educational attainment polygenic scores are associated with cortical total surface area and regions important for language and memory. *NeuroImage*, 116691 (2020).

71. Lett, T. A. *et al.*, Cortical Surfaces Mediate the Relationship Between Polygenic Scores for Intelligence and General Intelligence. *Cerebral Cortex* **30**, 2708-2719 (2019).

72. Vuoksimaa, E. *et al.*, The Genetic Association Between Neocortical Volume and General Cognitive Ability Is Driven by Global Surface Area Rather Than Thickness. *Cerebral Cortex* **25**, 2127-2137 (2014).

73. Marek, S. *et al.*, Reproducible brain-wide association studies require thousands of individuals. *Nature* **603**, 654-660 (2022).

74. Loughnan, R. J. *et al.*, Gene-experience correlation during cognitive development: Evidence from the Adolescent Brain Cognitive Development (ABCD) Study. *bioRxiv* (2019).

75. Walhovd, K. B. *et al.*, Neurodevelopmental origins of lifespan changes in brain and cognition. *Proceedings of the National Academy of Sciences* **113**, 9357-9362 (2016).

76. Jernigan, T. L. & Brown, S. A., Introduction. *Developmental Cognitive Neuroscience* **32**, 1-3 (2018).

77. Auchter, A. M. *et al.*, A description of the ABCD organizational structure and communication framework. *Developmental Cognitive Neuroscience* **32**, 8-15 (2018).

78. Garavan, H. *et al.*, Recruiting the ABCD sample: Design considerations and procedures. *Developmental Cognitive Neuroscience* **32**, 16-22 (2018).

79. Compton, W. M., Dowling, G. J. & Garavan, H., Ensuring the Best Use of Data. *JAMA Pediatrics* **173**, 809 (2019).

80. Luciana, M. *et al.*, Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. *Developmental Cognitive Neuroscience* **32**, 67-79 (2018).

81. Casey, B. J. *et al.*, The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience* **32**, 43-54 (2018).

82. Hagler, D. J. *et al.*, Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *NeuroImage* **202**, 116091 (2019).

83. Bates, D., Mächler, M., Bolker, B. & Walker, S., Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**, 1-48 (2015).

84. Revelle, W., 2020.

85. Mowinckel, A. M. & Vidal-Piñeiro, D., Visualization of Brain Statistics With R Packages ggseg and ggseg3d. *Advances in Methods and Practices in Psychological Science* **3**, 466-483 (2020).

86. Mowinckel, A. M. & Vidal-Piñeiro, D., 2021.

87. Purcell, S. *et al.*, PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559-575 (2007).

88. Coleman, J. R. I. *et al.*, Quality control, imputation and analysis of genome-wide genotyping data from the Illumina HumanCoreExome microarray. *Briefings in Functional Genomics* **15**, 298-304 (2015).

89. Patterson, N., Price, A. L. & Reich, D., Population Structure and Eigenanalysis. *PLoS Genetics* **2**, e190 (2006).

90. Auton, A. *et al.*, A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).

91. Choi, S. W. & OReilly, P. F., PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience* **8** (2019).

92. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K., mediation:RPackage for Causal Mediation Analysis. *Journal of Statistical Software* **59** (2014).

93. Imai, K., Keele, L. & Tingley, D., A general approach to causal mediation analysis. *Psychological Methods* **15**, 309-334 (2010).

94. Smith, S. *et al.*, Structural Variability in the Human Brain Reflects Fine-Grained Functional Architecture at the Population Level. *The Journal of Neuroscience* **39**, 6136-6149 (2019).

95. Littlejohns, T. J. *et al.*, The UK Biobank imaging enhancement of 100,000 participants:0.167emrationale, data collection, management and future directions. *Nature Communications* **11** (2020).

96. Jednoróg, K. *et al.*, How reliable are gray matter disruptions in specific reading disability across multiple countries and languages? insights from a large-scale voxel-based morphometry study. *Human Brain Mapping* **36**, 1741-1754 (2015).

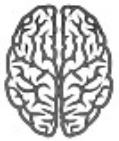
827

A. Which are robust brain correlates of reading?

1. Left hemisphere structural cortical measures associated to reading

Oral reading
sMRI

LMM



B. Does genetics influence those brain-reading associations?

2. Genetic architecture of reading-related cognitive and brain measures

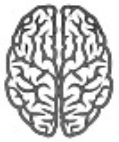
Herita-bility

Oral reading
sMRI
Genotyping

h^2_{SNP} (GREML)

sMRI
Cognitive

h^2_{SNP} (LDSC)



3. Shared genetic influences on brain and reading

Genetic correlation

sMRI
Cognitive

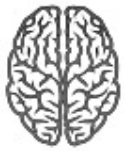
r_g (LDSC)

Polygenic scoring / Mediation

Cognitive

sMRI
Genotyping

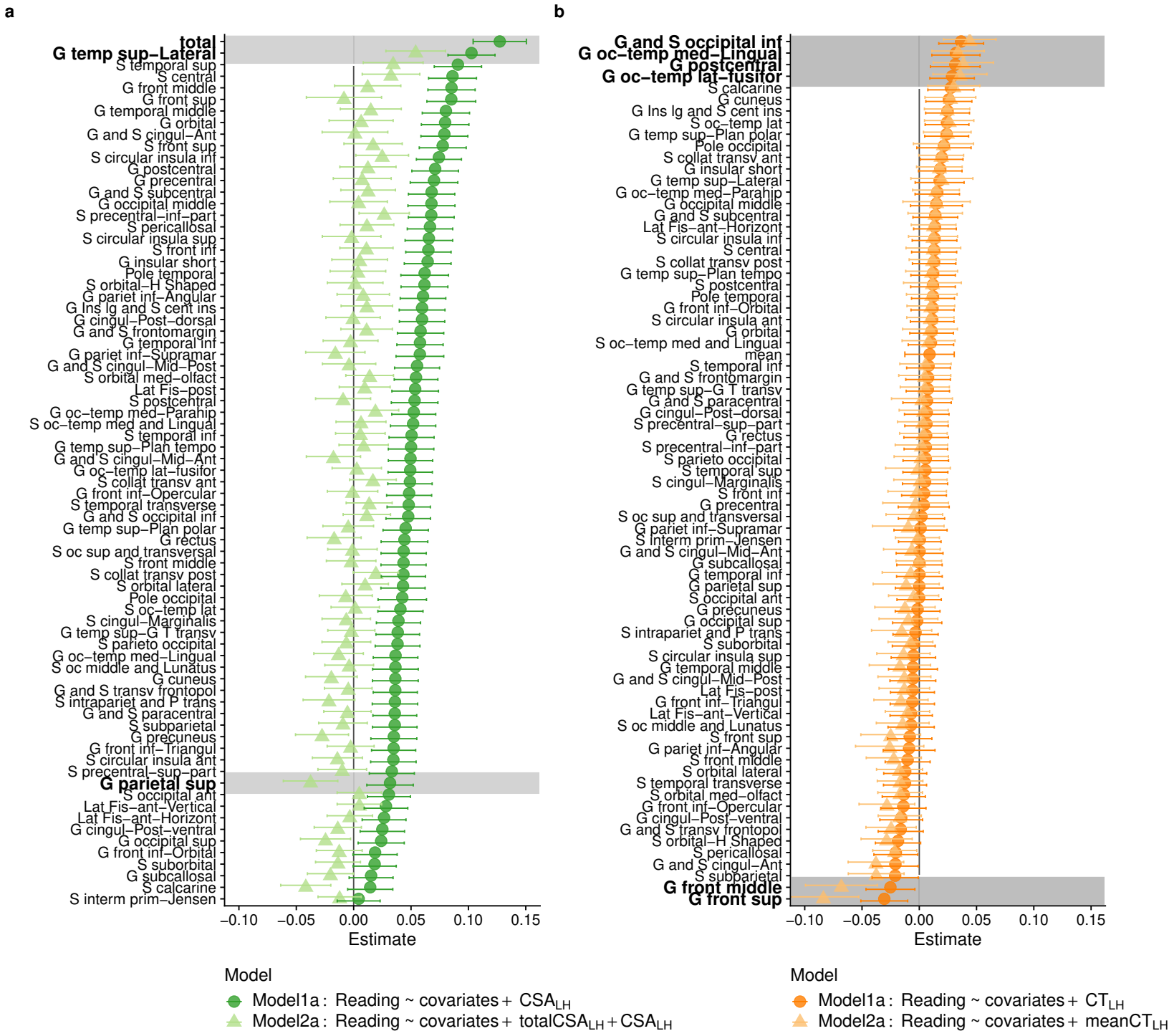
PGS

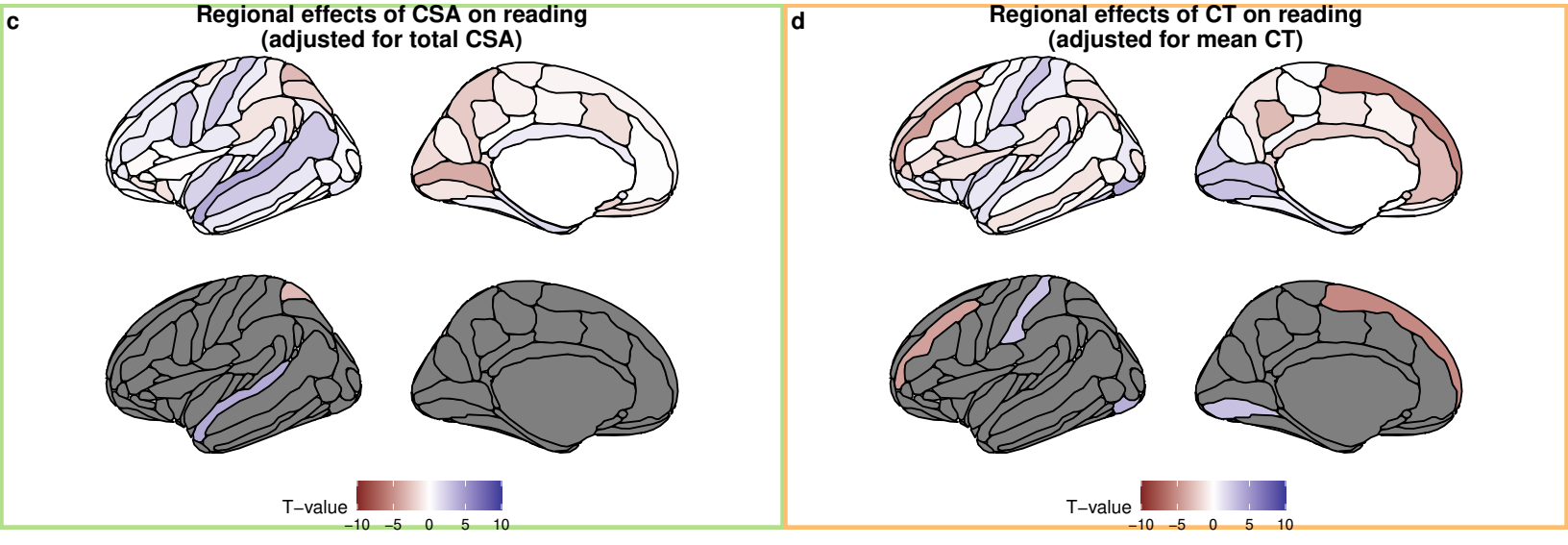


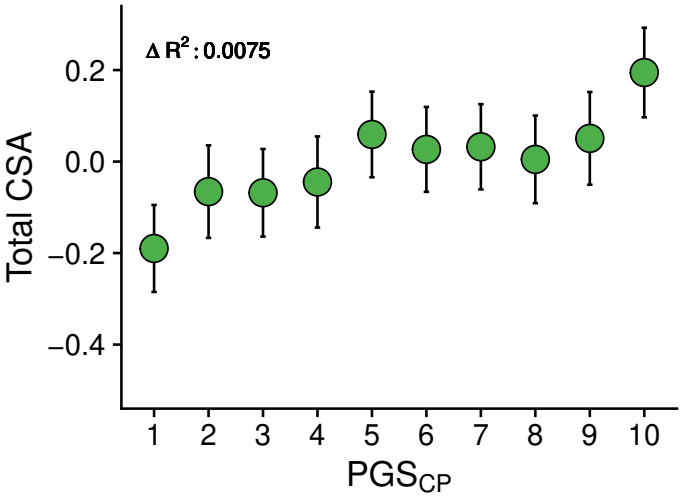
Data

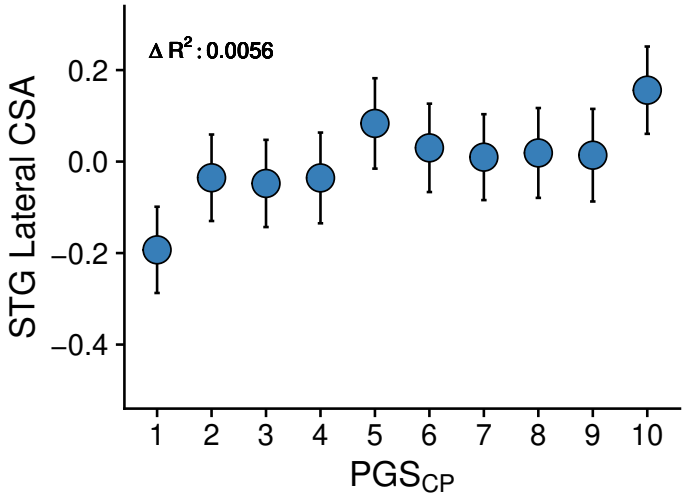
Individual level data
ABCD dataset

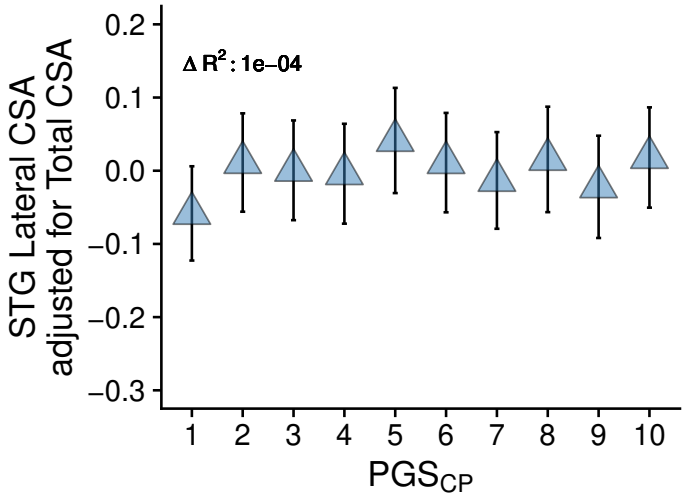
GWAS summary statistics
Table S3

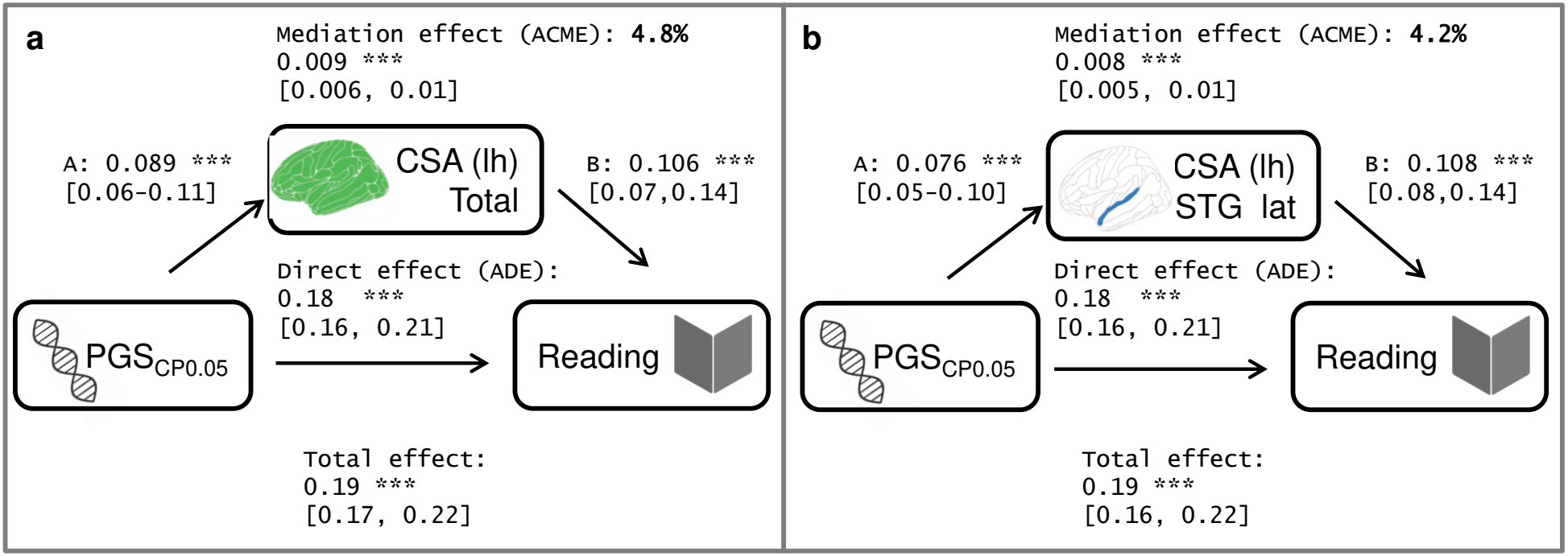

a
b
total
G temp sup−Lateral
G parietal sup
G and S occipital inf
G oc−temp med−Lingual
G postcentral
G oc−temp lat−fusifor
G front middle
G front sup
Estimate
−0.10
−0.05
0.00
0.05
0.10
0.15
Model
Model1a : Reading ~ covariates + CSA_LH
Model2a : Reading ~ covariates + totalCSA_LH + CSA_LH
Model1a : Reading ~ covariates + CT_LH
Model2a : Reading ~ covariates + meanCT_LH


c
Regional effects of CSA on reading
(adjusted for total CSA)
d
Regional effects of CT on reading
(adjusted for mean CT)
T−value
−10
−5
0
5
10


Δ R²: 0.0075
Total CSA
0.2
0.0
−0.2
−0.4
1 2 3 4 5 6 7 8 9 10
PGS_CP


Δ R²: 0.0056
STG Lateral CSA
0.2
0.0
−0.2
−0.4
1 2 3 4 5 6 7 8 9 10
PGS_CP


Δ R²: 1e−04
STG Lateral CSA
adjusted for Total CSA
0.2
0.1
0.0
−0.1
−0.2
−0.3
1 2 3 4 5 6 7 8 9 10
PGS_CP


a
Mediation effect (ACME): 4.8%
0.009 ***
[0.006, 0.01]
A: 0.089 ***
[0.06-0.11]
CSA (lh)
Total
B: 0.106 ***
[0.07,0.14]
Direct effect (ADE):
0.18 ***
[0.16, 0.21]
PGS CP0.05
Reading
Total effect:
0.19 ***
[0.17, 0.22]
b
Mediation effect (ACME): 4.2%
0.008 ***
[0.005, 0.01]
A: 0.076 ***
[0.05-0.10]
CSA (lh)
STG lat
B: 0.108 ***
[0.08,0.14]
Direct effect (ADE):
0.18 ***
[0.16, 0.21]
PGS CP0.05
Reading
Total effect:
0.19 ***
[0.16, 0.22]