Am. J. Hum. Genet. 76:581–591, 2005

Strong Evidence That *KIAA0319* on Chromosome 6p Is a Susceptibility Gene for Developmental Dyslexia

Natalie Cope,[1] Denise Harold,[1] Gary Hill,[1] Valentina Moskvina,[2] Jim Stevenson,[3] Peter Holmans,[2] Michael J. Owen,[1] Michael C. O'Donovan,[1] and Julie Williams[1]

[1]Department of Psychological Medicine and [2]Biostatistics and Bioinformatics Unit, Wales College of Medicine, Cardiff University, Cardiff; and [3]School of Psychology, University of Southampton, Southampton, United Kingdom

Linkage between developmental dyslexia (DD) and chromosome 6p has been replicated in a number of independent samples. Recent attempts to identify the gene responsible for the linkage have produced inconsistent evidence for association of DD with a number of genes in a 575-kb region of chromosome 6p22.2, including *VMP*, *DCDC2*, *KIAA0319*, *TTRAP*, and *THEM2*. We aimed to identify the specific gene or genes involved by performing a systematic, high-density (~2–3-kb intervals) linkage disequilibrium screen of these genes in an independent sample, incorporating family-based and case-control designs in which dyslexia was defined as an extreme representation of reading disability. Using DNA pooling, we first observed evidence for association with 17 single-nucleotide polymorphisms (SNPs), 13 of which were located in the *KIAA0319* gene ($P < .01$–$.003$). After redundant SNPs were excluded, 10 SNPs were individually genotyped in 223 subjects with DD and 273 controls. Those SNPs that were significant at $P \leq .05$ were next genotyped in a semi-independent sample of 143 trios of probands with DD and their parents, to control for possible population stratification. Six SNPs showed significant evidence of association in both samples ($P \leq .04$–$.002$), including a SNP (*rs4504469*) in exon 4 of the *KIAA0319* gene that changes an amino acid ($P = .002$; odds ratio 1.5). Logistic regression analysis showed that two SNPs (*rs4504469* and *rs6935076*) in the *KIAA0319* gene best explained DD status. The haplotype composed of these two markers was significantly associated with DD (global $P = .00001$ in the case-control sample; $P = .02$ in trios). This finding was largely driven by underrepresentation of the most common haplotype in cases ($P = .00003$ in the case-control sample; $P = .006$ in trios; 1–degree-of-freedom tests). Our data strongly implicate *KIAA0319* as a susceptibility gene for dyslexia. The gene product is expressed in brain, but its specific function is currently unknown.

Introduction

Developmental dyslexia (DD [MIM 600202]), or reading disability, is a relatively common, complex cognitive disorder that affects 5%–10% of school-aged children (Shaywitz et al. 1992). The disorder is characterized by an impairment of reading performance despite adequate motivational, educational, and intellectual opportunities and in the absence of sensory or neurological disability. Although the pathophysiology of DD is unknown, there is strong evidence that genes make a substantial contribution to individual variation in risk of DD, with twin studies reporting heritability estimates of up to 0.71 (Fisher 1905; Hinshelwood 1907; DeFries et al. 1987, 1991; Stevenson et al. 1987; Pennington et al. 1991; Schulte-Körne et al. 1996). The mode of transmission is unknown, but DD is almost certainly a complex genetic disorder in which multiple genes play a role (Hohnen and Stevenson 1999).

Genetic linkage and association studies have implicated a number of chromosomal regions that may harbor susceptibility genes for DD. Regions showing replicated evidence for a role in DD include chromosome 1p (Rabin et al. 1993; Grigorenko et al. 2001; Tzenova et al. 2004), 2p (Fagerheim et al. 1999; Francks et al. 2002; Petryshen et al. 2002; Kaminen et al. 2003; Chapman et al. 2004), 6p (Cardon et al. 1994, 1995; Grigorenko et al. 1997, 2000, 2003; Fisher et al. 1999; Gayán et al. 1999; Kaplan et al. 2002; Turic et al. 2003; Chapman et al. 2004), 15q (Grigorenko et al. 1997; Schulte-Körne et al. 1998; Morris et al. 2000), and 18p (Fisher et al. 2002; Marlow et al. 2003; Chapman et al. 2004).

We have sought to identify a gene(s) in the most consistently supported region on chromosome 6p that shows association with DD. The broadest evidence for linkage stretches from marker *D6S109* (Grigorenko et al. 1997) to marker *D6S291* (Fisher et al. 1999), a distance of ~16 Mb, with numerous studies implicating regions be-

Received December 7, 2004; accepted for publication January 21, 2005; electronically published February 16, 2005.

Address for correspondence and reprints: Dr. Michael O'Donovan, Department of Psychological Medicine, Henry Wellcome Building, Cardiff University, Academic Avenue, Heath Park, Cardiff CF14 4XN, United Kingdom. E-mail: odonovanmc@cardiff.ac.uk

0002-9297/2005/7604-0006$15.00

tween these markers (*D6S105–TNFB* [Cardon et al. 1994, 1995], *D6S109–D6S306* [Grigorenko et al. 1997] *D6S276–D6S105* [Gayán et al. 1999] *D6S464–D6S273* [Grigorenko et al. 2000], *D6S109–JA01, D6S299–D6S1621,* and *D6S105–D6S265* [Grigorenko et al. 2003]). A region between *D6S461* and *D6S105* shows greatest overlap between studies, spanning ~4.2 Mb.

Recently, Deffenbacher and colleagues (2004) examined 10 candidate genes (*VMP, DCDC2, MRS2L, GPLD1, ALDH5A1, KIAA0319, TTRAP, THEM2, C6orf62,* and *GMNN*) in the ~680 kb between markers *D6S276* and *D6S1554,* a region that yielded maximal evidence for linkage in their sample. Thirty-one SNPs were selected at a marker density of ~1 marker per 20 kb; 13 of the SNPs provided some evidence for association with at least one of the phenotypes analyzed (discrepancy score, phoneme awareness, phoneme deletion, word reading, and orthographic coding). These SNPs were located in five genes: two in *VMP,* eight in *DCDC2,* and one each in *KIAA0319, TTRAP,* and *THEM2*. Francks and colleagues (2004) examined eight genes (*ALDH5A1, KIAA0319, TTRAP, THEM2, C6orf32, SCGN, BTN3A1,* and *BTN2A1*) on chromosome 6p as candidates for DD on the basis of known brain expression. Fifty-seven SNPs were analyzed in 89 U.K. families (sample 1). Association was detected in a 77-kb region spanning *TTRAP* and the first four exons of *KIAA0319*. In a second U.K. sample of 175 families (sample 2), 20 of the SNPs were analyzed; weaker evidence of association was found. It is noteworthy that the effect was most pronounced in a sample of families with probands representing the lower end of the reading-ability spectrum. Twenty-one SNPs were also analyzed in a U.S. sample of 159 families (sample 3). Again, association was observed in selected families representing the lower end of the reading-ability spectrum. A three-marker haplotype was significantly associated in both the U.S. sample and the combined U.K. sample, with different alleles forming the most significant haplotype in each sample.

Our study focused on the 575-kb region of chromosome 6p22.2, which was selected to include all genes implicated in recent candidate-gene association studies (Deffenbacher et al. 2004; Francks et al. 2004) and is situated within a region that shows the most consistent evidence of linkage (Deffenbacher et al. 2004; Francks et al. 2004). Our analysis employs a high-density (137 SNPs in seven genes, at 2–3-kb intervals) screen for linkage disequilibrium (LD) and uses both case-control and family-based designs to take account of possible population stratification.

The targeted genes comprise vesicular membrane protein p24 (*VMP*), doublecortin domain–containing 2 (*DCDC2*), kidney-associated antigen 1 (*KAAG1*), magnesium homeostasis factor (*MRS2L*), *KIAA0319*, TRAF and TNF receptor associated protein (*TTRAP*), thioesterase superfamily member 2 (*THEM2*), and chromosome 6 ORF 62 (*C6orf62*).

VMP is a neuron-specific vesicular membrane protein that is thought to play a role in vesicular organelle transport and neurotransmission (Cheng et al. 2002). *DCDC2* is a ubiquitously expressed gene with a doublecortin-homology domain. Doublecortin itself has been implicated as a cause of X-linked lissencephaly (Gleeson et al. 1998) and is involved in neuronal migration in the CNS, including in the cortex (Gleeson et al. 1999).

KAAG1 is encoded on the strand that is opposite to—and overlaps with—*DCDC2*. It is a kidney antigen–associated gene, found in numerous tumors and normal testis and kidney (Van Den Eynde et al. 1999). Although it has no known CNS function, it was included in our analysis since it is encoded on the strand that is opposite to the *DCDC2* gene and since it was covered by the SNP grid encompassing that gene. *MRS2L* is a ubiquitously expressed gene that is thought to encode a magnesium-transporter protein (Zsurka et al. 2001).

KIAA0319 is a protein of unknown function that is highly expressed in brain (Londin et al. 2003). The four polycystic kidney disease (PKD) domains found in KIAA0319 show homology to the extracellular domains of the PKD protein PKD1, which are involved in cell-adhesive functions (Streets et al. 2003).

TTRAP encodes a tumor necrosis factor receptor–associated protein. It has been shown to inhibit nuclear factor-κB (NF-κB) activation and subsequent downstream activation of transcription (Pype et al. 2000). NF-κB transcription has been shown to play a role in long-term potentiation and synaptic plasticity associated with learning and memory. In mice, inhibition of NF-κB has been shown to result in neurodegenerative-like phenotypes (Fridmacher et al. 2003). TTRAP can also interact with the cytoplasmic TNF receptor–associated factors (TRAFs) and with cytoplasmic domains of some members of the TNF-receptor superfamily (Pype et al. 2000).

THEM2 encodes an uncharacterized hypothalamus protein and is part of the thioesterase superfamily. The thioesterase superfamily catalyzes the hydrolysis of long-chain fatty acyl-CoA thioesters. It has been suggested that abnormal fatty-acid metabolism plays a role in DD (Richardson and Ross 2000; Richardson et al. 2000; Taylor and Richardson 2000).

C6orf62 is a gene with unknown function that is expressed ubiquitously, including in brain. Although not part of the present study, aldehyde dehydrogenase 5 family, member A1 (*ALDH5A1* [succinate-semialdehyde dehydrogenase]), and glycosylphosphatidylinositol-specific phospholipase D1 (*GPLD1*) on 6p22.2-p22.3, were previously examined using a direct gene-analysis approach that is based on de novo polymorphism discovery and

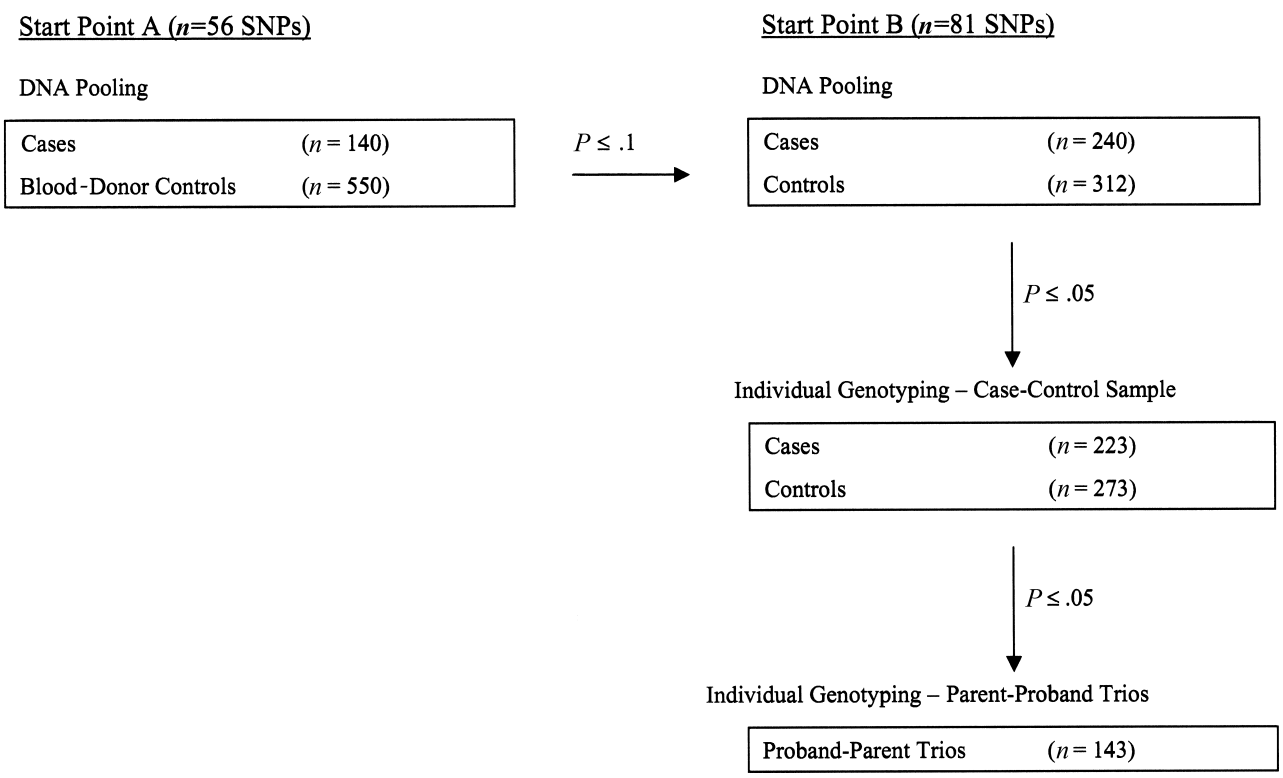

Figure 1 Flow diagram of the samples used at each stage of the analysis

analysis of all detected variants. Extensive analysis of both genes failed to provide evidence for association with DD (authors' unpublished data).

This study thus provides a systematic, high-density LD screen spanning putative functional candidates in a region that shows the most consistent evidence of linkage to DD. Our strategy employed both case-control and family-based designs in which dyslexia was defined, to capture the extreme end of the reading ability/disability continuum.

Material and Methods

Ethical approval was obtained from local ethics committees in the United Kingdom; appropriate and informed written consent was obtained from subjects' parents. All participants were of white U.K. origin. Children with DD—and, if available, their parents and siblings—were ascertained in the United Kingdom through contacts with local education authorities and schools specializing in the education of children with reading difficulties. The inclusion criteria for probands were an IQ of ⩾85 and a reading age ⩾2.5 years behind that expected from chronological age. No age-IQ discrepancy measures were employed. Four subtests from the WISC-III UK were used to provide a prorated, full-scale IQ score: Vocabulary, Similarities, Block Design, and Picture Completion (Wechsler 1992). The accuracy score from the Neale (1989) analysis of reading ability was used to determine reading age, except when probands were aged >13 years, in which case, we used the accuracy score of British Ability Scale (BAS) single-word reading (Elliot 1983).

Initially, controls were adult white U.K. blood donors. Subsequently, control children matched for age and sex were ascertained from the same schools as the children with DD. Children classed as controls were required to have an IQ of ⩾85 and a reading delay (RD) of no more than 6 mo. Control children were assessed using the accuracy score of the Neale (1989) analysis of reading ability and/or BAS single-word reading (Elliot 1983) to calculate reading age; the Vocabulary, Similarities, Block Design, and Picture Completion subsets of WISC-III UK enabled calculation of prorated IQ.

As a first pass, SNPs were analyzed in DNA pools. Because of sample availability, early analyses ($n = 56$ SNPs) used case pools containing 140 unrelated probands with DD (mean age [± SD] 13.22 ± 2.3 years; mean RD [± SD] −5.07 ± 1.76; mean IQ [± SD] 100.13 ± 11.03; 116 males, 24 females) and 550 adult blood-donor controls (mean age 41.39 ± 12.5 years; 391 male, 159 female) (fig. 1, *Start Point A*). Later,

pooled analyses were based on an extension of the original sample of subjects with DD ($n = 240$ unrelated probands with DD; mean age 13.17 ± 2.18 years; mean IQ 98.88 ± 18.38; mean RD −4.93 ± 1.87; 204 males, 36 females) and pools containing 312 age-matched and screened controls (mean age 11.98 ± 2.39 years; mean IQ 103.35 ± 11.97; mean RD +1.14 ± 1.45; 178 males, 134 females) (fig. 1, *Start Point B*).

DNA was extracted from venous blood or 25 ml saline mouthwashes by use of standard procedures (Morris et al. 2000). SNPs were selected from Ensembl or SNPper (CHIP Bioinformatics Tools) for each of the eight positional/functional candidate genes, at intervals of ~2–3 kb between each SNP. SNPs were also chosen for analysis if they showed association in the study by either Deffenbacher et al. (2004) or Francks et al. (2004), excluding SNPs in LD ($r^2 \geq 0.80$) with those we had already typed on the basis of HapMap data. A total of 137 SNPs were analyzed across the region. With the exception of *DCDC2,* SNP grids extended from 3 kb upstream to the predicted start of transcription across all exons and introns. Introns 2, 7, and 8 of *DCDC2* would have required 60–90 SNPs. Therefore, for pragmatic reasons, we restricted our analysis at these introns to 3 kb of flanking sequence on either side of each of the exons.

Genotyping of DNA pools was undertaken using the SNaPshot (Applied Biosystems) primer-extension method described by Norton and colleagues (2002). Pools were carefully constructed by a serial dilution method, each stage accompanied by quantitation of DNA concentration by use of the PicoGreen method, as we have described in detail elsewhere (Norton et al. 2004). Markers were selected for individual genotyping if pooled analysis revealed evidence for association with DD at an estimated level of $P \leq .05$ in the pooled samples of 240 cases and 312 controls. However, to allow for the smaller sample of cases ($n = 140$) in the early analysis, those markers showing only a trend for association ($P \leq .1$) were regenotyped in the larger case pools and were selected for individual genotyping if the trend was confirmed at $P \leq .05$ (see fig. 1). The pooling data presented in our tables reflect the analysis that is based on the larger of the samples in which it was conducted. SNPs selected for individual genotyping were examined, and those subjects for whom we had sufficient DNA (223 DD cases and 273 controls) were included in the pooling experiments. Those markers for which individual genotyping in this case-control sample confirmed the pooling data ($P \leq .05$) were then genotyped in a semi-independent sample of 143 parent-proband trios to ensure that the results were not attributable to population stratification. Mean age of probands in the trio sample was 13.17 ± 2.08 years; mean IQ was 104.01 ± 11.88; mean RD was −5.06 ± 1.76. All but 25 of the probands in the parent-proband trios were also included in the case-control sample. When genotypes were available for these 25 individuals, they were included in the final analysis of case-control association.

Individual genotyping was undertaken using a proprietary Amplifluor (Serologicals) genotyping method. Amplifluor reactions were performed in 5-μl reactions containing 50 ng DNA, in accordance with manufacturer instructions. Primers were designed using Amplifluor AssayArchitect and were obtained from Sigma-Genosys. PCR reactions were performed under standard conditions, with an initial denaturation stage of 96°C for 4 min, then 19 cycles at 96°C for 10 s, at 58°C or 60°C for 5 s, and at 72°C for 10 s; followed by 22 or 27 cycles at 96°C for 10 s, 20 s at 55°C, and 40 s at 72°C; and a final extension step at 72°C for 3 min. Genotypes were read on an LJL Biosystems Analyst. When we were unable to optimize Amplifluor, RFLP analysis of PCR products was undertaken. PCR reactions were performed under standard conditions in 12-μl reaction volumes with 32 ng of genomic DNA. Digests were undertaken in 17-μl reactions by use of the appropriate restriction enzyme (New England Biolabs), in accordance with manufacturer instructions. Products were visualized on 3% agarose gels stained with ethidium bromide.

All genotypes were tested for Hardy-Weinberg equilibrium with a χ^2 goodness-of-fit test (see the Simple Interactive Statistical Analysis Web site). Analysis of LD between markers (r^2 and D') was performed using Haploview. Standard contingency tables were used for single-marker case-control analysis. Trios were analyzed using UNPHASED (Dudbridge 2003) (see the Rosalind Franklin Centre for Genomics Research Web site). Haplotypes were analyzed using EHPlus (Zhao et al. 2000) and UNPHASED (Dudbridge 2003). Logistic and conditional logistic regression analyses were performed on case-control and trio data, respectively.

Results

Of the 137 SNPs analyzed in pools, 17 yielded evidence for association at $P \leq .05$, and 13 of these were located within *KIAA0319* (see table 1 for results of the pooled genotyping). Of these SNPs, 15 were then typed for 42 subjects with DD and 48 controls, to identify redundant markers on the basis of marker-marker LD (the two remaining SNPs, *rs1555090* and *rs926529,* were known to be in LD with other SNPs analyzed on the basis of

Table 1

All Pooled Genotyping Data

The table is available in its entirety in the online edition of *The American Journal of Human Genetics.*

Table 2

LD between SNPs Showing Significance ($P < .05$) in DNA Pools

Gene and SNP[a]	LD Value for SNP Pair															
	rs2793422	*rs4504469*	*rs6911855*	*rs6939068*	*rs7751357*	*rs2179515*	*rs6456622*	*rs9358783*	*rs9358784*	*rs7755579*	*rs6456624*	*rs6935076*	*rs2038137*	*rs2143340*	*rs3777664*	*rs1053598*
MRS2L:																
rs2793422		.2	.01	.06	.22	.18	.24	.24	.21	.03	.03	.04	.06	0	.15	.07
KIAA0319:																
rs4504469	0		***1***	***1***	***.82***	***.83***	***.85***	***.85***	***.86***	.62	.68	.59	.75	.15	.51	.57
rs6911855	.01	.02		***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	.52	***1***	.02
rs6939068	0	.02	.79		***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	.63	***1***	.22
rs7751357	.01	.51	.02	.02		***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	.63	.68	.66
rs2179515	.01	.55	.02	.02	***1***		***1***	***1***	***1***	***1***	***1***	***1***	***1***	.68	.69	.68
rs6456622	.02	.54	.02	.02	***1***	***1***		***1***	***1***	***1***	***1***	***1***	***1***	.64	.67	.69
rs9358783	.02	.54	.02	.02	***1***	***1***	***1***		***1***	***1***	***1***	***1***	***1***	.66	.69	.67
rs9358784	.01	.54	.02	.02	***1***	***1***	***1***	***1***		***1***	***1***	***1***	***1***	.65	.67	.67
rs7755579	0	.38	.02	.02	***.80***	.79	.79	***.80***	.78		***1***	***1***	***1***	.58	.65	.71
rs6456624	0	.42	.02	.02	***.85***	***.83***	***.84***	***.84***	***.82***	***.95***		***1***	***1***	.74	.64	.68
rs6935076	.02	.14	.02	.03	.35	.3	.33	.33	.33	.43	.41		***1***	.76	.69	***.93***
rs2038137	.01	.52	.02	.02	***.89***	***.89***	***.91***	***.91***	***.89***	***.90***	***.92***	.33		.73	.65	.7
TTRAP:																
rs2143340	.01	.05	.01	0	.03	.05	.03	.03	.03	.03	.05	.09	.08		***1***	***1***
THEM2:																
rs3777664	0	.21	0	0	.43	.37	.42	.42	.41	.36	.36	.11	.37	.08		***.94***
Intergenic:																
rs1053598	.01	.23	.03	.01	.41	.41	.45	.45	.42	.38	.35	.22	.41	.07	.77	

Note.—Intermarker LD values, as measured by D' (above the diagonal) and r^2 (below the diagonal). Values ⩾.8 are set in bold italics.

[a] Underlined SNPs were individually genotyped in the case-control sample.

HapMap data). In an attempt to replicate the most significant haplotype of Francks and colleagues (2004), we also genotyped *rs2143340,* although this SNP did not show significant evidence for association in DNA pools. Perfect LD ($r^2 = 1$) was noted for several of the markers in *KIAA0319* (see table 2). Genotyped markers showing LD, as assessed by r^2, of at least 0.8 were dropped. On this basis, a minimal set of 10 markers was chosen for individual genotyping in the case-control sample (see table 3).

All genotypes were in Hardy-Weinberg equilibrium for cases and controls and probands and parents, except for SNP *rs2038137*) (located near the exon 1/intron 1 boundary of *KIAA0319*), which showed slight distortion in the controls ($P = .03$). Our most significant results ($P \leq .01$) were found in *KIAA0319, MRS2L,* and *THEM2* in the case-control sample (see table 3). To ensure that these data did not arise from population stratification, the seven markers that were significant in the case-control sample at $P \leq .05$ were then individually genotyped in a sample of 143 parent-proband trios (table 4). This also provides for a degree of independence, since the control (nontransmitted) alleles are independent of the controls in the case-control study. After this step, six SNPs remained significant—three in the *KIAA0319* gene (family-based, $P = .04$–$.002$; case-control, $P = .007$–$.001$), one in *MRS2L* (family-based, $P = .04$; case-control, $P = .003$), and two in or flanking the *THEM2* gene (family-based, $P = .03$–$.01$; case-control, $P = .03$–$.008$). Two of these SNPs (*rs4504469*

Table 3

Genotype and Allele Counts for Selected SNPs in the Case-Control Sample

Gene, SNP, and Sample	No. of Subjects with Genotype				No. (%) of Alleles			
	1-1	1-2	2-2	*P*	1	2	*P*	OR (95% CI)
MRS2L[a]:								
rs2793422:								
Cases	99	94	22	***.03***	304 (69)	138 (31)	***.003***	1.50 (1.14–1.96)
Controls	91	117	43		299 (60)	203 (40)		
KIAA0319:								
rs4504469[a]:								
Cases	101	117	22	***.002***	319 (66)	161 (34)	***.002***	1.51 (1.17–1.95)
Controls	88	124	52		300 (57)	228 (43)		
rs6911855:								
Cases	200	17	1	.17	417 (96)	19 (04)	.07	.51 (.24–1.06)
Controls	253	12	0		518 (98)	12 (02)		
rs6939068:								
Cases	180	19	1	.16	379 (95)	21 (05)	.06	.52 (.26–1.04)
Controls	234	14	0		482 (97)	14 (03)		
rs2179515[a]:								
Cases	116	100	16	***.008***	332 (72)	132 (28)	***.007***	1.45 (1.11–1.90)
Controls	109	108	40		326 (63)	88 (37)		
rs6935076[a]:								
Cases	65	131	35	***.006***	261 (56)	201 (44)	***.006***	.70 (.54–.90)
Controls	107	118	30		332 (65)	178 (35)		
rs2038137[a]:								
Cases	112	104	13	***.0001***	328 (72)	130 (28)	***.001***	1.57 (1.20–2.05)
Controls[b]	106	105	46		317 (62)	197 (38)		
TTRAP:								
rs2143340:								
Cases	140	56	7	.26	336 (83)	70 (17)	.32	.83 (.58–1.19)
Controls	179	68	3		426 (85)	74 (15)		
THEM2[a]:								
rs3777664:								
Cases	119	92	13	***.02***	330 (74)	118 (26)	***.008***	1.45 (1.10–1.92)
Controls	112	113	31		337 (66)	175 (34)		
Intergenic[a]:								
rs1053598:								
Cases	124	92	9	***.05***	340 (76)	110 (24)	***.03***	1.36 (1.03–1.81)
Controls	123	112	23		358 (69)	158 (31)		

Note.—Genotypic and allelic *P* values are given, as are ORs and 95% CIs. *P* values ≤.05 are indicated in bold italics.

[a] Case-control analysis includes 25 extra probands from the proband-parent trios.

[b] Genotypes deviate from Hardy-Weinberg equilibrium.

Table 4

Transmission of Alleles at Selected SNPs in the Parent-Proband Trios

Gene, SNP, and Transmission	No. (%) of Alleles 1	2	*P*	OR (95% CI)[a]
MRS2L:				
rs2793422:				
Transmitted	191 (71)	79 (29)	***.04***	1.47 (1.02–2.1)
Nontransmitted	168 (62)	102 (38)		
KIAA0319:				
rs4504469:				
Transmitted	166 (68)	78 (32)	***.04***	1.48 (1.02–2.14)
Nontransmitted	144 (59)	100 (41)		
rs2179515:				
Transmitted	184 (71)	77 (30)	***.04***	1.46 (1.01–2.10)
Nontransmitted	162 (62)	99 (38)		
rs6935076:				
Transmitted	154 (56)	120 (44)	***.002***	.57 (.41–.82)
Nontransmitted	189 (69)	85 (31)		
rs2038137:				
Transmitted	169 (69)	77 (31)	.11	1.36 (.94–1.97)
Nontransmitted	152 (62)	94 (38)		
THEM2:				
rs3777664:				
Transmitted	165 (74)	59 (26)	***.03***	1.55 (1.04–2.33)
Nontransmitted	144 (64)	80 (36)		
Intergenic:				
rs1053598:				
Transmitted	186 (76)	59 (24)	***.01***	1.67 (1.13–2.48)
Nontransmitted	160 (65)	85 (35)		

Note.—*P* values for association were calculated using UNPHASED (Rosalind Franklin Centre for Genomics Research). *P* values ⩽.05 are indicated in bold italics.

[a] ORs and 95% CIs refer to allele 1.

and *rs2179515*) were reported in the study by Francks et al. (2004) to be associated with a number of componential measures of DD in the combined U.K. sample. Both are located in the *KIAA0319* gene ($r^2 = 0.55$).

To determine which SNPs accounted for the association, stepwise logistic regression analyses were performed on the case-control and trio data on the basis of all SNPs for which we had individual genotyping data (see tables 3 and 4). For the case-control sample, the SNPs—*rs2793422* (*MRS2L*); *rs4504469, rs6911855, rs6939068, rs2179515, rs6935076,* and *rs2038137* (*KIAA0319*); *rs2143340* (*TTRAP*); *rs3777664* (*THEM2*); and Intergenic *rs1053598*—were initially submitted into the logistic regression model ($P = .029$; 10 df). The stepwise procedure reduced the number of SNPs to three—*rs2793422* (*MRS2L*) and *rs4504469* and *rs6935076* (*KIAA0319*)—that showed a highly significant fit ($P = .00002$; 3 df). For the proband-parent trios, the probands were considered as cases, and nontransmitted alleles were employed to create pseudocontrols (Cordell and Clayton 2002). These data were submitted into conditional logistic regression analyses ($P = .347$; 7 df). The best model was again identified by use of a stepwise procedure ($P = .02$; 2 df) that removed every SNP except SNPs *rs4504469* and *rs6935076*. The addition of *rs2793422* did not significantly improve the model ($P = .10$ [log-likelihood ratio test]). *rs4504469* is a nonsynonymous SNP in exon 4 (Ala→Thr), and *rs6935076* is located in intron 1 of the *KIAA0319* gene (see fig. 2).

On the basis of the results of the regression analysis, we analyzed the two-marker haplotype that consisted of the *KIAA0319* SNPs *rs4504469* and *rs6935076* in the case-control and trio samples (see tables 5 and 6). Significant evidence for association was obtained on the basis of the global test ($P = .0001$ in the case-control sample; $P = .02$ in trios). In each sample, the 1-2 haplotype was associated with DD, but more striking is the significant underrepresentation of haplotype 2-1 in the cases based on the case-control ($P = .00003$; odds ratio [OR] 0.53; 95% CI 0.40–0.70) and family-based ($P = .006$; OR 0.57; 95% CI 0.39–0.84) analyses. (Fig. 2 summarizes our results.)

We also analyzed the three-marker haplotype that consisted of *rs4504469, rs2038137,* and *rs2143340,* which was reported as significantly associated with DD by Francks and colleagues (2004). This haplotype did not yield global evidence for association in our sample (see table 7). Two individual haplotypes did, however, show evidence for association with DD. The 1-1-1 haplotype was more frequent in subjects with DD than in control individuals ($P = .03$). The 2-2-1 haplotype, which was significantly associated with the READ phenotype in the combined U.K. sample of the study by Francks et al. (2004), also displayed evidence of association with DD in our case-control sample ($P = .01$) and showed the same direction of effect (in their study, the 2-2-1 haplotype was associated with better performance; in our sample, this haplotype was more frequent in control individuals). In our sample, this haplotype is, in fact, perfectly defined by the first two SNPs (since there was no observation of the 2-2-2 haplotype). We therefore excluded *rs2143340* and looked at the two-marker haplotype (2-2) that consisted of the other two SNPs in our family-based sample, but, although it was undertransmitted to the probands, this was not significant ($P = .10$). It should be noted that, since *rs4504469* shows

Table 5

Analysis of Haplotypes in *KIAA0319* Comprising SNPs *rs4504469* and *rs6935076* in 248 Subjects with DD and 273 Controls

Allele at SNP		Frequency in		Haplotype
rs4504469	*rs6935076*	Cases	Controls	*P*[a]
1	1	.31	.27	.26
1	2	.35	.30	.02
2	1	.25	.39	.00003
2	2	.09	.05	.17

[a] Global $P = .0001$.

Table 6

Analysis of Haplotypes in *KIAA0319* Comprising SNPs *rs4504469* and *rs6935076* in a Sample of 143 Families with DD

Allele at SNP		Frequency		Haplotype
rs4504469	*rs6935076*	Transmitted	Nontransmitted	P[a]
1	1	.32	.32	.95
1	2	.35	.27	.03
2	1	.24	.36	.006
2	2	.09	.05	.22

[a] Global $P = .02$.

more significance individually than does the 2-2-1 haplotype, no extra information was obtained from this haplotype in our sample. Moreover, the 1-1-2 haplotype that was reported to show association with componential measures of DD in both the U.S. sample and the combined U.K. sample in the study by Francks et al. (2004) was not significantly associated with DD in our sample ($P = .21$).

Discussion

Previous linkage and association studies of DD and chromosome 6p have implicated a region between markers *D6S461* and *D6S105*. More recently, following other positional candidate-gene studies, *VMP, DCDC2, KIAA0319, TTRAP,* and *THEM2* have been suggested as possible susceptibility genes within this region (Deffenbacher et al. 2004; Francks et al. 2004). Our study tested for association with each of these genes (*VMP, DCDC2, KAAG1, MRS2L, KIAA0319, TTRAP, THEM2,* and *C6orf62*) by use of a high-density SNP map and an independent sample. Initially, we genotyped DNA pools from subjects and controls and followed up those findings with individual genotyping in a case-control sample and a nested family-based association sample. In both samples, we observed evidence for association with three SNPs in *KIAA0319* (*rs4504469,* $P = .002$; *rs2179515,* $P = .007$; and *rs6935076,* $P = .006$), with one SNP in *MRS2L* (*rs2793422,* $P = .003$) and in *THEM2* (*rs3777664,* $P = .008$), and with an intergenic SNP (*rs1053598,* $P = .02$). Two of these SNPs, *rs4504469* and *rs2179515* (both located in *KIAA0319*), have been reported elsewhere to display significant association with a number of componential measures of DD in a U.K. sample (Francks et al. 2004). Our results support existing data (Deffenbacher et al. 2004; Francks et al. 2004) that implicate genes in this region in DD, and our results extend the previous findings by demonstrating that the source of the signal is likely to be variation in *KIAA0319*. The study by Francks and colleagues (2004) implicated a region containing *KIAA0319, TTRAP,* and *THEM2,* whereas that of Deffenbacher and colleagues (2004) implicated *KIAA0319, DCDC2, VMP, TTRAP,* and *THEM2*. Combining their data with our own produces a pattern of evidence that implicates *KIAA0319* as a susceptibility gene for DD.

This is compatible with the logistic regression and conditional logistic regression analyses in our case-control and proband-parent trio samples, respectively. Case-control data analyses with the use of a stepwise procedure revealed three SNPs that account for the association observed: *rs2793422* in *MRS2L* and *rs4504469* and *rs6935076* in *KIAA0319* ($P = .00002$). Analysis of the proband-parent trios identified two SNPs that account for the association observed: *rs4504469* and *rs6935076* ($P = .03$). The results, therefore, are consistent with crude inspection of the genes showing overlap between the studies and provide strong evidence that *KIAA0319* SNPs *rs4504469* and *rs6935076* are responsible for the association with DD observed in this study. A haplotype comprising these two SNPs is highly significantly associated with DD in both the case-control sample ($P = .00003$) and the trio sample ($P = .006$). This effect is largely driven by haplotype 2-1 as a "protective" haplotype.

Although the logistic regression analyses imply that, of the markers tested, *rs4504469* and *rs6935076* can account for the association signal, this does not imply that they are the direct susceptibility alleles, per se. However, interestingly, *rs4504469* (one of the two SNPs that makes up our most significant haplotype) is a nonsynonymous SNP (Ala→Thr), which suggests the possibility that this might, in part, contribute directly to the association. However, in our own sample, the threonine at this locus that is present on the protective *rs4504469/rs6935076* 2-1 haplotype is also present on the 2-2 haplotype (tables 5 and 6), which is more common in cases, albeit not significantly more so. This suggests that, if the nonsynonymous change at *rs4504469* can influence risk of DD directly, then its effects can be modified by a second susceptibility allele in the gene. Given that the SNP showed the same pattern of allelic association in the U.K. sample of Francks and colleagues

Table 7

Haplotype Analysis Spanning *KIAA0319* and *TTRAP*

Allele at SNP			Frequency in		Haplotype
rs4504469	*rs2038137*	*rs2143340*	Cases	Controls	P[a]
1	1	1	.47	.41	.03
1	1	2	.15	.12	.21
1	2	1	.04	.05	.66
2	1	1	.07	.08	.67
2	1	2	.02	.02	.67
2	2	1	.25	.33	.01

Note.—Analysis of haplotypes comprising SNPs *rs4504469, rs2038137,* and *rs2143340* in 223 subjects with DD and 273 controls. The 1-1-2 haplotype was observed by Francks et al. (2004) to be significantly associated with a number of reading-related measures but is not significant in our sample.

[a] Global $P = .10$.

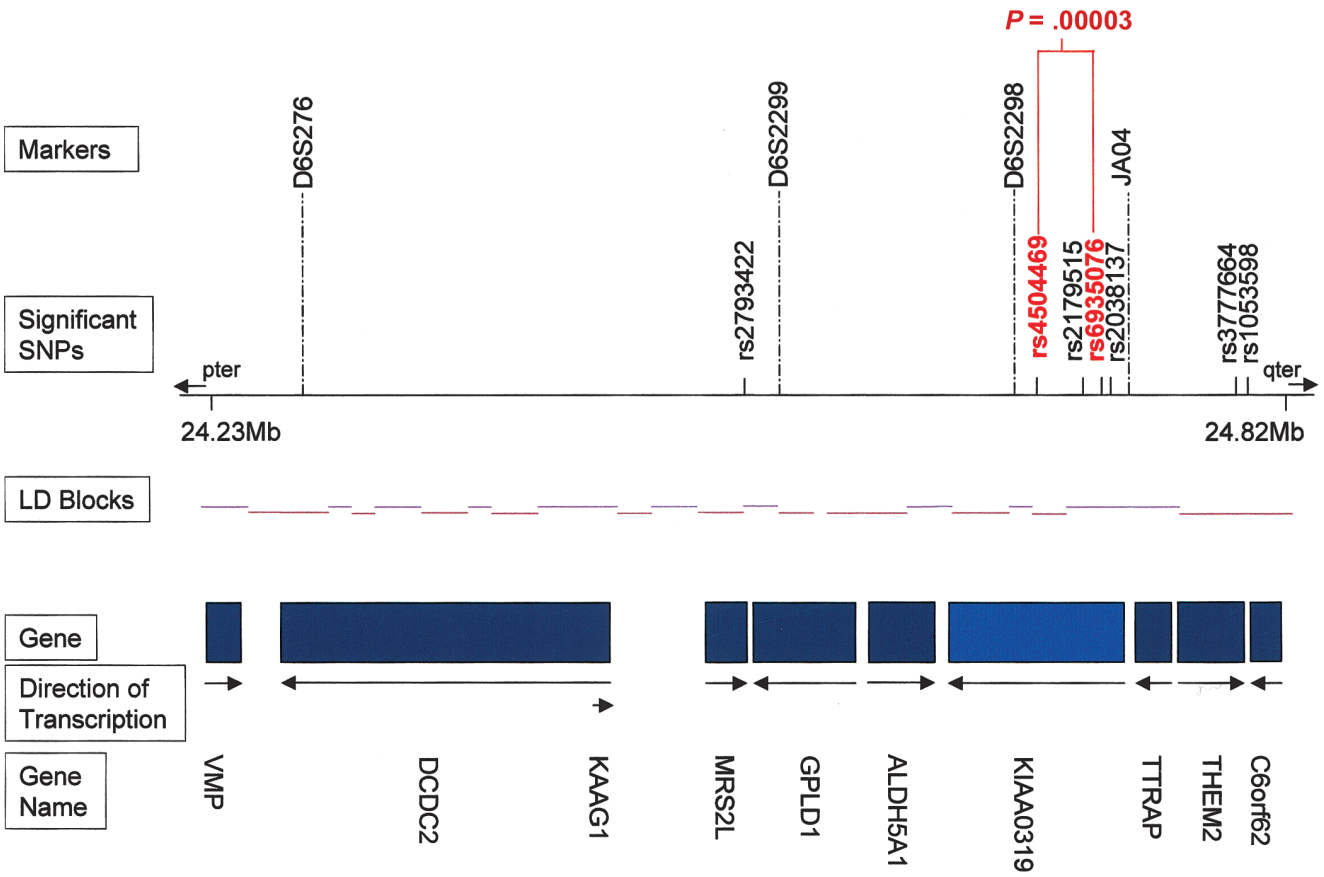

Figure 2 Location of candidate genes on chromosome 6p. The location of SNPs found to be significant ($P \leq .05$) in our case-control sample are shown relative to nearby markers. The direction of transcription is shown for each gene. LD blocks across the region are based on data from HapMap. The *P* value refers to the most significant haplotype (2-1) comprising the two SNPs indicated. An asterisk (*) indicates the amino acid–changing SNP in exon 4.

(2004) but not their U.S. sample, perhaps a more likely explanation is that this SNP does not directly influence susceptibility to DD. Thus, although our study provides strong evidence that variation in the *KIAA0319* gene is associated with increased risk of developing dyslexia, the true susceptibility alleles remain to be identified.

Acknowledgments

We thank all the parents and children who took part in this study, and we are grateful to the Health Foundation (reference number 2263/1921), the Cardiff University Ph.D. Fund, and the U.K. Medical Research Council Cooperative Group Grant (APP1485) for funding this research. We also thank the Welsh Assembly Government for supporting the Biostatistics and Bioinformatics Unit, in which V.M. is employed.

Electronic-Database Information

The URLs for data presented herein are as follows:

Amplifluor AssayArchitect https://apps.serologicals.com/AAA/
CHIP Bioinformatics Tools, http://snpper.chip.org/ (for SNPper)
Ensembl Genome Browser, http://www.ensembl.org/
Haploview, http://www.broad.mit.edu/mpg/haploview/index.php
International HapMap Project, http://www.hapmap.org/ (for LD data)
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for DD)
Rosalind Franklin Centre for Genomics Research, http://www.hgmp.mrc.ac.uk/ (for the UNPHASED application)
Sigma-Genosys, http://orders.sigma-genosys.eu.com
Simple Interactive Statistical Analysis, http://home.clara.net/sisa/ (for χ^2 tests of association)

References

Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC (1994) Quantitative trait locus for reading disability on chromosome 6. Science 266:276–279
——— (1995) Quantitative trait locus for reading disability: correction. Science 268:1553
Chapman NH, Igo RP, Thomson JB, Matsushita M, Brkanac Z, Holzman T, Berninger VW, Wijsman EM, Raskind WH

(2004) Linkage analyses of four regions previously implicated in dyslexia confirmation of a locus on chromosome 15q. Am J Med Genet B 131:67–75

Cheng C, Xu J, Ye X, Dai J, Wu Q, Zeng L, Wang L, Zhao W, Ji C, Gu S, Xie Y, Mao Y (2002) Cloning, expression and characterization of a novel human *VMP* gene. Mol Biol Rep 29:281–286

Cordell HJ, Clayton DG (2002) A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to *HLA* in type 1 diabetes. Am J Hum Genet 70: 124–141

Deffenbacher KE, Kenyon JB, Hoover DM, Olson RK, Pennington BF, DeFries JC, Smith SD (2004) Refinement of the 6p21.3 quantitative trait locus influencing dyslexia: linkage and association analyses. Hum Genet 115:128–138

DeFries JC, Fulker DW, Labuda MC (1987) Reading disability in twins: evidence for a genetic aetiology. Nature 329:537–539

DeFries JC, Olson R, Pennington BF, Smith SD (1991) Colorado reading project: past, present, and future. Learn Disabil 2:37–46

Dudbridge F (2003) Pedigree disequilibrium tests for multilocus haplotypes. Genet Epidemiol 25:115–221

Elliot CD (1983) British ability scales. NFER-Nelson, Windsor, United Kingdom

Fagerheim T, Raeymaekers P, Tønnessen FE, Pedersen M, Tranebjærg L, Lubs HA (1999) A new gene (*DYX3*) for dyslexia is located on chromosome 2. J Med Genet 36:664–669

Fisher JH (1905) Case of congenital word blindness (inability to learn to read). Ophthal Rev 24:315

Fisher SE, Francks C, Marlow AJ, MacPhie IL, Newbury DF, Cardon LR, Ishikawa-Brush Y, Richardson AJ, Talcott JB, Gayán J, Olson RK, Pennington BF, Smith SD, DeFries JC, Stein JF, Monaco AP (2002) Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. Nat Genet 30:86–91

Fisher SE, Marlow AJ, Lamb J, Maestrini E, Williams DF, Richardson AJ, Weeks DE, Stein JF, Monaco AP (1999) A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. Am J Hum Genet 64:146–156

Francks C, Fisher SE, Olson RK, Pennington BF, Smith SD, DeFries JC, Monaco AP (2002) Fine mapping of the chromosome 2p12-16 dyslexia susceptibility locus: quantitative association analysis and positional candidate genes *SEMA4F* and *OTX1*. Psychiatr Genet 12:35–41

Francks C, Paracchini S, Smith SD, Richardson AJ, Scerri TS, Cardon LR, Marlow AJ, MacPhie IL, Walter J, Pennington BF, Fisher SE, Olson RK, DeFries JC, Stein JF, Monaco AP (2004) A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. Am J Hum Genet 75:1046–1058

Fridmacher V, Kaltschmidt B, Goudeau B, Ndiaye D, Rossi FM, Pfeiffer J, Kaltschmidt C, Israel A, Memet S (2003) Forebrain-specific neuronal inhibition of nuclear factor-κB activity leads to loss of neuroprotection. J Neurosci 23:9403–9408

Gayán J, Smith SD, Cherny SS, Cardon LR, Fulker DW, Brower AM, Olson RK, Pennington BF, DeFries JC (1999) Quantitative-trait locus for specific language and reading deficits on chromosome 6p. Am J Hum Genet 64:157–164

Gleeson JG, Allen KM, Fox JW, Lamperti ED, Berkovic S, Scheffer I, Cooper EC, Dobyns WB, Minnerath SR, Ross ME, Walsh CA (1998) Doublecortin, a brain-specific gene mutated in human X-linked lissencephaly and double cortex syndrome, encodes a putative signaling protein. Cell 92:63–72

Gleeson JG, Lin PT, Flanagan LA, Walsh CA (1999) Doublecortin is a microtubule-associated protein and is expressed widely by migrating neurons. Neuron 23:257–271

Grigorenko EL, Wood FB, Golovyan L, Meyer M, Romano C, Pauls D (2003) Continuing the search for dyslexia genes on 6p. Am J Med Genet B Neuropsychiatr Genet 118:89–98

Grigorenko EL, Wood FB, Meyer MS, Hart LA, Speed WC, Shuster A, Pauls DL (1997) Susceptibility loci for distinct components of developmental dyslexia on chromosome 6 and 15. Am J Hum Genet 60:27–39

Grigorenko EL, Wood FB, Meyer MS, Pauls DL (2000) Chromosome 6p influences on different dyslexia-related cognitive processes: further confirmation. Am J Hum Genet 66:715–723

Grigorenko EL, Wood FB, Meyer MS, Pauls JED, Hart LA, Pauls DL (2001) Linkage studies suggest a possible locus for developmental dyslexia on chromosome 1p. Am J Med Genet 105:120–129

Hinshelwood J (1907) Four cases of hereditary word-blindness occurring in the same family. Brit Med J 2:1229–1232

Hohnen B, Stevenson J (1999) The structure of genetic influences on general cognitive, language, phonological, and reading abilities. Dev Psychol 35:590–603

Kaminen N, Hannula-Jouppi K, Kestilä M, Lahermo P, Muller K, Kaaranen M, Myllyluoma B, Voutilainen A, Lyytinen H, Nopola-Hemmi J, Kere J (2003) A genome scan for developmental dyslexia confirms linkage to chromosome 2p11 and suggests a new locus on 7q32. J Med Genet 40:340–345

Kaplan DE, Gayán J, Ahn J, Won T-W, Pauls D, Olson RK, DeFries JC, Wood F, Pennington BF, Page GP, Smith SD, Gruen JR (2002) Evidence for linkage and association with reading disability, on 6p21.3–22. Am J Hum Genet 70:1287–1298

Londin ER, Meng H, Gruen JR (2003) A transcription map of the 6p22.3 reading disability locus identifying candidate genes. BMC Genomics 4:25

Marlow AJ, Fisher SE, Francks C, MacPhie IL, Cherny SS, Richardson AJ, Talcott JB, Stein JF, Monaco AP, Cardon LR (2003) Use of multivariate linkage analysis for dissection of a complex cognitive trait. Am J Hum Genet 72:561–570

Morris DW, Robinson L, Turic D, Duke M, Webb V, Milham C, Hopkin E, Pound K, Fernando S, Easton M, Hamshere M, Williams N, McGuffin P, Stevenson J, Krawczak M, Owen MJ, O'Donovan MC, Williams J (2000) Family-based association mapping provides evidence for a gene for reading disability on chromosome 15q. Hum Mol Genet 9:843–848

Neale MD (1989) Analysis of reading ability, revised British edition (British adaptation and standardisation by Una Christophers and Chris Whetton). NFER-Nelson, Windsor, United Kingdom

Norton N, Williams NM, O'Donovan MC, Owen MJ (2004)

DNA pooling as a tool for large-scale association studies in complex traits. Ann Med 36:146–152

Norton N, Williams NM, Williams HJ, Spurlock G, Kirov G, Morris DW, Hoogendoorn B, Owen MJ, O'Donovan MC (2002) Universal, robust, highly quantitative SNP allele frequency measurement in DNA pools. Hum Genet 110:471–478

Pennington BF, Gilger JW, Pauls D, Smith SA, Smith SD, DeFries JC (1991) Evidence for a major gene transmission of developmental dyslexia. JAMA 266:1527–1534

Petryshen TL, Kaplan BJ, Hughes ML, Tzenova J, Field LL (2002) Supportive evidence for the *DYX3* dyslexia susceptibility gene in Canadian families. J Med Genet 39:125–126

Pype S, Declercq W, Ibrahimi A, Michiels C, Rietschoten JGV, Dewu N, Boer M, Vandenabeele P, Huylebroeck D, Remacle JE (2000) TTRAP, a novel protein that associates with CD40, tumour necrosis factor (TNF) receptor-75 and TNF receptor-associated factors (TRAFs) and that inhibits nuclear factor-κB activation. J Biol Chem 275:18586–18593

Rabin M, Wen XL, Hepburn M, Lubs HA (1993) Suggestive linkage of developmental dyslexia to chromosome 1p34-p36. Lancet 342:178

Richardson AJ, Calvin CM, Clisby C, Schoenheimer DR, Montgomery P, Hall JA, Hebb G, Westwood E, Talcott JB, Stein JF (2000) Fatty acid deficiency signs predict the severity of reading and related difficulties in dyslexic children. Prostaglandins Leukot Essent Fatty Acids 63:69–74

Richardson AJ, Ross MA (2000) Fatty acid metabolism in neurodevelopmental disorder: a new perspective on associations between attention-deficit/hyperactivity disorder, dyslexia, dyspraxia and the autistic spectrum. Prostaglandins Leukot Essent Fatty Acids 63:1–9

Schulte-Körne G, Deimel W, Muller K, Gutenbrunner C, Remschmidt H (1996) Familial aggregation of spelling disability. J Child Psychol Psych 37:817–822

Schulte-Körne G, Grimm T, Nöthen MM, Müller-Myhsok B, Cichon S, Vogt IR, Propping P, Remschmidt H (1998) Evidence for linkage of spelling disability to chromosome 15. Am J Hum Genet 63:279–282

Shaywitz BA, Fletcher JM, Holahan JM, Shaywitz SE (1992) Discrepancy compared to low achievement definitions of reading disability. J Learn Disabil 25:639–648

Stevenson J, Graham P, Fredman G, McLoughlin V (1987) A twin study of genetic influences on reading and spelling ability and disability. J Child Psychol Psychiatry 28:229–247

Streets AJ, Newby LJ, O'Hare MJ, Bukanov NO, Ibraghimov-Beskrovnaya O, Ang AC (2003) Functional analysis of PKD1 transgenic lines reveals a direct role for polycystin-1 in mediating cell-cell adhesion. J Am Soc Nephrol 14:1804–1815

Taylor KE, Richardson AJ (2000) Visual function, fatty acids and dyslexia. Prostaglandins Leukot Essent Fatty Acids 63: 89–93

Turic D, Robinson L, Duke M, Morris DW, Webb V, Hamshere M, Milham C, Hopkin E, Pound K, Fernando S, Grierson A, Easton M, Williams N, Van Den Bree M, Chowdhury R, Gruen J, Stevenson J, Krawczak M, Owen MJ, O'Donovan MC, Williams J (2003) Linkage disequilibrium mapping provides further evidence of a gene for reading disability on chromosome 6p21.3-22. Mol Psychiatry 8:176–185

Tzenova J, Kaplan BJ, Petryshen TL, Field LL (2004) Confirmation of a dyslexia susceptibility locus on chromosome 1p34-p36 in a set of 100 Canadian families. Am J Med Genet B Neuropsychiatr Genet 127:117–124

Van Den Eynde BJ, Gaugler B, Probst-Kepper M, Michaux L, Devuyst O, Lorge F, Weynants P, Boon T (1999) A new antigen recognized by cytolytic T lymphocytes on a human kidney tumor results from reverse strand transcription. J Exp Med 190:1793–1799

Wechsler D (1992) WISC-III UK, 3rd ed. The Psychological Corporation, Harcourt Brace, London

Zhao JH, Curtis D, Sham PC (2000) Model-free analysis and permutation tests for allelic associations. Hum Hered 50: 133–139

Zsurka G, Gregan J, Schweyen RJ (2001) The human mitochondrial MRS2 protein functionally substitutes for its yeast homologue, a candidate magnesium transporter. Genomics 72:158–168