Hypothesis-driven genome-wide association studies provide novel insights into genetics of reading disabilities

Kaitlyn M. Price1,2,3, Karen G. Wigg1, Else Eising4, Yu Feng1, Kirsten Blokland2, Margaret Wilkinson2, Elizabeth N. Kerr5,6, Sharon L. Guger5, Quantitative Trait Working Group of the GenLang Consortium*, Simon E. Fisher4,7, Maureen W. Lovett2,6, Lisa J. Strug8,9 and Cathy L. Barr1,2,3,5,3

© The Author(s) 2022

Reading Disability (RD) is often characterized by difficulties in the phonology of the language. While the molecular mechanisms underlying it are largely undetermined, loci are being revealed by genome-wide association studies (GWAS). In a previous GWAS for word reading (Price, 2020), we observed that top single-nucleotide polymorphisms (SNPs) were located near to or in genes involved in neuronal migration/axon guidance (NM/AG) or loci implicated in autism spectrum disorder (ASD). A prominent theory of RD etiology posits that it involves disturbed neuronal migration, while potential links between RD-ASD have not been extensively investigated. To improve power to identify associated loci, we up-weighted variants involved in NM/AG or ASD, separately, and performed a new Hypothesis-Driven (HD)–GWAS. The approach was applied to a Toronto RD sample and a meta-analysis of the GenLang Consortium. For the Toronto sample (n = 624), no SNPs reached significance; however, by gene-set analysis, the joint contribution of ASD-related genes passed the threshold (p = 1.45 × 10^-5, threshold = 2.5 × 10^-8). For the GenLang Cohort (n = 26,558), SNPs in DOCK7 and CDH4 showed significant association for the NM/AG hypothesis (sFDR q = 1.02 × 10^-3). To make the GenLang dataset more similar to Toronto, we repeated the analysis restricting to samples selected for reading/language deficits (n = 4152). In this GenLang selected subset, we found significant association for a locus intergenic between BTG3-C21orf91 for both hypotheses (sFDR q < 9.00 × 10^-10). This study contributes candidate loci to the genetics of word reading. Data also suggest that, although different variants may be involved, alleles implicated in ASD risk may be found in the same genes as those implicated in word reading. This finding is limited to the Toronto sample suggesting that ascertainment influences genetic associations.

Translational Psychiatry (2022) 12:495; https://doi.org/10.1038/s41398-022-02250-z

INTRODUCTION

Reading Disability (RD), also known as developmental dyslexia, is a neurodevelopmental disorder affecting children globally. In North America alone, it affects 5–7% of individuals [1–6]. RD is characterized by difficulties with word reading and spelling, despite typical intelligence and motivation to learn [7]. Affected children often have comorbid neurodevelopmental disorders, including language or speech impairments, or attention-deficit/hyperactivity disorder (ADHD) [8]. These factors increase social difficulties, decrease self-esteem, and hinder academic/occupational success [9–12]. RD, therefore, represents a major public health concern.

The genetics and underlying mechanisms of RD are not fully known. Twin and family studies initially demonstrated genetic contributions to RD by examining heritability within families [3, 13]. Researchers went on to identify specific chromosomal regions and genes implicated in RD via linkage analysis followed by fine-mapping association studies. These linked regions and genes were supported, to varying degrees, by independent studies [14] and meta-analyses [15–20]; however, sample sizes were small by current standards yielding low power and elevated risk of false-positive findings. Moreover, some results could not be replicated [16, 21, 22]. Therefore, researchers called into question the robustness of the genes as candidates for involvement in RD. Despite these caveats, a number of candidate genes identified from linkage/fine-mapping studies (KIAA0319, KIAA0319L, DCDC2, DNAAF4 (previously called DSY1C1 and EKN1), and ROBO1) were linked to neuronal migration, suggesting a potential molecular mechanism (but for a critical review see [23]). These associations were pertinent as previous postmortem brain studies (n < 10) found heterotopias and cortical dysplasias, signatures of altered migration, in RD-affected individuals [24, 25]. It was theorized that disrupted neuronal migration (DNM) may be involved in RD etiology [26, 27].
The first evidence to implicate allelic variation of these RD candidate genes in DNM came from studies finding that protein motifs/domains encoded by the genes were predicted to function in migration [28–32]. Further evidence emerged from in utero knockdown experiments of the genes (KIAA0319, KIAA0319L, DCDC2, and DNAFA4/DYX1C1) in the developing brains of mice or rats. When gene expression was reduced, neural cells did not migrate as expected from the lower ventricular zone to the higher cortical plate [33–38]. Instead, most cells remained in the ventricular or intermediate zones, indicating disrupted migration, albeit with different patterns of disruption for different genes [33–38]. Experiments in which the candidate genes were over-expressed also resulted in neural phenotypes, including aberrant neurite outgrowth (Dcc) [39], delayed radial migration (Kiaa0319) [40], and altered axon growth and regeneration (Kiaa0319) [41].

There was, however, skepticism regarding the proposed roles of these genes in migration, when independent knockout experiments did not replicate the results of the knockdown studies [23, 40, 42–44]. Discrepancies between knockout and knockdown findings may be due to developmental compensation by altered gene expression [45, 46]. For example, data from Dcc knockdown mice supports partial redundancy with Dcx, a member of the same gene family, which functionally compensates for the loss of Dcc [43]. Overall, it remains unclear at this time whether the candidate genes in question are indeed implicated in RD and if this involves effects on neuronal migration [23].

To further identify genes implicated in RD, and as the available techniques advanced, the field moved away from linkage analysis in families to genome-wide association studies (GWAS) — a more powerful approach for complex traits in which effect sizes of individual risk variants are small. In the last few years, GWAS for RD, reading performance, and/or reading-related tasks have begun to yield results with genome-wide statistical significance (SNP-based analysis (p<10^-8) [47–50]; gene-based analysis (p<10^-5) [51–53]). Across recent significant and previous non-significant GWAS findings, a number of the top genes are thought to be involved in neuronal migration/axon pathfinding, potentially supporting the DNM hypothesis. For example, in a GWAS by Price et al., 2020, using two samples, the Toronto (n = 624) and Philadelphia Neurodevelopmental Cohort (n = 4430), the most significant SNP (p < 5 x 10^-8) in the Toronto sample was located in an intron of ARHGRAP23, a gene involved in actin cytoskeleton polymerization/reorganization, neuronal development, and other growth cone/axon related processes [54]. Across both samples in that study, when top-ranked SNPs, at a less stringent p-value threshold (p<10^-5), were mapped into or near close proximity genes, additional genes were found to have been previously related to neuronal migration/axon guidance (NM-AG) (as well as growth cone formation which is encompassed within this term). For example, ASTN2, CNTN4, TUBB3, NRCAM, DSCAM, UNCSD, and GAP43 [53]. Larger GWAS studies also provided weak support for the DNM hypothesis. A meta-analysis of 22 samples (n ~34,000) by the GenLang Consortium, identified genome-wide significant variants in DOCK7 associated with word reading [47]. DOCK7 is critical for cortical neurogenesis, axon formation, and neural polarization [55]. In the largest GWAS study to date, analyzing the 23andMe cohort, the authors identified 42 significant loci associated with self-reported dyslexia (n_cases~51,000) of which genes had been previously related to NM/AG [48] (Nature Genetics, in press https://doi.org/10.10110/2021.08.20.2126334). However, it should be noted, a systematic targeted gene-set analysis in that sample found significant overrepresentation of genes related to axon guidance but not for those involved in neuronal migration [48].

The Price et al. study also identified, at the less stringent p-value threshold (p<10^-5), variants in or near genes implicated in neurodevelopmental disorders, particularly autism spectrum disorder (ASD) [53]. For example, ANKS1B, CNTN4, RBFOX1, CSM1D, and ASTN2. The study of the self-reported dyslexia in 23andMe observed top associated SNPs in ASD-related genes. Although there is little evidence to support phenotypic comorbidity between ASD and RD, both are neurodevelopmental disorders and ASD involves deficits in language and communication skills [56, 57]. Further, there is some preliminary evidence of shared genetic risk factors: rare and de novo variants identified in ASD cases for genes that have been independently associated with RD (PRTG, ROBO1, and KIAA0319L) [58, 59], and altered neuronal migration has been suggested as an etiological mechanism in each condition [60–64]. While putative links between RD and DNM have received much attention in prior literature, few investigations to date have explored potential overlaps in neurobiological foundations of RD and ASD.

Given GWAS-based observations of associated SNPs in genes previously implicated in ASD and genes involved in NM/AG, along with investigations of DNM in previous candidate gene and postmortem studies, we wished to leverage this information to improve power to identify significant variants. Most existing GWAS samples are modest in size, with the exception of [47, 48], and relatively few findings reach statistical significance at the genome-wide level. To leverage power within available samples, we used Hypothesis-Driven (HD)--GWAS, which up-weights or prioritizes variants based on previously established genetic or biological hypotheses [65]. Statistical corrections are performed independently on the up-weighted and non-up-weighted groups, reducing the threshold for significance and increasing power. The HD-GWAS was primarily conducted on the Toronto sample [53]. We also wanted to examine the issues in the context of a larger dataset, so secondarily we used the meta-analysis of the GenLang Consortium [47]. A tertiary analysis was conducted using only those GenLang samples that were selected based on reading and/or language difficulties, to make it more comparable to the Toronto samples.

For our HD-GWAS of word reading, we formulated two separate hypotheses based on the results from previous GWAS ("conventional", hypothesis-free GWAS) [53] and the literature. Specifically, we hypothesized that variants in (1) genes implicated in NM/AG and (2) genes implicated in ASD, would contribute to word reading.

**METHODS**

**Study populations and measures**

To complete the HD-GWAS analyses, summary statistics from the Toronto conventional GWAS [53] and the meta-analysis of the GenLang Consortium were used [47]. The gene-set analyses made use of the raw genotypes of the Toronto sample and the summary statistics of GenLang. For both samples, quantitative measures of word reading were used as the phenotype.

**Toronto sample.** The Toronto sample has previously been described and is part of an ongoing genetic study of RD-selected families recruited at the Hospital for Sick Children [53, 56, 66]. At the time of analysis, the sample consisted of children identified with reading difficulties (n = 453) and both their unaffected and affected siblings (n = 171). Children were of European ancestry, as determined by PCA analysis. This choice aimed to reduce variation and create a more homogeneous population; however, it does not fully encompass the complex nature of allelic structure [67]. Children were excluded if there was evidence of other neurodevelopmental disorders, including ASD, apraxia, dyspraxia, central auditory processing disorders, stuttering, and psychiatric disorders, as well as medical conditions that would interfere with reading. Information on psychiatric and neurodevelopmental disorders was obtained using a structured parent interview [68] and a semi-structured teacher interview [69]. Children with ADHD, mild anxiety disorders, and speech/language difficulties were included.

Translational Psychiatry (2022) 12:495
The GenLang Consortium. The GenLang Consortium is a large international effort to study genetic contributions to traits related to speech, language, and reading (https://www.genlang.org). It does so through meta-analyses of these traits in population-based samples, as well as family-based and case-control cohorts [47]. For the purposes of the current study, the Toronto sample was analyzed separately, because the overlap between RD and neuronal migration/ASD was originally observed in this sample and formed the basis of the hypotheses being tested in this work.

The GenLang meta-analysis dataset used in this study consisted of 17 samples (n = 26,588) with individuals of European ancestry, as determined by PCA analysis (Table S1) [47]. Data quality control and filtering were performed by each individual sample before the meta-analysis [47]. We refer to this sample collectively as "The GenLang Cohort".

In the GenLang Cohort, word reading was measured using a variety of standardized tests (depending on the individual sample) [47]. These phenotypic data were aligned across samples prior to the meta-analysis, as described in [47]. Consent was obtained from all participants in the GenLang Consortium and each individual sample's institution-approved data use.

The GenLang selected subset. Five samples of the GenLang Cohort, selected for reading or language difficulties, were also examined (n = 4152, Table S1) as a subset. With the exception of the SLIC cohort, all samples were selected for reading difficulties and participants were between the ages of 7 to 18. We refer to this collectively as the "GenLang selected subset". In the family-based samples, phenotypic data were available both from probands and their siblings, regardless of affection status; therefore, these samples included a range of quantitative variation.

Data processing and GWAS of Toronto. Genotyping and quality control of the Toronto sample were previously described [53]. Briefly, DNA from each participant was genotyped on the OmniExpress platform and unobserved variants were imputed using the Michigan imputation server [71]. Quality control was conducted over numerous steps. Sex was checked using the heterozygosity of markers on the X chromosome. Sibling relationships were confirmed genetically and individuals with cryptic relationship were removed. In addition, variants with low imputation quality (R² < 0.3), variants out of Hardy-Weinberg equilibrium (p < 0.00001), and variants with minor allele frequencies less than 5% were removed as well as samples with >2% missing genotypes and call rates <98%. After this filtering, 5.3 million SNPs were included in the analysis.

Because the Toronto sample included sibling pairs, the GWAS analysis was conducted using a linear mixed model in the R package ‘nlme’ to include a random effects term for family relationship [53]. Covariates for population structure (principal components) were also included as fixed effects. Only self-reported European Caucasian individuals were included in the analysis; verified by genotype. Principal components were generated in the program KING [72] and a Tracy-Widom statistic (EIGENSOFT Program) was used to determine significance [53].

Meta-analysis of the GenLang Cohort. Genotyping and quality control of the GenLang Cohort samples were previously described [47]. Individual cohort association analyses were performed with different tools, including SNPTEST [73], GEMMA [74], and PLINK [75] and included covariates for population structure (principal components) and family relationship.

A meta-analysis was performed on the samples using the program METAL [76], again using only individuals of European ancestry as determined by principal components. Effect size estimates were weighted with the inverse of the standard errors and genomic control on [47]. The GenLang data were included in the same meta-analysis process. Summary statistics for the GenLang Cohorts were provided for use in this study after review and approval of the project by the coordinating board of the GenLang Consortium.

The Manhattan and quantile-quantile plots were generated using FUMA [77]. The regional association plot was examined in LocusFocus (https://locusfocus.research.sickkids.ca) [78]. For both the GenLang Cohort and GenLang selected sample, only SNPs with a MAF ≥ 5%, and variants present in ≤50% of the total sample were used.

HD-GWAS. HD-GWAS serves as a powerful approach to GWAS by incorporating genetic or biological hypotheses based on the previously conducted research. This technique was developed and then tested by [65, 79–81]. Variants are divided into two strata: a stratum where it is hypothesized that variants are associated with the trait and a stratum where they are not. An estimated false discovery rate (FDR) is then calculated in separate on each stratum. This prioritization leads to a less stringent type 1 error cut-off and increases the power to detect associated SNPs.

Although the Toronto sample did not originally meet SNP-based significance in the conventional GWAS, it did produce SNP level p-values of 10⁻² and significant evidence by gene-based analysis. Thus, FDR correction was appropriate to increase power. The GenLang sample was a powerful sample with significant findings by conventional GWAS.

HD-GWAS was run using the stratified False Discovery Rate (sFDR) framework and program (http://www.utstat.toronto.edu/sun/Software/sFDR/index.html) [65, 79, 80]. As input, the program requires a SNP identifier, the p-value, and the weighting status (1– not up-weighted (not prioritized), 2– up-weighted (prioritized)). R (https://www.r-project.org) and the command "merge 1" were then used to incorporate up-weighting information with summary statistics. Variants that were not in the up-weighted group formed the control stratum. The sFDR commands “assoc” and “sFDR” were used. The output included the FDR q-value and a sFDR q-value.

To test the first hypothesis, we up-weighted variants in genes implicated in NI/AG as well as growth cone formation. The gene list was made with AmiGO (https://amigo.geneontology.org), which uses the Gene Ontology (GO) database to annotate genes. The following search terms were used: GO:0001764 neuron migration, GO:0007411 axon guidance, and GO:0030426 growth cone. We broadened neuronal migration to include axon guidance and growth cone formation because many axon guidance molecules are pleiotropic, with diverse functions in multiple tissues and in the brain, including neuronal migration in the developing brain [82, 83] and growth cones are at the tips of the leading processes of migrating neurons and elongating axons [82, 83].

Genes in all three pathways have been implicated in disorders of neuronal migration, including periventricular nodular heterotopia [84], a neuronal migration disorder in which cortical development is compromised, leading to epilepsy and RD [85]. A total of ~115,000 variants in 351 genes were tested (Table S2). Within this gene list, we included the original RD-linked candidate genes that have been implicated in neuronal migration (KIAA0319, DCDC2, DYS1C1, and ROBO1).

To test the second hypothesis, we up-weighted variants in genes implicated in ASD using the gene list from the Simons Foundation Autism Research Initiative (SFARI) database (https://gene.sfari.org/databse/human-gene/). More than half of these genes have been implicated in ASD through rare de novo mutations, or copy number variants (CNVs, syndromic or functional), but we also included those identified via genome-wide association studies. All categories were included, which at the time of the analysis consisted of 990 genes (Table S3) (~370,000 variants).

Gene set analysis. The joint contribution of the genes annotated to each hypothesis was tested using gene-set analysis in MAGMA [86]. For the NI/AG and ASD hypotheses, the AMIGO and the SFARI gene lists, respectively, were used. Gene-set analysis involves three steps, which were completed in MAGMA (https://ctg.cnrc.nl/software/magama). For the first step genes were annotated to SNPs. Input for this step was raw genotype data for the Toronto sample and the referred data of the 1000 genomes project for the GenLang [87]. Second, individual gene analysis was completed to determine the association between each gene and the phenotype. For both samples, this step was performed using linear regression to compute a p-value for each gene. The input data were as follows. For the Toronto sample, the predictor variables were gene annotations from the previous step and principal components for population structure. The outcome variable was the word reading phenotype. For the GenLang datasets, the predictor variables were the gene annotations from the previous step and the outcome was the summary statistic p-values. Lastly, gene annotations were aggregated to their set and tested as a unit to see if they affected the trait. MAGMA took into consideration gene size, gene density, and allele count. The null hypothesis was that the genes tested as a group showed no greater association than a random set of genes.

Threshold for significance. The threshold for significance for the HD-GWAS and gene-set analyses was set at 2.5 × 10⁻² (critical threshold 0.05/2 hypotheses). The Toronto sample formed the basis of our hypotheses and...
RESULTS

Conventional GWAS

The conventional GWAS for the Toronto and the GenLang Cohort have previously been published [47, 53]. The HD-GWAS analyses utilized the GenLang Cohort without the Toronto sample because the potential overlap between reading and neuronal migration/ASD was originally observed in that sample [53].

The GWAS of the GenLang selected subset, including only those GenLang samples that were recruited via probands with RD or language disorder, is a novel analysis that has not previously been described. The characteristics of cohorts in the selected subset are shown in Figs. S1–2. The p values depart from the expected line in the Q-Q plot at a p value of 10^{-4}. The λ statistic was a value of 1.02.

For the conventional GWAS of the GenLang selected subset (no hypothesis tested), the top associated locus was on chromosome 21q21.1 in the intergenic region between genes BTG3 and C21orf91. The most significant SNP was rs4818369 (Table 1, p = 2.37 × 10^{-10}, threshold p ≤ 5 × 10^{-8}; results at p < 10^{-6} shown in Table S4). The regional association plot is depicted in Fig. S3. Rs4818369 was not found to correlate with altered splicing or eQTLs (Table 2). Rs1168041 is an eQTL and splice quantitative trait locus (sQTL) for DOCK7, as are SNPs in LD (r^2 > 0.3) with this marker (GTEx v8, Table S9). This SNP is in LD (r^2 = .33) with the top SNP (rs11208009) in the original GenLang meta-analysis (22 samples, including the Toronto sample) located ~45 kb from DOCK7 [47]. Rs11208009 is also an eQTL and sQTL for DOCK7 (Eising et al. [2022] Supplementary) [47]. Rs6089259, intronic to CDH4, is not correlated with altered splicing or eQTLs according to available data.

For the gene-set analysis of the GenLang, no gene-sets passed the threshold for significance (threshold p ≤ 2.5 × 10^{-2}; Table S11).

HD-GWAS for the GenLang selected subset

For the HD-GWAS of the GenLang selected subset, the same locus and SNP (21q21.1, rs4818369) as the conventional GWAS passed the threshold for significance for both hypotheses (Table 4, p = 2.37 × 10^{-10}, sFDR q < 0.05 Tables S12–S13). This SNP was the top-ranked SNP by sFDR and not in the prioritized group, reflecting the robustness of the HD-GWAS [79]. Within the HD-GWAS and conventional GWAS of the GenLang selected subset, 14 and 18 SNPs, respectively, were previously identified with p < 10^{-6} in a prior GWAS analysis of word reading by Gialluisi et al. [50] (Tables S14–S15). The GenLang selected subset includes four cohorts which were included in the earlier Gialluisi et al. study [50], although at the time of that study the aforementioned cohorts were smaller in sample size than presently. The GenLang selected subset included one extra cohort (SLC). The GenLang selected subset comprised 4152 individuals while Gialluisi included 3468 individuals.

For the gene-set analysis of the GenLang selected subset, no gene-sets passed the threshold for significance (threshold p ≤ 2.5 × 10^{-2}; Table S16).

DISCUSSION

Until recently, GWAS investigations of reading skills have identified few associated loci that passed the threshold for genome-wide significance. However, with the organization of large-scale collaborations, such loci are beginning to be found [47–50]. The number and size of cohorts characterized for reading-related traits has been a limiting factor. Therefore, methods that improve power to capitalize on existing samples are necessary to move the field forward in the short term. Overlap with top loci and genes known to be related to NM/AG and ASD susceptibility were observed in previous GWAS findings [47, 48, 53]. To test these observations, we used HD-GWAS, prioritizing variants in genes implicated in NM/AG or ASD susceptibility. We also tested the joint contribution of the genes and therefore the prioritization hypotheses themselves.

For the hypothesis testing the relationship to ASD, we did not identify significant SNPs by HD-GWAS. However, gene-set analysis determined that the hypothesis itself tested as the joint contribution of ASD-related genes was significant in the Toronto sample. There was no relationship in the GenLang meta-analysis. Previous GWAS studies that examined the relationship between ASD and RD [47, 48, 53] did not find genetic correlations using polygenic risk scores or LD Score Regression (LDSC). This may be

---

Table 1. Conventional GWAS GenLang selected subset only.

| SNP     | Position | A1 | Freq | Gene     | Effect | SE   | P        | N       |
|---------|----------|----|------|----------|--------|------|----------|---------|
| rs4818369 | 21:19055075 | T  | 0.60 | BTG3-C21orf91 | -0.15  | 2.43×10^{-2} | 2.37×10^{-10} | 4152 |

---

Table 2. Gene-set results for Toronto.

| Hypothesis                  | Ngenes | Beta    | SE   | P       |
|----------------------------|--------|---------|------|---------|
| Neuronal Migration/ Axon Guidance | 327    | 0.066   | 0.048| 0.08584 |
| SFARI ASD                   | 890    | 0.066   | 0.030| 0.01448 |

Ngenes number of genes, SE Standard Error, P P value. Significance threshold p ≤ 2.5 × 10^{-2}.

was corrected separately from the GenLang meta-analyses. The Toronto sample had no participant overlap with the GenLang Cohort.
because the GWAS for ASD to date are relatively small or because the cohorts were composed of diverse neurodevelopmental phenotypes as previously suggested [48]. Further, the PRS/LDSC analyses depend on summary statistics from GWAS analysis, which use only common variants. The SFARI dataset contains genes implicated in ASD by rare variants or CNV analyses, as well as genes identified from association studies of common variation. Thus, ASD-reading trait overlap may not be detectable via PRS methods because rare variants were not imputed in the GWAS analyses. Another possibility is that there may indeed be shared genes involved, but that the specific risk alleles are different for ASD and reading and are not in LD. HD-GWAS prioritizes genes irrespective of the contributing genetic variants and allows us to more formally quantify word reading–ASD associations in the Toronto sample, which was previously an observation [53].

For the Toronto sample, we excluded children with known or suspected ASD, or with a first-degree relative with ASD, or other global/intellectual disabilities. Overlap between reading- and ASD-associated genes likely stemmed from shared genetic risk for neurodevelopmental disorders, particularly those contributing to language-related difficulties, as opposed to global delays [56, 57].

For the NM/AG hypothesis, HD-GWAS using the GenLang Consortium data identified two associated loci with the top SNPs located in DOCK7 and CDH4. DOCK7, Dedicator of Cytokinesis 7, is involved in axon formation and neuronal polarization [91]. The top marker, rs1168041, is an eQTL and sQTL for both genes related to those disorders. Alternatively, clinical studies may screen for, and exclude participants with, comorbid or medical conditions or environmental factors that would interfere with reading acquisition. These exclusions could alter the composition of the sample compared to population-based samples and possibly influence gene findings.

In summary, through an HD-GWAS framework, we identified significant associations with reading skills. We also found that genes related to ASD risk contribute to RD in the Toronto sample. Our findings support two core features of the HD-GWAS framework. First, this framework is robust to stratifying misspecification of up-weighted variants (i.e., less than ideal hypotheses [65]). We demonstrated this feature when using HD-GWAS we identified the same chromosomal 21 SNPs from conventional GWAS, even

Table 3. HD-GWAS results for the GenLang Cohort $q < 0.05$.

| Hypothesis | SNP      | Position   | Gene       | P GWAS | qSFDR |
|------------|----------|------------|------------|--------|-------|
| Neuronal Migration/Axon Guidance | rs1168041 | 1:62960250 | DOCK7     | 6.61 × 10$^{-7}$ | 1.02 × 10$^{-2}$ |
|            | rs6089259 | 2:60246390 | CDH4      | 7.03 × 10$^{-6}$ | 1.02 × 10$^{-2}$ |
|            | rs17158413 | 15:83235408 | CPEB1    | 4.42 × 10$^{-5}$ | 2.92 × 10$^{-2}$ |

Total SNPs: 5,552,103.
Upweighted SNPs: Neuronal migration 115,448.

Table 4. HD-GWAS results for GenLang selected subset only $q < 0.05$.

| Hypothesis | SNP      | Position   | Gene       | P GWAS | qSFDR |
|------------|----------|------------|------------|--------|-------|
| Neuronal Migration/Axon Guidance | rs4818369 | 21:19055075 | BTG3-C21orf91** | 2.37 × 10$^{-10}$ | 9.00 × 10$^{-4}$ |
|            | rs6090818 | 20:46883131 | LINCO0494** | 1.82 × 10$^{-7}$ | 3.67 × 10$^{-2}$ |
|            | rs6865160 | 5:31766737 | PDZD2      | 2.82 × 10$^{-10}$ | 4.10 × 10$^{-2}$ |
| SFARI ASD  | rs4818369 | 21:19055075 | BTG3-C21orf91** | 2.37 × 10$^{-10}$ | 8.00 × 10$^{-4}$ |
|            | rs6090818 | 20:46883131 | LINCO0494** | 1.82 × 10$^{-7}$ | 3.50 × 10$^{-2}$ |
|            | rs6865160 | 5:31766737 | PDZD2      | 2.82 × 10$^{-10}$ | 3.92 × 10$^{-2}$ |

Total SNPs: ~5,610,600.
Upweighted SNPs: Neuronal migration, 116,829; SFARI ASD 371,158.

SNP Single Nucleotide Polymorphism, Position chromosome: base pair (hg19), P GWAS P-value in GWAS, qSFDR q-value from SFDR program.

Additional information including LD SNPs or SNPs in same gene region, rank and up-weighted status in the supplementary material.

Significance threshold $q \leq 2.5 \times 10^{-2}$.

**Intergenic and associated with word reading in Gialluisi et al., (6.79 × 10$^{-6}$/3.14 × 10$^{-7}$).
though they were not in the up-weighted group for the GenLang selected subset. Second, this framework increases power to identify unstratified approaches. We illustrated this feature when we found that the ASD-related gene-set contributed to reading and identified loci upweighted in the NM/AG hypothesis. Future studies involving larger GWAS samples ascertainment through reading and language disorders may help to elucidate shared genetic mechanisms between RD and ASD.

**DATA AVAILABILITY**
Summary statistics for the Toronto sample and the GenLang sample used in this study are available upon application to the GenLang Consortium (http://www.genlang.org) and review of the proposal. To download summary statistics for the entire GenLang (not the sample specific to this study), use http://www.genlang.org or the public GWAS catalogue.

**REFERENCES**
1. Katusic SK, Colligan RC, Barbaresi WJ, Scheck DI, Jacobsen SJ. Incidence of reading disability in a population-based birth cohort, 1976–1982, Rochester, Minn. Mayo Clin Proc. 2001;76:1081–92.
2. Snowling MJ, Melby-Lervag M. Oral language deficits in familial dyslexia: A meta-analysis and review. Psychological Bull. 2016;142:498–545.
3. Shaywitz SE, Shaywitz BA. Dyslexia (specific reading disability). Biol Psychiatry. 2005;57:1301–9.
4. Wagner RK, Zirps FA, Edwards AA, Wood SG, Joyner RE, Becker BJ, et al. The Prevalence of Dyslexia: A New Approach to Its Estimation. J Learn disabilities. 2020;53:334–65.
5. Peterson BI, Pennington BF. Developmental dyslexia. Lancet 2012;379:1997–2007.
6. Di Folco C, Guez A, Peyre H, Ramus F. Epidemiology of reading disability: A comparison of DSM-5 and IC(D-11 criteria. Sci Studies Reading. 2021;26:1–19.
7. Lyon GR. Part I De
tification of reading disorders. Ann dyslexia. 2003;53:1–14.
8. Hendren RL, Haft SL, Black JM, White NC, Hoeft F. Recognizing Psychiatric Comorbidity With Reading Disorders. Front psychiatry. 2018;9:101.
9. Daniel SS, Walsh AK, Goldston DB, Arnold EM, Reussbin SA, Wood FB. Suicidality, school dropout, and reading problems among adolescents. J Learn disabilities. 2006;39:507–14.
10. Smart D, Youssef GA, Sano A, Prior M, Toumbourou JW, Olsson CA. Con-
trolling reading and language disorders may help to elucidate shared genetic mechanisms between RD and ASD.
11. Morgan PL, Farkas G, Wu Q. Do Poor Readers Feel Angry, Sad, and Unpopular? Sci Stud Read. 2012;16:360–81.
12. Hossain B, Bent S, Hendren R. The association between anxiety and academic achievement and employment in early adulthood. Br J Educ Psychol. 2016;86:212–33.
13. Fisher SE, DeFries JC. Developmental dyslexia: genetic dissection of a complex
trait. Nat Rev Neurosci. 2002;3:767–80.
14. Bates TC, Luciano M, Castle S, Coltheart M, Wright NJ, Martin NG. Replication of
effects of reading and language disorders may help to elucidate shared genetic mechanisms between RD and ASD.
15. Deng K-G, Zhao H, Zuo P-X. Association between KIAA0319 SNPs and risk of
teachers and identified loci upweighted in the NM/AG hypothesis. Future studies involving larger GWAS samples ascertainment through reading and language disorders may help to elucidate shared genetic mechanisms between RD and ASD.
16. Zou L, Chen W, Shao S, Sun Z, Zhong R, Shi J, et al. Genetic variant in KIAA0319, but not in DYX1C1, is associated with risk of dyslexia: an integrated meta-analysis. Am J Med Genet Part B, Neuropsychiatr Genet. 2012;159B:970–8.
17. Shao S, Niu Y, Zhang X, Kong R, Wang J, Liu L, et al. Opposite Associations between Individual KIAA0319 Polymorphisms and Developmental Dyslexia Risk across Populations: A Stratified Meta-Analysis by the Study Population. Sci Rep. 2016;6:30454.
18. Zhong R, Yang B, Tang H, Zou L, Song R, Zhu LQ, et al. Meta-analysis of the association between DCDC2 polymorphisms and risk of dyslexia. Mol Neurobiol. 2013;47:435–42.
19. Muller B, Wilcke A, Czepezauer I, Ahnert P, Boltze J, Kirsten H, et al. Association, characterisation and meta-analysis of SNPs linked to general reading ability in a German dyslexia case-control cohort. Sci Rep. 2016;6:27901.
20. Schumacher J, Hoffmann P, Schmal C, Schulte-Korne G, Nothen MM. Genetics of
dyslexia: the evolving landscape. J Med Genet. 2007;44:289–97.
21. Bellini G, Bravaccio C, Calamoneri F, Donatalia Cocuzza M, Fillion P, Gagliano A, et al. No evidence for association between dyslexia and DYX1C1 functional
variants in a group of children and adolescents from Southern Italy. J Mol Neu-
rosci. 2005;27:311–4.
22. Scerri TS, Macpherson E, Martellini A, Wa WC, Monaco AP, Stein J, et al. The
effect of reading and language disorders may help to elucidate shared genetic mechanisms between RD and ASD.
23. Guidi LG, Velasys-Baeza A, Martinez-Garay I, Monaco AP, Paracchini S, Bishop DVM, et al. The neuronal migration hypothesis of dyslexia: A critical evaluation 30 years on. Eur J Neurosci. 2018;48:3212–33.
24. Galaburda AM, Sherman GF, Rosen GD, Abotiz F, Geschwind N. Developmental
dyslexia: four consecutive patients with cortical anomalies. Ann Neurol. 1985;18:222–33.
25. Humphreys P, Kaufmann WE, Galaburda AM. Developmental dyslexia in women:
neuropsychological findings in three patients. Ann Neurol. 1990;28:727–38.
26. Ramus F. Neurobiology of dyslexia: a reinterpretation of the data. Trends Neu-
rosci. 2004;27:720–6.
27. Galaburda AM, LoTurco J, Ramus F, Fitch RH, Rosen GD. From genes to behavior in development dyslexia. Nat Neurosci. 2006;9:1213–7.
28. Taijade M, Kaminen N, Nopola-Hemmi J, Halta M, Myllyluoma B, Lytinen H, et al. A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. Proc Natl Acad Sci USA. 2003;100:11553–8.
29. Raskind WH, Peter B, Richards T, Eckert MM, Berninger VW. The genetics of reading
disabilities: from phenotypes to candidate genes. Front Psychol. 2012;3:601.
30. Poelmans G, Buitelaar JK, Paalens DL, Frank B. A theoretical molecular network for
dyslexia: integrating available genetic findings. Mol psychiatry. 2011;16:365–82.
31. Velasys-Baeza A, Toma C, Paracchini S, Monaco AP. The dyslexia-associated gene
KIAA0319 encodes highly N- and O-glycosylated plasma membrane and secreted isoforms. Hum Mol Genet. 2008;17:859–71.
32. Kidd T, Brose K, Mitchell R, Fetter RD, Tessier-Lavigne M, Goodman CS, et al. Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors. Cell 1992;98:205–15.
33. Meng H, Smith GD, Hager K, Hu M, Liu J, Olsson R, et al. DCDC2 is associated
effectively understanding dyslexia. Translational Psychiatry (2022) 12:495
65. Sun L, Craiu RV, Paterson AD, Bull SB. Strati
66. Tran C, Wigg KG, Zhang K, Cate-Carter TD, Kerr E, Field LL, et al. Association of the
63. Stoner R, Chow ML, Boyle MP, Sunkin SM, Mouton PR, Roy S, et al. Patches of
57. Eicher JD, Gruen JR. Language impairment and dyslexia genes in
56. Price KM, Wigg KG, Misener VL, Clarke A, Yeung N, Blokland K, et al. Language
54. Martin-Vilchez S, Whitmore L, Asmussen H, Zareno J, Horwitz R, Newell-Litwa K.
51. Gialluisi A, Andlauer TFM, Mirza-Schreiber N, Moll K, Becker J, Hoffmann P, et al.
46. Rossi A, Kontarakis Z, Gerri C, Nolte H, Holper S, Kruger M, et al. Genetic com-
40. Slotkin TA, Egan GF, Parnell C, Yolken RH. The cholinergic hypothesis of autism
39. Paton JY, Parnell C, Yolken RH, Grant S, Rosier M, Parnell L, et al. Cholinergic hypoth-
38. Ramírez-Falostas DJ, milkman LA, Schueren B, et al. Association of a new
37. Nishiyama S, Izumi K, Imai T, et al. Association of a single nucleotide polymor-
36. Gazzaniga MS, Galaburda AM. Split brain. In: Geschwind N, ed. The Cerebral Cor-
35. Gross WD, Plomin R, et al. Association of childhood depression and introversion with
34. R Development Core Team. R: A language and environment for statistical comput-
33. Behar DM, Zegura WE, Hamshere ML, et al.-improved prediction of the
32. Furlong WW, Elston RC, et al. One-year observation of the model of
31. Willer CJ, Li Y, Abecasis GR. METAL: fast and ef-
30. Antczak-Marach D, et al. Comprehensive genomic analysis of patients with dis-
29. Willer CJ, Li Y, Abecasis GR. METAL: fast and ef-
28. Yoo YJ, Bull SB, Paterson AD, Waggott D, Sun L. Were genome-wide linkage
27. Willer CJ, Li Y, Abecasis GR. METAL: fast and ef-
26. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
25. Westerberg AH, Karlsson F, et al. The impact of genome-wide association studies on
24. Furlong WW, Elston RC, et al. One-year observation of the model of
23. Willer CJ, Li Y, Abecasis GR. METAL: fast and ef-
22. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
21. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
20. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
19. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
18. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
17. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
16. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
15. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
14. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
13. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
12. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
11. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
10. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
9. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
8. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
7. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
6. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
5. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
4. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
3. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
2. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
1. Addington DN, Kremen WS, et al. Studies of neuropsychiatric traits in a twin cohort
K.M. Price et al.

Translational Psychiatry (2022)12:495

SPRINGERNATURE
Kirsten Blokland, Milene Bonte, Dorret I. Boomsma, Thomas Bourgeron, Daniel Brandeis, Manuel Carreiras, Angela Martinelli, Urs Maurer, Jacob J. Michaelson, Nazanin Mirza-Schreiber, Kristina Moll, Anthony P. Monaco, Angela T. Morgan, Bertram Müller-Myhsok, Dianne F. Newbury, Markus M. Nöthen, Richard K. Olson, Karin Landerl, Gabriel T. Leonard, Zhijie Liao, Maureen W. Lovett, Heikki Lyytinen, Nicholas G. Martin, Alessandro Gialluisi, Scott D. Gordon, Jeffrey R. Gruen, Sharon L. Guger, Marianna E. Hayiou-Thomas, Juan Hernández-H.T., is supported by the Netherlands Organization for Scientific Research (NWO) and ANR-11-BSV4-014-01 and European Commission Grant LSHM-CT-2005-018696. H.T. is supported by the Netherlands Organization for Scientific Research (NWO) and Netherlands Organisation for Health and Research and Development (ZonMW) Grant VICI 016.VICI170.200. J.C.D. was supported by NICHD Grant P50 HD 27802. J.M.R., J.B.T., and T.K. were supported by NICHD Grant R01 DC014489. K.M.P. was supported by the Hospital for Sick Children Research Program (Brampton). K.K. is supported by a Sir Henry Wellcome Postdoctoral Fellowship (213514/1/Z/18/). M.J.S. is supported by Wellcome Trust Grant WT082032MA. S.P. and F.A. are supported by Royal Society Grants UF150663 and RGF/EA/180141. T.B. is supported by Institut Pasteur, the Bettencourt-Schueller Foundation, and Université de Paris. The Adolescent Brain Cognitive Development Study is supported by the NIH and additional federal partners (NIGMS Grants U01DA014048, U01DA050989, U01DA051016, U01DA051022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041166, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041105, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, and U24DA041147). The Aston Cohort was supported by funding from European Union (EU) Horizon 2020 Programme 641652 and Waterloo Foundation Grant 797/17290. The St. Andrews Bioinformatics Unit is funded by Wellcome Trust Grant 105621/Z/14/Z and 204821/Z/16/2. ALSPAC is supported by UK Medical Research Council and Wellcome Grant 217065/Z/19/Z and the University of Bristol. A comprehensive list of grant funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf). The Basque Center on Cognition, Brain and Language (BCBL) cohort was supported by the Basque Government through the Basic Excellence Research Centre program and the Agencia Estatal de Investigación through BCBL Severo Ochoa excellence accreditation. The Brisbane Adolescent Twin Sample was supported by Australian Research Council Grants A7960034, A79906588, A79801419, DP0221016, and DP0343921, with genotyping funded by the NHMRC Grant 389891. The Colorado Learning Disabilities Research Center cohort was supported by NICHD Grant P50 HD 27802. The Early Language in Victoria Study was supported by NHMRC Grant 369595. The Familial Influences on Literacy Abilities cohort is supported by the University of Amsterdam, the Max Planck Institute Nijmegen, and NWO Grants Rubicon 446-12-005 and VENI 451-15-017. The Iowa study was funded by DC00496 and DC027476 from the National Institute of Deafness and Other Communication Disorders (NIDCD). The GiDaSt study was supported by the Manton Foundation, NHMRC Grants P50-HD070922 and K99-HD094902, and the Lambert Family. NeuroDys was funded by an EU Sixth Framework Program grant to the NeuroDys Consortium, Swiss National Science Foundation Grant 32-108130, and Austrian Science Fund Grant 18351-002. The Netherlands Twin Register is funded by NWO Grants 480-04-004, 481-08-011, 056-32-010, 024.001.003, 480-15-001/674, 184.021.007, 184.033.111, and 56-464-14192; ZonMW Grants 911-09-032 and 912-10-020; the Amsterdam Public Health and Amsterdam Reproduction and Development Research Institutes; European Scientific Council Grant ERC Advanced 230374; EU Seventh Framework Programm (FP7) Grant 260665; and National Institute of Mental Health (NIMH) Grants U24 MH068457-06, R01 MH58799-03, and 1RC2 MH089995; and the Avera Institute for Human Genetics. The Pediatric Imaging, Neurocognition, and Genetics cohort is funded by NHMRC Grant RC2DA029475, the National Institute on Drug Abuse, and the Eunice Kennedy Shriver NICHD. The Philadelphia Neurodevelopmental Cohort is funded by NHMRC Grants RC2MH089883 and RC2MH089824, an institutional development award to the Center for Applied Genomics from The Children’s Hospital of Philadelphia, and a donation from Adelle and Daniel Kubert and thanks the NIH data repository. The Raine study was supported by long-term funding from NHMRC Grants 572613, 403981, 1059711, 6344445, 634509, and 1021105 and Canadian Institutes of Health Research (CIHR) Grant MOP-82893. Funding was also provided by the University of Western Australia, Curtin University, the Women and Infants Research Foundation, the Telethon Kids Institute, Edith Cowan University, Murdoch University, the University of Notre Dame Australia, and the Raine Medical Research Foundation. The Raine study analyses were supported by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia. The Saguenay Youth Study is supported by the CIHR, the Heart and Stroke Foundation of Quebec, and the Canadian Foundation for Innovation. The SLI Consortium was funded by Wellcome Trust Grant 076566 and UK Medical Research Council Grant G1000569. The Twins Early Development Study is supported by UK Medical Research Council Grants MR/ V012878/1 and MR/A201475/1, NHMRC AG046938, and the EU FP7 grant FP7/ 2007-2013/1St-602768. Toronto was supported by CIHR Grant MOP-133440. UK Dyslexia was supported by Wellcome Trust Grants 076566/20/Z/05 and 075491/Z/04, Waterloo Foundation Grant 797–1720, EU Grant 018869, and Royal Society Grant UF100463. The York cohort was funded by Wellcome Trust Grant 082036/B/07/2. 

### AUTHOR CONTRIBUTIONS

KMP, JJS, and CLB designed research; KMP, KGW, YF, KB, MW, ENK, SLG, and EE performed research and analyzed data; MWL, ENK, SLG, MW, KB and CLB collected the Toronto cohort; The Quantitative Trait Working Group of the GenLang Consortium collected and analysed the data on the cohorts (led by EE and SEF); KMP and CLB wrote the paper with input from all authors.

### COMPETING INTERESTS

The authors declare no competing interests.

### ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41398-022-02250-z.

**Correspondence** and requests for materials should be addressed to Cathy L. Barr.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

---

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

© The Author(s) 2022

---

**QUANTITATIVE TRAIT WORKING GROUP OF THE GENLANG CONSORTIUM**

Filippo Abbondanza, Andrea G. Allegrini, Till F. M. Andlauer, Cathy L. Barr, Kenneth Blokland, Milene Bonte, Dorret I. Boomsma, Thomas Bourgeron, Daniel Brandeis, Manuel Carreiras, Fabiola Ceroni, Valéria Céspe, Philip S. Dale, John C. DeFries, Peter F. de Jong, Jean François Démonet, Eveline L. de Zeeuw, Elise Eising, Yulong Feng, Simon E. Fisher, Marie-Christine J. Franken, Clyde Franks, Marjorie G. Gerrits, Alessandro Gialluisi, Scott D. Gordon, Jeffrey R. Gruen, Sharon L. Guger, Marianna E. Hayiou-Thomas, Juan Hernández-Cabrera, Jouke-Jan Hottenga, Charles Hulme, Philip R. Jansen, Juha Kere, Elizabeth N. Kerr, Tanner Koomar, Karin Landerl, Gabriel T. Leonard, Zhijie Liao, Maureen W. Lovett, Michelle Luciano, Heikki Lyytinen, Nicholas G. Martin, Angela Martellini, Urs Maurer, Jacob J. Michaelson, Nazanin Mirza-Schreiber, Kristina Moll, Anthony P. Monaco, Angela T. Morgan, Bertram Müller-Mythso, Dianne F. Newbury, Markus M. Nöthen, Richard K. Olson.
Silvia Paracchini10, Tomas Pau16,67,68,9 Zdenka Pau30,15,70 Craig E. Pennell1,7,2,73 Bruce F. Pennington74, Robert J. Plomin11, Kaitlyn M. Price1,2,3, Franck Ramus25, Sheena Reilly62,76, Louis Richer77, Kaili Rimfeld11,78, Gerd Schulte-Körne60, Chin Yang Shapland79,80, Nuala H. Simpson81, Shelley D. Smith82, Margaret J. Snowling81,83, Beate St Pourcain4,7,8, John F. Stein84, Lisa J. Strug8,9, Joel B. Talcott85, Henning Tiemeier47,86, J. Bruce Tomblin87, Dongnhu T. Truong43, Elsje van Bergen17,18,88, Marc P. van der Schroeff89, Marjolein Van Donkelaar4, Ellen Verhoef3, Kate E. Watkins81, Andrew J. O. Whitehouse86, Karen G. Wigg87,43,5, Margaret Wilkinson2, Margaret J. Wright41 and Gu Zhu20

10School of Medicine, University of St Andrews, KY16 9TF St Andrews, Scotland, United Kingdom. 11Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, UK. 12Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany. 13Department of Psychology, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany. 14Department of Psychology, University of Edinburgh, Edinburgh EH9 3JZ, UK. 15Department of Physiology, University of Toronto, Toronto, Ontario M5S 1A1, Canada. 16Department of Cognitive Neuroscience and Maastricht Brain Imaging Center, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands. 17Netherlands Twin Register, Amsterdam, The Netherlands. 18Department of Biological Psychology, Vrije Universiteit Amsterdam, 1081 BT Amsterdam, The Netherlands. 19Amsterdam Reproduction and Development (AR&D) Research Institute, Amsterdam, UMC, The Netherlands. 20Human Genetics and Cognitive Function Unit, Institut Pasteur, Paris, France. 21Centre National Recherche Scientifique (CNRS) UMR 3571, Université de Paris, Paris, France. 22Department of Child and Adolescent Psychiatry and Psychotherapy, Psychiatric University Hospital Zurich, University of Zurich, 8032 Zurich, Switzerland. 23Zurich Center for Integrative Human Physiology, University of Zurich and ETH Zurich, 8057 Zurich, Switzerland. 24Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, 68159 Mannheim, Germany. 25Basque Center on Cognition, Brain and Language (BCBL), Donostia-San Sebastian, Gipuzkoa, Spain. 26IKERBASQUE, Basque Foundation for Science, Bilbao, Vizcaya, Spain. 27Lengua Vasca y Comunicación, University of the Basque Country (UPV/EHU), Bilbao, Vizcaya, Spain. 28Department of Pharmacy and Biotechnology, University of Bologna, 40126 Bologna, Italy. 29Faculty of Health and Life Sciences, Oxford Brookes University, Oxford OX3 0BP, UK. 30Brain Imaging Centre, Research Centre for Natural Sciences, Budapest 1117, Hungary. 31Multilingualism Doctoral School, Faculty of Modern Philology and Social Sciences, University of Pannonia, Veszprém 8200, Hungary. 32Department of Speech & Hearing Sciences, University of New Mexico, Albuquerque, NM 87131, USA. 33Institute for Behavioral Genetics, University of Colorado, Boulder, CO 80309-0447, USA. 34Department of Psychology and Neuroscience, University of Colorado, Boulder, CO 80309-0447, USA. 35Department of Child Development and Education, University of Amsterdam, 1012 WX Amsterdam, The Netherlands. 36Leenaards Memory Centre, Department of Clinical Neurosciences, Lausanne University Hospital (CHUV), University of Lausanne, CH-1011 Lausanne, Switzerland. 37Department of Otolaryngology, Head and Neck Surgery, Erasmus University Medical Center, 3015 GD Rotterdam, The Netherlands. 38Department of Human Genetics, Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands. 39Department of Epidemiology and Prevention, IRCCS Istituto Neurologico Mediterraneo Neuromed, Pozzilli, Italy. 40EPIMED Research Center, Department of Medicine and Surgery, University of Insurbia, Varese, Italy. 41Genetic Epidemiology Laboratory, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia. 42Departments of Pediatrics and Genetics, Yale Medical School, New Haven, CT 06510, USA. 43Department of Psychology, University of York, York YO10 5DD, UK. 44Department of Informatics, Medizinische Universität Innsbruck, 6020 Innsbruck, Austria. 45Department of Psychology, University of York, York YO10 5DD, UK. 46Department of Psychology, University of Oxford, Oxford OX2 6PY, UK. 47Department of Child and Adolescent Psychiatry/Psychology, Erasmus University Medical Center, 3000 CB Rotterdam, The Netherlands. 48Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam, Amsterdam, The Netherlands. 49Department of Human Genetics, VU Medical Center, Amsterdam University Medical Center, 1081 BT Amsterdam, The Netherlands. 50Department of Biosciences and Nutrition, Karolinska Institutet, 171 77 Stockholm, Sweden. 51Stem Cells and Metabolism Research Program, University of Helsinki and Folkhälsoan Research Center, 00180 Helsinki, Finland. 52Department of Psychiatry, University of Iowa, Iowa City, IA 52242, USA. 53Institute of Psychology, Karl-Franzens-Universität Graz, 8010 Graz, Austria. 54BioTechMed-Graz, Graz, Austria. 55Cognitive Neuroscience Neurology and Neurosurgery, McGill University, Montreal, QC H3A 1G1, Canada. 56Department of Psychology, University of Toronto, Toronto, ON, Canada. 57Department of Psychology, University of Jyväskylä, 40014 Jyväskylä, Finland. 58Department of Psychology, The Chinese University of Hong Kong, Hong Kong, China. 59Institute of Neurogenomics, Helmholtz Zentrum München, Munich, Germany. 60Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, Ludwig-Maximilians-Institute University Hospital Munich, Munich 80336, Germany. 61Office of the President, Tufts University, Medford, MA 02155, USA. 62Speech and Language, Murdoch Children’s Research Institute, Melbourne, VIC 3052, Australia. 63Department of Audiology and Speech Pathology, University of Melbourne, Melbourne, VIC 3052, Australia. 64Speech Pathology Department, Royal Children’s Hospital, Melbourne, VIC 3052, Australia. 65Department of Health Science, University of Liverpool, Liverpool L69 7ZX, UK. 66Institute of Human Genetics, University Hospital of Bonn, 53127 Bonn, Germany. 67Department of Psychiatry and Neurology and Centre Hospitalier Universitaire Sainte Justine, University of Montreal, Montreal, QC H3T 1J4, Canada. 68Department of Psychology, University of Toronto, Toronto, ON, Canada. 69Department of Psychology, University of Toronto, Toronto, ON, Canada. 70Translational Medicine, Hospital for Sick Children, Toronto, ON, Canada. 71School of Medicine and Public Health, College of Health, Medicine and Wellbeing, University of Newcastle, Newcastle, NSW 2308, Australia. 72Mothers and Babies Research Program, Hunter Medical Research Institute, Newcastle, NSW 2305, Australia. 73Maternity and Gynaecology, John Hunter Hospital, Newcastle, New South Wales, Australia. 74Department of Psychology, University of Denver, Denver, CO, USA. 75Laboratoire de Sciences Cognitives et Psycholinguistiques, École Normale Supérieure, Paris Sciences & Lettres University, École des Hautes Études en Sciences Sociales (EHESS), Centre National de la Recherche Scientifique (CNRS), Paris 75005, France. 76The Health Group, Griffith University, Gold Coast, Queensland, Australia. 77Department of Health Sciences, Université du Québec à Chicoutimi, Chicoutimi, QC, Canada. 78Department of Psychology, Royal Holloway, University of London, Egham TW20 0EY, UK. 79MRC Integrative Epidemiology Unit, University of Bristol, Bristol BS8 2BN, UK. 80Population Health Sciences, University of Bristol, Bristol BS8 2PS, UK. 81Department of Experimental Psychology, University of Oxford, Oxford OX2 6GG, UK. 82Department of Neurological Sciences, College of Medicine, University of Nebraska Medical Center, Omaha, NE 68198, USA. 83St John’s College, University of Oxford, Oxford OX1 3JP, UK. 84Department of Physiology, Anatomy and Genetics, Oxford University, Oxford OX1 3PT, UK. 85Institute of Health and Neurodevelopment, Aston University, Birmingham B4 7ET, UK. 86T.H. Chan School of Public Health, Harvard, Boston, MA 02115, USA. 87Communication Sciences and Disorders, University of Iowa, Iowa City, IA, USA. 88Research Institute LEARN, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. 89Generation R Study Group, Erasmus MC, Rotterdam, The Netherlands. 90Telethon Kids Institute, The University of Western Australia, Perth, WA 6009, Australia. 91Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia.

Translational Psychiatry (2022) 12:495

SPRINGER NATURE