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

24 13. Department of Biostatistics, University of Washington, Seattle, Washington, USA

25 14. Department of Psychiatry & Behavioral Sciences, University of Washington, Seattle, WA,

26 USA

27 #Current address: Pacific Northwest Research Institute. 720 Broadway, Seattle, WA 98122.

28 ##Professor Emerita

29 * Corresponding authors: Wendy Raskind: wendyrun@uw.edu

30 Ellen Wijsman: wijsman@uw.edu

31

32 **Short title:** Dyslexia candidate gene MIP seq

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.

Key Words: Dyslexia genetics, molecular inversion probe capture, dyslexia-risk candidate genes

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

96 multiple candidate susceptibility genes have been nominated from genomic regions of interest
97 (ROIs) identified by linkage analyses [30, 31], genome wide association studies, GWAS, [32-
98 34], copy number scan [35], structural chromosome rearrangements [36, 37], or whole genome
99 sequencing [27]
100
101 Of the multiple reported ROIs for dyslexia risk, a small number have the strongest support.
102 DYX1 on chromosome 15q [38-44] and DYX2 on chromosome 6p [39, 45-48] are the loci with
103 the greatest replication across samples. The University of Washington (UW) and Hospital for
104 Sick Children (SickKids, SK) samples have implicated several additional ROIs, with one of the
105 most convincing on chromosome 12q [49]. Candidate genes that have been most studied in the
106 ROIs on chromosomes 15q and 6p include Dynein Axonemal Assembly Factor 4 (*DNAAF4*,
107 previously named *DYX1C1*/*EKN1*) on chromosome 15q [50-52], and Double Cortin Domain
108 Containing 2 (*DCDC2*) and *KIAA0319* on chromosome 6p [53-61]. Support for involvement of
109 these and other candidate risk genes has been reported from a variety of association and linkage
110 studies in humans, and functional studies of brain development in rodents, but evidence is
111 inconsistent [62-68]. There have been failures to detect linkage [69-71] or association [72-78],
112 reports of association to opposite alleles [51, 52, 72], and demonstration of functional
113 competence of the putative risk allele [79]. Meta-analyses have not resolved these conflicts [76,
114 80-82]. Nor have GWAS, which have provided at most weak support for the loci [34, 83-88].
115 This includes a recent large GWAS that did not detect significant evidence of association with
116 any reported candidate dyslexia risk gene [32]. Variability in conclusions across the different
117 study designs is not surprising: genome-wide linkage analyses and GWAS both allow location
118 scans, but linkage analyses are most sensitive to relatively rare variation of large effect and

119 GWAS are most sensitive to common, smaller-effect variation. Neither approach queries all the
120 DNA variation, which requires more-expensive DNA sequencing of (at least) the regions of
121 interest.
122
123 The putative effect of candidate genes on neuronal migration has been used to bolster their
124 credibility [89, 90], given early reports of cortical brain abnormalities in people who were
125 thought to have had dyslexia [91, 92]. However, although cortical abnormalities have been
126 observed with knockdown of the rat orthologues of *Dnaaf4* [79], *Kiaa0319* [62], or *Dcdc2* [93],
127 this is not observed in knockout mice [94-96], and the cortical migration hypothesis remains
128 unproven [65]. Observations that dyslexia candidate genes seem to have a role in ciliogenesis
129 [97, 98], synaptic transmission [99], or axonal growth [100], have led to alternative hypotheses
130 of pathogenesis.
131
132 Although such issues are to be expected in a complex disorder, to date, no causative pathogenic
133 variants have been confirmed. Some possible explanations for this failure include: (1) genetic
134 and/or phenotypic heterogeneity that masks detection in samples ascertained and phenotyped
135 with different criteria; (2) the risk element(s) may alter expression of the protein but not its
136 amino acid sequence; (3) the risk element may escape recognition, but affect splicing; and (4) the
137 number of samples sequenced comprehensively has been too small to have the power to detect
138 variants of modest effect size.
139
140 Recent advancements in DNA sequencing methods now enable the larger scale sequencing
141 efforts that are necessary to evaluate genetic variation in ROIs more comprehensively than was

142 possible earlier. This technology allows us, in this current multi-site study, to investigate the
143 potential role of variants of smaller effect size, non-coding variants, and sample heterogeneity as
144 possible explanations for previous variable results in ROIs implicated in dyslexia. To search for
145 variants that show evidence of association with dyslexia phenotypes, we report the results of
146 genomic sequencing and association analyses in a collection of >2000 participants in families
147 with dyslexia and shared phenotypic measures enrolled at three institutions. We carried out a
148 comprehensive analysis of the coding region and some regulatory element motifs of five putative
149 dyslexia risk genes and found additional evidence for a role in dyslexia risk for *DCDC2*,
150 *KIAA0319*, *GRIN2B* and *CYP19A1*, but not for *DNAAF4*.
151

152 **METHODS**

153 **Participants and Phenotypes Across Sites**

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

165 spelling outcomes or that assess processes related to reading and spelling achievement such as
166 verbal reasoning or phonological memory. To maximize sample size, only test measures
167 administered by more than one of the three cohort sites described below were evaluated. For all
168 cohorts, individuals with evidence of intellectual disabilities, neurological or severe psychiatric
169 disorder, or known genetic disorder associated with language impairment were excluded.
170

171 *University of Washington (UW) Cohort*

172 Recruitment and evaluation of probands and their multigenerational family members were done
173 under a protocol approved by the Institutional Review Board of the UW and are described
174 elsewhere in detail [8, 11] and briefly summarized in the Supporting Information document. The
175 full data set consists of 2079 individuals in 284 families. Of the available subjects, 1347
176 individuals from 278 families provided quality DNA samples for sequencing. Phenotypic data
177 were available for 96.8% of the 1333 samples that passed quality control (QC) testing. Family
178 sizes ranged from 3 to 51 individuals, with a median family size of 12 in 2-4 generations. Self-
179 reported ethnicities were as follows: non-Hispanic White (90%), Asian (2.1%), Native American
180 (2.0%), African American (1.1%), Hispanic (0.8%), and Pacific Islander (0.2%).
181

182 *The SickKids (SK) Cohort*

183 Details of the ascertainment, assessment, and inclusion/exclusion criteria have been described
184 previously [101, 102] and a brief description is provided in the Supportive Information
185 document. Written informed consent and/or assent was obtained from all participants under
186 protocols approved by the Hospital for Sick Children and University Health Network Research
187 Ethics Boards. The SK cohort comprised 816 participants from 245 families (155 with the

188 proband and one or both parents and 90 with two or more children and one or both parents).
189 Phenotypes were available for 99% of children (349 of 351), and 85% of parents (394 of
190 465). The phenotype battery in parents was limited to TOWRE measures of single word
191 reading efficiency and phonological decoding efficiency. Self-reported ancestry was
192 available for 185 (76%) of the families. Of these, 180 (73.5%) reported European or
193 European-Canadian ancestry. The remaining families reported small amounts of indigenous
194 ancestry (3 families, 1.6%), African ancestry (1 individual) and Mexican ancestry (1
195 individual).
196
197 *The University of Houston (UH) Cohort*
198 The subset of families of probands with dyslexia used here is a component of the ongoing
199 collection supported by the Florida Learning Disabilities Research Center (FLDRC). Written
200 informed consent was obtained from all participants under protocols approved by the Research
201 Ethics Boards of Yale University or the University of Houston. The probands were children aged
202 8–17 who had problems reading and were native English speakers. They were registered with
203 FLDRC and recruited through their registration. The probands and their siblings in the same age
204 range were administered a battery of tests including IQ and reading abilities. Designation of
205 affected status required the Wechsler Abbreviated Scale of Intelligence, WASI [103], score ≥80
206 and a score at least 1.5 SD below the mean on 2 of 3 measures of single real- or non-word
207 reading, or 1 SD below the mean on all 3 indicators. Self-reported ethnicities were as follows:
208 African American (13.5%) and Hispanic (86.5%). This cohort consisted of 49 participants (37
209 children and 12 parents) from 15 nuclear families. Parents provided blood samples but were not
210 phenotyped.

211

212 **Molecular Methods**

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

219

220 *Genomic regions investigated.* We targeted DNA sequence in three previously published regions
221 of interest (ROIs) for comprehensive evaluation. Two of the ROIs have been widely
222 investigated, with both positive linkage and association analyses reported by more than one
223 group: the DYX1 locus on chromosome 15q and the DYX2 locus on chromosome 6p. Previous
224 linkage analyses in the UW cohort provided support for DYX1 as a dyslexia candidate region
225 [52, 70] and similar studies in subsets of the SK cohort supported both DYX1 and DYX2 [101,
226 104, 105]. From each of these ROIs, we selected for study two genes based on prior publications
227 supporting a role in dyslexia: in DYX1, *DNAAF*4 [40, 44, 50-52] and C*YP19A1* [37, 106, 107];
228 and in DYX2, *KIAA0319* [54, 62, 77, 108, 109] and *DCDC2* [81, 110, 111]. Intron 2 of *DCDC2*
229 [93] contains READ1 – a complex compound STR polymorphism that was previously proposed
230 as the functional dyslexia-risk component in the DYX2 region [112-114] and was included as
231 part of our capture-design. We also selected the glutamate ionotropic receptor NMDA type
232 subunit 2B *(GRIN2B)*, in a third ROI, on chromosome 12p. This region was among those with
233 the strongest evidence of linkage from the UW family studies, with support across several test-

234 battery items [49, 115, 116]. *GRIN2B* has also been implicated as a dyslexia risk factor [117-
235 119]. For analysis of these five genes, we developed a set of custom capture probes to enable
236 comprehensive evaluation of potential regulatory and splice region sequence variants in addition
237 to coding region variants, as described below.
238
239 *Sample preparation.* For most samples, genomic DNA was extracted from peripheral blood
240 mononuclear cells or Epstein–Barr virus-transformed B-lymphoblastoid cell lines. When only
241 saliva samples were available, DNA was extracted using a DNAGenotek OGR-500 kit
242 (DNAGenotek Inc, Ontario, Canada) according to the manufacturer's instructions.
243
244 *Single molecule molecular inversion probe (smMIP) targeted capture and sequencing.* The
245 capture target consisted of all potentially functional sequences and variants in each ROI around
246 and including each of the five selected genes. This included a total of 277 potential regulatory
247 regions. We used smMIPs to capture targeted DNA with methods described elsewhere [120,
248 121], followed by multiplex sequencing. Additional details are provided in the Supporting
249 Information document.
250
251 We used the UW pipeline [122] to design smMIPs to capture all exons and 10-20bp of flanking
252 intron sequences. This approach ensured capture of the splice site branch A points (RefSeq,
253 hg19/GrCH37 build [121]). All analysis was on this genome build. For the non-coding
254 regulatory regions, we used the ATAC-seq data in Brain Open Chromatin Atlas (BOCA) [123]
255 and ENCODE Consortium for the brain specific (including fetal brain) DNAseI hypersensitive
256 sites [124] to identify chromatin accessible regions (CARs) from 80 kilobases (kb) upstream of

257 the transcription start site (TSS) to the same distance downstream of the 3′ untranslated region
258 (3′ UTR) in the genes of interest. We placed emphasis on the BOCA data that contain CARs for
259 neuronal and non-neuronal cell types across 14 brain regions [125] to identify SNPs associated
260 with neural-specific regulatory elements. The brain regions and abbreviations are described in
261 the Supporting Information document. The smMIPs were designed to minimally overlap each
262 ~200 bp DHS or ATAC-seq site with an additional 50 bp of flanking sequences. The resulting
263 1574 smMIPs and a 55 probe smMIP fingerprinting collection were pooled, tested on a set of
264 control DNAs, and rebalanced, resulting in a final pool of 1569 smMIPs (Supporting Information
265 - Methods and Table S1).
266
267 *Multiplexed next generation sequencing.* Libraries were prepared [120, 126] and pooled for
268 sequencing in batches of 384. Each pool was sequenced using standard paired-end (100bp) rapid
269 run chemistry in a single lane on a HiSeq 2500 (Illumina, San Diego, CA). The final batch
270 contained repeats from previous batches. Using a quality control benchmark requiring that each
271 sample have a minimum of 80% of target bases covered with a depth of at least 10, 2040 (92%),
272 2190 (99%), and 2176 (98%) of the 2209 samples prepared passed the benchmark on
273 chromosomes 6, 12, and 15, respectively. For de-multiplexing, generation of FASTQ files, and
274 annotation of sequence data, we used the same in-house pipeline as for MIP design (see
275 Supporting Information). After QC steps, the sample sizes were 1333, 782 and 46 for the UW,
276 SK and UH samples, respectively, with average genotype completion rates of 98.7%, 99.2% and
277 98.9%, respectively. Variants were annotated using the 1000 Genomes Project (1KGP) and
278 Ensembl Variant Effect Predictor (VEP) [127].
279

280 **Statistical and Bioinformatic Analyses**

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

286

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

294

295 *Phenotypes and adjustments.* Reading-related phenotypes used for our analyses that were
296 directly comparable across the three data sets included word identification (WID) and word
297 attack (WA) from the *Woodcock Reading Mastery Test* (*WRMT*; [129], single word reading
298 efficiency (SWE) and phonological decoding efficiency (PDE) from the *Test of Word Reading*
299 *Efficiency* (*TOWRE;* [130]). The UW and SK cohorts also included spelling (SP) from the *Wide*
300 *Range Achievement Test – Revised* (*WRAT3-R;* [131]), and nonword repetition (NWR) from the
301 *Comprehensive Test of Phonological Processing* [132]. For all reading-related phenotypes
302 considered here, lower scores indicate more impairment on the measure. In addition, only the

303 UW and SK cohorts included verbal IQ (VIQ), which was used as a covariate for some analyses.
304 Two different phenotype adjustments were considered: (1) adjustment for age and sex only
305 (UNADJ), and (2) adjustment for age, sex and VIQ (VIQADJ). The second set of adjustments
306 represents traits that can be interpreted as free from VIQ effects but could only be performed on
307 the UW sample and children in the SK sample, because of the availability of VIQ. For brevity,
308 when discussing results, we use the format Trait:Adjustment (*e.g.*, SWE:UNADJ and SWE:ADJ)
309 to refer to the phenotype without or with VIQ adjustment.
310
311 *Association testing*. Phenotypes were adjusted for covariates within each data set using linear
312 regression, and the residuals were used as the response variable in the association testing.
313 Adjustments were done within data sets both because different editions of tests were used across
314 data sets, and because there were differences in ascertainment. Association testing using
315 GENESIS [133] was done using the combined data sets, with distinct means and variances
316 modeled. Data set-specific variances were used to allow for differences between data sets.
317 Family relationships were accounted for by using the expected pedigree-defined kinship in the
318 covariance matrix via a mixed model. For all SNPs observed in two or more copies in the
319 combined data set, we regressed the phenotype of interest against the dose of the minor allele,
320 yielding a single-variant test. Because SNPs with minor allele frequency (MAF) > 0.01 should
321 result in approximately 40 copies in our dataset, we chose this value as the threshold over which
322 statistical analyses should be reliable. All SNPs, including those with lower MAF, were included
323 in aggregate analyses, with grouping according to their location relative to each candidate gene,
324 defined as 5′ region, 3′ region, exons, and introns. The SKAT-O [134] aggregate test was
325 performed in GENESIS, using weights following a Beta distribution with parameters (1,25) and

326 dependent on the MAF [134]. This choice of weight distribution more heavily weights rare
327 variants, but still allows for a contribution from common variants. The SKAT-O test optimizes
328 power by finding the maximal weighted average of the Burden test (more powerful when most
329 variants are causal and effects are in the same direction) and SKAT test (more powerful when
330 most variants are not causal and effects can be in either direction). Correcting for multiple testing
331 is difficult in these data because there are strong correlations between the phenotypes used, and
332 the focus is on small regions of the genome where we expect substantial linkage disequilibrium
333 (LD). As an approximate correction, we require $p < 0.01$ for significant evidence of association,
334 which corresponds to a p-value of 0.05 divided by 5, the number of candidate genes being
335 considered.
336
337 *Haplotype Estimation and Testing*. When multiple SNPs in LD with one another achieved p
338 <0.01, we used Beagle 5.4 [135] with the 1KGP European reference population (EUR) [136] to
339 obtain phased genotypes, thus providing pairs of phased haplotypes for each subject. For a locus
340 with n common haplotypes, we fit n additive models for each phenotype, where the ith model
341 estimates the dose effect of haplotype i relative to the other haplotypes pooled. GENESIS
342 allowed us to correct for relationships by using the pedigree-defined kinship in the covariance
343 matrix of a mixed model. When more than one haplotype in a region showed significant
344 association in a dose-model with a phenotype, we fitted a model including dose effects for all the
345 significant haplotypes and allowed for interaction between them.
346
347 *Annotating Non-coding Variants in CARs*. We explored the potential impact of all non-coding
348 variants with MAF > 0.01 and $p < 0.01$ on at least one single-variant association test for any of

349 the UNADJ and VIQADJ phenotypes. We annotated variants using the JASPAR tracks on the
350 UCSC Genome Browser [137] as well as the JASPAR database [138]. We considered four
351 characteristics of non-coding variants that together are suggestive of a regulatory effect: 1) the
352 variant is in the peak signal (~200bp) in either ATAC-seq [125] or DNAseI-seq brain profiles
353 [139] in any brain region; 2) it overlaps with a known TF motif, as found in JASPAR [138]; 3)
354 the change disrupts a conserved position in the motif, as assessed by the position frequency
355 matrices in JASPAR; and 4) the TF whose motif is disrupted has an open promoter in the same
356 brain region(s) as the variant [140-142].

357

358 **RESULTS**

359 **Sample Characteristics**

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

370

371 *Phenotypes*. In the tables and text that follow, these abbreviations are used for the tests and the
372 processes they assess: **WID** (*WRMT-R Word Identification* for accuracy of oral reading of real
373 words), **WA** (*WRMT-R Word Attack* for accuracy of oral reading of nonwords), **SP** (*WRAT-3* or
374 *WRAT-R Spelling* for written spelling of orally dictated words), **SWE** (*TOWRE* speed of oral
375 reading of real words), **PDE** (*TOWRE* speed of oral reading of nonwords), **NWR** (*CTOPP*
376 Nonword Repetition NWR for phonological memory). Table 1 shows sample sizes for UNADJ
377 and VIQADJ phenotypes for each dataset and the combined dataset. The variability in sample
378 numbers included in the analyses reflects differences in the phenotyping protocols at the three
379 institutions. VIQ scores were obtained for parents and children in the UW cohort, only for
380 children in the SK cohort, and were not obtained for the UH cohort; therefore, the VIQADJ
381 samples include only the UW and SK cohorts. For SWE and PDE, the UNADJ dataset is
382 substantially larger than the VIQADJ dataset, because VIQ was not available for SK parents.
383 Results presented here in the main text focus on the larger, UNADJ, dataset except when
384 findings are substantially different for VIQADJ.
385

386 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**

387 1. Word identification (WID) and word attack (WA) subtests of the *Woodcock Reading Mastery Test, WRMT* [129];
388 spelling (SP) subtest of the *Wide Range Achievement Test – Revised, WRAT3-R* [131]; single word reading
389 efficiency (SWE) and phonological decoding efficiency (PDE) subtests of the *Test of Word Reading Efficiency,*
390 *TOWRE* [130]; and non-word repetition (NWR) subtest of the *Comprehensive Test of Phonological Processing*
391 [132]. Minor differences in samples size reflect occasional missing values.
392 * Includes scores for parents
393
394 Supplemental Tables S2 and S3 contain demographic data for the samples used in the UNADJ
395 and VIQADJ analyses with means and standard deviations for the phenotypes in each group. The
396 average VIQ score in the UW data set is almost a standard deviation higher than in the SK data
397 set, as might be expected from the difference in sample selection between the two samples. This
398 is supported by noting that the average score in the UW data set of children (109.7) is not
399 significantly greater than that expected (106.4) from restricting enrollment to VIQ > 90 in a
400 random sample. All the phenotypes have means around zero because they are the residuals from
401 a linear model. The means are not exactly zero because the adjustments were done on a larger
402 data set than only the genotyped participants. Consideration of the SD column demonstrates that
403 the phenotypes fall into two categories: WID, WA and SP where the pre-adjustment value was a
404 standard score, and SWE, PDE and NWR where the pre-adjustment value was a z-score. This
405 difference is reflected in the magnitude of the effect sizes estimated for phenotypes in each
406 category.

407

408 **Association Analyses**

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

415

416 Table 2: Association (p<0.01) of common (MAF > 0.01) variant(s) with UNADJ and VIQADJ
417 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*	*	-	-	-	-	-	*	-	-	*	*	*

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

419

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.

423 *DYX1 on chromosome 15*. Of the two genes investigated in DYX1, only *CYP19A1* shows
424 evidence of contribution of common variants to the traits analyzed at a significance threshold of
425 $p < 0.01$ whether in the larger UNADJ sample or the smaller VIQADJ sample (Tables 2, S7, and
426 S8). Variants in *CYP19A1* provide significant evidence of association to multiple phenotypes
427 that is robust to presence or absence of adjustment for VIQ (Table 2). In the UNADJ analysis,
428 three SNPs each provide evidence of association with one or more phenotype (Table 4). One
429 SNP, rs55712458, provides highly significant *p*-values for association with four of the six traits,
430 as well as providing the strongest evidence for association across all analyses. This result was
431 robust to phenotypic adjustment, UNADJ vs. VIQADJ (Tables 4 and S4). For three of the four
432 phenotypes with significant evidence of association (WID, WA, and SP), the strength of the
433 evidence is considerably stronger than required to pass our initial screen, particularly in the
434 UNADJ analysis. These results are also stronger than thresholds used in conventional GWAS
435 scans, after factoring in the number of genes evaluated here. A second SNP (rs12594395),
436 located very close to rs55712458, also provides significant evidence for association with
437 WID:UNADJ and has a nominal $p < 0.05$ for WID:VIQADJ (Table S9). Finally, the rare allele
438 of rs141853734 is significantly associated with decreased performance on both WA:UNADJ and
439 WA:VIQADJ.

440 Table 4: Common variants (MAF ≥ 0.01) in *CYP19A1*, UNADJ phenotypes. Effect size (*p*-value) for dose of the rarer

441 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

443 the alternate allele.

444 2. Build GRCh37/hg19

445 3. DS: downstream, I: intronic, E: exonic, US: upstream

446 4. VEP build38

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

455

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

470 respectively) affect residues that are immediate neighbors in an exon predicted to code a

471 Fibronectin Type III domain [144].

472

473 Table 5: Common variants (MAF ≥ 0.01) in *KIAA0319*, UNADJ phenotypes. Effect size (*p*-value) for dose of

474 the rarer allele with adjustment for non-European ancestry. Variants appear in this table if $p < 0.01$ for at least

475 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)

476 1. Phenotypes as described in Table 1. Bold indicates SNP/phenotype combinations where $p < 0.01$. – indicates *p*-values ≥ 0.05.

477 2. Build GRCh37/hg19

478 3. Freq refers to the alternate allele.

479 4. DS: downstream, I: intronic, E: exonic, US: upstream

480 5. VEP build38

481

482 The common variants implicated in *KIAA0319* display sensitivity to covariate adjustment and,
483 except for the two missense variants in the common variant cluster, are all in non-coding regions.
484 Aggregate tests also implicate non-coding DNA in the downstream region of *KIAA0319*, and
485 common SNPs with significant association with SWE:UNADJ are in the downstream region
486 (rs114526784 and rs114979321, see Table 5). Finally, there are some variants that only show
487 significant evidence for association with WA:VIQADJ, including rs9295627 and rs9461045
488 (Table S5). These two SNPs provide nominal evidence for WA:UNADJ (Table S7). SNP
489 rs9461045, here associated with WA:VIQADJ and PDE:VIQADJ (Table S5), was previously
490 shown to affect expression of *KIAA0319* in a different sample [62]. However, WA:UNADJ
491 provides only nominal evidence of association (Table S7), demonstrating moderate sensitivity to
492 the covariate adjustment for this particular SNP. All the SNPs implicated by these analyses are in
493 the large 77kb region that has been previously shown to be associated with reading related
494 phenotypes [53].
495
496 Association testing for SNPs in *DCDC2* (Tables 2 and 3) showed sensitivity to covariate
497 adjustment. Table 6 shows two neighboring upstream variants (rs142310124 and rs116652616)
498 significantly associated with SP:UNADJ and SWE:UNADJ, with suggestive evidence for
499 association with WA:UNADJ and WID:UNADJ. In this gene, with adjustment of phenotypes for
500 VIQ, additional intronic variants met criteria for significance for association with SWE:VIQADJ
501 (Table 6). All these additional variants show suggestive evidence for association with
502 SWE:UNADJ at $p < 0.05$ (Table S7). Similarities in minor allele frequencies and estimated
503 effect sizes suggest the presence of at least three haplotype blocks that underlie the region: Block
504 A, rs72830851- rs68182041; Block B, rs12196880 – rs7763196; and Block C, rs77743903 –

505 rs116652616 excluding rs4599626. Block C contains the variants that are also significant with
506 VIQ adjustment (Table 6). Examination of LD among these 9 SNPs in 1KGP EUR [143] shows
507 strong disequilibrium (D′ > 0.94) within blocks A, B, and C, and between blocks A and B. The
508 same pattern is seen within 1KGP AFR [143] with the addition of strong disequilibrium between
509 all three blocks. Due to the strong disequilibrium in each block, we cannot statistically
510 distinguish the effects of individual variants; however, bioinformatics prioritization (described
511 below) provides further insights.
512

513 Table 6: Common variants (MAF ≥ 0.01) in *DCDC2*, UNADJ and VIQADJ phenotypes. Effect size (*p*-value) for

514 dose of the rarer allele with adjustment for non-European ancestry. Variants appear in this table if $p < 0.01$ for at

515 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.

522 Evaluation of haplotype effects in *DCDC2* provides evidence for more than one important
523 variant in the region. Table 7 shows the six most common haplotypes defined by the nine SNPs
524 in blocks A, B and C, along with haplotype dosage effects as estimated in six separate linear
525 models (Table 7, left). Neither the dosage of the most common haplotype (1: ATC-CGA-AAG)
526 nor haplotypes 3, 5, or 6 are associated with SWE:VIQADJ, whereas haplotypes 2 (CAT-CGA-
527 AAG) and 4 (ATC-TTG-GCA) are associated with significant effects on SWE:VIQADJ that are
528 in opposite directions. Haplotype 2 carries the rarer of two common haplotypes in block A (**CAT**
529 in Table 7) and is associated with an increase in SWE:VIQADJ. Conversely, the rarer of two
530 common haplotypes in block C (**GCA** in Table 7), on haplotype 4, is associated with a decrease
531 in SWE:VIQADJ. The right side of Table 7 shows the final model in which doses of haplotypes
532 2 and 4 are both included. The estimated effect sizes are very similar to those found in the
533 individual models, and there is no evidence ($p = 0.16$, model not shown) of interaction between
534 haplotypes 2 and 4. These results indicate that the rare haplotype in block A is associated with an
535 increase in SWE:VIQADJ, and the rare haplotype in block C is independently associated with a
536 decrease in SWE:VIQADJ. The region spanned by these haplotypes includes READ1 (a repeat
537 element previously proposed as the causal variant in this region [112-114], which is located
538 between blocks B and C. Due to poor performance of probes targeting READ1 (Figure S1), we
539 were not able to investigate READ1 directly.

540 Table 7: Haplotype models in *DCDC2,* associated with SWE:VIQADJ. The left side shows
541 effect estimates are from six individual haplotype dosage models. The rarer haplotype in each
542 block is in **bold**. The final model on the right shows effect estimates when dose effects are
543 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					

544

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

559 association in the VIQADJ phenotypes as in the stronger test results from the UNADJ
560 phenotypes.
561
562 *GRIN2B on chromosome 12*. In *GRIN2B*, several common variants show significant, robust
563 evidence of association, primarily with NWR. Four variants in the downstream region and an
564 intronic variant are associated with NWR:UNADJ (Table 8). Three of the downstream variants
565 (rs11055486, rs10772689, and rs4764003) also meet the criterion for association with
566 NWR:VIQADJ (Table S6) and the fourth (rs7307659) gives suggestive evidence of linkage. A
567 single intronic variant (rs61914322) is significantly associated with poorer performance on both
568 WID:UNADJ and SWE:UNADJ, and gives suggestive evidence for WID:VIQADJ and
569 SWE:VIQADJ. Aggregate variant testing in *GRIN2B* shows significant results only for the SP
570 phenotype. Results in Table 3 show significant evidence of association between variants in the
571 downstream region of *GRIN2B* and SP:UNADJ. Variants in intronic regions show significant
572 evidence of association with SP:VIQADJ and provide suggestive evidence of association with
573 SP:UNADJ. There is only suggestive support for a role of variants in the exonic regions for
574 SP:VIQADJ. All such variants are in the 3′ or 5′ UTR regions or lead to synonymous changes,
575 and none provide statistical support for a role on the phenotype.

576 Table 8: Common variants (MAF ≥ 0.01) in *GRIN2B*, UNADJ phenotypes. Effect size (*p*-value)
577 for dose of the rarer allele with adjustment for non-European ancestry. Variants appear in this
578 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)**	-	-

579 1. Phenotypes as described in Table 1. Bold indicates SNP/phenotype
580 combinations where $p < 0.01$. – indicates *p*-values ≥ 0.05.
581 2. Build GRCh37/hg19
582 3. Freq refers to the alternate allele.
583 4. DS: downstream, I: intronic, E: exonic, US: upstream
584 5. VEP build38
585

586 *Non-coding variants in CARs*. Bioinformatic investigation of JASPAR, an open access database
587 of non-redundant TFs [138] identified ten variants that are within the peak of a CAR and are
588 predicted to disrupt a TF binding site in brain and/or or neuronal tissue (Table 9). These potential
589 regulatory elements are in four of the five candidate genes, adding further support for a possible
590 role in brain development and/or functioning that leads to dyslexia. Of the total 27 non-coding
591 variants identified by association tests (Table S9), 14 fail to meet both requirements of falling in
592 a peak profile identified by ATAC-seq or DNAaseI-seq and overlapping with a known TF motif
593 and were not considered further. The remaining 13 variants reside in CARs and collectively

594 overlap 50 TF motifs. Nineteen of these variant-motif pairs involve a change in a conserved base

595 of the motif, suggesting a likely impact on TF binding (Table 9).

596 Table 9: Non-coding variants that are in a CAR and are predicted to disrupt binding of the named transcription

597 factor. Bold in the right-hand column indicates a tissue in which both the SNP and the promotor of the TF are in a

598 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.

606 In *CYP19A1* (Table 4), two of the three non-coding variants potentially have a regulatory role
607 (Table 9). Variant rs12594395 is located 16.1 kb downstream of the 3´UTR in an occipital lobe
608 specific CAR with a predicted loss-of-binding of the TF PRDM1. PRDM1's promoter is in open
609 chromatin in the same tissue, suggesting it is active in this brain region [140-142] and its
610 function may be disrupted by rs12594395. Variant rs141853734 simultaneously affects two
611 overlapping motifs for TFs KLF9 and ZSCAN4 in neuronal cells of the primary motor cortex
612 (PMC), nucleus accumbens (NAC) and putamen (PUT). KLF9 has an active/poised promoter in
613 these brain regions while ZSCAN4 does not.
614
615 In *KIAA0319*, one of two non-coding variants that are not part of the haplotype that carries
616 missense variants (Table 5) may affect TF binding. One (rs114526784) is predicted to disrupt a
617 TF binding site (Table 9). The other (rs114526784) is located on overlapping motifs for TFs
618 ZNF384 and LIN54 in neuronal cells of the PUT. ZNF384 is unperturbed by the nucleotide
619 substitution, but the variant changes the LIN54 motif in a highly conserved fourth residue of the
620 motif that predicts a loss of TF binding. LIN54 has been shown to be a member of the DREAM
621 complex that functions as either a transcription activator or repressor depending on the context
622 [145, 146]. The second SNP, rs11497321, is in a CAR that is present in fetal brain and neuronal
623 cells of 11 adult brain regions, but the resulting nucleotide substitution in NrfI and
624 ELK1:HOXA1 is not predicted to significantly affect binding of either TF.
625
626 Several non-coding variants throughout *DCDC2* (Table 6) have the potential for a regulatory
627 role. In the long region defined by the associated haplotype blocks A, B and C described above,
628 there is at least one non-coding variant in each that is predicted to interfere with TF binding. In

629 block A, all three variants are in CARs in some brain tissues, but only rs72830859 in a CAR in
630 mediodorsal thalamus (MDT) results in loss-of-binding mutations predicted to affect TFs
631 PHOX2B, NFIL3 and HLF. Both NFIL3 and HLF have active/poised promoters in all brain
632 regions including MDT. Block B contains a cluster of three variants that span 289 bp and reside
633 within a single CAR that is present in fetal brain and neuronal cells of eight different brain
634 regions. Two of these variants disrupt TF motifs. Variant rs12198320 in Block B is centered
635 within the CAR and disrupts TF motifs for TCF7L1, TCF7L2, and Hnf1A. Hnf1A is inactive in
636 CNS, but TCF7L1 has an active/poised promoter in the neuronal cells of PUT, and TCF7L2 has
637 an active/poised promoter in many neuronal cell types of the brain. Variant rs7763196 in Block B
638 disrupts motifs for TFs CDX4 and HOXD10 in neuronal cells from five brain regions. CDX4 is
639 not expressed in the central nervous system, but HOXD10 has an active/poised promoter in
640 bipolar neurons. Block C variants span 80 kb from the second intron of *DCDC2* to the ninth
641 intron of the downstream gene *MRS2*. One variant, rs142310124, is in a CAR that is specific for
642 neuronal cells in NAC and PUT and is predicted to disrupt motifs for four different TFs
643 (POU2F3, POU3F3, PHOX2B and LIN54). Of these four TFs only LIN54 has an active/poised
644 promoter in PUT. Additional variants rs9379643 and rs61232212 in *DCDC2* are also predicted
645 by JASPAR to disrupt TF binding sites, but neither is present in any neuronal cell types.
646
647 Three non-coding variants in *GRIN2B* are predicted to disrupt TF binding. Variants rs11055486
648 and rs4764003 are both in neuronal-cell specific CARs in MDT and affect the TF motif of Six4
649 and MZF1, respectively. Both Six4 and MZF1 also have open promoters in MDT. Variant
650 rs12184498 is in CARs present in seven different brain regions and affects eight different TF
651 motifs for KLF9, KLF1, PRDM9, SPIB, EHF, Spi1, SP3 and ZNF667. KLF9, PRDM9, SP3 and

652 ZNF667. All these TFs have open promoters in the relevant tissues, suggesting the availability of
653 these TF for binding. Variant rs10772689 is located within a CAR in some neuronal cells but
654 does not overlap with any known TF motifs. Variant rs7307659 is upstream of a neuronal-
655 specific PUT CAR and was captured in the smMIP experiments, but because it is outside of a
656 peak signal, we are unable to speculate regarding its role.
657

658 **DISCUSSION**

659 Here we provided results of a comprehensive investigation of underlying genomic variation in
660 and surrounding five genes with prior evidence for an inherited effect on endophenotypes of
661 dyslexia risk. The MIP sequencing approach that we used allowed inclusion of many more
662 participants and variants than have been previously considered and provided an agnostic
663 approach for identifying underlying causal variants. Reliance of previous studies on detectable
664 linkage disequilibrium between a causal variant and a small number of nearby genotyped
665 polymorphisms is a possible cause of conflicting results across laboratories [51, 52, 72]. In
666 contrast, in the current study, we evaluated much of the gene neighborhoods, focusing on
667 sequence that had the greatest potential for bioinformatic interpretation: protein coding regions
668 and potential regulatory sites upstream and downstream of the candidate genes. Variants
669 evaluated span a wide allele frequency range and fall in both coding and non-coding DNA.
670

671 The variants that met our threshold for association and further bioinformatic consideration
672 represent primarily non-coding DNA, with no clearly pathogenic coding variants. Although
673 limited to a small number of selected genes and gene neighborhoods, the results provide an
674 initial prediction of the types of variation that are likely to be more broadly identified through

675 evaluation of DNA sequence in genome-scale studies of specific learning disabilities such as
676 dyslexia. We speculate that variation in coding sequence that results in dramatic alteration of
677 protein structure or function, as is typical of Mendelian disorders, is unlikely to play a role in
678 dyslexia. Such protein-coding variants generally are rare and subject to negative selection, with
679 large impacts on the phenotype. Dyslexia is a phenotype that only is recognized in the presence
680 of widespread literacy, and non-coding variants with subtle effects on gene expression or control
681 are more likely to be relevant. Selection against such variants, with their weak effects on the
682 phenotype, would have been relatively ineffective in the small populations that were typical until
683 very recently in human history. Instead, stochastic effects, such as genetic drift, would have had
684 a role in driving changes in allele frequencies.

685

686 Of the variants implicated herein, common variants in *CYP19A1* provide the most significant
687 support for association with all the reading- and spelling-related phenotypes we studied.
688 Bioinformatics evaluation suggests a possible role for regulatory variation that may alter the
689 amount of mRNA, and thus protein, produced. The strongest statistical result was for a variant
690 (rs55712458) in the downstream region of the gene. This variant does not overlap any TF motifs,
691 but it is less than 300 bp away from rs12594395, which occupies an occipital lobe specific CAR,
692 and is expected to affect the binding of the PRDM1 TF. Another variant in *CYP19A1* affects a
693 motif for KLF9, which is a neuronal TF with an active/poised promoter in all brain regions,
694 including fetal brain. In addition to these individual-SNP results, aggregate testing suggests that
695 some probably rare, intronic variants may also be involved. Finally, *CYP19A1* was also
696 previously nominated as a dyslexia-risk gene through identification of the breakpoint of a
697 t(2;15)(p12;q21) translocation that disrupted the promoter region of the gene in a person with

698 dyslexia [37]. *CYP19A1* encodes the estrogen synthase CP19A that converts C19 androgens to
699 C18 estrogens and is responsible for local synthesis of estrogens outside of the reproductive
700 system. In the brain it is expressed from prenatal stages to adulthood [147] in multiple cell types
701 where it regulates synaptic plasticity and plays a role in cognition, memory and language, and
702 many other functions [148, 149].

703

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

717

718 Finally, common variants downstream of *GRIN2B* are associated with decreased performance on
719 phonological non-word memory, the only phenotype on which this gene had an effect. None of

720 the other studied genes showed an association with this phenotype. One *GRIN2B* variant affects
721 the TF motif of a homeobox protein that may have a role in the differentiation, maturation, and
722 survival of neuronal cells [153-155]. *GRIN2B* was selected for the current study because it lies in
723 a region with evidence of linkage for a phonological non-word memory phenotype in the UW
724 cohort [49] and was associated with verbal memory phenotypes in two European dyslexia
725 datasets [117, 118]. *GRIN2B* encodes GluN2B, one of the glutamate-binding subunits of the
726 tetrameric N-methyl-D-aspartate ionotropic glutamate receptors (NMDARs) that are important
727 for neuronal development and plasticity [156]. GluN2B is highly expressed prenatally in the
728 brain where it is involved in learning and working memory via its role in synaptic plasticity and
729 enhanced long-term potentiation [157, 158]. Pathogenic coding variants and deletions in
730 *GRIN2B* also cause a spectrum of neurodevelopmental disorders [159-161]. Non-coding variants
731 in *GRIN2B* have also been associated with short term and working memory, intelligence quotient
732 and cognitive impairments in dyslexia [117, 118] and with other cognitive and behavioral traits
733 [162-164].

734

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

741 mean [165], with consequent loss of power to detect the effect of the predictor – the SNP in this
742 case.

743

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

758

759 In summary, we provide evidence that variants in or near *DCDC2*, *KIAA0319*, *CYP19A1*, and
760 *GRIN2B* play a role in risk for dyslexia. Given the large sample size, these results argue strongly
761 against the causative involvement of large-effect coding variants in any of the five candidate
762 genes. Instead, these results support a potential role in transcriptional regulation that alters the
763 quantity of RNA produced. The non-coding variant/transcription factor/brain-region

764 combinations we identified provide a starting point for future work on the role of non-coding
765 variants in dyslexia. These results also illustrate some of the challenges that the field will face in
766 identifying causal variants that may act through gene regulation rather than alteration of protein
767 sequences. Use of whole-genome sequence (WGS), especially long-read, would capture
768 regulatory elements with fewer complications, including detection of alterations in repeat
769 sequences that might reside in deep intronic or intergenic regions. However, the WGS approach
770 adds significant cost that could critically limit the number of samples used. The most feasible
771 approach to corroboration of variants and haplotypes of interest discussed herein will therefore
772 require evaluation in other dyslexia sample sets followed by functional studies in an appropriate
773 cell model to begin to determine biological relevance [166], an endeavor that is well beyond the
774 scope of the current analysis.

775

776 **ACKNOWLEDGEMENTS**

777 We are grateful to the family members who volunteered their time to participate in the research.
778 John Wolff, Hiep Nguyen, and Edith PA Fuerte provided excellent technical, computational, and
779 bioinformatics assistance. We thank the many graduate student assistants who administered the
780 test batteries.

781

782 **LITERATURE CITED**

783 1. Lyon GR, Shaywitz SE, Shaywitz B. A definition of dyslexia. Annals of Dyslexia.
784 2003;53:1-14.
785 2. Katusic SK, Colligan RC, Barbaresi WJ, Schaid DJ, Jacobsen SJ. Incidence of reading
786 disability in a population-based birth cohort, 1976-1982, Rochester, Minn. Mayo Clin Proc.
787 2001;76(11):1081-92.
788 3. Cai L, Chen Y, Hu X, Guo Y, Zhao X, Sun T, et al. An epidemiological study of Chinese
789 children with developmental dyslexia. J Dev Behav Pediatr. 2020;41(3):203-11.
790 4. Wagner RK, Zirps FA, Edwards AA, Wood SG, Joyner RE, Becker BJ, et al. The
791 prevalence of dyslexia: A new approach to its estimation. J Learn Disabil. 2020;53(5):354-
792 65.
793 5. Yang L, Li C, Li X, Zhai M, An Q, Zhang Y, et al. Prevalence of developmental dyslexia in
794 primary school children: A systematic review and meta-analysis. Brain Sci. 2022;12(2).
795 6. Barbiero C, Montico M, Lonciari I, Monasta L, Penge R, Vio C, et al. The lost children: The
796 underdiagnosis of dyslexia in Italy. A cross-sectional national study. PloS one.
797 2019;14(1):e0210448.
798 7. Shaywitz SE, Shaywitz JE, Shaywitz BA. Dyslexia in the 21st century. Curr Opin
799 Psychiatry. 2021;34(2):80-6.
800 8. Raskind W, Hsu L, Berninger V, Thomson J, Wijsman E. Familial aggregation of dyslexia
801 phenotypes. Behav Genet. 2000;30:385-95.
802 9. Flannery KA, Liederman J, Daly L, Schultz J. Male prevalence for reading disability is
803 found in a large sample of Black and White children free from ascertainment bias. J Int
804 Neuropsychological Soc. 2000;6(4):433-42.

805 10. Quinn JM, Wagner RK. Gender differences in reading impairment and in the identification
806 of impaired readers: results from a large-scale study of at-risk readers. J Learn Disabil.
807 2015;48(4):433-45.
808 11. Berninger V, Abbott R, thomson JB, Raskind WH. Language phenotype for reading and
809 writing disability: a family approach. Sci Stud Read. 2001;5:59-105.
810 12. Hatcher J, Snowling MJ, Griffiths YM. Cognitive assessment of dyslexic students in higher
811 education. Br J Educ Psychol. 2002;72(Pt 1):119-33.
812 13. Wilson A, Lesaux N. Persistence of phonological processing deficits in college students
813 with dyslexia who have age-appropriate reading skills. J Learn Disabil. 2001;34:394-400.
814 14. Morris D, Turnbull P. A survey-based exploration of the impact of dyslexia on career
815 progression of UK registered nurses. J Nurs Manag. 2007;15(1):97-106.
816 15. Gerber PJ. The impact of learning disabilities on adulthood: a review of the evidenced-based
817 literature for research and practice in adult education. J Learn Disabil. 2012;45(1):31-46.
818 16. McLaughlin MJ, Speirs KE, Shenassa ED. Reading disability and adult attained education
819 and income: evidence from a 30-year longitudinal study of a population-based sample. J
820 Learn Disabil. 2014;47(4):374-86.
821 17. Andreola C, Mascheretti S, Belotti R, Ogliari A, Marino C, Battaglia M, et al. The
822 heritability of reading and reading-related neurocognitive components: A multi-level meta-
823 analysis. Neurosci Biobehav Rev. 2021;121:175-200.
824 18. van Bergen E, de Jong PF, Plakas A, Maassen B, vdL, A. Child and parental literacy levels
825 within families with a history of dyslexia. J Child Psychol Psychiatry. 2012;53(1):23-36.

826 19. Kirkpatrick RM, Legrand LN, Iacono WG, McGue M. A twin and adoption study of reading
827 achievement: Exploration of shared-environmental and gene-environment-interaction
828 effects. Learn Individ Differ. 2011;21(4):368-75.
829 20. Erbeli F, Rice M, Paracchini S. Insights into dyslexia genetics research from the last two
830 decades. Brain Sci. 2022;12(1):27.
831 21. DeFries J, Fulker D, LaBuda M. Evidence for a genetic aetiology in reading disability of
832 twins. Nature. 1987;329:537-9.
833 22. Gayan J, Olson R. Reading disability: evidence for a genetic etiology. Eur Child Adolesc
834 Psychiatry. 1999;8(Suppl):52-5.
835 23. Fagerheim T, Raeymaekers P, Tonnessen FE, Pedersen M, Tranebjaerg L, Lubs HA. A new
836 gene (DYX3) for dyslexia is located on chromosome 2. J Med Genet. 1999;36(9):664-9.
837 24. Nopola-Hemmi J, Myllyluoma B, Haltia T, Taipale M, Ollikainen V, Ahonen T, et al. A
838 dominant gene for developmental dyslexia on chromosome 3. J Med Genet.
839 2001;38(10):658-64.
840 25. de Kovel CGF, Hol FA, Heister J, Willemen J, Sandkuijl LA, Franke B, et al. Genomewide
841 scan identifies susceptibility locus for dyslexia on Xq27 in an extended Dutch family. J Med
842 Genet. 2004;41(9):652-7.
843 26. Grimm T, Garshasbi M, Puettmann L, Chen W, Ullmann R, Muller-Myhsok B, et al. A
844 novel locus and candidate gene for familial developmental dyslexia on chromosome 4q. Z
845 Kinder Jugendpsychiatr Psychother. 2020;48(6):478-89.
846 27. Carrion-Castillo A, Estruch SB, Maassen B, Franke B, Francks C, Fisher SE. Whole-
847 genome sequencing identifies functional noncoding variation in SEMA3C that cosegregates
848 with dyslexia in a multigenerational family. Hum Genet. 2021;140(8):1183-200.

849 28. Einarsdottir E, Svensson I, Darki F, Peyrard-Janvid M, Lindvall JM, Ameur A, et al.
850 Mutation in CEP63 co-segregating with developmental dyslexia in a Swedish family. Hum
851 Genet. 2015;134(11-12):1239-48.
852 29. Georgitsi M, Dermitzakis I, Soumelidou E, Bonti E. The polygenic nature and complex
853 genetic architecture of specific learning disorder. Brain Sci. 2021;11(5).
854 30. Fisher SE, Francks C, Marlow AJ, MacPhie IL, Newbury DF, Cardon LR, et al.
855 Independent genome-wide scans identify a chromosome 18 quantitative-trait locus
856 influencing dyslexia. Nat Genet. 2002;30(1):86-91.
857 31. Hannula-Jouppi K, Kaminen-Ahola N, Taipale M, Eklund R, Nopola-Hemmi J, Kaariainen
858 H, et al. The axon guidance receptor gene ROBO1 is a candidate gene for developmental
859 dyslexia. PLoS Genet. 2005;1(4):467-74.
860 32. Doust C, Fontanillas P, Eising E, Gordon SD, Wang Z, Alagoz G, et al. Discovery of 42
861 genome-wide significant loci associated with dyslexia. Nat Genet. 2022;54(11):1621-9.
862 33. Gialluisi A, Andlauer TFM, Mirza-Schreiber N, Moll K, Becker J, Hoffmann P, et al.
863 Genome-wide association scan identifies new variants associated with a cognitive predictor
864 of dyslexia. Transl Psychiatry. 2019;9(1):77.
865 34. Price KM, Wigg KG, Feng Y, Blokland K, Wilkinson M, He G, et al. Genome-wide
866 association study of word reading: Overlap with risk genes for neurodevelopmental
867 disorders. Genes Brain Behav. 2020;19(6):e12648.
868 35. Veerappa AM, Saldanha M, Padakannaya P, Ramachandra NB. Genome-wide copy number
869 scan identifies disruption of PCDH11X in developmental dyslexia. Am J Med Genet B
870 Neuropsychiatr Genet. 2013;162B(8):889-97.

871 36. Nopola-Hemmi J, Taipale M, Haltia T, Lehesjoki AE, Voutilainen A, Kere J. Two
872 translocations of chromosome 15q associated with dyslexia. J Med Genet. 2000;37(10):771-
873 5.
874 37. Anthoni H, Sucheston LE, Lewis BA, Tapia-Paez I, Fan X, Zucchelli M, et al. The
875 aromatase gene CYP19A1: several genetic and functional lines of evidence supporting a
876 role in reading, speech and language. Behav Genet. 2012;42(4):509-27.
877 38. Bates TC, Luciano M, Castles A, Coltheart M, Wright MJ, Martin NG. Replication of
878 reported linkages for dyslexia and spelling and suggestive evidence for novel regions on
879 chromosomes 4 and 17. Eur J Hum Genet. 2007;15(2):194-203.
880 39. Grigorenko EL, Wood FB, Meyer MS, Hart LA, Speed WC, Shuster A, et al. Susceptibility
881 loci for distinct components of developmental dyslexia on chromosomes 6 and 15. Am J
882 Hum Genet. 1997;60:27-9.
883 40. Marino C, Citterio A, Giorda R, Facoetti A, Menozzi G, Vanzin L, et al. Association of
884 short-term memory with a variant within DYX1C1 in developmental dyslexia. Genes Brain
885 Behav. 2007;6(7):640-6.
886 41. Morris DW, Robinson L, Turic D, Duke M, Webb V, Milham C, et al. Family-based
887 association mapping provides evidence for a gene for reading disability on chromosome
888 15q. Hum Molec Genet. 2000;9(5):843-8.
889 42. Nöthen MM, Schulte-Korne G, Grimm T, Cichon S, Vogt IR, Muller-Myhsok B, et al.
890 Genetic linkage analysis with dyslexia: evidence for linkage of spelling disability to
891 chromosome 15. Eur Child Adolesc Psychiatry. 1999;8 Suppl 3:56-9.

892 43. Smith SD, Kimberling WJ, Pennington BF, Lubs HA. Specific reading disability:
893 identification of an inherited form through linkage analysis. Science. 1983;219(4590):1345-
894 7.
895 44. Venkatesh SK, Siddaiah A, Padakannaya P, Ramachandra NB. Association of SNPs of
896 DYX1C1 with developmental dyslexia in an Indian population. Psychiatr Genet.
897 2014;24(1):10-20.
898 45. Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC.
899 Quantitative trait locus for reading disability: correction. Science. 1995;268:1553.
900 46. Fisher SE, Marlow AJ, Lamb J, Maestrini E, Williams DF, Richardson AJ, et al. A
901 quantitative-trait locus on chromosome 6p influences different aspects of developmental
902 dyslexia. Am J Hum Genet. 1999;64:146-56.
903 47. Grigorenko EL, Wood FB, Golovyan L, Meyer M, Romano C, Pauls D. Continuing the
904 search for dyslexia genes on 6p. Am J Med Genet B Neuropsychiatr Genet.
905 2003;118B(1):89-98.
906 48. Turic D, Robinson L, Duke M, Morris DW, Webb V, Hamshere M, et al. Linkage
907 disequilibrium mapping provides further evidence of a gene for reading disability on
908 chromosome 6p21.3-22. Mol Psychiatry. 2003;8(2):176-85.
909 49. Brkanac Z, Chapman N, Igo RP, Matsushita M, Nielsen K, Berninger V, et al. Linkage
910 analysis of a nonword repetition endophenotype in families with dyslexia: initial findings
911 and replication. Behav Genet. 2008;38(5):462-75.
912 50. Taipale M, Kaminen N, Nopola-Hemmi J, Haltia T, Myllyluoma B, Lyytinen H, et al. A
913 candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat

914 domain protein dynamically regulated in brain. Proc. Natl. Acad. Sci. U.S.A.
915 2003;100:11553-8.
916 51. Wigg K, Couto J, Feng Y, Anderson B, Cate-Carter T, Macciardi F, et al. Support for EKN1
917 as the susceptibility locus for dyslexia on 15q21. Molec Psychiatry. 2004;9:1111-21.
918 52. Brkanac Z, Chapman N, Matsushita M, Chun L, Nielsen K, E C, et al. Evaluation of
919 candidate genes for DYX1 and DYX2 in families with dyslexia. Am J Med Genet
920 (Neuropsychiatr Genet). 2007;144B:556-60.
921 53. Francks C, Paracchini S, Smith S, Richardson A, Scerri T, Cardon L, et al. A 77-kilobase
922 region of chromosome 6p22.2 is associated with dyslexia in families from the United
923 Kingdom and from the United States. Am J Hum Genet. 2004;75:1046-58.
924 54. Cope N, Harold D, Hill G, Moskvina V, Stevenson J, Holmans P, et al. Strong evidence that
925 KIA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. Am J Hum
926 Genet. 2005;76(4):581-91.
927 55. Harold D, Paracchini S, Scerri T, Dennis N, Cope H, Hill G, et al. Further evidence that the
928 *KIAA0319* gene confers susceptibility to developmental dyslexia. Molec Psychiatry.
929 2006;11:1085-91.
930 56. Scerri TS, Morris AP, Buckingham LL, Newbury DF, Miller LL, Monaco AP, et al.
931 DCDC2, KIAA0319 and CMIP are associated with reading-related traits. Biol Psychiatry.
932 2011;70(3):237-45.
933 57. Venkatesh SK, Siddaiah A, Padakannaya P, Ramachandra NB. Analysis of genetic variants
934 of dyslexia candidate genes KIAA0319 and DCDC2 in Indian population. J Hum Genet.
935 2013;58(8):531-8.

936 58. Eicher JD, Powers NR, Miller LL, Mueller KL, Mascheretti S, Marino C, et al.
937 Characterization of the DYX2 locus on chromosome 6p22 with reading disability, language
938 impairment, and IQ. Hum Genet. 2014;133(7):869-81.
939 59. Matsson H, Huss M, Persson H, Einarsdottir E, Tiraboschi E, Nopola-Hemmi J, et al.
940 Polymorphisms in DCDC2 and S100B associate with developmental dyslexia. J Hum
941 Genet. 2015;60(7):399-401.
942 60. Zhao H, Chen Y, Zhang BP, Zuo PX. KIAA0319 gene polymorphisms are associated with
943 developmental dyslexia in Chinese Uyghur children. J Hum Genet. 2016;61(8):745-52.
944 61. Trezzi V, Forni D, Giorda R, Villa M, Molteni M, Marino C, et al. The role of READ1 and
945 KIAA0319 genetic variations in developmental dyslexia: testing main and interactive
946 effects. J Hum Genet. 2017;62(11):949-55.
947 62. Paracchini S, Thomas AC, Castro S, Lai C, Paramasivam M, Wang Y, et al. The
948 chromosome 6p22 haplotype associated with dyslexia reduces the expression of *KIAA0319*,
949 a novel gene involved in neuronal migration. Hum Molec Genet. 2006;15:1659-66.
950 63. Dennis MY, Paracchini S, Scerri TS, Prokunina-Olsson L, Knight JC, Wade-Martins R, et
951 al. A common variant associated with dyslexia reduces expression of the KIAA0319 gene.
952 PLoS Genet. 2009;5(3):e1000436.
953 64. Carrion-Castillo A, Franke B, Fisher SE. Molecular genetics of dyslexia: an overview.
954 Dyslexia. 2013;19(4):214-40.
955 65. Guidi LG, Velayos-Baeza A, Martinez-Garay I, Monaco AP, Paracchini S, Bishop DVM, et
956 al. The neuronal migration hypothesis of dyslexia: A critical evaluation 30 years on. Eur J
957 Neurosci. 2018;48(10):3212-33.

958 66. Riva V, Mozzi A, Forni D, Trezzi V, Giorda R, Riva S, et al. The influence of DCDC2 risk
959 genetic variants on reading: Testing main and haplotypic effects. Neuropsychologia.
960 2019;130:52-8.
961 67. Deng KG, Zhao H, Zuo PX. Association between KIAA0319 SNPs and risk of dyslexia: a
962 meta-analysis. J Genet. 2019;98(1).
963 68. Bieder A, Yoshihara M, Katayama S, Krjutskov K, Falk A, Kere J, et al. Dyslexia candidate
964 gene and ciliary gene expression dynamics during human neuronal differentiation. Mol
965 Neurobiol. 2020;57(7):2944-58.
966 69. Petryshen TL, Kaplan BJ, Liu MF, Field LL. Absence of significant linkage between
967 phonological coding dyslexia and chromosome 6p23-21.3, as determined by use of
968 quantitative-trait methods: confirmation of qualitative analyses. Am J Hum Genet.
969 2000;66:708-14.
970 70. Chapman N, Igo RJ, Thomson J, Matsushita M, Brkanac Z, Holzman T, et al. Linkage
971 analyses of four regions previously implicated in dyslexia: confirmation of a locus on
972 chromosome 15q. Am J Med Genet B Neuropsychiatr Genet. 2004;131B:67-75.
973 71. de Kovel CG, Franke B, Hol FA, Lebrec JJ, Maassen B, Brunner H, et al. Confirmation of
974 dyslexia susceptibility loci on chromosomes 1p and 2p, but not 6p in a Dutch sib-pair
975 collection. Am J Med Genet B Neuropsychiatr Genet. 2008;147(3):294-300.
976 72. Scerri T, Fisher S, Francks C, MacPhie I, Paracchini S, Richardson A, et al. Putative
977 functional alleles of DYX1C1 are not associated with dyslexia susceptibility in a large
978 sample of sibling pairs from the UK. J Med Genet. 2004;41:853-7.

979 73. Bellini G, Bravaccio C, Calamoneri F, Cocuzza MD, Fiorillo P, Gagliano A, et al. No
980 evidence for association between dyslexia and DYX1C1 functional variants in a group of
981 children and adolescents from Southern Italy. J Mol Neurosci. 2005;27(3):311-4.
982 74. Cope N, Hill G, van den Bree M, Harold D, Moskvina V, Green E, et al. No support for
983 association between Dyslexia Susceptibility 1 Candidate 1 and developmental dyslexia.
984 Molec Psychiatry. 2005;10(3):237-8.
985 75. Marino C, Giorda R, Lorusso M, Vanzin L, Salandi N, Nobile M, et al. A family-based
986 association study does not support DYX1C1 on chromosome 15q21.3 as a candidate gene in
987 developmental dyslexia. Eur J Hum Genet. 2005;13(4):491-9.
988 76. Tran C, Gagnon F, Wigg KG, Feng Y, Gomez L, Cate-Carter TD, et al. A family-based
989 association analysis and meta-analysis of the reading disabilities candidate gene DYX1C1.
990 Am J Med Genet B Neuropsychiatr Genet. 2013;162B(2):146-56.
991 77. Carrion-Castillo A, Maassen B, Franke B, Heister A, Naber M, van der Leij A, et al.
992 Association analysis of dyslexia candidate genes in a Dutch longitudinal sample. Eur J Hum
993 Genet. 2017;25(4):452-60.
994 78. Sharma P, Sagar R, Deep R, Mehta M, Subbiah V. Assessment for familial pattern and
995 association of polymorphisms in KIAA0319 gene with specific reading disorder in children
996 from North India visiting a tertiary care centre: A case-control study. Dyslexia.
997 2020;26(1):104-14.
998 79. Wang Y, Paramasivam M, Thomas A, Bai J, Kaminen-Ahola N, Kere J, et al. DYX1C1
999 functions in neuronal migration in developing neocortex. Neuroscience. 2006;143(2):515-
1000 22.

1001 80. Zou L, Chen W, Shao S, Sun Z, Zhong R, Shi J, et al. Genetic variant in KIAA0319, but not
1002 in DYX1C1, is associated with risk of dyslexia: an integrated meta-analysis. Am J Med
1003 Genet B Neuropsychiatr Genet. 2012;159B(8):970-6.
1004 81. Zhong R, Yang B, Tang H, Zou L, Song R, Zhu LQ, et al. Meta-analysis of the association
1005 between DCDC2 polymorphisms and risk of dyslexia. Mol Neurobiol. 2013;47(1):435-42.
1006 82. Shao S, Niu Y, Zhang X, Kong R, Wang J, Liu L, et al. Opposite Associations between
1007 Individual KIAA0319 Polymorphisms and developmental dyslexia risk across populations:
1008 A stratified meta-analysis by the study Ppopulation. Sci Rep. 2016;6:30454.
1009 83. Meaburn EL, Harlaar N, Craig IW, Schalkwyk LC, Plomin R. Quantitative trait locus
1010 association scan of early reading disability and ability using pooled DNA and 100K SNP
1011 microarrays in a sample of 5760 children. Mol Psychiatry. 2008;13(7):729-40.
1012 84. Field LL, Shumansky K, Ryan J, Truong D, Swiergala E, Kaplan BJ. Dense-map genome
1013 scan for dyslexia supports loci at 4q13, 16p12, 17q22; suggests novel locus at 7q36. Genes
1014 Brain Behav. 2013;12(1):56-69.
1015 85. Luciano M, Evans DM, Hansell NK, Medland SE, Montgomery GW, Martin NG, et al. A
1016 genome-wide association study for reading and language abilities in two population cohorts.
1017 Genes Brain Behav. 2013;12(6):645-52.
1018 86. Gialluisi A, Newbury DF, Wilcutt EG, Olson RK, DeFries JC, Brandler WM, et al.
1019 Genome-wide screening for DNA variants associated with reading and language traits.
1020 Genes Brain Behav. 2014;13(7):686-701.
1021 87. Gialluisi A, Andlauer TFM, Mirza-Schreiber N, Moll K, Becker J, Hoffmann P, et al.
1022 Genome-wide association study reveals new insights into the heritability and genetic
1023 correlates of developmental dyslexia. Mol Psychiatry. 2021;26(7):3004-17.

1024 88. Eising E, Mirza-Schreiber N, de Zeeuw EL, Wang CA, Truong DT, Allegrini AG, et al.
1025 Genome-wide analyses of individual differences in quantitatively assessed reading- and
1026 language-related skills in up to 34,000 people. Proc Natl Acad Sci U S A.
1027 2022;119(35):e2202764119.
1028 89. Galaburda AM, LoTurco J, Ramus F, Fitch RH, Rosen GD. From genes to behavior in
1029 developmental dyslexia. Nat Neurosci. 2006;9(10):1213-7.
1030 90. Paracchini S, Scerri T, Monaco AP. The genetic lexicon of dyslexia. Annu Rev Genomics
1031 Hum Genet. 2007;8:57-79.
1032 91. Galaburda AM, Kemper TL. Cytoarchitectonic abnormalities in developmental dyslexia: a
1033 case study. Ann Neurol. 1979;6(2):94-100.
1034 92. Kaufmann WE, Galaburda AM. Cerebrocortical microdysgenesis in neurologically normal
1035 subjects: a histopathologic study. Neurology. 1989;39(2 Pt 1):238-44.
1036 93. Meng H, Smith SD, Hager K, Held M, Liu J, Olson RK, et al. DCDC2 is associated with
1037 reading disability and modulates neuronal development in the brain. Proc Natl Acad Sci U S
1038 A. 2005;102(47):17053-8.
1039 94. Martinez-Garay I, Guidi LG, Holloway ZG, Bailey MA, Lyngholm D, Schneider T, et al.
1040 Normal radial migration and lamination are maintained in dyslexia-susceptibility candidate
1041 gene homolog Kiaa0319 knockout mice. Brain Struct Funct. 2017;222(3):1367-84.
1042 95. Rendall AR, Tarkar A, Contreras-Mora HM, LoTurco JJ, Fitch RH. Deficits in learning and
1043 memory in mice with a mutation of the candidate dyslexia susceptibility gene Dyx1c1.
1044 Brain Lang. 2017;172:30-8.

1045 96. Wang Y, Yin X, Rosen G, Gabel L, Guadiana SM, Sarkisian MR, et al. Dcdc2 knockout
1046 mice display exacerbated developmental disruptions following knockdown of doublecortin.
1047 Neuroscience. 2011;190:398-408.
1048 97. Massinen S, Hokkanen ME, Matsson H, Tammimies K, Tapia-Paez I, Dahlstrom-Heuser V,
1049 et al. Increased expression of the dyslexia candidate gene DCDC2 affects length and
1050 signaling of primary cilia in neurons. PloS one. 2011;6(6):e20580.
1051 98. Tarkar A, Loges NT, Slagle CE, Francis R, Dougherty GW, Tamayo JV, et al. DYX1C1 is
1052 required for axonemal dynein assembly and ciliary motility. Nat Genet. 2013;45(9):995-
1053 1003.
1054 99. Che A, Truong DT, Fitch RH, LoTurco JJ. Mutation of the dyslexia-associated gene Dcdc2
1055 enhances glutamatergic synaptic transmission between layer 4 neurons in mouse neocortex.
1056 Cereb Cortex. 2016;26(9):3705-18.
1057 100. Franquinho F, Nogueira-Rodrigues J, Duarte JM, Esteves SS, Carter-Su C, Monaco AP, et
1058 al. The dyslexia-susceptibility protein KIAA0319 inhibits axon growth through Smad2
1059 signaling. Cereb Cortex. 2017;27(3):1732-47.
1060 101. Couto JM, Livne-Bar I, Huang K, Xu Z, Cate-Carter T, Feng Y, et al. Association of
1061 reading disabilities with regions marked by acetylated H3 histones in KIAA0319. Am J Med
1062 Genet B Neuropsychiatr Genet. 2010;153B(2):447-62.
1063 102. Tran C, Wigg KG, Zhang K, Cate-Carter TD, Kerr E, Field LL, et al. Association of the
1064 ROBO1 gene with reading disabilities in a family-based analysis. Genes Brain Behav.
1065 2014;13(4):430-8.
1066 103. Wechsler D. Wechsler abbreviated scale of intelligence. San Antonio; 1999.

1067 104. Couto JM, Gomez L, Wigg K, Cate-Carter T, Archibald J, Anderson B, et al. The
1068 KIAA0319-like (KIAA0319L) gene on chromosome 1p34 as a candidate for reading
1069 disabilities. J Neurogenet. 2008;22(4):295-313.
1070 105. Elbert A, Lovett MW, Cate-Carter T, Pitch A, Kerr EN, Barr CL. Genetic variation in the
1071 KIAA0319 5' region as a possible contributor to dyslexia. Behav Genet. 2011;41(1):77-89.
1072 106. Kravitz HM, Meyer PM, Seeman TE, Greendale GA, Sowers MR. Cognitive functioning
1073 and sex steroid hormone gene polymorphisms in women at midlife. Am J Med. 2006;119(9
1074 Suppl 1):S94-S102.
1075 107. Luciano M, Gow AJ, Pattie A, Bates TC, Deary IJ. The influence of dyslexia candidate
1076 genes on reading skill in old age. Behav Genet. 2018;48(5):351-60.
1077 108. Couto JM, Gomez L, Wigg K, Ickowicz A, Pathare T, Malone M, et al. Association of
1078 attention-deficit/hyperactivity disorder with a candidate region for reading disabilities on
1079 chromosome 6p. Biol Psychiatry. 2009;66(4):368-75.
1080 109. Muller B, Wilcke A, Czepezauer I, Ahnert P, Boltze J, Kirsten H, et al. Association,
1081 characterisation and meta-analysis of SNPs linked to general reading ability in a German
1082 dyslexia case-control cohort. Sci Rep. 2016;6:27901.
1083 110. Schumacher J, Anthoni H, Dahdouh F, Konig IR, Hillmer AM, Kluck N, et al. Strong
1084 genetic evidence of DCDC2 as a susceptibility gene for dyslexia. Am J Hum Genet.
1085 2006;78(1):52-62.
1086 111. Marino C, Meng H, Mascheretti S, Rusconi M, Cope N, Giorda R, et al. DCDC2 genetic
1087 variants and susceptibility to developmental dyslexia. Psychiatr Genet. 2012;22(1):25-30.
1088 112. Meng H, Powers NR, Tang L, Cope NA, Zhang PX, Fuleihan R, et al. A dyslexia-associated
1089 variant in DCDC2 changes gene expression. Behav Genet. 2011;41(1):58-66.

1090 113. Powers NR, Eicher JD, Butter F, Kong Y, Miller LL, Ring SM, et al. Alleles of a
1091 polymorphic ETV6 binding site in DCDC2 confer risk of reading and language impairment.
1092 Am J Hum Genet. 2013;93(1):19-28.
1093 114. Powers NR, Eicher JD, Miller LL, Kong Y, Smith SD, Pennington BF, et al. The regulatory
1094 element READ1 epistatically influences reading and language, with both deleterious and
1095 protective alleles. J Med Genet. 2016;53(3):163-71.
1096 115. Igo RP, Chapman NH, Berninger V, Matsushita M, Brkanac Z, Rothstein J, et al. Genome
1097 wide scan for real-word reading subphenotypes of dyslexia: Novel chromosome 13 locus
1098 and genetic complexity. Am J Med Genet B Neuropsychiatric Genetics. 2006;141(1):15-27.
1099 116. Rubenstein KB, Raskind WH, Berninger VW, Matsushita MM, Wijsman EM. Genome scan
1100 for cognitive trait loci of dyslexia: Rapid naming and rapid switching of letters, numbers,
1101 and colors. Am J Med Genet B Neuropsychiatr Genet. 2014;165B(4):345-56.
1102 117. Ludwig KU, Roeske D, Herms S, Schumacher J, Warnke A, Plume E, et al. Variation in
1103 GRIN2B contributes to weak performance in verbal short-term memory in children with
1104 dyslexia. Am J Med Genet B Neuropsychiatr Genet. 2010;153B(2):503-11.
1105 118. Mascheretti S, Facoetti A, Giorda R, Beri S, Riva V, Trezzi V, et al. GRIN2B mediates
1106 susceptibility to intelligence quotient and cognitive impairments in developmental dyslexia.
1107 Psychiatr Genet. 2015;25(1):9-20.
1108 119. Liu Q, Zhu B, Xue Q, Xie X, Zhou Y, Zhu K, et al. The associations of zinc and GRIN2B
1109 genetic polymorphisms with the risk of dyslexia. Environ Res. 2020;191:110207.
1110 120. O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, et al. Multiplex targeted
1111 sequencing identifies recurrently mutated genes in autism spectrum disorders. Science.
1112 2012;338(6114):1619-22.

1113 121. Boyle EA, O'Roak BJ, Martin BK, Kumar A, Shendure J. MIPgen: optimized modeling and
1114 design of molecular inversion probes for targeted resequencing. Bioinformatics.
1115 2014;30(18):2670-2.
1116 122. Hildebrand MS, Myers CT, Carvill GL, Regan BM, Damiano JA, Mullen SA, et al. A
1117 targeted resequencing gene panel for focal epilepsy. Neurology. 2016;86(17):1605-12.
1118 123. Brain Open Chromatin Atlas (BOCA) [Internet]. Available from:
1119 https://labs.icahn.mssm.edu/roussos-lab/boca/.
1120 124. ENCODE Available from: https://www.encodeproject.org.
1121 125. Fullard JF, Hauberg ME, Bendl J, Egervari G, Cirnaru MD, Reach SM, et al. An atlas of
1122 chromatin accessibility in the adult human brain. Genome Res. 2018;28(8):1243-52.
1123 126. Hiatt JB, Pritchard CC, Salipante SJ, O'Roak BJ, Shendure J. Single molecule molecular
1124 inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome
1125 Res. 2013;23(5):843-54.
1126 127. Ensembl Variant Effect Predictor (VEP) [Internet]. Available from:
1127 http://ensembl.org/Homo_sapiens/Tools/VEP.
1128 128. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship
1129 inference in genome-wide association studies. Bioinformatics. 2010;26(22):2867-73.
1130 129. Woodcock R. Woodcock reading master tests - revised (WRMT-R). Circle Pines, MN:
1131 American Guidance Service; 1987.
1132 130. Torgesen J, Wagner R, Reshotte C. Test of word reading efficiency (TOWRE). Austin: Pro-
1133 Ed; 1999.
1134 131. Wilkinson G. Wide range achievement tests - revised (WRAT-R). Wilmington, DE: Wide
1135 Range, Inc.; 1993.

1136 132. Wagner R, Torgesen J, Rashotte C. Comprehensive Test of Phonological Processing
1137 (CTOPP). Austin, TX: PRO-ED; 1999.
1138 133. Gogarten SM, Sofer T, Chen H, Yu C, Brody JA, Thornton TA, et al. Genetic association
1139 testing using the GENESIS R/Bioconductor package. Bioinformatics. 2019;35(24):5346-8.
1140 134. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for
1141 sequencing data with the sequence kernel association test. Am J Hum Genet. 2011;89(1):82-
1142 93.
1143 135. Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale
1144 sequence data. Am J Hum Genet. 2021;108(10):1880-90.
1145 136. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A
1146 global reference for human genetic variation. Nature. 2015;526(7571):68-74.
1147 137. Nassar LR, Barber GP, Benet-Pages A, Casper J, Clawson H, Diekhans M, et al. The UCSC
1148 Genome Browser database: 2023 update. Nuc Scids Res. 2023;51(D1):D1188-D95.
1149 (available at http://genome.ucsc.edu)
1150 138. Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Lemma RB, Turchi L, Blanc-
1151 Mathieu R, et al. JASPAR 2022: the 9th release of the open-access database of transcription
1152 factor binding profiles. Nuc Acids Res. 2022;50(D1):D165-D73. (available at
1153 https://jaspar.genereg.net/).
1154 139. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated
1155 encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.
1156 (available at https://www.encodeproject.org/)
1157 140. Wu C. The 5' ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase
1158 I. Nature. 1980;286(5776):854-60.

1159 141. Keene MA, Corces V, Lowenhaupt K, Elgin SCR. DNase I Hypersensitive sites in
1160 Drosophila chromatin occur at the 5′ ends of regions of transcription. Proc Natl Acad Sci U
1161 S A Biol Sci. 1981;78(1):143-6.
1162 142. McGhee JD, Wood WI, Dolan M, Engel JD, Felsenfeld G. A 200 base pair region at the 5′
1163 end of the chicken adult beta-globin gene is accessible to nuclease digestion. Cell.
1164 1981;27(1):45-55.
1165 143. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-
1166 specific haplotype structure and linking correlated alleles of possible functional variants.
1167 Bioinfiormatics. 2015;31(21):3555-7.
1168 144. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, et al.
1169 Ensembl 2022. Nucl Acids Res. 2022;50(D1):D988-D95.
1170 145. Schmit F, Korenjak M, Mannefeld M, Schmitt K, Franke C, von Eyss B, et al. LINC, a
1171 human complex that is related to pRB-containing complexes in invertebrates regulates the
1172 expression of G2/M genes. Cell Cycle. 2007;6(15):1903-13.
1173 146. Litovchick L, Sadasivam S, Florens L, Zhu X, Swanson SK, Velmurugan S, et al.
1174 Evolutionarily conserved multisubunit RBL2/p130 and E2F4 protein complex represses
1175 human cell cycle-dependent genes in quiescence. Mol Cell. 2007;26(4):539-51.
1176 147. Azcoitia I, Yague JG, Garcia-Segura LM. Estradiol synthesis within the human brain.
1177 Neuroscience. 2011;191:139-47.
1178 148. Lu Y, Sareddy GR, Wang J, Wang R, Li Y, Dong Y, et al. Neuron-derived estrogen
1179 regulates synaptic plasticity and memory. J Neurosci. 2019;39(15):2792-809.
1180 149. Azcoitia I, Mendez P, Garcia-Segura LM. Aromatase in the human brain. Androg Clin Res
1181 Ther. 2021;2(1):189-202.

1182 150. Dalton CJ, Lemmon CA. Fibronectin: Molecular structure, fibrillar structure and
1183 mechanochemical signaling. Cells. 2021;10(9).
1184 151. Muniz AJ, Topal T, Brooks MD, Sze A, Kim DH, Jordahl J, et al. Engineered extracellular
1185 matrices facilitate brain organoids from human pluripotent stem cells. Ann Clin Transl
1186 Neurol. 2023.
1187 152. Mahajan D, Kancharla S, Kolli P, Sharma AK, Singh S, Kumar S, et al. Role of fibulins in
1188 embryonic stage development and their involvement in various diseases. Biomolecules.
1189 2021;11(5).
1190 153. Konishi Y, Ikeda K, Iwakura Y, Kawakami K. Six1 and Six4 promote survival of sensory
1191 neurons during early trigeminal gangliogenesis. Brain Res. 2006;1116(1):93-102.
1192 154. Garcez RC, Le Douarin NM, Creuzet SE. Combinatorial activity of Six1-2-4 genes in
1193 cephalic neural crest cells controls craniofacial and brain development. Cell Mol Life Sci.
1194 2014;71(11):2149-64.
1195 155. Chen R, Hou Y, Connell M, Zhu S. Homeodomain protein Six4 prevents the generation of
1196 supernumerary Drosophila type II neuroblasts and premature differentiation of intermediate
1197 neural progenitors. PLoS Genet. 2021;17(2):e1009371.
1198 156. Sanz-Clemente A, Nicoll RA, Roche KW. Diversity in NMDA receptor composition: many
1199 regulators, many consequences. Neuroscientist. 2013;19(1):62-75.
1200 157. Kim JI, Kim JW, Park S, Hong SB, Lee DS, Paek SH, et al. The GRIN2B and GRIN2A
1201 gene variants are associated with continuous performance test variables in ADHD. J Atten
1202 Disord. 2020;24(11):1538-46.
1203 158. Hayashi Y. Molecular mechanism of hippocampal long-term potentiation - Towards
1204 multiscale understanding of learning and memory. Neurosci Res. 2022;175:3-15.

1205 159. Endele S, Rosenberger G, Geider K, Popp B, Tamer C, Stefanova I, et al. Mutations in
1206 GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable
1207 neurodevelopmental phenotypes. Nat Genet. 2010;42(11):1021-6.
1208 160. Dimassi S, Andrieux J, Labalme A, Lesca G, Cordier MP, Boute O, et al. Interstitial
1209 12p13.1 deletion involving GRIN2B in three patients with intellectual disability. Am J Med
1210 Genet A. 2013;161A(10):2564-9.
1211 161. Platzer K, Yuan H, Schutz H, Winschel A, Chen W, Hu C, et al. GRIN2B encephalopathy:
1212 novel findings on phenotype, variant clustering, functional consequences and treatment
1213 aspects. J Med Genet. 2017;54(7):460-70.
1214 162. Dorval KM, Wigg KG, Crosbie J, Tannock R, Kennedy JL, Ickowicz A, et al. Association
1215 of the glutamate receptor subunit gene GRIN2B with attention-deficit/hyperactivity
1216 disorder. Genes Brain Behav. 2007;6(5):444-52.
1217 163. Pan Y, Chen J, Guo H, Ou J, Peng Y, Liu Q, et al. Association of genetic variants of
1218 GRIN2B with autism. Sci Rep. 2015;5:8296.
1219 164. Jiang Y, Lin MK, Jicha GA, Ding X, McIlwrath SL, Fardo DW, et al. Functional human
1220 GRIN2B promoter polymorphism and variation of mental processing speed in older adults.
1221 Aging (Albany NY). 2017;9(4):1293-306.
1222 165. Spearman C. The proof and measurement of association between two things. Amer J
1223 Psychol. 1904;15(1):72-101.
1224 166. Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, et al.
1225 Determination and inference of eukaryotic transcription factor sequence specificity. Cell.
1226 2014;158(6):1431-43.
1227