Targeted analysis of dyslexia-associated regions on chromosomes 6, 12 and 15 in large multigenerational cohorts

Nicola H. Chapman, PhD[1], Patrick Navas, PhD[1], Michael O. Dorschner, PhD[1], Michele Mehaffey, PhD[2#], Karen G. Wigg, BSc[3], Kaitlyn M. Price, PhD[3,4,5], Oxana Y. Naumova, PhD[6], Elizabeth N. Kerr[4,5], Sharon L. Guger[4], Maureen W. Lovett[3,5], Elena L. Grigorenko, PhD[6,7], Virginia Berninger, PhD[8##], Cathy L. Barr, PhD[3,9-12], Ellen M. Wijsman, PhD[1,13,##*], Wendy H. Raskind, MD, PhD[##1,14,*]

1. Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, Washington, USA
2. Department of Pediatrics, Division of Pediatric Genetics, Pediatric Genetics, University of Washington, Seattle, Washington, USA
3. Program in Neuroscience and Mental Health, Hospital for Sick Children, Toronto, Ontario, Canada
4. Department of Psychology, Hospital for Sick Children, Toronto, Ontario, Canada
5. Department of Pediatrics, University of Toronto, Toronto, Ontario, Canada
6. Department of Psychology, University of Houston, Houston, Texas, USA
7. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, USA
8. Department of Educational Psychology, University of Washington, Seattle, WA, USA
9. Division of Experimental and Translational Neuroscience, Krembil Research Institute, University Health Network, Toronto, Ontario, Canada
10. Department of Physiology, University of Toronto, Toronto, Ontario, Canada
11. Institute of Medical Sciences, Toronto, Ontario, Canada
12. Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada
13. Department of Biostatistics, University of Washington, Seattle, Washington, USA
14. Department of Psychiatry & Behavioral Sciences, University of Washington, Seattle, WA, USA

#Current address: Pacific Northwest Research Institute. 720 Broadway, Seattle, WA 98122.
##Professor Emerita
* Corresponding authors: Wendy Raskind: wendyrun@uw.edu
Ellen Wijsman: wijsman@uw.edu

ABSTRACT
Dyslexia is a common specific learning disability with a strong genetic basis that affects word reading and spelling. An increasing list of loci and genes have been implicated, but analyses to-date investigated only limited genomic variation within each locus with no confirmed pathogenic variants. In a collection of >2000 participants in families enrolled at three independent sites, we performed targeted capture and comprehensive sequencing of all exons and some regulatory elements of five candidate dyslexia risk genes (*DNAAF4*, *CYP19A1*, *DCDC2*, *KIAA0319* and *GRIN2B*) for which prior evidence of association exists from more than one sample. For each of six dyslexia-related phenotypes we used both individual-single nucleotide polymorphism (SNP) and aggregate testing of multiple SNPs to evaluate evidence for association. We detected no promoter alterations and few potentially deleterious variants in the coding exons, none of which showed evidence of association with any phenotype. All genes except *DNAAF4* provided evidence of association, corrected for the number of genes, for multiple non-coding variants with one or more phenotypes. Results for a variant in the downstream region of *CYP19A1* and a haplotype in *DCDC2* yielded particularly strong statistical significance for association. This haplotype and another in *DCDC2* affected performance of real word reading in opposite directions. In *KIAA0319*, two missense variants annotated as tolerated/benign associated with poor performance on spelling. Ten non-coding SNPs likely affect transcription factor binding. Findings were similar regardless of whether phenotypes were adjusted for verbal IQ. Our findings from this large-scale sequencing study complement those from genome-wide association studies (GWAS), argue strongly against the causative involvement of large-effect coding variants in these five candidate genes, support an oligogenic etiology, and suggest a role of transcriptional regulation.

Author Summary
Family studies show that genes play a role in dyslexia and a small number of genomic regions have been implicated to date. However, it has proven difficult to identify the specific genetic variants in those regions that affect reading ability by using indirect measures of association with evenly spaced polymorphisms chosen without regard to likely function. Here, we use recent advances in DNA sequencing to examine more comprehensively the role of genetic variants in five previously nominated candidate dyslexia risk genes on several dyslexia-related traits. Our analysis of more than 2000 participants in families with dyslexia provides strong evidence for a contribution to dyslexia risk for the non-protein coding genetic variant rs9930506 in the *CYP19A1* gene on chromosome 15 and excludes the *DNAAF4* gene on the same chromosome. We identified other putative causal variants in genes *DCDC2* and *KIAA0319* on chromosome 6 and *GRIN2B* on chromosome 12. Further studies of these DNA variants, all of which were non-coding, may point to new biological pathways that affect susceptibility to dyslexia. These findings are important because they implicate regulatory variation in this complex trait that affects ability of individuals to effectively participate in our increasingly informatic world.

INTRODUCTION

Dyslexia is a complex disorder of neurobiological origin that can be defined as unexpectedly low accuracy and/or rate of oral reading of single words or pronounceable pseudowords, or low accuracy of spelling [1]. It manifests as difficulty in learning to read and spell despite adequate instruction and is not attributable to general cognitive impairment, primary sensory or motor impairment, psychiatric or other neurologic disorder or delays in aural or oral language. The stated prevalence of dyslexia varies depending on ascertainment schemes, exclusion criteria, tests included in diagnostic assessment, and thresholds used for a categorical diagnosis. In school-aged children, most estimates of dyslexia fall between 5-12% [2-5] but have been as low as 3.5% [6] and as high as 20% [7]. In almost all past studies, including our own, males are at greater risk than females for both presence of dyslexia and its severity [2, 5, 8-10]. Even with educational intervention, many aspects of dyslexia can persist into adulthood, including slow reading speed and poor spelling-related writing abilities [11-13] leaving lasting impacts on self-esteem, educational opportunities, and occupational choices [14-16].

Multiple lines of evidence, including twin [17], familial aggregation [8, 18], adoption [19], and linkage and/or association studies [20], have led to the consensus that there is a substantial genetic contribution to dyslexia and its component phenotypes. Heritability estimates are as high as 50-70% [21, 22]. Although rare families have been described in which dyslexia appears to be transmitted as a single gene disorder [23-28] studies in the general population show that, like most complex traits, dyslexia and its correlated underlying processes are genetically heterogeneous and likely involve the influence of variation in multiple genes [29]. Such heterogeneity complicates identification of underlying genes, regardless of the study design, but multiple candidate susceptibility genes have been nominated from genomic regions of interest (ROIs) identified by linkage analyses [30, 31], genome wide association studies, GWAS, [32-34], copy number scan [35], structural chromosome rearrangements [36, 37], or whole genome sequencing [27]

Of the multiple reported ROIs for dyslexia risk, a small number have the strongest support. DYX1 on chromosome 15q [38-44] and DYX2 on chromosome 6p [39, 45-48] are the loci with the greatest replication across samples. The University of Washington (UW) and Hospital for Sick Children (SickKids, SK) samples have implicated several additional ROIs, with one of the most convincing on chromosome 12q [49]. Candidate genes that have been most studied in the ROIs on chromosomes 15q and 6p include Dynein Axonemal Assembly Factor 4 (*DNAAF4*, previously named *DYX1C1*/*EKN1*) on chromosome 15q [50-52], and Double Cortin Domain Containing 2 (*DCDC2*) and *KIAA0319* on chromosome 6p [53-61]. Support for involvement of these and other candidate risk genes has been reported from a variety of association and linkage studies in humans, and functional studies of brain development in rodents, but evidence is inconsistent [62-68]. There have been failures to detect linkage [69-71] or association [72-78], reports of association to opposite alleles [51, 52, 72], and demonstration of functional competence of the putative risk allele [79]. Meta-analyses have not resolved these conflicts [76, 80-82]. Nor have GWAS, which have provided at most weak support for the loci [34, 83-88]. This includes a recent large GWAS that did not detect significant evidence of association with any reported candidate dyslexia risk gene [32]. Variability in conclusions across the different study designs is not surprising: genome-wide linkage analyses and GWAS both allow location scans, but linkage analyses are most sensitive to relatively rare variation of large effect and

GWAS are most sensitive to common, smaller-effect variation. Neither approach queries all the DNA variation, which requires more-expensive DNA sequencing of (at least) the regions of interest.

The putative effect of candidate genes on neuronal migration has been used to bolster their credibility [89, 90], given early reports of cortical brain abnormalities in people who were thought to have had dyslexia [91, 92]. However, although cortical abnormalities have been observed with knockdown of the rat orthologues of *Dnaaf4* [79], *Kiaa0319* [62], or *Dcdc2* [93], this is not observed in knockout mice [94-96], and the cortical migration hypothesis remains unproven [65]. Observations that dyslexia candidate genes seem to have a role in ciliogenesis [97, 98], synaptic transmission [99], or axonal growth [100], have led to alternative hypotheses of pathogenesis.

Although such issues are to be expected in a complex disorder, to date, no causative pathogenic variants have been confirmed. Some possible explanations for this failure include: (1) genetic and/or phenotypic heterogeneity that masks detection in samples ascertained and phenotyped with different criteria; (2) the risk element(s) may alter expression of the protein but not its amino acid sequence; (3) the risk element may escape recognition, but affect splicing; and (4) the number of samples sequenced comprehensively has been too small to have the power to detect variants of modest effect size.

Recent advancements in DNA sequencing methods now enable the larger scale sequencing efforts that are necessary to evaluate genetic variation in ROIs more comprehensively than was possible earlier. This technology allows us, in this current multi-site study, to investigate the potential role of variants of smaller effect size, non-coding variants, and sample heterogeneity as possible explanations for previous variable results in ROIs implicated in dyslexia. To search for variants that show evidence of association with dyslexia phenotypes, we report the results of genomic sequencing and association analyses in a collection of >2000 participants in families with dyslexia and shared phenotypic measures enrolled at three institutions. We carried out a comprehensive analysis of the coding region and some regulatory element motifs of five putative dyslexia risk genes and found additional evidence for a role in dyslexia risk for *DCDC2*, *KIAA0319*, *GRIN2B* and *CYP19A1*, but not for *DNAAF4*.

METHODS

Participants and Phenotypes Across Sites

Sample and phenotype selection strategy. We selected participants enrolled in studies of dyslexia at three institutions that had overlapping phenotype batteries: University of Washington (UW), The Hospital for Sick Children (SickKids; SK), and the University of Houston (UH). Phenotypes were measured by standardized normed tests administered by more than one site. This strategy provided a large total number of participants screened with an essentially equivalent test-battery allowing for joint analysis across the three cohorts while minimizing introduction of excess phenotypic heterogeneity. The sample is largely separate from other cohorts analyzed for association of dyslexia with the ROIs investigated in the current study, thus providing an independent evaluation. Under the assumption that genetic heterogeneity in dyslexia may be reflected in phenotypic heterogeneity, the focus was on individual subtests or index scores based on multiple subtests in standardized, nationally normed tests that are predictive of reading or

spelling outcomes or that assess processes related to reading and spelling achievement such as verbal reasoning or phonological memory. To maximize sample size, only test measures administered by more than one of the three cohort sites described below were evaluated. For all cohorts, individuals with evidence of intellectual disabilities, neurological or severe psychiatric disorder, or known genetic disorder associated with language impairment were excluded.

University of Washington (UW) Cohort

Recruitment and evaluation of probands and their multigenerational family members were done under a protocol approved by the Institutional Review Board of the UW and are described elsewhere in detail [8, 11] and briefly summarized in the Supporting Information document. The full data set consists of 2079 individuals in 284 families. Of the available subjects, 1347 individuals from 278 families provided quality DNA samples for sequencing. Phenotypic data were available for 96.8% of the 1333 samples that passed quality control (QC) testing. Family sizes ranged from 3 to 51 individuals, with a median family size of 12 in 2-4 generations. Self-reported ethnicities were as follows: non-Hispanic White (90%), Asian (2.1%), Native American (2.0%), African American (1.1%), Hispanic (0.8%), and Pacific Islander (0.2%).

The SickKids (SK) Cohort

Details of the ascertainment, assessment, and inclusion/exclusion criteria have been described previously [101, 102] and a brief description is provided in the Supportive Information document. Written informed consent and/or assent was obtained from all participants under protocols approved by the Hospital for Sick Children and University Health Network Research Ethics Boards. The SK cohort comprised 816 participants from 245 families (155 with the proband and one or both parents and 90 with two or more children and one or both parents). Phenotypes were available for 99% of children (349 of 351), and 85% of parents (394 of 465). The phenotype battery in parents was limited to TOWRE measures of single word reading efficiency and phonological decoding efficiency. Self-reported ancestry was available for 185 (76%) of the families. Of these, 180 (73.5%) reported European or European-Canadian ancestry. The remaining families reported small amounts of indigenous ancestry (3 families, 1.6%), African ancestry (1 individual) and Mexican ancestry (1 individual).

The University of Houston (UH) Cohort

The subset of families of probands with dyslexia used here is a component of the ongoing collection supported by the Florida Learning Disabilities Research Center (FLDRC). Written informed consent was obtained from all participants under protocols approved by the Research Ethics Boards of Yale University or the University of Houston. The probands were children aged 8–17 who had problems reading and were native English speakers. They were registered with FLDRC and recruited through their registration. The probands and their siblings in the same age range were administered a battery of tests including IQ and reading abilities. Designation of affected status required the Wechsler Abbreviated Scale of Intelligence, WASI [103], score ≥80 and a score at least 1.5 SD below the mean on 2 of 3 measures of single real- or non-word reading, or 1 SD below the mean on all 3 indicators. Self-reported ethnicities were as follows: African American (13.5%) and Hispanic (86.5%). This cohort consisted of 49 participants (37 children and 12 parents) from 15 nuclear families. Parents provided blood samples but were not phenotyped.

Molecular Methods

Overview of approach. The targeted capture approach used here was designed to produce a comprehensive assessment of sequence variants relevant to the subset of dyslexia candidate genes evaluated. In the large, combined sample, this included all coding variants, as well as non-coding variants potentially involved in some aspects of gene regulation. For this investigation, the focus was on the potential impact on RNA processing and/or transcription-factor (TF) motifs in tissue-dependent open chromatin regions that control gene expression.

Genomic regions investigated. We targeted DNA sequence in three previously published regions of interest (ROIs) for comprehensive evaluation. Two of the ROIs have been widely investigated, with both positive linkage and association analyses reported by more than one group: the DYX1 locus on chromosome 15q and the DYX2 locus on chromosome 6p. Previous linkage analyses in the UW cohort provided support for DYX1 as a dyslexia candidate region [52, 70] and similar studies in subsets of the SK cohort supported both DYX1 and DYX2 [101, 104, 105]. From each of these ROIs, we selected for study two genes based on prior publications supporting a role in dyslexia: in DYX1, *DNAAF*4 [40, 44, 50-52] and C*YP19A1* [37, 106, 107]; and in DYX2, *KIAA0319* [54, 62, 77, 108, 109] and *DCDC2* [81, 110, 111]. Intron 2 of *DCDC2* [93] contains READ1 – a complex compound STR polymorphism that was previously proposed as the functional dyslexia-risk component in the DYX2 region [112-114] and was included as part of our capture-design. We also selected the glutamate ionotropic receptor NMDA type subunit 2B *(GRIN2B)*, in a third ROI, on chromosome 12p. This region was among those with the strongest evidence of linkage from the UW family studies, with support across several test-battery items [49, 115, 116]. *GRIN2B* has also been implicated as a dyslexia risk factor [117-119]. For analysis of these five genes, we developed a set of custom capture probes to enable comprehensive evaluation of potential regulatory and splice region sequence variants in addition to coding region variants, as described below.

Sample preparation. For most samples, genomic DNA was extracted from peripheral blood mononuclear cells or Epstein–Barr virus-transformed B-lymphoblastoid cell lines. When only saliva samples were available, DNA was extracted using a DNAGenotek OGR-500 kit (DNAGenotek Inc, Ontario, Canada) according to the manufacturer's instructions.

Single molecule molecular inversion probe (smMIP) targeted capture and sequencing. The capture target consisted of all potentially functional sequences and variants in each ROI around and including each of the five selected genes. This included a total of 277 potential regulatory regions. We used smMIPs to capture targeted DNA with methods described elsewhere [120, 121], followed by multiplex sequencing. Additional details are provided in the Supporting Information document.

We used the UW pipeline [122] to design smMIPs to capture all exons and 10-20bp of flanking intron sequences. This approach ensured capture of the splice site branch A points (RefSeq, hg19/GrCH37 build [121]). All analysis was on this genome build. For the non-coding regulatory regions, we used the ATAC-seq data in Brain Open Chromatin Atlas (BOCA) [123] and ENCODE Consortium for the brain specific (including fetal brain) DNAseI hypersensitive sites [124] to identify chromatin accessible regions (CARs) from 80 kilobases (kb) upstream of the transcription start site (TSS) to the same distance downstream of the 3′ untranslated region (3′ UTR) in the genes of interest. We placed emphasis on the BOCA data that contain CARs for

neuronal and non-neuronal cell types across 14 brain regions [125] to identify SNPs associated with neural-specific regulatory elements. The brain regions and abbreviations are described in the Supporting Information document. The smMIPs were designed to minimally overlap each ~200 bp DHS or ATAC-seq site with an additional 50 bp of flanking sequences. The resulting 1574 smMIPs and a 55 probe smMIP fingerprinting collection were pooled, tested on a set of control DNAs, and rebalanced, resulting in a final pool of 1569 smMIPs (Supporting Information - Methods and Table S1).

Multiplexed next generation sequencing. Libraries were prepared [120, 126] and pooled for sequencing in batches of 384. Each pool was sequenced using standard paired-end (100bp) rapid run chemistry in a single lane on a HiSeq 2500 (Illumina, San Diego, CA). The final batch contained repeats from previous batches. Using a quality control benchmark requiring that each sample have a minimum of 80% of target bases covered with a depth of at least 10, 2040 (92%), 2190 (99%), and 2176 (98%) of the 2209 samples prepared passed the benchmark on chromosomes 6, 12, and 15, respectively. For de-multiplexing, generation of FASTQ files, and annotation of sequence data, we used the same in-house pipeline as for MIP design (see Supporting Information). After QC steps, the sample sizes were 1333, 782 and 46 for the UW, SK and UH samples, respectively, with average genotype completion rates of 98.7%, 99.2% and 98.9%, respectively. Variants were annotated using the 1000 Genomes Project (1KGP) and Ensembl Variant Effect Predictor (VEP) [127].

Statistical and Bioinformatic Analyses

Overview. We carried out association analyses on all MIPs sequence variants and samples that passed QC. A full range of variant frequencies including some low-frequency causal risk variants were considered. Low frequency variants were specifically evaluated because each ROI investigated was first identified through linkage analysis, which is most sensitive to low frequency segregating variation.

Ancestry adjustment. Self-reported continental ancestry was available for most samples. Because almost all samples were of European origin, either by self-report or KING estimation (Supporting Information), a simple European/non-European indicator assigned by self-report was used to adjust for ancestry in all analyses as a potential nuisance covariate. SNP-based ancestry estimates conflicted with self-reported ancestry in only six individuals. SNP-based ancestry was used in these individuals because self-reported ancestry may reflect cultural affiliation rather than genetic ancestry [128].

Phenotypes and adjustments. Reading-related phenotypes used for our analyses that were directly comparable across the three data sets included word identification (WID) and word attack (WA) from the *Woodcock Reading Mastery Test* (*WRMT*; [129], single word reading efficiency (SWE) and phonological decoding efficiency (PDE) from the *Test of Word Reading Efficiency* (*TOWRE;* [130]). The UW and SK cohorts also included spelling (SP) from the *Wide Range Achievement Test – Revised* (*WRAT3-R;* [131]), and nonword repetition (NWR) from the *Comprehensive Test of Phonological Processing* [132]. For all reading-related phenotypes considered here, lower scores indicate more impairment on the measure. In addition, only the UW and SK cohorts included verbal IQ (VIQ), which was used as a covariate for some analyses. Two different phenotype adjustments were considered: (1) adjustment for age and sex only (UNADJ), and (2) adjustment for age, sex and VIQ (VIQADJ). The second set of adjustments

represents traits that can be interpreted as free from VIQ effects but could only be performed on the UW sample and children in the SK sample, because of the availability of VIQ. For brevity, when discussing results, we use the format Trait:Adjustment (*e.g.*, SWE:UNADJ and SWE:ADJ) to refer to the phenotype without or with VIQ adjustment.

Association testing. Phenotypes were adjusted for covariates within each data set using linear regression, and the residuals were used as the response variable in the association testing. Adjustments were done within data sets both because different editions of tests were used across data sets, and because there were differences in ascertainment. Association testing using GENESIS [133] was done using the combined data sets, with distinct means and variances modeled. Data set-specific variances were used to allow for differences between data sets. Family relationships were accounted for by using the expected pedigree-defined kinship in the covariance matrix via a mixed model. For all SNPs observed in two or more copies in the combined data set, we regressed the phenotype of interest against the dose of the minor allele, yielding a single-variant test. Because SNPs with minor allele frequency (MAF) > 0.01 should result in approximately 40 copies in our dataset, we chose this value as the threshold over which statistical analyses should be reliable. All SNPs, including those with lower MAF, were included in aggregate analyses, with grouping according to their location relative to each candidate gene, defined as 5′ region, 3′ region, exons, and introns. The SKAT-O [134] aggregate test was performed in GENESIS, using weights following a Beta distribution with parameters (1,25) and dependent on the MAF [134]. This choice of weight distribution more heavily weights rare variants, but still allows for a contribution from common variants. The SKAT-O test optimizes power by finding the maximal weighted average of the Burden test (more powerful when most variants are causal and effects are in the same direction) and SKAT test (more powerful when most variants are not causal and effects can be in either direction). Correcting for multiple testing is difficult in these data because there are strong correlations between the phenotypes used, and the focus is on small regions of the genome where we expect substantial linkage disequilibrium (LD). As an approximate correction, we require $p < 0.01$ for significant evidence of association, which corresponds to a p-value of 0.05 divided by 5, the number of candidate genes being considered.

Haplotype Estimation and Testing. When multiple SNPs in LD with one another achieved $p < 0.01$, we used Beagle 5.4 [135] with the 1KGP European reference population (EUR) [136] to obtain phased genotypes, thus providing pairs of phased haplotypes for each subject. For a locus with n common haplotypes, we fit n additive models for each phenotype, where the ith model estimates the dose effect of haplotype i relative to the other haplotypes pooled. GENESIS allowed us to correct for relationships by using the pedigree-defined kinship in the covariance matrix of a mixed model. When more than one haplotype in a region showed significant association in a dose-model with a phenotype, we fitted a model including dose effects for all the significant haplotypes and allowed for interaction between them.

Annotating Non-coding Variants in CARs. We explored the potential impact of all non-coding variants with MAF > 0.01 and $p < 0.01$ on at least one single-variant association test for any of the UNADJ and VIQADJ phenotypes. We annotated variants using the JASPAR tracks on the UCSC Genome Browser [137] as well as the JASPAR database [138]. We considered four characteristics of non-coding variants that together are suggestive of a regulatory effect: 1) the variant is in the peak signal (~200bp) in either ATAC-seq [125] or DNAseI-seq brain profiles

[139] in any brain region; 2) it overlaps with a known TF motif, as found in JASPAR [138]; 3) the change disrupts a conserved position in the motif, as assessed by the position frequency matrices in JASPAR; and 4) the TF whose motif is disrupted has an open promoter in the same brain region(s) as the variant [140-142].

RESULTS

Sample Characteristics

Ancestry. In the SK data set, all samples with SNP data (all 251 children) had estimated proportion of EUR ancestry greater than 95%. Therefore, the SK data set (including parents) was assumed to be 100% European in genetic background. For the 532 people with SNPs in the UW data set, 504, 15 and 2 individuals had EUR, AFR and East Asian (EAS) ancestry proportion greater than 95%, respectively. Eleven people were admixed (8 EUR/EAS and 3 EUR/AFR) and were counted as non-Europeans. Considering both self-reported ethnicity and SNP-estimated ancestry, 1208 people in the UW data set were assigned to the European category and 100 people to the non-European category. Twenty-five people categorized as unknown because data were unavailable were dropped from the analysis. The UH cohort had 5 African American individuals and 32 white Hispanic individuals as determined by self-report.

Phenotypes. In the tables and text that follow, these abbreviations are used for the tests and the processes they assess: **WID** (*WRMT-R Word Identification* for accuracy of oral reading of real words), **WA** (*WRMT-R Word Attack* for accuracy of oral reading of nonwords), **SP** (*WRAT-3* or *WRAT-R Spelling* for written spelling of orally dictated words), **SWE** (*TOWRE* speed of oral reading of real words), **PDE** (*TOWRE* speed of oral reading of nonwords), **NWR** (*CTOPP* Nonword Repetition NWR for phonological memory). Table 1 shows sample sizes for UNADJ and VIQADJ phenotypes for each dataset and the combined dataset. The variability in sample numbers included in the analyses reflects differences in the phenotyping protocols at the three institutions. VIQ scores were obtained for parents and children in the UW cohort, only for children in the SK cohort, and were not obtained for the UH cohort; therefore, the VIQADJ samples include only the UW and SK cohorts. For SWE and PDE, the UNADJ dataset is substantially larger than the VIQADJ dataset, because VIQ was not available for SK parents. Results presented here in the main text focus on the larger, UNADJ, dataset except when findings are substantially different for VIQADJ.

Table 1: Sample size by data set for phenotypes analyzed.

	Phenotype[1]											
Sample	**WID**		**WA**		**SP**		**SWE**		**PDE**		**NWR**	
	UN ADJ	**VIQ ADJ**	**UN ADJ**	**VIQ ADJ**	**UN ADJ**	**VIQ ADJ**	**UN ADJ**	**VIQ ADJ**	**UN ADJ**	**VIQ ADJ**	**UN ADJ**	**VIQ ADJ**
UW	1315*	1315*	1315*	1315*	1313*	1313*	1305*	1305*	1304*	1304*	1302*	1302*
SK	335	331	336	332	338	338	713*	332	711*	332	337	333
UH	34	0	34	0	0	0	34	0	34	0	0	0
Total	**1684**	**1646**	**1685**	**1647**	**1651**	**1651**	**2052**	**1637**	**2049**	**1636**	**1639**	**1635**

1. Word identification (WID) and word attack (WA) subtests of the *Woodcock Reading Mastery Test, WRMT* [129]; spelling (SP) subtest of the *Wide Range Achievement Test – Revised, WRAT3-R* [131]; single word reading efficiency (SWE) and phonological decoding efficiency (PDE) subtests of the *Test of Word Reading Efficiency, TOWRE* [130]; and non-word repetition (NWR) subtest of the *Comprehensive Test of Phonological Processing* [132]. Minor differences in samples size reflect occasional missing values.

* Includes scores for parents

Supplemental Tables S2 and S3 contain demographic data for the samples used in the UNADJ and VIQADJ analyses with means and standard deviations for the phenotypes in each group. The average VIQ score in the UW data set is almost a standard deviation higher than in the SK data set, as might be expected from the difference in sample selection between the two samples. This is supported by noting that the average score in the UW data set of children (109.7) is not significantly greater than that expected (106.4) from restricting enrollment to VIQ > 90 in a random sample. All the phenotypes have means around zero because they are the residuals from a linear model. The means are not exactly zero because the adjustments were done on a larger data set than only the genotyped participants. Consideration of the SD column demonstrates that the phenotypes fall into two categories: WID, WA and SP where the pre-adjustment value was a standard score, and SWE, PDE and NWR where the pre-adjustment value was a z-score. This difference is reflected in the magnitude of the effect sizes estimated for phenotypes in each category.

Association Analyses

Tables 2 and 3 summarize results for single variant testing (for variants with MAF > 0.01 only) and aggregate testing, respectively, over all five candidate regions for both the UNADJ and VIQADJ phenotypes. Below we present detailed results primarily on the UNADJ analysis, as it represents the larger data set. Summary tables for the remaining VIQADJ analyses are provided in Tables S4-S6. Complete analysis results for all individual sequence variants for both the UNADJ and VIQADJ phenotypes are also provided in Tables S7 and S8, respectively.

Table 2: Association (p<0.01) of common (MAF > 0.01) variant(s) with UNADJ and VIQADJ phenotypes by candidate gene region.

		Phenotype[1]											
Locus	**Gene**	**WID**		**WA**		**SP**		**SWE**		**PDE**		**NWR**	
		UN ADJ	**VIQ ADJ**	**UN ADJ**	**VIQ ADJ**	**UN ADJ**	**VIQ ADJ**	**UN ADJ**	**VIQ ADJ**	**UN ADJ**	**VIQ ADJ**	**UN ADJ**	**VIQ ADJ**
DYX1	*DNAAF4*	-	-	-	-	-	-	-	-	-	-	-	-
	CYP19A1	*	*	*	*	*	*	-	-	*	*	-	-
DYX2	*DCDC2*	*	*	-	*	*	*	*	*	-	-	-	-
	KIAA0319	*	*	*	*	*	-	*	*	-	*	-	-
Other	*GRIN2B*	*	-	-	-	-	-	*	-	-	*	*	*

1. Phenotypes as described in Table 1. *: at least one common variant associated at p<0.01.

Table 3: SKAT-O p-values for UNADJ, VIQADJ phenotypes, with adjustment for non-European ancestry.

			Phenotype[1]											
Locus	**Gene[2]**	**Sites**	**WID**		**WA**		**SP**		**SWE**		**PDE**		**NWR**	
			UN ADJ	**VIQ ADJ**	**UN ADJ**	**VIQ ADJ**	**UN ADJ**	**VIQ ADJ**	**UN ADJ**	**VIQ ADJ**	**UN ADJ**	**VIQ ADJ**	**UN ADJ**	**VIQ ADJ**
	DNAAF4													
	DS	5	-	-	-	-	-	-	-	-	-	-	-	-
DYX1	Exons	15	-	-	-	-	-	-	-	-	-	-	-	-
	Introns	37	-	-	-	-	-	-	-	-	-	-	-	0.045
	US	20	-	-	-	-	-	-	-	-	-	-	-	-
	CYP19A1													
	DS	31	-	-	-	-	-	-	-	-	-	-	-	-
DYX1	Exons	27	-	-	-	-	-	-	-	-	-	-	-	-
	Introns	210	-	0.027	-	-	-	-	-	-	-	-	-	-
	US	105	-	-	-	-	-	-	-	-	-	-	-	-
	DCDC2													
	DS	34	-	-	-	-	-	-	-	-	-	-	-	-
DYX2	Exons	54	-	-	-	-	-	-	-	-	-	-	-	-
	Introns	112	-	-	0.045	0.048	0.020	0.030	-	-	-	-	-	-
	US	46	**0.007**	0.011	0.019	0.036	0.011	0.023	0.046	0.015	-	-	-	-
	KIAA0319													
	DS	42	**0.003**	0.028	-	-	0.031	-	**0.003**	**0.004**	-	-	-	-
DYX2	Exons	63	0.013	-	**0.007**	0.026	**0.007**	0.036	-	-	-	-	-	-
	Introns	145	-	-	0.033	-	-	-	-	-	-	-	-	-
	US	13	-	-	-	-	-	-	-	-	-	-	-	-
	GRIN2B													
	DS	85	-	-	-	-	**0.003**	-	-	-	-	-	-	-
Other	Exons	50	-	-	-	-	-	0.018	-	-	-	-	-	-
	Introns	303	-	-	-	-	0.024	**0.008**	-	-	-	-	-	-
	US	80	-	-	-	-	-	-	-	-	-	-	-	-

1. Phenotypes as described in Table 1. **Bold** indicates region/phenotype combinations where $p < 0.01$. - indicates a p-value ≥ 0.05.
2. DS: downstream, US: upstream.

DYX1 on chromosome 15. Of the two genes investigated in DYX1, only *CYP19A1* shows evidence of contribution of common variants to the traits analyzed at a significance threshold of $p < 0.01$ whether in the larger UNADJ sample or the smaller VIQADJ sample (Tables 2, S7, and S8). Variants in *CYP19A1* provide significant evidence of association to multiple phenotypes that is robust to presence or absence of adjustment for VIQ (Table 2). In the UNADJ analysis, three SNPs each provide evidence of association with one or more phenotype (Table 4). One SNP, rs55712458, provides highly significant *p*-values for association with four of the six traits, as well as providing the strongest evidence for association across all analyses. This result was robust to phenotypic adjustment, UNADJ vs. VIQADJ (Tables 4 and S4). For three of the four phenotypes with significant evidence of association (WID, WA, and SP), the strength of the evidence is considerably stronger than required to pass our initial screen, particularly in the UNADJ analysis. These results are also stronger than thresholds used in conventional GWAS scans, after factoring in the number of genes evaluated here. A second SNP (rs12594395), located very close to rs55712458, also provides significant evidence for association with WID:UNADJ and has a nominal $p < 0.05$ for WID:VIQADJ (Table S9). Finally, the rare allele of rs141853734 is significantly associated with decreased performance on both WA:UNADJ and WA:VIQADJ.

Table 4: Common variants (MAF ≥ 0.01) in *CYP19A1*, UNADJ phenotypes. Effect size (*p*-value) for dose of the rarer allele with adjustment for non-European ancestry. Variants appear in this table if $p < 0.01$ for at least one phenotype.

rsID	Position[2]	REF/ALT (Freq)	Region[3]	Function[4]	Phenotype[1]: UNADJ					
					WID	WA	SP	SWE	PDE	NWR
rs12594395	15:51,483,707	C/G (0.384)	DS	NA	**1.45 (0.003)**	1.10 (0.014)	1.13 (0.016)	-	-	-
rs55712458	15:51,483,996	G/C (0.198)	DS	NA	**2.82 (2.7×10^{-6})**	**2.45 (1.0×10^{-5})**	**2.43 (2.0×10^{-5})**	0.10 (0.012)	**0.11 (0.002)**	-
rs141853734	15:51,627,969	C/T (0.013)	I	Regulatory	-4.38 (0.041)	**-5.36 (0.006)**	-4.12 (0.04)	-	-	-

1. Phenotypes as described in Table 1. Bold indicates SNP/phenotype combinations where $p < 0.01$. – indicates *p*-values ≥ 0.05. Freq refers to the alternate allele.
2. Build GRCh37/hg19
3. DS: downstream, I: intronic, E: exonic, US: upstream
4. VEP build38

No aggregate tests in *CYP19A1* or *DNAAF4* yields $p < 0.01$; thus, there is no support for an effect on any of the traits of rare variants in either gene (Table 3). For both genes, only the class of intronic variants provide even suggestive evidence ($p < 0.05$) of association and only for NWR in *DNAAF4* and WID in *CYP19A1*) (Table 3). These aggregate tests were sensitive to the type of phenotypic adjustment, providing, at best, marginal evidence for association only in the VIQADJ and not the UNADJ analysis. Finally, detailed exploration of exonic variants for both genes in DYX1 failed to implicate any variants, common or rare, that could explain more than a tiny fraction of the observed variation in the phenotypes (Tables S7 and S8).

DYX2 on chromosome 6. Both common and rare variants in *KIAA0319* show significant evidence for association to multiple phenotypes (Tables 2 and 3). For both common variants and the aggregate tests that weight rare variants more heavily, the phenotypes implicated are primarily WA, WID, SP and SWE, with weaker evidence obtained with PDE and NWR. Two nearly co-located missense variants in *KIAA0319*, both annotated as tolerated/benign, are strongly associated with poor performance on SP:UNADJ (Table 5). In fact, a cluster of four variants (rs2817191, rs2744550, rs2760164 and rs2760163) co-occur in a tiny region of 40 base pairs, and all four variants are strongly associated with a decrease in performance on SP:UNADJ (Table 5). In continental populations of the 1KGP, the variant alleles of these four SNPs are always found together on a single haplotype that occurs with frequency 0.8% in EUR and 15.3% in AFR [143]. The neighboring SNP rs2744549 is in very tight but not complete disequilibrium with the cluster; the slight differences in test statistics and *p*-values for rs2744549 in our data set are explained by missing genotypes in 6 individuals who were typed for the other four variants. The two missense variants (rs2817191 and rs2744550; valine to alanine and serine to glycine, respectively) affect residues that are immediate neighbors in an exon predicted to code a Fibronectin Type III domain [144].

Table 5: Common variants (MAF ≥ 0.01) in *KIAA0319*, UNADJ phenotypes. Effect size (*p*-value) for dose of the rarer allele with adjustment for non-European ancestry. Variants appear in this table if $p < 0.01$ for at least one phenotype

rsID	Position[2]	REF/ALT (Freq)[3]	Region[4]	Function[5]	Phenotype[1]: UNADJ					
					WID	WA	SP	SWE	PDE	NWR
rs114526784	6:24,519,996	T/A (0.053)	DS	NA	-	-	-	**0.22 (0.003)**	0.163 (0.016)	-
rs114979321	6:24,544,140	A/G (0.031)	DS	Regulatory	-2.96 (0.028)	-	-3.15 (0.021)	**-0.32 (0.007)**	-	-
rs2744549	6:24,563,552	C/T (0.021)	I	NA	**-4.47 (0.006)**	**-4.14 (0.005)**	**-4.78 (0.003)**	-0.24 (0.038)	-0.21 (0.046)	-0.26 (0.024)
rs2817191	6:24,564,540	A/G (0.021)	E	missense	-4.00 (0.016)	-3.80 (0.012)	**-4.25 (0.009)**	-	-	-0.23 (0.049)
rs2744550	6:24,564,544	T/C (0.021)	E	missense	-4.00 (0.016)	-3.80 (0.012)	**-4.25 (0.009)**	-	-	-0.23 (0.049)
rs2760164	6:24,564,573	G/A (0.021)	I	splice	-4.00 (0.016)	-3.80 (0.012)	**-4.25 (0.009)**	-	-	-0.23 (0.049)
rs2760163	6:24,564,579	A/G (0.021)	I	splice	-4.00 (0.016)	-3.80 (0.012)	**-4.25 (0.009)**	-	-	-0.23 (0.049)

1. Phenotypes as described in Table 1. Bold indicates SNP/phenotype combinations where $p < 0.01$. – indicates *p*-values ≥ 0.05.
2. Build GRCh37/hg19
3. Freq refers to the alternate allele.
4. DS: downstream, I: intronic, E: exonic, US: upstream
5. VEP build38

The common variants implicated in *KIAA0319* display sensitivity to covariate adjustment and, except for the two missense variants in the common variant cluster, are all in non-coding regions. Aggregate tests also implicate non-coding DNA in the downstream region of *KIAA0319*, and common SNPs with significant association with SWE:UNADJ are in the downstream region (rs114526784 and rs114979321, see Table 5). Finally, there are some variants that only show significant evidence for association with WA:VIQADJ, including rs9295627 and rs9461045 (Table S5). These two SNPs provide nominal evidence for WA:UNADJ (Table S7). SNP rs9461045, here associated with WA:VIQADJ and PDE:VIQADJ (Table S5), was previously shown to affect expression of *KIAA0319* in a different sample [62]. However, WA:UNADJ provides only nominal evidence of association (Table S7), demonstrating moderate sensitivity to the covariate adjustment for this particular SNP. All the SNPs implicated by these analyses are in the large 77kb region that has been previously shown to be associated with reading related phenotypes [53].

Association testing for SNPs in *DCDC2* (Tables 2 and 3) showed sensitivity to covariate adjustment. Table 6 shows two neighboring upstream variants (rs142310124 and rs116652616) significantly associated with SP:UNADJ and SWE:UNADJ, with suggestive evidence for association with WA:UNADJ and WID:UNADJ. In this gene, with adjustment of phenotypes for VIQ, additional intronic variants met criteria for significance for association with SWE:VIQADJ (Table 6). All these additional variants show suggestive evidence for association with SWE:UNADJ at $p < 0.05$ (Table S7). Similarities in minor allele frequencies and estimated effect sizes suggest the presence of at least three haplotype blocks that underlie the region: Block A, rs72830851- rs68182041; Block B, rs12196880 – rs7763196; and Block C, rs77743903 – rs116652616 excluding rs4599626. Block C contains the variants that are also significant with VIQ adjustment (Table 6). Examination of LD among these 9 SNPs in 1KGP EUR [143] shows strong disequilibrium ($D' > 0.94$) within blocks A, B, and C, and between blocks A and B. The same pattern is seen within 1KGP AFR [143] with the addition of strong disequilibrium between all three blocks. Due to the strong disequilibrium in each block, we cannot statistically distinguish the effects of individual variants; however, bioinformatics prioritization (described below) provides further insights.

Table 6: Common variants (MAF ≥ 0.01) in *DCDC2*, UNADJ and VIQADJ phenotypes. Effect size (*p*-value) for dose of the rarer allele with adjustment for non-European ancestry. Variants appear in this table if $p < 0.01$ for at least one phenotype.

rsID	Position[2]	REF/ALT (Freq)[3]	Region[4]	Function[5]	Phenotype[1]: UNADJ					
					WID	WA	SP	SWE	PDE	NWR
rs77743903	6:24,332,778	A/G (0.021)	I	Regulatory	**-4.74 (0.007)**	-3.26 (0.042)	-4.03 (0.018)	**-0.41 (0.001)**	-	-
rs142310124*	6:24,421,582	A/C (0.023)	US	NA	-4.15 (0.010)	-2.95 (0.042)	**-4.29 (0.008)**	**-0.30 (0.0098)**	-	-
rs116652616	6:24,421,659	G/A (0.023)	US	NA	-4.15 (0.010)	-2.95 (0.042)	**-4.29 (0.008)**	**-0.30 (0.0098)**	-	-
					Phenotype[1]: VIQADJ					
					WID	**WA**	**SP**	**SWE**	**PDE**	**NWR**
rs9379643	6:24,105,447	G/A (0.231)	DS	NA	**1.37 (0.006)**	-	-	0.11 (0.012)	-	-
rs59336318	6:24,223,585	C/T (0.043)	I	Regulatory	-	**-2.61 (0.0099)**	-	-	-	-0.15 (0.041)
rs61232212	6:24,223,611	A/T (0.043)	I	Regulatory	-	**-2.61 (0.0099)**	-	-	-	-0.15 (0.041)
rs72830851	6:24,261,168	A/C (0.105)	I	NA	-	-	-	**0.17 (0.003)**	0.12 (0.012)	-
rs72830859*	6:24,269,173	T/A (0.104)	I	NA	-	-	-	**0.16 (0.004)**	0.12 (0.012)	-
rs68182041	6:24,283,345	C/T (0.110)	I	Regulatory	-	-	-	**0.17 (0.003)**	0.11 (0.024)	-
rs12196880	6:24,313,747	C/T (0.092)	I	Regulatory	-	-	-1.51 (0.042)	**-0.16 (0.008)**	-	-
rs12198320*	6:24,313,828	G/T (0.092)	I	Regulatory	-	-	-1.50 (0.042)	**-0.16 (0.008)**	-	-
rs7763196*	6:24,314,065	A/G (0.092)	I	Regulatory	-	-	-1.48 (0.047)	**-0.16 (0.009)**	-	-
rs77743903	6:24,332,778	A/G (0.021)	I	Regulatory	-3.50 (0.018)	-	-3.23 (0.036)	**-0.46 (0.0005)**	-	-
rs4599626	6:24,341,357	C/A (0.242)	I	Regulatory	-	-	-	**0.12 (0.005)**	0.08 (0.033)	-
rs142310124*	6:24,421,582	A/C (0.023)	US	NA	**-4.65 (0.0009)**	-3.36 (0.017)	**-3.76 (0.0099)**	**-0.44 (0.0004)**	-0.22 (0.041)	-

rs116652616	6:24,421,659	G/A (0.023)	US	NA	**-4.65 (0.0009)**	-3.36 (0.017)	**-3.76 (0.0099)**	**-0.44 (0.0004)**	-0.22 (0.041)	-

1. Phenotypes as described in Table 1. Bold indicates SNP/phenotype combinations where $p < 0.01$. – indicates p-values ≥ 0.05. * Alters brain TF binding site
2. Build GRCh37/hg19
3. Freq refers to the alternate allele.
4. DS: downstream, I: intronic, E: exonic, US: upstream
5. VEP build38.

Evaluation of haplotype effects in *DCDC2* provides evidence for more than one important variant in the region. Table 7 shows the six most common haplotypes defined by the nine SNPs in blocks A, B and C, along with haplotype dosage effects as estimated in six separate linear models (Table 7, left). Neither the dosage of the most common haplotype (1: ATC-CGA-AAG) nor haplotypes 3, 5, or 6 are associated with SWE:VIQADJ, whereas haplotypes 2 (CAT-CGA-AAG) and 4 (ATC-TTG-GCA) are associated with significant effects on SWE:VIQADJ that are in opposite directions. Haplotype 2 carries the rarer of two common haplotypes in block A (**CAT** in Table 7) and is associated with an increase in SWE:VIQADJ. Conversely, the rarer of two common haplotypes in block C (**GCA** in Table 7), on haplotype 4, is associated with a decrease in SWE:VIQADJ. The right side of Table 7 shows the final model in which doses of haplotypes 2 and 4 are both included. The estimated effect sizes are very similar to those found in the individual models, and there is no evidence ($p = 0.16$, model not shown) of interaction between haplotypes 2 and 4. These results indicate that the rare haplotype in block A is associated with an increase in SWE:VIQADJ, and the rare haplotype in block C is independently associated with a decrease in SWE:VIQADJ. The region spanned by these haplotypes includes READ1 (a repeat element previously proposed as the causal variant in this region [112-114], which is located between blocks B and C. Due to poor performance of probes targeting READ1 (Figure S1), we were not able to investigate READ1 directly.

Table 7: Haplotype models in *DCDC2,* associated with SWE:VIQADJ. The left side shows effect estimates are from six individual haplotype dosage models. The rarer haplotype in each block is in **bold**. The final model on the right shows effect estimates when dose effects are modeled simultaneously.

Individual haplotype dosage models						Final model		
Haplotype/ Model	**SNP alleles**	**EUR freq.**	**Obs. freq.**	**Effect Estimate**	**p-value**	**Term**	**Effect Estimate**	**p-value**
Hap 1	ATC-CGA-AAG	0.778	0.784	-0.010	0.91	Intercept	-0.040	0.38
Hap 2	**CAT**-CGA-AAG	0.109	0.105	0.174	0.002	Non-Cauc.	-0.19	0.11
Hap 3	ATC-**TTG**-AAG	0.073	0.076	-0.071	0.29	UW group	0.022	0.70
Hap 4	ATC-**TTG-GCA**	0.011	0.014	-0.579	0.0001	Hap 2 dose	0.17	0.004
Hap 5	ATT-CGA-AAG	0.004	0.006	0.044	0.84	Hap 4 dose	-0.56	0.0002
Hap 6	ATC-CGA-ACA	0.001	0.006	-0.249	0.28			
Other			0.009					

Aggregate tests including all variants provide evidence for association in both *KIAA0319* and *DCDC2* on subsets of the traits (Table 3). Variants in *KIAA0319* show evidence for association in aggregate with either downstream or exonic variants for each of the UNADJ real-word phenotypes (p = 0.003-0.007), as well as for one VIQADJ real-word phenotype, SWE:VIQADJ (p = 0.004). Although aggregate tests do not identify the specific variants of interest, 9 of the 14 variants in the 3′ UTR exonic region of *KIAA0319* have estimated negative contribution to the trait values for both WA:UNADJ and SP:UNADJ (the two phenotypes with significant aggregate test values for exonic variants). Variants upstream of *DCDC2* show significant evidence for association with WID:UNADJ in aggregate testing (p=0.007), and nearly meet our threshold with WID:VIQADJ (p = 0.011). Fourteen of the 18 variants in the upstream gene region of *DCDC2* have negative estimated contribution to WID:UNADJ. Like results for single variant testing, evidence for association is usually stronger in the UNADJ phenotypes than the VIQADJ phenotype analyses. However, despite fewer associations that meet the nominal threshold of interest, the same combination of traits and aggregate classes show at least suggestive ($p < 0.05$) association in the VIQADJ phenotypes as in the stronger test results from the UNADJ phenotypes.

GRIN2B on chromosome 12. In *GRIN2B*, several common variants show significant, robust evidence of association, primarily with NWR. Four variants in the downstream region and an intronic variant are associated with NWR:UNADJ (Table 8). Three of the downstream variants (rs11055486, rs10772689, and rs4764003) also meet the criterion for association with NWR:VIQADJ (Table S6) and the fourth (rs7307659) gives suggestive evidence of linkage. A single intronic variant (rs61914322) is significantly associated with poorer performance on both WID:UNADJ and SWE:UNADJ, and gives suggestive evidence for WID:VIQADJ and SWE:VIQADJ. Aggregate variant testing in *GRIN2B* shows significant results only for the SP phenotype. Results in Table 3 show significant evidence of association between variants in the downstream region of *GRIN2B* and SP:UNADJ. Variants in intronic regions show significant evidence of association with SP:VIQADJ and provide suggestive evidence of association with SP:UNADJ. There is only suggestive support for a role of variants in the exonic regions for SP:VIQADJ. All such variants are in the 3′ or 5′ UTR regions or lead to synonymous changes, and none provide statistical support for a role on the phenotype.

Table 8: Common variants (MAF ≥ 0.01) in *GRIN2B*, UNADJ phenotypes. Effect size (*p*-value) for dose of the rarer allele with adjustment for non-European ancestry. Variants appear in this table if $p < 0.01$ for at least one phenotype.

rsID	Position[2]	REF/ALT (Freq)[3]	Region[4]	Function[5]	Phenotype[1]: UNADJ					
					WID	WA	SP	SWE	PDE	NWR
rs11055486	12:13,656,999	T/C (0.256)	DS	Regulatory	-	-	-	-	-	**-0.12 (0.002)**
rs10772689	12:13,657,684	T/C (0.529)	DS	Regulatory	-	-	-	-	-	**-0.11 (0.001)**
rs4764003	12:13,659,564	C/T (0.371)	DS	NA	-	-	-	-	-	**0.10 (0.003)**
rs7307659	12:13,676,366	C/T (0.072)	DS	Regulatory	-2.08 (0.024)	-	-	-	-	**-0.18 (0.005)**
rs12184498	12:13,823,248	T/C (0.038)	I	NA	-2.56 (0.043)	-	-	-	-	**-0.26 (0.002)**
rs61914322	12:14,101,470	G/A (0.100)	I	NA	**-2.26 (0.005)**	-	-	**-0.14 (0.008)**	-	-

1. Phenotypes as described in Table 1. Bold indicates SNP/phenotype combinations where $p < 0.01$. – indicates *p*-values ≥ 0.05.
2. Build GRCh37/hg19
3. Freq refers to the alternate allele.
4. DS: downstream, I: intronic, E: exonic, US: upstream
5. VEP build38

Non-coding variants in CARs. Bioinformatic investigation of JASPAR, an open access database of non-redundant TFs [138] identified ten variants that are within the peak of a CAR and are predicted to disrupt a TF binding site in brain and/or or neuronal tissue (Table 9). These potential regulatory elements are in four of the five candidate genes, adding further support for a possible role in brain development and/or functioning that leads to dyslexia. Of the total 27 non-coding variants identified by association tests (Table S9), 14 fail to meet both requirements of falling in a peak profile identified by ATAC-seq or DNAaseI-seq and overlapping with a known TF motif and were not considered further. The remaining 13 variants reside in CARs and collectively overlap 50 TF motifs. Nineteen of these variant-motif pairs involve a change in a conserved base of the motif, suggesting a likely impact on TF binding (Table 9).

Table 9: Non-coding variants that are in a CAR and are predicted to disrupt binding of the named transcription factor. Bold in the right-hand column indicates a tissue in which both the SNP and the promotor of the TF are in a CAR.

Gene	SNP	Neuronal tissues[1] with an ATAC-seq CAR	Brain tissue[2] with a DNAseI-seq CAR	Likely disrupted TF	Brain tissues with TF promoter in CAR
CYP19A1	rs12594395		Ol	PRDM1	**Ol**
	rs141853734	PMC, NAC, PUT		KLF9	**PMC**, **NAC**, **PUT** and others
KIAA0319	rs114526784	PUT		LIN54	**PUT** and others
DCDC2	rs72830859	MDT		NFIL3	**MDT** and others
	rs72830859	MDT		HLF	**MDT** and others
	rs12198320	OFC, VLPFC, ACC, STC, ITC, PMC, AMY, HIPP	fb, ast, aoc, aoh, aosc, bme, cpe, ol	TCF7L1	PUT
	rs12198320	OFC, VLPFC, ACC, STC, ITC, PMC, AMY, HIPP	fb, ast, aoc, aoh, aosc, bme, cpe, ol	TCF7L2	**OFC**, **VLPFC**, **ACC**, **STC**, **ITC**, **PMC**, **AMY**, **HIPP**
	rs7763196	OFC, VLPFC, ACC, PVC, HIPP	fb	HOXD10	Bpn
	rs142310124	NAC, PUT		LIN54	**NAC**, **PUT** and others
GRIN2B	rs11055486	MDT	Fb	Six4	**MDT** and others
	rs4764003	MDT	Ol	MZF1	**MDT** and others
	rs12184498	OFC, VLPFC, ACC, STC, PMC, NAC, PUT	Fb	KLF9	**OFC**, **VLPFC**, **ACC**, **STC**, **PMC**, **NAC**, **PUT**
	rs12184498	OFC, VLPFC, ACC, STC, PMC, NAC, PUT	Fb	KLF1	Cer
	rs12184498	OFC, VLPFC, ACC, STC, PMC, NAC, PUT	Fb	SPIB	aoh, fc
	rs12184498	OFC, VLPFC, ACC, STC, PMC, NAC, PUT	Fb	PRDM9	**OFC**, **VLPFC**, **ACC**, **STC**, **PMC**, **PUT**
	rs12184498	OFC, VLPFC, ACC, STC, PMC, NAC, PUT	Fb	EHF	HIPP, fb, aoh
	rs12184498	OFC, VLPFC, ACC, STC, PMC, NAC, PUT	Fb	Spi1	bpn, cer, fc
	rs12184498	OFC, VLPFC, ACC, STC, PMC, NAC, PUT	Fb	SP3	**OFC**, **VLPFC**, **ACC**, **STC**, **PMC**, **NAC**, **PUT**
	rs12184498	OFC, VLPFC, ACC, STC, PMC, NAC, PUT	Fb	ZNF667	**NAC**, **PUT**

1. ATAC-seq neuronal tissues: OFC, Orbitofrontal cortex; VLPFC, Ventrolatelar prefrontal cortex; ACC, Anterior cingulate cortex; STC, Superior temporal cortex; ITC, Inferior temporal cortex; PMC, Primary motor cortex; PVC, Primary visual cortex; AMY, Amygdala; HIPP, Hippocampus, MDT, Mediodorsal thalamus; NAC, Nucleus accumbens; PUT, Putamen.
2. DNAseI-seq tissue: fb, fetal brain (day 72, 80, 85, 96, 101, 104, 105, 109, 112, 117, 122, 142); ast, astrocyte; aoc, astrocyte of cerebellum; aoh, astrocyte of hippocampus; aosc, astrocyte of spinal cord; bpn, bipolar neuron; bme, brain microvascular endothelium; per, pericyte; cer, cerebellum; cpe, choroid plexus epithelial cells; fc, frontal cortex; ol, occipital lobe.

In *CYP19A1* (Table 4), two of the three non-coding variants potentially have a regulatory role (Table 9). Variant rs12594395 is located 16.1 kb downstream of the 3´UTR in an occipital lobe specific CAR with a predicted loss-of-binding of the TF PRDM1. PRDM1's promoter is in open chromatin in the same tissue, suggesting it is active in this brain region [140-142] and its function may be disrupted by rs12594395. Variant rs141853734 simultaneously affects two overlapping motifs for TFs KLF9 and ZSCAN4 in neuronal cells of the primary motor cortex (PMC), nucleus accumbens (NAC) and putamen (PUT). KLF9 has an active/poised promoter in these brain regions while ZSCAN4 does not.

In *KIAA0319*, one of two non-coding variants that are not part of the haplotype that carries missense variants (Table 5) may affect TF binding. One (rs114526784) is predicted to disrupt a TF binding site (Table 9). The other (rs114526784) is located on overlapping motifs for TFs ZNF384 and LIN54 in neuronal cells of the PUT. ZNF384 is unperturbed by the nucleotide substitution, but the variant changes the LIN54 motif in a highly conserved fourth residue of the motif that predicts a loss of TF binding. LIN54 has been shown to be a member of the DREAM complex that functions as either a transcription activator or repressor depending on the context [145, 146]. The second SNP, rs11497321, is in a CAR that is present in fetal brain and neuronal cells of 11 adult brain regions, but the resulting nucleotide substitution in NrfI and ELK1:HOXA1 is not predicted to significantly affect binding of either TF.

Several non-coding variants throughout *DCDC2* (Table 6) have the potential for a regulatory role. In the long region defined by the associated haplotype blocks A, B and C described above, there is at least one non-coding variant in each that is predicted to interfere with TF binding. In block A, all three variants are in CARs in some brain tissues, but only rs72830859 in a CAR in mediodorsal thalamus (MDT) results in loss-of-binding mutations predicted to affect TFs PHOX2B, NFIL3 and HLF. Both NFIL3 and HLF have active/poised promoters in all brain regions including MDT. Block B contains a cluster of three variants that span 289 bp and reside within a single CAR that is present in fetal brain and neuronal cells of eight different brain regions. Two of these variants disrupt TF motifs. Variant rs12198320 in Block B is centered within the CAR and disrupts TF motifs for TCF7L1, TCF7L2, and Hnf1A. Hnf1A is inactive in CNS, but TCF7L1 has an active/poised promoter in the neuronal cells of PUT, and TCF7L2 has an active/poised promoter in many neuronal cell types of the brain. Variant rs7763196 in Block B disrupts motifs for TFs CDX4 and HOXD10 in neuronal cells from five brain regions. CDX4 is not expressed in the central nervous system, but HOXD10 has an active/poised promoter in bipolar neurons. Block C variants span 80 kb from the second intron of *DCDC2* to the ninth intron of the downstream gene *MRS2*. One variant, rs142310124, is in a CAR that is specific for neuronal cells in NAC and PUT and is predicted to disrupt motifs for four different TFs (POU2F3, POU3F3, PHOX2B and LIN54). Of these four TFs only LIN54 has an active/poised promoter in PUT. Additional variants rs9379643 and rs61232212 in *DCDC2* are also predicted by JASPAR to disrupt TF binding sites, but neither is present in any neuronal cell types.

Three non-coding variants in *GRIN2B* are predicted to disrupt TF binding. Variants rs11055486 and rs4764003 are both in neuronal-cell specific CARs in MDT and affect the TF motif of Six4 and MZF1, respectively. Both Six4 and MZF1 also have open promoters in MDT. Variant rs12184498 is in CARs present in seven different brain regions and affects eight different TF motifs for KLF9, KLF1, PRDM9, SPIB, EHF, Spi1, SP3 and ZNF667. KLF9, PRDM9, SP3 and

ZNF667. All these TFs have open promoters in the relevant tissues, suggesting the availability of these TF for binding. Variant rs10772689 is located within a CAR in some neuronal cells but does not overlap with any known TF motifs. Variant rs7307659 is upstream of a neuronal-specific PUT CAR and was captured in the smMIP experiments, but because it is outside of a peak signal, we are unable to speculate regarding its role.

DISCUSSION

Here we provided results of a comprehensive investigation of underlying genomic variation in and surrounding five genes with prior evidence for an inherited effect on endophenotypes of dyslexia risk. The MIP sequencing approach that we used allowed inclusion of many more participants and variants than have been previously considered and provided an agnostic approach for identifying underlying causal variants. Reliance of previous studies on detectable linkage disequilibrium between a causal variant and a small number of nearby genotyped polymorphisms is a possible cause of conflicting results across laboratories [51, 52, 72]. In contrast, in the current study, we evaluated much of the gene neighborhoods, focusing on sequence that had the greatest potential for bioinformatic interpretation: protein coding regions and potential regulatory sites upstream and downstream of the candidate genes. Variants evaluated span a wide allele frequency range and fall in both coding and non-coding DNA.

The variants that met our threshold for association and further bioinformatic consideration represent primarily non-coding DNA, with no clearly pathogenic coding variants. Although limited to a small number of selected genes and gene neighborhoods, the results provide an initial prediction of the types of variation that are likely to be more broadly identified through evaluation of DNA sequence in genome-scale studies of specific learning disabilities such as dyslexia. We speculate that variation in coding sequence that results in dramatic alteration of protein structure or function, as is typical of Mendelian disorders, is unlikely to play a role in dyslexia. Such protein-coding variants generally are rare and subject to negative selection, with large impacts on the phenotype. Dyslexia is a phenotype that only is recognized in the presence of widespread literacy, and non-coding variants with subtle effects on gene expression or control are more likely to be relevant. Selection against such variants, with their weak effects on the phenotype, would have been relatively ineffective in the small populations that were typical until very recently in human history. Instead, stochastic effects, such as genetic drift, would have had a role in driving changes in allele frequencies.

Of the variants implicated herein, common variants in *CYP19A1* provide the most significant support for association with all the reading- and spelling-related phenotypes we studied. Bioinformatics evaluation suggests a possible role for regulatory variation that may alter the amount of mRNA, and thus protein, produced. The strongest statistical result was for a variant (rs55712458) in the downstream region of the gene. This variant does not overlap any TF motifs, but it is less than 300 bp away from rs12594395, which occupies an occipital lobe specific CAR, and is expected to affect the binding of the PRDM1 TF. Another variant in *CYP19A1* affects a motif for KLF9, which is a neuronal TF with an active/poised promoter in all brain regions, including fetal brain. In addition to these individual-SNP results, aggregate testing suggests that

some probably rare, intronic variants may also be involved. Finally, *CYP19A1* was also previously nominated as a dyslexia-risk gene through identification of the breakpoint of a t(2;15)(p12;q21) translocation that disrupted the promoter region of the gene in a person with dyslexia [37]. *CYP19A1* encodes the estrogen synthase CP19A that converts C19 androgens to C18 estrogens and is responsible for local synthesis of estrogens outside of the reproductive system. In the brain it is expressed from prenatal stages to adulthood [147] in multiple cell types where it regulates synaptic plasticity and plays a role in cognition, memory and language, and many other functions [148, 149].

We also obtained support for involvement in multiple measures of reading/spelling for the two genes evaluated in DYX2: *KIAA0319* and *DCDC2.* In *KIAA0319*, a haplotype that includes missense variants in a predicted fibronectin type III domain [144] is associated with several dyslexia phenotypes. Fibronectin is a protein that has important roles in cell adhesion, migration, differentiation, and proliferation, which has implications for development, among other functions [150-152]. Additional associations were found for common variants, including one (rs114526784) predicted to disrupt a binding site of TF LIN54 in neuronal cells of PUT. For *DCDC2* we detected common intronic variants and both protective and deleterious haplotypes that are significantly associated with dyslexia phenotypes. In particular, two common haplotypes affected timed real word reading scores in opposite directions. Variants within these and other haplotypes are predicted to disrupt binding sites for several TFs. For both genes, therefore, as for *CYP19A1*, some associated variants implicate transcriptional regulation as a possible mechanism of modulation of dyslexia risk.

Finally, common variants downstream of *GRIN2B* are associated with decreased performance on phonological non-word memory, the only phenotype on which this gene had an effect. None of the other studied genes showed an association with this phenotype. One *GRIN2B* variant affects the TF motif of a homeobox protein that may have a role in the differentiation, maturation, and survival of neuronal cells [153-155]. *GRIN2B* was selected for the current study because it lies in a region with evidence of linkage for a phonological non-word memory phenotype in the UW cohort [49] and was associated with verbal memory phenotypes in two European dyslexia datasets [117, 118]. *GRIN2B* encodes GluN2B, one of the glutamate-binding subunits of the tetrameric N-methyl-D-aspartate ionotropic glutamate receptors (NMDARs) that are important for neuronal development and plasticity [156]. GluN2B is highly expressed prenatally in the brain where it is involved in learning and working memory via its role in synaptic plasticity and enhanced long-term potentiation [157, 158]. Pathogenic coding variants and deletions in *GRIN2B* also cause a spectrum of neurodevelopmental disorders [159-161]. Non-coding variants in *GRIN2B* have also been associated with short term and working memory, intelligence quotient and cognitive impairments in dyslexia [117, 118] and with other cognitive and behavioral traits [162-164].

Given diverse opinions about the importance of taking VIQ into account, we note that conclusions in the current study were not strongly affected by adjustment, or not, for verbal IQ. If anything, results were slightly stronger in the absence of VIQ adjustment. Two possible explanations for this observation are (1) that IQ-adjusted analysis had lower power to detect association because of a smaller sample size of participants with the necessary IQ data, and/or (2) use of a covariate with measurement error, as is the case for VIQ, leads to regression to the mean [165], with consequent loss of power to detect the effect of the predictor – the SNP in this case.

There are, of course, some limitations to our study. Although we were able to generate more sequence data on a larger sample than has previously been evaluated for dyslexia, the tradeoff was that the sequencing strategy was limited to a subset of short DNA segments within a small number of previously implicated regions. We therefore cannot comment on genes and genomic regions that fall outside of the regions investigated or sequence alterations such as structural rearrangements or copy number variants (*e.g.*, the READ1 polymorphism [112-114]), that would likely be missed by this short-read technique. Our data in the regions investigated allowed evaluation of most DNA positions in the regions for which current understanding of molecular mechanisms allows bioinformatics interpretation about effects on initiation of transcription. Even so, current knowledge about normal human variation in the regulome is still incomplete, and we acknowledge that transcription factor families share binding motifs, making the definitive identification of specific transcription factors difficult. It is also possible that variants in other regulatory motifs, which can be quite distant from the coding portions of the genes, may hold the causative DNA alterations.

In summary, we provide evidence that variants in or near *DCDC2*, *KIAA0319*, *CYP19A1*, and *GRIN2B* play a role in risk for dyslexia. Given the large sample size, these results argue strongly against the causative involvement of large-effect coding variants in any of the five candidate genes. Instead, these results support a potential role in transcriptional regulation that alters the quantity of RNA produced. The non-coding variant/transcription factor/brain-region combinations we identified provide a starting point for future work on the role of non-coding variants in dyslexia. These results also illustrate some of the challenges that the field will face in identifying causal variants that may act through gene regulation rather than alteration of protein sequences. Use of whole-genome sequence (WGS), especially long-read, would capture regulatory elements with fewer complications, including detection of alterations in repeat sequences that might reside in deep intronic or intergenic regions. However, the WGS approach adds significant cost that could critically limit the number of samples used. The most feasible approach to corroboration of variants and haplotypes of interest discussed herein will therefore require evaluation in other dyslexia sample sets followed by functional studies in an appropriate cell model to begin to determine biological relevance [166], an endeavor that is well beyond the scope of the current analysis.

ACKNOWLEDGEMENTS
We are grateful to the family members who volunteered their time to participate in the research. John Wolff, Hiep Nguyen, and Edith PA Fuerte provided excellent technical, computational, and

bioinformatics assistance. We thank the many graduate student assistants who administered the test batteries.

LITERATURE CITED

1. Lyon GR, Shaywitz SE, Shaywitz B. A definition of dyslexia. Annals of Dyslexia. 2003;53:1-14.
2. Katusic SK, Colligan RC, Barbaresi WJ, Schaid DJ, Jacobsen SJ. Incidence of reading disability in a population-based birth cohort, 1976-1982, Rochester, Minn. Mayo Clin Proc. 2001;76(11):1081-92.
3. Cai L, Chen Y, Hu X, Guo Y, Zhao X, Sun T, et al. An epidemiological study of Chinese children with developmental dyslexia. J Dev Behav Pediatr. 2020;41(3):203-11.
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 Disabil. 2020;53(5):354-65.
5. Yang L, Li C, Li X, Zhai M, An Q, Zhang Y, et al. Prevalence of developmental dyslexia in primary school children: A systematic review and meta-analysis. Brain Sci. 2022;12(2).
6. Barbiero C, Montico M, Lonciari I, Monasta L, Penge R, Vio C, et al. The lost children: The underdiagnosis of dyslexia in Italy. A cross-sectional national study. PloS one. 2019;14(1):e0210448.
7. Shaywitz SE, Shaywitz JE, Shaywitz BA. Dyslexia in the 21st century. Curr Opin Psychiatry. 2021;34(2):80-6.
8. Raskind W, Hsu L, Berninger V, Thomson J, Wijsman E. Familial aggregation of dyslexia phenotypes. Behav Genet. 2000;30:385-95.
9. Flannery KA, Liederman J, Daly L, Schultz J. Male prevalence for reading disability is found in a large sample of Black and White children free from ascertainment bias. J Int Neuropsychological Soc. 2000;6(4):433-42.
10. Quinn JM, Wagner RK. Gender differences in reading impairment and in the identification of impaired readers: results from a large-scale study of at-risk readers. J Learn Disabil. 2015;48(4):433-45.
11. Berninger V, Abbott R, thomson JB, Raskind WH. Language phenotype for reading and writing disability: a family approach. Sci Stud Read. 2001;5:59-105.
12. Hatcher J, Snowling MJ, Griffiths YM. Cognitive assessment of dyslexic students in higher education. Br J Educ Psychol. 2002;72(Pt 1):119-33.
13. Wilson A, Lesaux N. Persistence of phonological processing deficits in college students with dyslexia who have age-appropriate reading skills. J Learn Disabil. 2001;34:394-400.
14. Morris D, Turnbull P. A survey-based exploration of the impact of dyslexia on career progression of UK registered nurses. J Nurs Manag. 2007;15(1):97-106.
15. Gerber PJ. The impact of learning disabilities on adulthood: a review of the evidenced-based literature for research and practice in adult education. J Learn Disabil. 2012;45(1):31-46.
16. McLaughlin MJ, Speirs KE, Shenassa ED. Reading disability and adult attained education and income: evidence from a 30-year longitudinal study of a population-based sample. J Learn Disabil. 2014;47(4):374-86.
17. Andreola C, Mascheretti S, Belotti R, Ogliari A, Marino C, Battaglia M, et al. The heritability of reading and reading-related neurocognitive components: A multi-level meta-analysis. Neurosci Biobehav Rev. 2021;121:175-200.

18. van Bergen E, de Jong PF, Plakas A, Maassen B, vdL, A. Child and parental literacy levels within families with a history of dyslexia. J Child Psychol Psychiatry. 2012;53(1):23-36.
19. Kirkpatrick RM, Legrand LN, Iacono WG, McGue M. A twin and adoption study of reading achievement: Exploration of shared-environmental and gene-environment-interaction effects. Learn Individ Differ. 2011;21(4):368-75.
20. Erbeli F, Rice M, Paracchini S. Insights into dyslexia genetics research from the last two decades. Brain Sci. 2022;12(1):27.
21. DeFries J, Fulker D, LaBuda M. Evidence for a genetic aetiology in reading disability of twins. Nature. 1987;329:537-9.
22. Gayan J, Olson R. Reading disability: evidence for a genetic etiology. Eur Child Adolesc Psychiatry. 1999;8(Suppl):52-5.
23. Fagerheim T, Raeymaekers P, Tonnessen FE, Pedersen M, Tranebjaerg L, Lubs HA. A new gene (DYX3) for dyslexia is located on chromosome 2. J Med Genet. 1999;36(9):664-9.
24. Nopola-Hemmi J, Myllyluoma B, Haltia T, Taipale M, Ollikainen V, Ahonen T, et al. A dominant gene for developmental dyslexia on chromosome 3. J Med Genet. 2001;38(10):658-64.
25. de Kovel CGF, Hol FA, Heister J, Willemen J, Sandkuijl LA, Franke B, et al. Genomewide scan identifies susceptibility locus for dyslexia on Xq27 in an extended Dutch family. J Med Genet. 2004;41(9):652-7.
26. Grimm T, Garshasbi M, Puettmann L, Chen W, Ullmann R, Muller-Myhsok B, et al. A novel locus and candidate gene for familial developmental dyslexia on chromosome 4q. Z Kinder Jugendpsychiatr Psychother. 2020;48(6):478-89.
27. Carrion-Castillo A, Estruch SB, Maassen B, Franke B, Francks C, Fisher SE. Whole-genome sequencing identifies functional noncoding variation in SEMA3C that cosegregates with dyslexia in a multigenerational family. Hum Genet. 2021;140(8):1183-200.
28. Einarsdottir E, Svensson I, Darki F, Peyrard-Janvid M, Lindvall JM, Ameur A, et al. Mutation in CEP63 co-segregating with developmental dyslexia in a Swedish family. Hum Genet. 2015;134(11-12):1239-48.
29. Georgitsi M, Dermitzakis I, Soumelidou E, Bonti E. The polygenic nature and complex genetic architecture of specific learning disorder. Brain Sci. 2021;11(5).
30. Fisher SE, Francks C, Marlow AJ, MacPhie IL, Newbury DF, Cardon LR, et al. Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. Nat Genet. 2002;30(1):86-91.
31. Hannula-Jouppi K, Kaminen-Ahola N, Taipale M, Eklund R, Nopola-Hemmi J, Kaariainen H, et al. The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. PLoS Genet. 2005;1(4):467-74.
32. Doust C, Fontanillas P, Eising E, Gordon SD, Wang Z, Alagoz G, et al. Discovery of 42 genome-wide significant loci associated with dyslexia. Nat Genet. 2022;54(11):1621-9.
33. Gialluisi A, Andlauer TFM, Mirza-Schreiber N, Moll K, Becker J, Hoffmann P, et al. Genome-wide association scan identifies new variants associated with a cognitive predictor of dyslexia. Transl Psychiatry. 2019;9(1):77.
34. Price KM, Wigg KG, Feng Y, Blokland K, Wilkinson M, He G, et al. Genome-wide association study of word reading: Overlap with risk genes for neurodevelopmental disorders. Genes Brain Behav. 2020;19(6):e12648.

35. Veerappa AM, Saldanha M, Padakannaya P, Ramachandra NB. Genome-wide copy number scan identifies disruption of PCDH11X in developmental dyslexia. Am J Med Genet B Neuropsychiatr Genet. 2013;162B(8):889-97.
36. Nopola-Hemmi J, Taipale M, Haltia T, Lehesjoki AE, Voutilainen A, Kere J. Two translocations of chromosome 15q associated with dyslexia. J Med Genet. 2000;37(10):771-5.
37. Anthoni H, Sucheston LE, Lewis BA, Tapia-Paez I, Fan X, Zucchelli M, et al. The aromatase gene CYP19A1: several genetic and functional lines of evidence supporting a role in reading, speech and language. Behav Genet. 2012;42(4):509-27.
38. Bates TC, Luciano M, Castles A, Coltheart M, Wright MJ, Martin NG. Replication of reported linkages for dyslexia and spelling and suggestive evidence for novel regions on chromosomes 4 and 17. Eur J Hum Genet. 2007;15(2):194-203.
39. Grigorenko EL, Wood FB, Meyer MS, Hart LA, Speed WC, Shuster A, et al. Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. Am J Hum Genet. 1997;60:27-9.
40. Marino C, Citterio A, Giorda R, Facoetti A, Menozzi G, Vanzin L, et al. Association of short-term memory with a variant within DYX1C1 in developmental dyslexia. Genes Brain Behav. 2007;6(7):640-6.
41. Morris DW, Robinson L, Turic D, Duke M, Webb V, Milham C, et al. Family-based association mapping provides evidence for a gene for reading disability on chromosome 15q. Hum Molec Genet. 2000;9(5):843-8.
42. Nöthen MM, Schulte-Korne G, Grimm T, Cichon S, Vogt IR, Muller-Myhsok B, et al. Genetic linkage analysis with dyslexia: evidence for linkage of spelling disability to chromosome 15. Eur Child Adolesc Psychiatry. 1999;8 Suppl 3:56-9.
43. Smith SD, Kimberling WJ, Pennington BF, Lubs HA. Specific reading disability: identification of an inherited form through linkage analysis. Science. 1983;219(4590):1345-7.
44. Venkatesh SK, Siddaiah A, Padakannaya P, Ramachandra NB. Association of SNPs of DYX1C1 with developmental dyslexia in an Indian population. Psychiatr Genet. 2014;24(1):10-20.
45. Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC. Quantitative trait locus for reading disability: correction. Science. 1995;268:1553.
46. Fisher SE, Marlow AJ, Lamb J, Maestrini E, Williams DF, Richardson AJ, et al. A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. Am J Hum Genet. 1999;64:146-56.
47. Grigorenko EL, Wood FB, Golovyan L, Meyer M, Romano C, Pauls D. Continuing the search for dyslexia genes on 6p. Am J Med Genet B Neuropsychiatr Genet. 2003;118B(1):89-98.
48. Turic D, Robinson L, Duke M, Morris DW, Webb V, Hamshere M, et al. Linkage disequilibrium mapping provides further evidence of a gene for reading disability on chromosome 6p21.3-22. Mol Psychiatry. 2003;8(2):176-85.
49. Brkanac Z, Chapman N, Igo RP, Matsushita M, Nielsen K, Berninger V, et al. Linkage analysis of a nonword repetition endophenotype in families with dyslexia: initial findings and replication. Behav Genet. 2008;38(5):462-75.
50. Taipale M, Kaminen N, Nopola-Hemmi J, Haltia T, Myllyluoma B, Lyytinen H, et al. A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat

domain protein dynamically regulated in brain. Proc. Natl. Acad. Sci. U.S.A. 2003;100:11553-8.
51. Wigg K, Couto J, Feng Y, Anderson B, Cate-Carter T, Macciardi F, et al. Support for EKN1 as the susceptibility locus for dyslexia on 15q21. Molec Psychiatry. 2004;9:1111-21.
52. Brkanac Z, Chapman N, Matsushita M, Chun L, Nielsen K, E C, et al. Evaluation of candidate genes for DYX1 and DYX2 in families with dyslexia. Am J Med Genet (Neuropsychiatr Genet). 2007;144B:556-60.
53. Francks C, Paracchini S, Smith S, Richardson A, Scerri T, Cardon L, et al. A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. Am J Hum Genet. 2004;75:1046-58.
54. Cope N, Harold D, Hill G, Moskvina V, Stevenson J, Holmans P, et al. Strong evidence that KIA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. Am J Hum Genet. 2005;76(4):581-91.
55. Harold D, Paracchini S, Scerri T, Dennis N, Cope H, Hill G, et al. Further evidence that the *KIAA0319* gene confers susceptibility to developmental dyslexia. Molec Psychiatry. 2006;11:1085-91.
56. Scerri TS, Morris AP, Buckingham LL, Newbury DF, Miller LL, Monaco AP, et al. DCDC2, KIAA0319 and CMIP are associated with reading-related traits. Biol Psychiatry. 2011;70(3):237-45.
57. Venkatesh SK, Siddaiah A, Padakannaya P, Ramachandra NB. Analysis of genetic variants of dyslexia candidate genes KIAA0319 and DCDC2 in Indian population. J Hum Genet. 2013;58(8):531-8.
58. Eicher JD, Powers NR, Miller LL, Mueller KL, Mascheretti S, Marino C, et al. Characterization of the DYX2 locus on chromosome 6p22 with reading disability, language impairment, and IQ. Hum Genet. 2014;133(7):869-81.
59. Matsson H, Huss M, Persson H, Einarsdottir E, Tiraboschi E, Nopola-Hemmi J, et al. Polymorphisms in DCDC2 and S100B associate with developmental dyslexia. J Hum Genet. 2015;60(7):399-401.
60. Zhao H, Chen Y, Zhang BP, Zuo PX. KIAA0319 gene polymorphisms are associated with developmental dyslexia in Chinese Uyghur children. J Hum Genet. 2016;61(8):745-52.
61. Trezzi V, Forni D, Giorda R, Villa M, Molteni M, Marino C, et al. The role of READ1 and KIAA0319 genetic variations in developmental dyslexia: testing main and interactive effects. J Hum Genet. 2017;62(11):949-55.
62. Paracchini S, Thomas AC, Castro S, Lai C, Paramasivam M, Wang Y, et al. The chromosome 6p22 haplotype associated with dyslexia reduces the expression of *KIAA0319*, a novel gene involved in neuronal migration. Hum Molec Genet. 2006;15:1659-66.
63. Dennis MY, Paracchini S, Scerri TS, Prokunina-Olsson L, Knight JC, Wade-Martins R, et al. A common variant associated with dyslexia reduces expression of the KIAA0319 gene. PLoS Genet. 2009;5(3):e1000436.
64. Carrion-Castillo A, Franke B, Fisher SE. Molecular genetics of dyslexia: an overview. Dyslexia. 2013;19(4):214-40.
65. Guidi LG, Velayos-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(10):3212-33.

66. Riva V, Mozzi A, Forni D, Trezzi V, Giorda R, Riva S, et al. The influence of DCDC2 risk genetic variants on reading: Testing main and haplotypic effects. Neuropsychologia. 2019;130:52-8.
67. Deng KG, Zhao H, Zuo PX. Association between KIAA0319 SNPs and risk of dyslexia: a meta-analysis. J Genet. 2019;98(1).
68. Bieder A, Yoshihara M, Katayama S, Krjutskov K, Falk A, Kere J, et al. Dyslexia candidate gene and ciliary gene expression dynamics during human neuronal differentiation. Mol Neurobiol. 2020;57(7):2944-58.
69. Petryshen TL, Kaplan BJ, Liu MF, Field LL. Absence of significant linkage between phonological coding dyslexia and chromosome 6p23-21.3, as determined by use of quantitative-trait methods: confirmation of qualitative analyses. Am J Hum Genet. 2000;66:708-14.
70. Chapman N, Igo RJ, Thomson J, Matsushita M, Brkanac Z, Holzman T, et al. Linkage analyses of four regions previously implicated in dyslexia: confirmation of a locus on chromosome 15q. Am J Med Genet B Neuropsychiatr Genet. 2004;131B:67-75.
71. de Kovel CG, Franke B, Hol FA, Lebrec JJ, Maassen B, Brunner H, et al. Confirmation of dyslexia susceptibility loci on chromosomes 1p and 2p, but not 6p in a Dutch sib-pair collection. Am J Med Genet B Neuropsychiatr Genet. 2008;147(3):294-300.
72. Scerri T, Fisher S, Francks C, MacPhie I, Paracchini S, Richardson A, et al. Putative functional alleles of DYX1C1 are not associated with dyslexia susceptibility in a large sample of sibling pairs from the UK. J Med Genet. 2004;41:853-7.
73. Bellini G, Bravaccio C, Calamoneri F, Cocuzza MD, Fiorillo 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 Neurosci. 2005;27(3):311-4.
74. Cope N, Hill G, van den Bree M, Harold D, Moskvina V, Green E, et al. No support for association between Dyslexia Susceptibility 1 Candidate 1 and developmental dyslexia. Molec Psychiatry. 2005;10(3):237-8.
75. Marino C, Giorda R, Lorusso M, Vanzin L, Salandi N, Nobile M, et al. A family-based association study does not support DYX1C1 on chromosome 15q21.3 as a candidate gene in developmental dyslexia. Eur J Hum Genet. 2005;13(4):491-9.
76. Tran C, Gagnon F, Wigg KG, Feng Y, Gomez L, Cate-Carter TD, et al. A family-based association analysis and meta-analysis of the reading disabilities candidate gene DYX1C1. Am J Med Genet B Neuropsychiatr Genet. 2013;162B(2):146-56.
77. Carrion-Castillo A, Maassen B, Franke B, Heister A, Naber M, van der Leij A, et al. Association analysis of dyslexia candidate genes in a Dutch longitudinal sample. Eur J Hum Genet. 2017;25(4):452-60.
78. Sharma P, Sagar R, Deep R, Mehta M, Subbiah V. Assessment for familial pattern and association of polymorphisms in KIAA0319 gene with specific reading disorder in children from North India visiting a tertiary care centre: A case-control study. Dyslexia. 2020;26(1):104-14.
79. Wang Y, Paramasivam M, Thomas A, Bai J, Kaminen-Ahola N, Kere J, et al. DYX1C1 functions in neuronal migration in developing neocortex. Neuroscience. 2006;143(2):515-22.
80. 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 B Neuropsychiatr Genet. 2012;159B(8):970-6.

81. 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(1):435-42.
82. 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 Ppopulation. Sci Rep. 2016;6:30454.
83. Meaburn EL, Harlaar N, Craig IW, Schalkwyk LC, Plomin R. Quantitative trait locus association scan of early reading disability and ability using pooled DNA and 100K SNP microarrays in a sample of 5760 children. Mol Psychiatry. 2008;13(7):729-40.
84. Field LL, Shumansky K, Ryan J, Truong D, Swiergala E, Kaplan BJ. Dense-map genome scan for dyslexia supports loci at 4q13, 16p12, 17q22; suggests novel locus at 7q36. Genes Brain Behav. 2013;12(1):56-69.
85. Luciano M, Evans DM, Hansell NK, Medland SE, Montgomery GW, Martin NG, et al. A genome-wide association study for reading and language abilities in two population cohorts. Genes Brain Behav. 2013;12(6):645-52.
86. Gialluisi A, Newbury DF, Wilcutt EG, Olson RK, DeFries JC, Brandler WM, et al. Genome-wide screening for DNA variants associated with reading and language traits. Genes Brain Behav. 2014;13(7):686-701.
87. Gialluisi A, Andlauer TFM, Mirza-Schreiber N, Moll K, Becker J, Hoffmann P, et al. Genome-wide association study reveals new insights into the heritability and genetic correlates of developmental dyslexia. Mol Psychiatry. 2021;26(7):3004-17.
88. Eising E, Mirza-Schreiber N, de Zeeuw EL, Wang CA, Truong DT, Allegrini AG, et al. Genome-wide analyses of individual differences in quantitatively assessed reading- and language-related skills in up to 34,000 people. Proc Natl Acad Sci U S A. 2022;119(35):e2202764119.
89. Galaburda AM, LoTurco J, Ramus F, Fitch RH, Rosen GD. From genes to behavior in developmental dyslexia. Nat Neurosci. 2006;9(10):1213-7.
90. Paracchini S, Scerri T, Monaco AP. The genetic lexicon of dyslexia. Annu Rev Genomics Hum Genet. 2007;8:57-79.
91. Galaburda AM, Kemper TL. Cytoarchitectonic abnormalities in developmental dyslexia: a case study. Ann Neurol. 1979;6(2):94-100.
92. Kaufmann WE, Galaburda AM. Cerebrocortical microdysgenesis in neurologically normal subjects: a histopathologic study. Neurology. 1989;39(2 Pt 1):238-44.
93. Meng H, Smith SD, Hager K, Held M, Liu J, Olson RK, et al. DCDC2 is associated with reading disability and modulates neuronal development in the brain. Proc Natl Acad Sci U S A. 2005;102(47):17053-8.
94. Martinez-Garay I, Guidi LG, Holloway ZG, Bailey MA, Lyngholm D, Schneider T, et al. Normal radial migration and lamination are maintained in dyslexia-susceptibility candidate gene homolog Kiaa0319 knockout mice. Brain Struct Funct. 2017;222(3):1367-84.
95. Rendall AR, Tarkar A, Contreras-Mora HM, LoTurco JJ, Fitch RH. Deficits in learning and memory in mice with a mutation of the candidate dyslexia susceptibility gene Dyx1c1. Brain Lang. 2017;172:30-8.
96. Wang Y, Yin X, Rosen G, Gabel L, Guadiana SM, Sarkisian MR, et al. Dcdc2 knockout mice display exacerbated developmental disruptions following knockdown of doublecortin. Neuroscience. 2011;190:398-408.

97. Massinen S, Hokkanen ME, Matsson H, Tammimies K, Tapia-Paez I, Dahlstrom-Heuser V, et al. Increased expression of the dyslexia candidate gene DCDC2 affects length and signaling of primary cilia in neurons. PloS one. 2011;6(6):e20580.
98. Tarkar A, Loges NT, Slagle CE, Francis R, Dougherty GW, Tamayo JV, et al. DYX1C1 is required for axonemal dynein assembly and ciliary motility. Nat Genet. 2013;45(9):995-1003.
99. Che A, Truong DT, Fitch RH, LoTurco JJ. Mutation of the dyslexia-associated gene Dcdc2 enhances glutamatergic synaptic transmission between layer 4 neurons in mouse neocortex. Cereb Cortex. 2016;26(9):3705-18.
100. Franquinho F, Nogueira-Rodrigues J, Duarte JM, Esteves SS, Carter-Su C, Monaco AP, et al. The dyslexia-susceptibility protein KIAA0319 inhibits axon growth through Smad2 signaling. Cereb Cortex. 2017;27(3):1732-47.
101. Couto JM, Livne-Bar I, Huang K, Xu Z, Cate-Carter T, Feng Y, et al. Association of reading disabilities with regions marked by acetylated H3 histones in KIAA0319. Am J Med Genet B Neuropsychiatr Genet. 2010;153B(2):447-62.
102. Tran C, Wigg KG, Zhang K, Cate-Carter TD, Kerr E, Field LL, et al. Association of the ROBO1 gene with reading disabilities in a family-based analysis. Genes Brain Behav. 2014;13(4):430-8.
103. Wechsler D. Wechsler abbreviated scale of intelligence. San Antonio; 1999.
104. Couto JM, Gomez L, Wigg K, Cate-Carter T, Archibald J, Anderson B, et al. The KIAA0319-like (KIAA0319L) gene on chromosome 1p34 as a candidate for reading disabilities. J Neurogenet. 2008;22(4):295-313.
105. Elbert A, Lovett MW, Cate-Carter T, Pitch A, Kerr EN, Barr CL. Genetic variation in the KIAA0319 5' region as a possible contributor to dyslexia. Behav Genet. 2011;41(1):77-89.
106. Kravitz HM, Meyer PM, Seeman TE, Greendale GA, Sowers MR. Cognitive functioning and sex steroid hormone gene polymorphisms in women at midlife. Am J Med. 2006;119(9 Suppl 1):S94-S102.
107. Luciano M, Gow AJ, Pattie A, Bates TC, Deary IJ. The influence of dyslexia candidate genes on reading skill in old age. Behav Genet. 2018;48(5):351-60.
108. Couto JM, Gomez L, Wigg K, Ickowicz A, Pathare T, Malone M, et al. Association of attention-deficit/hyperactivity disorder with a candidate region for reading disabilities on chromosome 6p. Biol Psychiatry. 2009;66(4):368-75.
109. 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.
110. Schumacher J, Anthoni H, Dahdouh F, Konig IR, Hillmer AM, Kluck N, et al. Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. Am J Hum Genet. 2006;78(1):52-62.
111. Marino C, Meng H, Mascheretti S, Rusconi M, Cope N, Giorda R, et al. DCDC2 genetic variants and susceptibility to developmental dyslexia. Psychiatr Genet. 2012;22(1):25-30.
112. Meng H, Powers NR, Tang L, Cope NA, Zhang PX, Fuleihan R, et al. A dyslexia-associated variant in DCDC2 changes gene expression. Behav Genet. 2011;41(1):58-66.
113. Powers NR, Eicher JD, Butter F, Kong Y, Miller LL, Ring SM, et al. Alleles of a polymorphic ETV6 binding site in DCDC2 confer risk of reading and language impairment. Am J Hum Genet. 2013;93(1):19-28.

114. Powers NR, Eicher JD, Miller LL, Kong Y, Smith SD, Pennington BF, et al. The regulatory element READ1 epistatically influences reading and language, with both deleterious and protective alleles. J Med Genet. 2016;53(3):163-71.
115. Igo RP, Chapman NH, Berninger V, Matsushita M, Brkanac Z, Rothstein J, et al. Genome wide scan for real-word reading subphenotypes of dyslexia: Novel chromosome 13 locus and genetic complexity. Am J Med Genet B Neuropsychiatric Genetics. 2006;141(1):15-27.
116. Rubenstein KB, Raskind WH, Berninger VW, Matsushita MM, Wijsman EM. Genome scan for cognitive trait loci of dyslexia: Rapid naming and rapid switching of letters, numbers, and colors. Am J Med Genet B Neuropsychiatr Genet. 2014;165B(4):345-56.
117. Ludwig KU, Roeske D, Herms S, Schumacher J, Warnke A, Plume E, et al. Variation in GRIN2B contributes to weak performance in verbal short-term memory in children with dyslexia. Am J Med Genet B Neuropsychiatr Genet. 2010;153B(2):503-11.
118. Mascheretti S, Facoetti A, Giorda R, Beri S, Riva V, Trezzi V, et al. GRIN2B mediates susceptibility to intelligence quotient and cognitive impairments in developmental dyslexia. Psychiatr Genet. 2015;25(1):9-20.
119. Liu Q, Zhu B, Xue Q, Xie X, Zhou Y, Zhu K, et al. The associations of zinc and GRIN2B genetic polymorphisms with the risk of dyslexia. Environ Res. 2020;191:110207.
120. O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science. 2012;338(6114):1619-22.
121. Boyle EA, O'Roak BJ, Martin BK, Kumar A, Shendure J. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. Bioinformatics. 2014;30(18):2670-2.
122. Hildebrand MS, Myers CT, Carvill GL, Regan BM, Damiano JA, Mullen SA, et al. A targeted resequencing gene panel for focal epilepsy. Neurology. 2016;86(17):1605-12.
123. Brain Open Chromatin Atlas (BOCA) [Internet]. Available from: https://labs.icahn.mssm.edu/roussos-lab/boca/.
124. ENCODE Available from: https://www.encodeproject.org.
125. Fullard JF, Hauberg ME, Bendl J, Egervari G, Cirnaru MD, Reach SM, et al. An atlas of chromatin accessibility in the adult human brain. Genome Res. 2018;28(8):1243-52.
126. Hiatt JB, Pritchard CC, Salipante SJ, O'Roak BJ, Shendure J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res. 2013;23(5):843-54.
127. Ensembl Variant Effect Predictor (VEP) [Internet]. Available from: http://ensembl.org/Homo_sapiens/Tools/VEP.
128. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010;26(22):2867-73.
129. Woodcock R. Woodcock reading master tests - revised (WRMT-R). Circle Pines, MN: American Guidance Service; 1987.
130. Torgesen J, Wagner R, Reshotte C. Test of word reading efficiency (TOWRE). Austin: Pro-Ed; 1999.
131. Wilkinson G. Wide range achievement tests - revised (WRAT-R). Wilmington, DE: Wide Range, Inc.; 1993.
132. Wagner R, Torgesen J, Rashotte C. Comprehensive Test of Phonological Processing (CTOPP). Austin, TX: PRO-ED; 1999.

133. Gogarten SM, Sofer T, Chen H, Yu C, Brody JA, Thornton TA, et al. Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics. 2019;35(24):5346-8.
134. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011;89(1):82-93.
135. Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale sequence data. Am J Hum Genet. 2021;108(10):1880-90.
136. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.
137. Nassar LR, Barber GP, Benet-Pages A, Casper J, Clawson H, Diekhans M, et al. The UCSC Genome Browser database: 2023 update. Nuc Scids Res. 2023;51(D1):D1188-D95. (available at http://genome.ucsc.edu)
138. Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Lemma RB, Turchi L, Blanc-Mathieu R, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nuc Acids Res. 2022;50(D1):D165-D73. (available at https://jaspar.genereg.net/).
139. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74. (available at https://www.encodeproject.org/)
140. Wu C. The 5' ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. Nature. 1980;286(5776):854-60.
141. Keene MA, Corces V, Lowenhaupt K, Elgin SCR. DNase I Hypersensitive sites in Drosophila chromatin occur at the 5′ ends of regions of transcription. Proc Natl Acad Sci U S A Biol Sci. 1981;78(1):143-6.
142. McGhee JD, Wood WI, Dolan M, Engel JD, Felsenfeld G. A 200 base pair region at the 5′ end of the chicken adult beta-globin gene is accessible to nuclease digestion. Cell. 1981;27(1):45-55.
143. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinfiormatics. 2015;31(21):3555-7.
144. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, et al. Ensembl 2022. Nucl Acids Res. 2022;50(D1):D988-D95.
145. Schmit F, Korenjak M, Mannefeld M, Schmitt K, Franke C, von Eyss B, et al. LINC, a human complex that is related to pRB-containing complexes in invertebrates regulates the expression of G2/M genes. Cell Cycle. 2007;6(15):1903-13.
146. Litovchick L, Sadasivam S, Florens L, Zhu X, Swanson SK, Velmurugan S, et al. Evolutionarily conserved multisubunit RBL2/p130 and E2F4 protein complex represses human cell cycle-dependent genes in quiescence. Mol Cell. 2007;26(4):539-51.
147. Azcoitia I, Yague JG, Garcia-Segura LM. Estradiol synthesis within the human brain. Neuroscience. 2011;191:139-47.
148. Lu Y, Sareddy GR, Wang J, Wang R, Li Y, Dong Y, et al. Neuron-derived estrogen regulates synaptic plasticity and memory. J Neurosci. 2019;39(15):2792-809.
149. Azcoitia I, Mendez P, Garcia-Segura LM. Aromatase in the human brain. Androg Clin Res Ther. 2021;2(1):189-202.
150. Dalton CJ, Lemmon CA. Fibronectin: Molecular structure, fibrillar structure and mechanochemical signaling. Cells. 2021;10(9).

151. Muniz AJ, Topal T, Brooks MD, Sze A, Kim DH, Jordahl J, et al. Engineered extracellular matrices facilitate brain organoids from human pluripotent stem cells. Ann Clin Transl Neurol. 2023.
152. Mahajan D, Kancharla S, Kolli P, Sharma AK, Singh S, Kumar S, et al. Role of fibulins in embryonic stage development and their involvement in various diseases. Biomolecules. 2021;11(5).
153. Konishi Y, Ikeda K, Iwakura Y, Kawakami K. Six1 and Six4 promote survival of sensory neurons during early trigeminal gangliogenesis. Brain Res. 2006;1116(1):93-102.
154. Garcez RC, Le Douarin NM, Creuzet SE. Combinatorial activity of Six1-2-4 genes in cephalic neural crest cells controls craniofacial and brain development. Cell Mol Life Sci. 2014;71(11):2149-64.
155. Chen R, Hou Y, Connell M, Zhu S. Homeodomain protein Six4 prevents the generation of supernumerary Drosophila type II neuroblasts and premature differentiation of intermediate neural progenitors. PLoS Genet. 2021;17(2):e1009371.
156. Sanz-Clemente A, Nicoll RA, Roche KW. Diversity in NMDA receptor composition: many regulators, many consequences. Neuroscientist. 2013;19(1):62-75.
157. Kim JI, Kim JW, Park S, Hong SB, Lee DS, Paek SH, et al. The GRIN2B and GRIN2A gene variants are associated with continuous performance test variables in ADHD. J Atten Disord. 2020;24(11):1538-46.
158. Hayashi Y. Molecular mechanism of hippocampal long-term potentiation - Towards multiscale understanding of learning and memory. Neurosci Res. 2022;175:3-15.
159. Endele S, Rosenberger G, Geider K, Popp B, Tamer C, Stefanova I, et al. Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. Nat Genet. 2010;42(11):1021-6.
160. Dimassi S, Andrieux J, Labalme A, Lesca G, Cordier MP, Boute O, et al. Interstitial 12p13.1 deletion involving GRIN2B in three patients with intellectual disability. Am J Med Genet A. 2013;161A(10):2564-9.
161. Platzer K, Yuan H, Schutz H, Winschel A, Chen W, Hu C, et al. GRIN2B encephalopathy: novel findings on phenotype, variant clustering, functional consequences and treatment aspects. J Med Genet. 2017;54(7):460-70.
162. Dorval KM, Wigg KG, Crosbie J, Tannock R, Kennedy JL, Ickowicz A, et al. Association of the glutamate receptor subunit gene GRIN2B with attention-deficit/hyperactivity disorder. Genes Brain Behav. 2007;6(5):444-52.
163. Pan Y, Chen J, Guo H, Ou J, Peng Y, Liu Q, et al. Association of genetic variants of GRIN2B with autism. Sci Rep. 2015;5:8296.
164. Jiang Y, Lin MK, Jicha GA, Ding X, McIlwrath SL, Fardo DW, et al. Functional human GRIN2B promoter polymorphism and variation of mental processing speed in older adults. Aging (Albany NY). 2017;9(4):1293-306.
165. Spearman C. The proof and measurement of association between two things. Amer J Psychol. 1904;15(1):72-101.
166. Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, et al. Determination and inference of eukaryotic transcription factor sequence specificity. Cell. 2014;158(6):1431-43.