Strong Genetic Evidence of *DCDC2* as a Susceptibility Gene for Dyslexia

Johannes Schumacher,[1] Heidi Anthoni,[3] Faten Dahdouh,[1] Inke R. König,[4] Axel M. Hillmer,[2] Nadine Kluck,[1] Malou Manthey,[1] Ellen Plume,[5] Andreas Warnke,[5] Helmut Remschmidt,[6] Jutta Hülsmann,[6] Sven Cichon,[2] Cecilia M. Lindgren,[3] Peter Propping,[1] Marco Zucchelli,[3] Andreas Ziegler,[4] Myriam Peyrard-Janvid,[3] Gerd Schulte-Körne,[6] Markus M. Nöthen,[2] and Juha Kere[3,7]

[1]Institute of Human Genetics, and [2]Department of Genomics, Life & Brain Center, University of Bonn, Bonn; [3]Department of Biosciences at Novum and Clinical Research Center, Karolinska Institute, Stockholm; [4]Institute of Medical Biometry and Statistics, University Hospital Schleswig-Holstein–Campus Lübeck, Lübeck, Germany; [5]Department of Child and Adolescent Psychiatry and Psychotherapy, University of Würzburg, Würzburg, Germany; [6]Department of Child and Adolescent Psychiatry and Psychotherapy, University of Marburg, Marburg, Germany; and [7]Department of Medical Genetics, University of Helsinki, Helsinki

We searched for linkage disequilibrium (LD) in 137 triads with dyslexia, using markers that span the most-replicated dyslexia susceptibility region on 6p21-p22, and found association between the disease and markers within the *VMP/DCDC2/KAAG1* locus. Detailed refinement of the LD region, involving sequencing and genotyping of additional markers, showed significant association within *DCDC2* in single-marker and haplotype analyses. The association appeared to be strongest in severely affected patients. In a second step, the study was extended to include an independent sample of 239 triads with dyslexia, in which the association—in particular, with the severe phenotype of dyslexia—was confirmed. Our expression data showed that *DCDC2*, which contains a doublecortin homology domain that is possibly involved in cortical neuron migration, is expressed in the fetal and adult CNS, which—together with the hypothesized protein function—is in accordance with findings in dyslexic patients with abnormal neuronal migration and maturation.

Dyslexia (MIM 600202) is a specific developmental disorder characterized by severe difficulties in learning to read and spell, despite adequate schooling, normal visual acuity, and a mental age that is within a normal range (ICD-10) (Dilling et al. 1991). This disorder is more frequent in boys than in girls (Rutter et al. 2004) and affects 5%–12% of school-aged children (Katusic et al. 2001). Although reading and spelling disorders characterize the core dyslexia phenotype, related abilities—for example, phoneme awareness (PA) (analyzing and discriminating phonemes), verbal memory, and orthographic processing—are correlates of a broader definition of the dyslexia phenotype.

In recent years, linkage studies have identified chromosomal regions likely to contain genes contributing to dyslexia. Nine regions (*DYX1–DYX9*) have been suggested to date and are listed by the Human Gene Nomenclature Committee. Of these loci, *DYX2* on chromosome 6p21-p22 should be considered one of the most promising candidate regions, since several groups have independently reported linkage between *DYX2* and dyslexia (Cardon et al. 1994; Grigorenko et al. 1997, 2000, 2003; Fisher et al. 1999; Gayán et al. 1999; Kaplan et al. 2002). According to National Center for Biotechnology Information (NCBI) Build 35, the region spans ~16.4 Mb between STR markers *D6S109* and *D6S291*. At the association level, however, the situation is less clear, and positive results have been reported for two independent gene clusters (Deffenbacher et al. 2004; Francks et al. 2004; Cope et al. 2005).

In the present study, we aimed to explore the contribution of the chromosome 6 locus to dyslexia and related phenotypes in the German population. In a first step, we searched for association in a sample of 137 triads (initial sample), employing a combination of STR and SNP-marker genotyping as well as sequencing. In a second step, the study was extended to include an independent sample of 239 triads with dyslexia (replication sample), to confirm the association obtained in the initial sample. Finally, extensive mutation and expression analysis was employed for further characterization of the susceptibility locus.

Material and Methods

Ascertainment of Families

All families of the two samples were of German descent and were recruited through the Departments of Child and Adolescent Psychiatry and Psychotherapy at the Universities of Marburg and Würzburg. All individuals (or the parents of children aged <14 years) gave written informed consent for participa-

Received August 2, 2005; accepted for publication October 7, 2005; electronically published November 17, 2005.

Address for correspondence and reprints: Dr. Johannes Schumacher, Institute of Human Genetics, University of Bonn, Wilhelmstrasse 31, D-53111 Bonn, Germany. E-mail: johannes.schumacher@uni-bonn.de

Am. J. Hum. Genet. 2006;78:000–000. 0002-9297/2006/7801-00XX$15.00

tion in the study. The study was approved by the ethics committees of the Universities of Marburg and Würzburg.

The diagnostic inclusion criteria and phenotypic measures have been described in detail elsewhere (Schulte-Körne et al. 1996, 2001; Schumacher et al. 2005; Ziegler et al. 2005) and are shown in table 1. In brief, potential probands who had difficulty learning to spell or who had received the diagnosis of dyslexia were referred to the investigators by parents, teachers, special educators, or practitioners. Because clinical studies of dyslexia in Germany usually base sample selection on spelling disorder and because our previous findings all rest on this selection criterion (see Schulte-Körne et al. 1996, 1998; Schumacher et al. 2005; Ziegler et al. 2005), the probands' spelling ability was selected for inclusion (for diagnostic criterion, see the following subsection). Since a spelling disorder cannot be reliably diagnosed at an earlier age, only those children attending a regular primary school (no special school, e.g., for learning-disabled children) who had reached at least the middle of second grade were included in the study. All children were investigated at one of the Departments of Child and Adolescent Psychiatry, by use of standardized and unstandardized tests as described below, and family and medical history were taken.

To exclude families in which the proband or a sibling showed symptoms of attention deficit–hyperactivity disorder (ADHD [MIM 143465]), a standardized clinical interview (Unnewehr et al. 1995) was performed with the mother. The reasons for exclusion of comorbid children with dyslexia and ADHD or of children who had siblings with ADHD were, first, that the traits might overlap (Willcutt et al. 2002) and, second, that symptoms of inattention and hyperactivity might influence child behavior during the neuropsychological examinations. Additional exclusionary criteria were a bilingual education, intelligence quotient (IQ) <85, and an uncorrected peripheral hearing or vision disorder or a psychiatric or neurological disorder influencing the development of reading and spelling ability.

Criteria for Dyslexia

The diagnosis of dyslexia was based on the spelling score, with use of the T distribution of the general population. For inclusion in the trial, the following discrepancy criterion had to be fulfilled by the proband: on the basis of an assumed correlation between the IQ and spelling of 0.4 (Schulte-Körne et al. 1996, 2001), an expected spelling score was estimated. The child was classified as "affected" if the discrepancy between the expected and the observed spelling score was $\geqslant 1\sigma$.

Spelling was measured using age-appropriate spelling tests (writing to dictation) that render T scores that are distributed as $N(50,100)$ in unaffected children (Schulte-Körne et al. 2001). The IQ was assessed using the Culture Fair Test (CFT-1) (Weiß and Osterland 1997) or CFT-20 (Weiß 1998), depending on the proband's age.

Phenotypic Measures

Probands fulfilling the inclusion criteria were assessed by use of several psychometric tests, none of which was administered to the parents. These tests targeted different aspects of the dyslexia phenotype, with single-word reading, PA, phonological decoding (PD), rapid naming (RN), and orthographic coding (OC).

Table 1

Sample Characteristics

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

All probands and their siblings performed a single-word and nonword reading test (Salzburger Lese- und Rechtschreibtest [Landerl et al. 1997]). This test also renders T scores that are distributed as $N(50,100)$ in unaffected children (Landerl et al. 1997). Because there are no standardized German reading or PD tests for children at or above the fifth grade, an unstandardized reading test was administered to the children in our study who had reached that grade (Schumacher et al. 2005). The test requires children to read a list of 48 words and 48 pronounceable nonwords as accurately and as quickly as possible. The dependent variables were the number of words and nonwords read correctly in 1 min. Population data and age corrections were not available for this test.

Three tests were administered to measure the spectrum of phonological awareness of study children in grades 2–4. The tests were presented verbally, and subjects were to respond orally. The tests included a phoneme-segmentation task, a phoneme-deletion task, and a phoneme-reversal task. For children at or above the fifth grade, a phoneme-segmentation test, a phoneme-reversal test, and a phoneme-binding and word-reversal test were administered.

A pseudohomophone test was administered, which assesses the ability to discriminate real words from pseudohomophones (OC). These pseudohomophones were generated by substituting or adding graphemes in a real word, which results in a pseudohomophone that sounds very similar to the real word but is orthographically wrong. This test is considered a measure of orthographic processing, because pseudowords and real words sound the same, and the phonological analysis of the word cannot discriminate between them. The children heard single words, with headphones, at a sound pressure of 70 dB. After this, a word or a pseudoword corresponding to the word or nonword presented via headphones appeared on the computer screen. The subjects were asked to press the right button for a pseudoword and the left button for a real word. Thirty-five words or pseudowords were presented one after another in a random order. The number of correct responses (i.e., number of real words that were recognized as correctly spelled, maximally 20) was recorded by the computer. The test was started after four practice trials.

The rapid-naming test used for this study was developed on the basis of the work of Denckla and Rudel (1974). Four trials—naming objects, number, letters, and colors—were conducted. Each trial consisted of items that were arrayed in consecutive rows. Each row consisted of five items that were repeated in a different order, for a total of 10 rows. The trials were printed on a sheet of paper, and children were asked to read them as quickly as they could without making mistakes. A stopwatch was used to measure the time taken by the child to name all stimuli on the entire list. A practice item was given before starting the tests. Color naming was measured by rectangles of five different colors (red, green, brown, blue, and black). Each color was presented in a random order, with the provision that no item appear twice in succession. Number

naming was measured with 1-digit numbers (e.g., 7, 2, 9, 6, 4) in the same way as color naming. Object naming was measured with colored line drawings of common objects (e.g., scissors, candle, comb, clock, key); for letter naming, single consonants or vowels (e.g., p, s, o, a, d) were presented.

STR- and SNP-Marker Genotyping

For STR-marker genotyping, one oligonucleotide of each primer pair was fluorescein labeled, and PCRs were performed on MJ Research thermocyclers. The resulting amplified products were separated on denaturing polyacrylamide gels on an automated DNA sequencer (Model 377 [Applied Biosystems]). Allele sizes were determined relative to an internal size standard in each lane by use of Genescan Analysis Version 2.1.1 and Genotyper Version 2.0 software (Applied Biosystems). All gels were scored independently by two individuals who were blind to the disease status. SNP-marker genotyping was performed using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (Sequenom). PCR assays and associated extension reactions were performed on MJ Research thermocyclers. Cleaned extension products were analyzed by a Mass ARRAY mass spectrometer (Bruker Daltonik), and peaks were identified using the SpectroTYPER RT 2.0 software (Sequenom). All genotypes (1) were scored independently by two individuals who were blind to the disease status and (2) were tested for Mendelian inheritance, in every triad, by use of PedCheck (O'Connell and Weeks 1998).

DNA Resequencing

PCR was performed on MJ Research thermocyclers. All products were cleaned of unincorporated primers and dNTPs by use of shrimp alkaline phosphatase and exonuclease I and were further sequenced using DYEnamic ET Dye terminator kit (Amersham Biosciences). Sequencing products were electrophoresed using a MegaBACE 1000 instrument and MegaBACE long-read matrix, were visualized using the Sequence Analyzer v3.0 software (Amersham Biosciences), and were further aligned using the Pregap and Gap4 software (Staden Package). In addition, a separate viewer compared each FASTA output from sequencing results with corresponding genomic sequences (GenBank accession number NT_007592) with use of Blast 2 sequences.

Northern-Blot and Expression Analysis

A *DCDC2* probe was generated, by touch-down PCR, with oligonucleotides DCDC2-probeF (5′-GCAAGTCGAGGAGATTCTGG-3′) and DCDC2-probeR (5′-CAAGTGGCAATTGTCTTCCA-3′) at annealing temperatures of 63°C–55°C. The PCR product was [^{32}P]-labeled by the use of Ready-to-Go DNA labeling beads (Amersham) and was purified by ProbeQuant G-50 Micro Columns (Amersham), in accordance with the manufacturer's recommendations. Hybridization of human brain tissue northern-blot panels II and V and human fetal tissues II (CLONTECH) was performed overnight in ExpressHyb Solution (CLONTECH) at 68°C. Blots were washed in accordance with standard protocol, and autoradiographs were exposed 3–8 d at −70°C.

The CLONTECH human multiple tissue cDNA panels I and II and human-blood fractions were used as templates for PCRs with primers DCDC2-2F (5′-CCTCATAAACCCAGCTTCTCG-3′) and DCDC2-2R (5′-CAGGTTGAGGTTCCAGTCGAT-3′) for the analysis of *DCDC2* "long," DCDC2-smallF (5′-ATTCAGCCACACGATGTCAC-3′) and DCDC2-probeR (5′-CAAGTGGCAATTGTCTTCCA-3′) for analysis of *DCDC2* "short," and KAAG1-F2 (5′-CCCAACTGCAACGAGAGTTT-3′) and KAAG1-R2 (5′-GTGTGTGTGCGTCCTCCTC-3′) for the analysis of *KAAG1*. PCR reactions were performed with annealing temperatures of 60°C (DCDC2-2F/-2R), 56°C (DCDC2-smallF/-probeR), and 63°C (KAAG1-F2/-R2) in 35 cycles. PCR products were separated by electrophoresis on 1.5% agarose gels and were visualized by UV and ethidium-bromide staining.

Statistical Analysis

In both samples, transmitted and untransmitted alleles were determined, and the transmission/disequilibrium test (TDT) was performed to test for genetic association in the triads. To allow for multiple alleles, we used the marginal homogeneity test (Spielman and Ewens 1996). If an asymptotic P value was <.1, an exact P value was calculated as implemented in S.A.G.E., v4.6. In the initial, replication, and pooled samples, genotypic relative risks (GRR) with 95% CIs were estimated for single loci on the basis of the methods of Scherag et al. (2002) and Franke et al. (2005). Since this assumes diallelic markers, STR alleles were summarized on the basis of their transmission frequencies. Two-locus haplotype frequencies were estimated using the expectation-maximization algorithm, as implemented in UNPHASED v2.403 (Dudbridge 2003). The respective GRRs with 95% CIs were obtained as indicated above. To test for association with additionally assessed quantitative component processes in the initial sample, the approach of Rabinowitz (1997), implemented in QTDT v2.4.6 (Abecasis et al. 2000), was utilized with estimation of P values from 10,000 permutations.

Results

Association Analysis and Characterization of the Linkage-Disequilibrium (LD) Region

To identify LD within the linkage region *DYX2*, we first analyzed 16 STR markers encompassing an ~24-Mb region between *D6S289* and *D6S1610* (see table 2). STR-marker positions and distances were extracted from the Marshfield map (Center for Medical Genetics) and from the UCSC Genome Browser (UCSC Genome Bioinformatics). In our initial sample, comprising 137 triads with dyslexia, we found most-significant association between the disease status and STR marker *D6S276* ($P = .004$) (see table 3). Alleles 5 and 6 were more frequently transmitted to the affected children (53 transmissions vs. 33

Table 2

Location of STR Markers Used for Systematic TDT Analysis of Chromosomal Region 6p21-p22

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

Table 3

Systematic TDT Analysis: Results of the Initial Dyslexia Sample (n = 137 Triads)

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

nontransmissions and 13 transmissions vs. 1 nontransmission, respectively). *D6S276* is located within the doublecortin-domain-containing-2 gene (*DCDC2* [MIM 605755]) and in close proximity to the kidney-associated-antigen-1 gene (*KAAG1* [MIM 608211]) and the vesicular-membrane-protein-p24 gene (*VMP*) (fig. 1*A*).

To characterize the LD region and to explore the genetic variation within these genes, the nonrepetitive DNA segments at the *VMP/DCDC2/KAAG1* locus were resequenced in three patients, two of whom carried one of the risk alleles at *D6S276*. In total, 170 kb of genomic DNA was analyzed (from 24213376 bp to 24486788 bp, according to the public-map contig NT_007592), including all coding and most intronic sequences, the putative promoters, and the 5′ and 3′ UTR regions. Comparison of the susceptibility sequence with the public sequence (NT_007592) revealed 165 SNPs and 12 deletion/insertion polymorphisms. Of the identified variants, 33 were previously unknown and are now submitted to dbSNP under accession numbers ss38319782–ss38319810 and ss38342944–ss38342947.

To refine the LD region and to determine its boundaries, we performed TDT analysis with 29 of the identified genetic variants, encompassing all three genes with constant intermarker distances. In total, 25 SNPs and 4 STR markers were genotyped in the initial dyslexia sample, with an average marker density of 8.8 kb (fig. 1 and table 4). Of the genotyped variants, significant transmission disequilibrium (53 transmissions vs. 30 nontransmissions; $P = .011$) was found at SNP marker *rs793862,* which is located within *DCDC2*. A trend toward association was observed for *rs807701* (69 transmissions vs. 50 nontransmissions; $P = .058$), located within *DCDC2,* and *DYX2_SNP5* (rs number pending) at the *VMP* locus (37 transmissions vs. 22 nontransmissions; $P = .053$). For the haplotype analysis, we studied marker combinations, using the four most-associated variants from the single-marker analysis (table 5). This analysis strengthened the observed association, with two marker combinations showing the strongest association. Both the haplotypes G-5/6 at *DYX2_SNP5-D6S276* and A-C at *rs793862–rs807701* were highly significantly overtransmitted (both $P < .0001$) to the affected probands.

Detailed Genotype-Phenotype Analysis

Since two of the previous association studies of chromosomal region 6p22 have reported only positive association in the subsets of their samples with a more severe phenotype of dyslexia (Deffenbacher et al. 2004; Francks et al. 2004), we performed additional analyses, with selection of families with more-severely affected children from the initial sample, and reanalyzed the three most-associated SNP markers. For this detailed genotype-phenotype study, varying criteria for severity of spelling disorder were used as the phenotypic trait. In short, on the basis of a spelling score expected from the IQ, the proband had to have a discrepancy of at least 1 SD between the observed and expected spelling scores. To select families of more-severely affected children in further analyses, probands had to have a discrepancy of 2–2.5 SDs. All 137 affected probands of our initial sample fulfilled the criteria of a difference of 1 SD, whereas 101 index cases were more severely affected (2 SD), and 47 probands showed a most-severe spelling disorder, defined by 2.5 SD. For both markers located in *DCDC2,* the GRRs increased continuously with the application of more-marked definitions of spelling disorder. The GRR increased for the genotype A-A at *rs793862* from 3.15 (95% CI 1.30–7.66; $P = .011$) with use of the 1-SD criterion to 3.62 (95% CI 1.27–10.30; $P = .001$) with use of 2 SD and to 5.40 (95% CI 1.27–23.01; $P = .002$) with use of the most-severe criteria of 2.5 SD. Accordingly, the GRR for the genotype C-C at *rs807701* increased from 1.88 (95% CI 0.89–3.97; $P = .058$) with use of 1 SD to 2.05 (95% CI 0.85–4.93; $P = .032$) with use of 2 SD and to 5.04 (95% CI 1.35–18.88; $P = .002$) with use of 2.5 SD (see table 6). Furthermore, with selection of more-severely affected families, one additional SNP with use of the 2-SD criterion and four additional SNPs with use of the 2.5-SD criterion appeared to be significantly associated with the patients (table 6). In the haplotype analysis of the *DCDC2* marker combination *rs793862–rs807701,* the GRR also increased for the homozygous haplotype A-C, from 4.11 (95% CI 2.77–6.08; $P < .0001$) with use of 1 SD to 4.81 with use of 2 SD (95% CI 3.24–7.13; $P < .0001$) and to 11.13 (95% CI 6.32–19.60; $P < .0001$) with use of 2.5 SD (table 5). In contrast, the GRR of *DYX2_SNP5,* which is located within *VMP* and ~60 kb more telomeric to the analyzed *DCDC2* variants, did not show even an increasing trend with selection of more-severely affected families (table 6). In addition, no other SNP marker at the *VMP* locus showed significant association with selection of the sample for severity of spelling disorder (table 6).

Finally, we investigated association in the initial sample using quantitative component processes, including PA, PD, RN, and OC. None of these various quantitative component measures showed evidence of association, neither for SNP markers at the *DCDC2* locus (*rs793862*; *rs807701*) nor for *DYX2_SNP5* within the *VMP* gene (data not shown). This was also true when the analysis was restricted to more-severely selected affected families

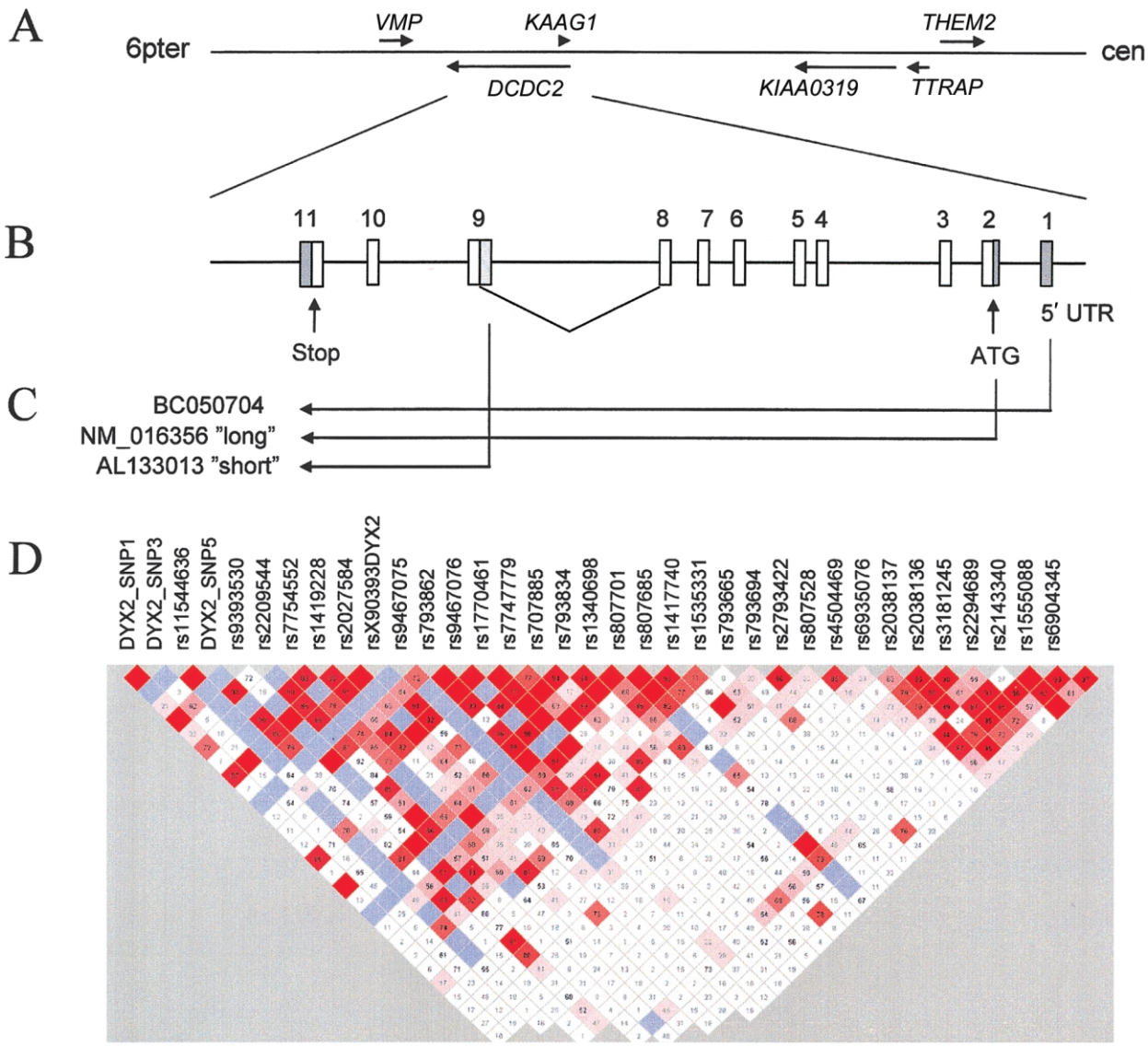

Figure 1 Genomic, transcript, and LD maps for the *DYX2* locus. *A*, The two *DYX2* gene clusters: the distal cluster with *VMP/DCDC2/KAAG1* and the more-proximal, with *KIAA0319/TTRAP/THEM2*. *B* and *C*, Genomic structure of the *DCDC2* gene implicated by association study. Transcripts produced by alternative initiation and splicing from *DCDC2*. Gray boxes denote untranslated exons. *D*, LD between markers genotyped in this study. The distal and proximal gene cluster are not in LD, consistent with their different levels of association to dyslexia. Also, *VMP* and *DCDC2* show only moderate LD.

(data not shown). At the phenotypic level, therefore, no component process characterizes the genetic effect of the *DCDC2* locus more precisely than does spelling disorder itself.

Replication Analysis of an Independent Sample

To confirm the association results, we extended our analysis and included an independent replication sample of 239 triads with dyslexia. In this sample, we genotyped all genetic *DCDC2* markers (*D6S276*, *rs793862*, and *rs807701*) that had been significant in the initial sample by performing single-marker or haplotype analysis, as well as by analyzing the genotype-phenotype relationship. In the single-marker analysis, none of the three variants showed association with the disease status of dyslexia (tables 4 and 6). However, the haplotype A-C at *rs793862–rs807701* again showed association, with overtransmission to the affected probands ($P = .001$) (table 5). The same haplotype was associated in the initial sample, and we therefore replicated our results, at the haplotypic level, in an independent sample.

Since our association results were strongest in the initial sample with selection of families with more-severe affection, we also analyzed our replication data set by

Table 4

TDT of the Initial Dyslexia Sample (137 Triads) with Use of Genetic Variants at the *VMP/DCDC2/KAAG1* Gene Locus

Gene and Marker	Position[a] (bp)	Intermarker Distances (bp)	Allele Variant[b]	TDT *P*[c]
VMP:				
DYX2_SNP1	24234089	…	G/A	1.000
DYX2_STR1	24234569	480	…	.542
DYX2_SNP3	24242396	7,827	C/T	.237
rs11544636	24253873	11,477	T/G	.066
DYX2_SNP5	24254738	865	G/T	.053
rs9393530	24255337	599	G/A	.847
DYX2_STR2	24258101	2,764	…	.617
rs2209544	24265354	7,253	T/G	.199
rs7754552	24267116	1,762	A/G	.617
DCDC2/KAAG1:				
DYX2_STR3	24281285	14,169	…	.396
rs1419228[d]	24286285	5,000	G/A	.515
D6S276[e]	24294148	7,863	…	***.005***
rs2027584	24299438	5,290	C/A	.376
rsX90393DYX2	24302564	3,125	C/T	.500
rs793862[d,e]	24315179	12,615	A/G	***.011***
rs9467076	24317234	2,055	C/T	.384
rs9460976	24321099	3,865	A/G	.274
rs1770461	24329784	8,685	T/G	.206
rs7747779	24331586	1,802	T/C	.896
rs707885	24332655	1,069	A/G	.695
rs793834	24342912	10,257	T/C	.204
rs1340698	24364705	21,793	G/A	.257
rs807701[e]	24381770	17,065	C/T	.058
D6S2439	24414964	33,194	…	.326
rs807685[d]	24418623	3,659	T/A	.127
rs1417740	24433606	14,983	G/T	.112
rs1535331	24450591	16,985	A/G	.123
rs793665	24471260	20,669	G/C	.248
rs793694	24483529	12,269	G/A	.182
DYX2_STR4	24490272	6,743	…	.431
rs2793422[f]	24526327	36,055	A/G	.104
KIAA0319/TTRAP/THEM2:				
rs807528	24652461	…	C/G	.183
rs4504469[f,g]	24696863	44,402	C/T	.460
rs6935076[f]	24752301	55,438	C/T	.485
rs2038137[f,g]	24753922	1,621	C/A	.738
rs2038136	24753926	4	C/G	.664
rs3181245	24759299	5,373	G/C	1.000
rs2294689[g]	24761252	1,953	G/C	.376
rs2143340[f,g]	24767050	5,798	T/C	.274
rs1555088[g]	24792775	25,725	A/G	.652
rs6904345	24805403	12,628	T/C	.344

[a] Positions are based on NCBI Build 35.

[b] The first allele is more frequently transmitted to the affected probands.

[c] Significant values are shown in bold italics.

[d] SNP analyzed in the study by Deffenbacher et al. (2004).

[e] Single-marker analysis of the independent triad sample—*D6S276*: global $P = .366$; *rs793862* allele A 84 transmitted/80 not transmitted ($P = .797$); *rs807701* allele C: 95 transmitted/94 not transmitted ($P = .934$). (See also table 6.)

[f] SNP analyzed in the study by Cope et al. (2005).

[g] SNP analyzed in the study by Francks et al. (2004).

Table 5

Haplotype Analysis of the Initial (137 Triads) and Replication (239 Triads) Dyslexia Samples, with Use of Varying Criteria for Severity of Spelling Disability

		Findings by Criterion					
		1 SD		2 SD		2.5 SD	
Sample and Haplotype	Alleles	*P*	GRR (95% CI)	*P*	GRR (95% CI)	*P*	GRR (95% CI)
Initial[a]:							
DYX2_SNP5–D6S276	G-5/6	<.0001	…	<.0001	…	.5000	…
rs793862–rs807701	A-C	<.0001	4.11 (2.77–6.08)	<.0001	4.81 (3.24–7.13)	<.0001	11.13 (6.32–19.60)
Replication[b]:							
rs793862–rs807701	A-C	<.0001	1.49 (1.16–1.93)	<.0001	2.36 (1.70–3.27)	<.0001	2.95 (1.76–3.27)

Note.—GRR given for the homozygous haplotypes with 95% CI. Two-sided *P* values are shown for both samples.

[a] For 1-SD sample, $n = 137$; for 2-SD sample, $n = 101$; for 2.5-SD sample, $n = 47$.

[b] For 1-SD sample, $n = 239$; for 2-SD sample, $n = 140$; for 2.5-SD sample, $n = 64$.

using the various SD criteria for spelling disorder. Of the 239 patients who all fulfilled the criteria for 1 SD, 140 probands were affected with use of the 2-SD criterion, and 64 patients showed the most-severe spelling disorder, with use of the 2.5-SD criterion. In the single-marker analysis, both SNPs (*rs793862* and *rs807701*) appeared to be significantly associated with the patients who fulfilled the 2.5-SD criterion (table 6). Accordingly, the GRR for the marker combination *rs793862–rs807701* increased for the homozygous haplotype A-C from 1.49 (95% CI 1.16–1.93; $P = .001$) with use of the 1-SD criterion to 2.36 with use of the 2-SD criterion (95% CI 1.70–3.27; $P < .0001$) to 2.95 (95% CI 1.76–3.27; $P < .0001$) with use of the 2.5-SD criterion (table 5). Thus, our results from both the initial and the replication sample independently suggest that genetic variation within the *DCDC2* locus contributes in particular to the development of severe spelling disorder.

Mutation and Expression Analysis

To further characterize the identified susceptibility locus, we performed extensive mutation and expression analysis. Since our association evidence was strongest for *DCDC2,* we were interested mainly in the mutation analysis and expression pattern of this gene. Since *KAAG1* and *DCDC2* overlap, with *KAAG1* extending from intron 1 to intron 2 of *DCDC2,* we also analyzed the expression pattern of *KAAG1.*

The mutation analysis aimed at both the detection of common-disease variants and the identification of rare mutations, which are often causative in patients with more-severe phenotypes. We therefore resequenced the coding as well as the 5′ and 3′ UTR regions of *DCDC2* and *KAAG1* (in total, 12 exons with flanking sequences; 4,330 bp) of 47 patients with dyslexia and 47 controls; both samples were randomly selected. In total, we identified 19 genetic variants, 9 of which had been previously unknown; we submitted the 9 to dbSNP (rs numbers pending). Four of the identified variants are responsible for an amino acid substitution, and all of them are located in *DCDC2*: *DCDC2-SNP1* in exon 4 (Pro152Ala), *rs2274305* in exon 5 (Ser221Gly), *DCDC2-SNP2* in exon 10 (His394Pro), and *rs9460973* in exon 11 (Asp456Lys). Of these variants, only one SNP seemed to be associated with disease risk: allele A of *rs2274305* occurred more frequently in patients than in controls (63.8% vs. 50%, respectively) (data not shown). However, it seems unlikely that this is the common risk allele, since allele A of this marker is not specific to the identified risk haplotype A-C at *rs793862–rs807701.*

We then investigated the expression patterns of *DCDC2* and *KAAG1* by RT-PCRs, using CLONTECH human–multiple-tissue cDNA panels I and II. In addition, both genes were analyzed by northern-blot hybridizations with use of CLONTECH human–brain-tissue northern-blot panels II and V. To detect and further characterize alternative transcripts, 5′ and 3′ RACE experiments were performed using Marathon-Ready cDNA (CLONTECH) from brain tissue. All expression analyses were performed in accordance with the manufacturers' protocols. For *DCDC2,* the expression of three alternative transcripts was analyzed. The transcript BC050704 consists of 11 exons, and an expression at low levels was found in kidney tissue only. The transcript NM_016356 (called the "long" transcript) ranges from exon 2 to exon 11, whereas the third transcript (AL133043) is the "short" *DCDC2* form and starts with a 5′ extended exon 9. Our 5′ RACE analysis with the "short" isoform showed that the alternative exon 9 consists of at least 1,646 bp and is substantially longer than indicated in the

Table 6

Selection for Severity of Spelling Disability: TDT Results for the Initial Dyslexia Sample, Replication Sample, and Pooled Sample, with Use of SNPs at the *VMP/DCDC2/KAAG1* Gene Locus

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

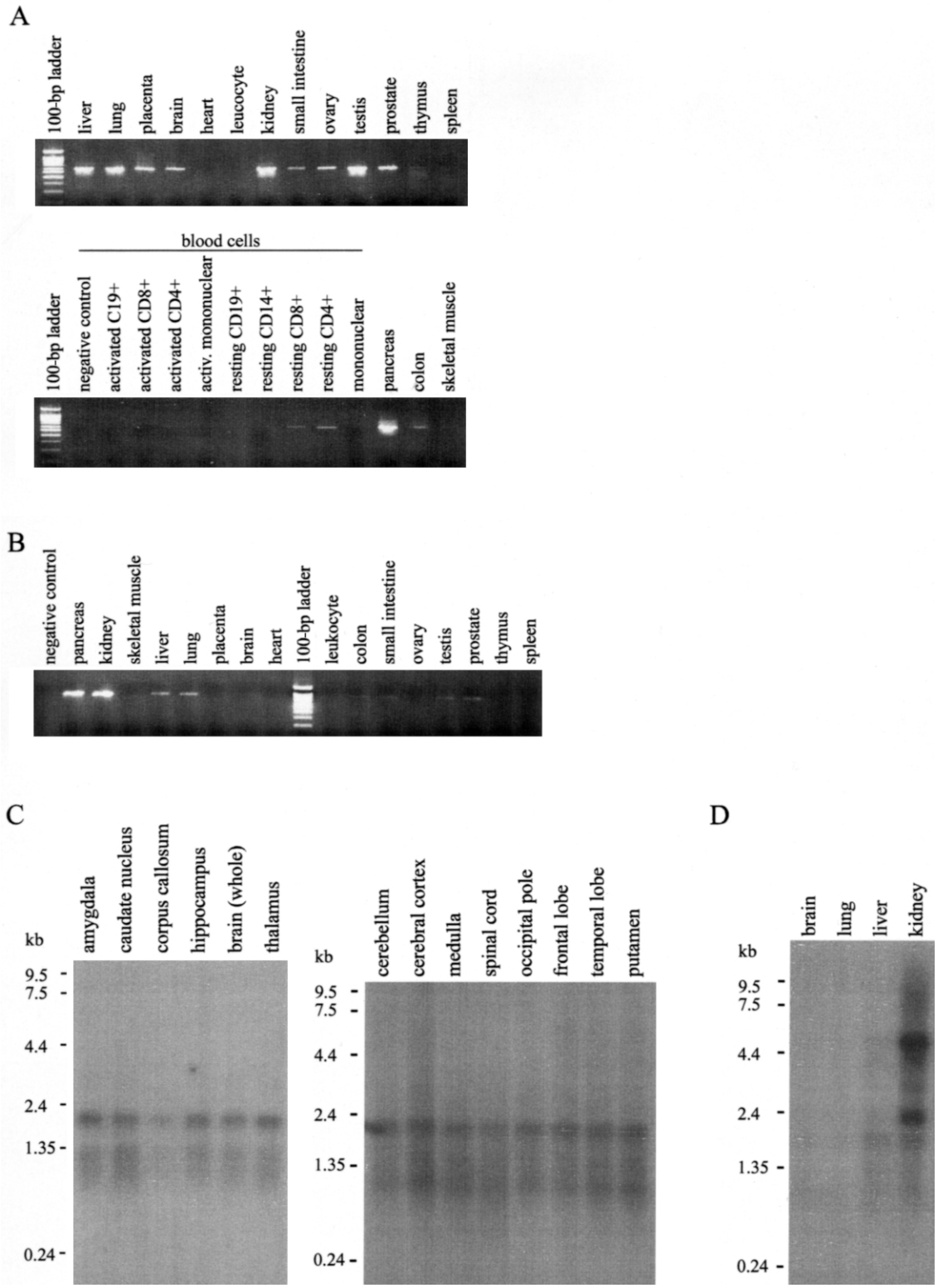

Figure 2 Expression of *DCDC2*. *A,* Expression analysis of *DCDC2* "long" by RT-PCR of human tissue samples with use of primers DCDC2-2F and DCDC2-2R located in exons 5 and 9, respectively. PCR products of 577 bp were separated by electrophoresis on agarose gels and were visualized by UV and ethidium-bromide staining. *B,* Expression analysis of *DCDC2* "short" by RT-PCR of human tissue samples with use of primers DCDC2-smallF and DCDC2-probeR located in the 5′ extension of exon 9 and in exon 11, respectively, which generate an 844-bp PCR product. *C,* Northern-blot analysis of the *DCDC2* transcript of human brain tissues, which indicate an expression of an ~2-kb transcript in most brain tissues. *D,* Northern-blot analysis of *DCDC2* of human fetal tissues, which shows strong expression of high molecular transcripts in fetal kidney.

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

Figure 3 Expression of *KAAG1.* The legend is available in its entirety in the online edition of *The American Journal of Human Genetics.*

GenBank entry. The RT-PCR analyses of cDNAs showed that the "long" isoform of *DCDC2* (NM_016356) is expressed at its highest levels in liver, lung, kidney, testis, and pancreas, and the "short" isoform (AL133043) is expressed at its highest levels in kidney and pancreas (fig. 2*A* and 2*B*). In brain tissue, only the "long" transcript (NM_016356) was expressed (fig. 2*A* and 2*B*). By performing RACE experiments with use of brain cDNA, we also found a weak expression in brain tissue of the "short" *DCDC2* transcript (AL133043). The northern-blot hybridization of different brain tissues showed that an ~2-kb messenger that corresponds to the "long" *DCDC2* transcript (NM_016356) is expressed to an equal degree in all brain regions, with the exception of the corpus callosum, in which the expression is weak (fig. 2*C*). Using northern-blot hybridization of fetal-tissue blot panels, we also found an expression of the "long" *DCDC2* transcript (NM_016356) in brain tissue, although the expression in the liver and kidney was stronger (fig. 2*D*). For *KAAG1,* we analyzed the expression of the transcript with AF181722 (GenBank). The expression of *KAAG1,* which consists of one exon, could not be detected in brain tissue by northern-blot hybridization (data not shown). This is consistent with our RT-PCR results, since no brain expression was observed with use of the multiple-tissue cDNA panels. The strongest expression of *KAAG1* was found in the pancreas; lower levels were found in the testis, liver, and kidney (see fig. 3).

The KIAA0319/TTRAP/THEM2 *Gene Cluster*

The *DYX2* linkage region harbors a second gene cluster, for which Francks et al. (2004) and Cope et al. (2005) reported evidence of association, using dyslexic subjects from the United Kingdom (fig. 1*A*). Within this cluster, which is separated from the *VMP/DCDC2/KAAG1* locus by ~185 kb, three genes are located—*KIAA0319* (MIM 609269), the TRAF-and-TNF-receptor-associated-protein gene (*TTRAP* [MIM 605764]), and the thioesterase-superfamily-member-2 gene (*THEM2*). To assess whether genetic variation at this locus may also contribute to dyslexia in the German population, we genotyped 10 SNPs encompassing this gene cluster. To achieve the maximal statistical power for the replication analysis, we used our pooled sample of 376 triads with dyslexia. None of the variants showed association in the overall sample (table 4). Only one variant showed a borderline significant finding in the most-strongly affected children with spelling disorder (table 7). Our study included the most-significantly associated markers reported by Francks et al. (2004) and Cope et al. (2005) and had an average power of >95% for detection of an effect at the threshold level of 5%. Power was estimated under the assumption of full dominance, with reduced penetrance and a population risk of 5%–10%. On the basis of our results, we conclude that genetic variation at the *KIAA0319/TTRAP/THEM2* gene cluster is not of major importance in the development of dyslexia in the German population.

Discussion

Together with the reported findings of Deffenbacher et al. (2004), who found evidence of an association at the *VMP/DCDC2/KAAG1* locus in a U.S. sample of dyslexia stratified for severity, our results provide compelling evidence that genetic variation within this locus confers susceptibility to the development of dyslexia. Although we found association in two triad samples that were not selected for disease severity, our results independently indicate that this locus contributes in particular to severe deficits in spelling ability. In the pooled sample with the criteria of ≥ 2.5 SD for severe spelling disability (111 families), we estimate that the genetic effect of *DCDC2* is associated with a GRR in the range of 4.88 (95% CI 3.32–7.15, on the basis of the homozygous presence of haplotype A-C at *rs793862–rs807701*).

Cope et al. (2005) employed a high-density screen encompassing the *VMP/DCDC2/KAAG1* locus, with SNPs at 2–3-kb intervals (with the exception of *DCDC2* introns 2, 7, and 8), and they failed to detect an association in their U.K. case-control sample using a DNA pooling approach. There are several explanations for these divergent findings. First, in the sample of Cope et al. (2005), no selection for disease severity was employed. Second, the power of their sample may have been too small to detect the genetic effect of the *VMP/DCDC2/KAAG1* locus, especially with use of a DNA-pooling approach in which no individual genotype were obtained and haplotype analysis could not be performed. Third, population-specific genetic heterogeneity within the *DYX2* region could explain the different association finding through different populations.

Table 7

Selection for Severity of Spelling Disability: TDT Results of the Pooled Sample with Use of SNPs at the *KIAA0319/TTRAP/THEM2* Gene Locus

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

The strongest association in the study performed by Deffenbacher et al. (2004) was found with four SNP markers (called SNP 6, 11, 12, and 13). These markers—one of which is *rs793862*—are located in the same genomic region within *DCDC2* in which we also observed the strongest association in our German sample (single-marker analysis with use of criteria of ⩾2.5 SD for severe spelling disability: significant association between *D6S276* and *rs1417740* and between intron 8 and intron 2, respectively). Therefore, this genomic interval of ~140 kb within *DCDC2* is most likely to harbor the risk variant(s) for disease susceptibility. This *DCDC2* genomic region is also implicated as a separate haplotype block when LD analysis is performed with our genotypic data (fig. 1*D*). Interestingly, *DCDC2* contains a doublecortin homology domain, which indicates a role in cortical neuron migration. This would be in accordance with our expression data, in which the "long" *DCDC2* transcript was found to be expressed in the fetal and adult CNS, and with postmortem findings, in which abnormal neuronal migration and maturation has been found in patients with dyslexia (summarized by Demonet et al. [2004]).

It is very difficult, at this stage, to integrate the findings for this region, obtained by us and others, into a coherent picture of what phenotype might be most closely associated with disturbances in this region. First of all, at the level of linkage, the effects of the two gene clusters in this region *(VMP/DCDC2/KAAG1* and *KIAA0319/TTRAP/THEM2)* cannot be separated. If these gene clusters independently contribute to different phenotype dimensions, linkage findings in this region are the result of a superimposed picture and become difficult to interpret with respect to the phenotypes involved. Inconsistent findings are also observed at the level of association findings for the individual gene clusters. This may be due to several factors. First, sample selection criteria differed between individual studies. Dyslexic individuals in our study were selected on the basis of a discrepancy between actual spelling and IQ. Cope et al. (2005), Francks et al. (2004), and Deffenbacher et al. (2004) selected dyslexic individuals on the basis of their low word-reading ability or the discrepancy between their IQ and word-reading ability. This could mean that we analyzed different subsamples of the dyslexia spectrum. One might speculate from this that the *DCDC2* gene has more relevance for spelling capability, whereas the *KIAA0319/TTRAP/THEM2* gene cluster (Francks et al. 2004; Cope et al. 2005) has more relevance for word reading. Other possible reasons for discrepancies in research findings include differences in sample size (and, consequently, of power) and differences in patterns of missing genotype-phenotype data (Francks et al. 2004). The basic task for dyslexia research is to identify which phenotype characteristics are associated with which gene cluster, which would thereby lead to an understanding of the genes and the functions of their encoded proteins. A definite answer to this question will require collaborative studies with sufficiently large sample sizes.

Acknowledgments

We are grateful to all investigated individuals for their cooperation in this study. This study was supported by the Deutsche Forschungsgemeinschaft. P.P. was supported by the National Genomic Network of the Bundesministerium für Bildung und Forschung. M.M.N. received support for this work from the Alfried Krupp von Bohlen und Halbach-Stiftung. M.P.-J. is a recipient of a research position from the Swedish Research Council (Vetenskpsrådet). J.K. is supported by grants from the Swedish Research Council, the Academy of Finland, the Sigrid Jusélius Foundation, and the Päivikki and Sakari Sohlberg Foundation. Some of the results were obtained by using the program package S.A.G.E., which is supported by U.S. Public Health Service Resource grant RR03655 from the National Center for Research Resources.

Web Resources

Accession numbers and URLs for data presented herein are as follows:

Blast 2, http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi/
Center for Medical Genetics, http://research.marshfieldclinic.org/genetics/ (for the Marshfield map)
dbSNP, http://www.ncbi.nlm.nih.gov/
GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for NT_007592 and AF181722)
Human Gene Nomenclature Committee, http://www.gene.ucl.ac.uk/nomenclature/
NCBI, http://www.ncbi.nlm.nih.gov/ (for Build 35)
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for dyslexia, ADHD, *DCDC2, KAAG1, KIAA0319,* and *TTRAP*)
Staden Package, http://staden.sourceforge.net/
UCSC Genome Bioinformatics, http://genome.ucsc.edu/ (for the UCSC Genome Browser)
UNPHASED, http://www.mrc-bsu.cam.ac.uk/personal/frank/software/unphased/

References

Abecasis GR, Cardon LR, Cookson WOC (2000) A general test of association for quantitative traits in nuclear families. Am J Hum Genet 66:279–292

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

Cope N, Harold D, Hill G, Moskvina V, Stevenson J, Holmans P, Owen MJ, O'Donovan MC, Williams J (2005) Strong evidence that *KIAA0319* on chromosome 6p is a susceptibility gene for developmental dyslexia. Am J Hum Genet 76:581–591

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

Demonet JF, Taylor MJ, Chaix Y (2004) Developmental dyslexia. Lancet 363:1451–1460

Denckla MB, Rudel RG (1974) Rapid automatized naming of pictured objects, colors, letters and numbers by normal children. Cortex 10: 186–202

Dilling H, Mombour W, Schmidt MH (1991) International classification of mental diseases, ICD-10, German edition. Hans Huber, Bern

Dudbridge F (2003) Pedigree disequilibrium tests for multilocus haplotypes. Genet Epidemiol 25:115–121

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, 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

Franke D, Philippi A, Tores F, Hager J, Ziegler A, König IR (2005) On confidence intervals for genotype relative risks and attributable risks from case parent trio design for candidate-gene studies. Hum Hered 60:81–88

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

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 chromosomes 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

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

Katusic SK, Colligan RC, Barbaresi WJ, Schaid DJ, Jacobsen SJ (2001) Incidence of reading disability in a population-based birth cohort, 1976–1982, Rochester, Minn. Mayo Clin Proc 76:1081–1092

Landerl K, Wimmer H, Moser E (1997) SLRT—Salzburger Lese-und Rechtschreibtest. Hans Huber, Bern

O'Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63:259–266

Rabinowitz D (1997) A transmission disequilibrium test for quantitative trait loci. Hum Hered 47:342–350

Rutter M, Caspi A, Fergusson D, Horwood LJ, Goodman R, Maughan B, Moffitt TE, Meltzer H, Carroll J (2004) Sex differences in developmental reading disability: new findings from 4 epidemiological studies. JAMA 291:2007–2012

Scherag A, Dempfle A, Hinney A, Hebebrand J, Schafer H (2002) Confidence intervals for genotype relative risks and allele frequencies from the case parent trio design for candidate-gene studies. Hum Hered 54:210–217

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

Schulte-Körne G, Deimel W, Remschmidt H (2001) Diagnosis of reading and spelling disorder. Z Kinder Jugendpsychiatr Psychother 29: 113–116

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

Schumacher J, König IR, Plume E, Propping P, Warnke A, Manthey M, Duell M, Kleensang A, Repsilber D, Preis M, Remschmidt H, Ziegler A, Nöthen MM, Schulte-Körne G (2005) Linkage analyses of chromosomal region 18p11-q12 in dyslexia. J Neural Transm (http://www.springerlink.com/(clxfmy45ptvarl553jtgalbs)/app/home/contribution.asp?referrer=parent&backto=searcharticlesresults,1,2;) (electronically published August 3, 2005; accessed November 16, 2005)

Spielman RS, Ewens WJ (1996) The TDT and other family-based tests for linkage disequilibrium and association. Am J Hum Genet 59: 983–989

Unnewehr S, Schneider S, Margraf J (1995) Kinder-DIPS—Diagnostisches Interview bei psychischen Störungen im Kindes-und Jugendalter. Hogrefe, Göttingen

Weiß RH (1998) Grundintelligenztest Skala 2 CFT 20. Hogrefe, Göttingen

Weiß RH, Osterland J (1997) Grundintelligenztest Skala 1 CFT 1. Hogrefe, Göttingen

Willcutt EG, Pennington BF, Smith SD, Cardon LR, Gayán J, Knopik VS, Olson RK, DeFries JC (2002) Quantitative trait locus for reading disability on chromosome 6p is pleiotropic for attention-deficit/hyperactivity disorder. Am J Med Genet 114:260–268

Ziegler A, König IR, Deimel W, Plume E, Nöthen MM, Propping P, Kleensang A, Müller-Myhsok B, Warnke A, Remschmidt H, Schulte-Körne G (2005) Developmental dyslexia—recurrence risk estimates from a German bi-center study using the single proband sib pair design. Hum Hered 59:136–143