DCDC2 is associated with reading disability and modulates neuronal development in the brain

Haiying Meng, Shelley D. Smith, Karl Hager, Matthew Held, Jonathan Liu, Richard K. Olson, Bruce F. Pennington, John C. DeFries, Joel Gelernter, Thomas O’Reilly-Polo, Stefan Somlo, Pawel Skudlarski, Sally E. Shaywitz, Bennett A. Shaywitz, Karen Marchione, Yu Wang, Murugan Paramasivam, Joseph J. LoTurco, Grier P. Page, and Jeffrey R. Gruen

Department of Pediatrics, Yale Child Health Research Center, Yale University School of Medicine, New Haven, CT 06520; Munroe Meyer Institute, University of Nebraska Medical Center, Omaha, NE 68198; M. Keck Foundation Biotechnology Resource Laboratory, Yale University School of Medicine, New Haven, CT 06511; SoftGenetics, State College, PA 16803; Department of Psychology and Institute of Behavioral Genetics, University of Colorado, Boulder, CO 80309; Department of Psychology, University of Denver, Denver, CO 80210; Department of Psychiatry, Veterans Affairs Medical Center, West Haven, CT 06516; Departments of Psychiatry and Internal Medicine, Yale University School of Medicine, New Haven, CT 06519; Department of Physiology and Neurobiology, University of Connecticut, Storrs, CT 06269; and Department of Bionistics, University of Alabama at Birmingham, Birmingham, AL 35294

Communicated by Sherman M. Weissman, Yale University School of Medicine, New Haven, CT, October 3, 2005 (received for review July 19, 2005)

DYX2 on 6p22 is the most replicated reading disability (RD) locus. By saturating a previously identified peak of association with single nucleotide polymorphism markers, we identified a large polymorphic deletion that encodes tandem repeats of putative brain-related transcription factor binding sites in intron 2 of DCDC2. Alleles of this compound repeat are in significant disequilibrium with multiple reading traits. RT-PCR data show that DCDC2 localizes to the regions of the brain where fluent reading occurs, and RNA interference studies show that down-regulation alters neuronal migration. The statistical and functional studies are complementary and are consistent with the latest clinical imaging data for RD. Thus, we propose that DCDC2 is a candidate gene for RD.

Reading disability (RD), or developmental dyslexia, is one of the most common of the complex neurobehavioral disorders, with prevalence rates ranging from 5% to 17% (1). It is characterized by an impairment of reading ability in subjects with normal intelligence and adequate educational opportunities. A range of neuroimaging studies, including diffusion tensor and functional magnetic resonance imaging, show that dyslexics have altered brain activation patterns compared to fluent readers when challenged with reading tasks (2). Partial remediation in language processing deficits results in improved reading, ameliorates disrupted function in brain regions associated with phonologic processing, and produces additional compensatory activation in other brain areas (3). These studies also implicate specific brain locations where genes integral to reading and language are expressed, and which likely are altered in RD.

Over the past 30 years, clinical studies have shown that up to 50% of children of dyslexic parents, 50% of siblings of dyslexics, and 50% of parents of dyslexic children are affected (4). Estimates of heritability range from 44% to 75% (5). The first RD susceptibility region, DYX1, was reported on chromosome 15 in 1983 (6). Additional susceptibility regions localized to either the left or right hemispheres. Six donors were amplified by multiple replacement amplification (Molecular Staging) (18).

RNA Samples. Total RNA samples from 18 areas of adult human brain were purchased from Ambion (see Fig. 4) and were procured from 10 white donors ranging in age from 45 to 79 years, with unknown handedness. RNA samples could not be localized to either the left or right hemispheres. Six donors were

Conflict of interest statement: No conflicts declared.
male. Seven donors died because of cardiac (e.g., congestive heart failure) or respiratory disease (e.g., respiratory failure), one had liver cancer, one had bladder cancer, and one was listed as unknown.

**Genotyping.** TaqMan Assay-on-Demand and Assay-by-Design probes (Applied Biosystems, Foster City, CA) were used to genotype 109 and 39 SNPs, respectively. Six SNPs failed web-based primer design for TaqMan and, consequently, were genotyped by pyrosequencing (Biotage, Upplands). The primers for these SNPs are presented in Table 2, which is published as supporting information on the PNAS web site.

**Deletion genotype.** The common 2,445-bp deletion was genotyped by allele-specific amplification with a combination of three primers in one reaction: universal forward primer (AGCCTGCTACCACACAGAGAA), reverse primer for nondeleted chromosomes (GGAA-CACCTTCAGAAATG), and reverse primer for deleted chromosomes (TGAAAACCCGTCCTCTACTGAG). Reaction products were resolved on 1.5% agarose gels. The deletion fusion fragment was 225 bp, and the nondeleted fragment was 550 bp.

**dbSTS ID no. 808238 genotype.** The compound STR, dbSTS ID no. 808238, was genotyped by sequencing PCR products generated with a forward primer (TGGTTGAATCCCCAGACACCAA) and reverse primer (ATCCCCCATGAAATGAAAAAGG). The sequencing method is described below. Sequence traces results were analyzed and alleles assigned with MUTATION SURVEYOR version 2.6 (SoftGenetics) by comparing samples to reference traces after alignment.

**Quantitative Real-Time RT-PCR.** TaqMan gene expression kits for eight genes in the candidate region (KIAA0319, DCDC2, MRS2L, GPLD1, ALDH5A1, TTRAP, THEM2, and GMNN) and six control genes (GAPDH, 18S, β-actin, HPRT1, PPLA, and PKG1) were purchased from Applied Biosystems. Details are described in Supporting Methods and Materials.

**Statistical Analysis.** GENETIC ANALYSIS SYSTEM (http://users.ox.ac.uk/~ayoun/gas.html) was used to assess the Mendelian transmission of alleles. Identity-by-descent probabilities were estimated with SIMWALK2 (19). We used QTDT (20) to simultaneously test for transmission disequilibrium in the presence of linkage by the orthogonal model (-ao) with variance components (-wega), and permutations for exact P values (-m1000 –1). Through different modeling within QTDT, we tested for parent of origin effects (-ot), the significance of polygenic effects (-weg), evidence for linkage without association (-vega), total association (-at), and population stratification (-ap). HAPLOVIEW (21) and GOLD (22) were used to examine the haplotype structure of the markers, to generate haplotype blocks, and to assess intermarker linkage disequilibrium. HaploTDT was analyzed by FBAT (23).

**In Utero RNA Interference (RNAi).** Plasmids were directly introduced into cells at the cerebral ventricular zone of living rat embryos by in utero electroporation as described in ref. 24. Cells were cotransfected with pCA-eGFP and DCDC2 small hairpin RNA (shRNA) plasmid or control shRNA plasmid. The shRNA plasmid directed against DCDC2 contained the hairpin sequence 5′-ccccccaagaacttttagaaaca(aca)gtgtgaaagttgggaggg-3′ and the control sequence was 5′-ccccagactcagatttt(aaa)ttttaatcaggatcgg-3′. The sequence was selected by its asymmetry and for the absence of any matches to rat genomic sequence in

<table>
<thead>
<tr>
<th>SNP ID</th>
<th>Gene</th>
<th>DISC</th>
<th>PTP</th>
<th>HCH</th>
<th>ENSEMBL location</th>
<th>Celera location</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>DCDC2 Int 6</td>
<td>0.0003</td>
<td>24386848</td>
<td>25512242</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>DCDC2 Int 1</td>
<td>0.0035</td>
<td>24463129</td>
<td>25588523</td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>GPLD1 Int 24</td>
<td>0.0018</td>
<td>24539037</td>
<td>25664490</td>
<td></td>
<td></td>
</tr>
<tr>
<td>117</td>
<td>Intergene</td>
<td>0.0077</td>
<td>24872844</td>
<td>25998238</td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>Intergene</td>
<td>0.0067</td>
<td>0.035</td>
<td>0.0811</td>
<td>25027795</td>
<td>26142106</td>
</tr>
</tbody>
</table>

PTP, phenome transposition; HCH, homonym choice.
the database. Four days after transfection, rat embryonic brains were fixed with 4% paraformaldehyde and sectioned with a vibratome (Leica VT1000S) at 60–80 μm. Nuclei were labeled with TOP-PRO-3 (Molecular Probes). Images were acquired with a Leica TCS SP2 confocal microscope system (0.5–1.0 μm optical section) and processed by using PHOTOSHOP 7.0 (Adobe Systems). For cumulative probability migration plots, the distance of each cell (200–1,400 cells in each analysis condition) from the VZ surface was determined 4 days after transfection. Migration distances were determined with automated particle analyses in IMAGEJ (National Institutes of Health, Bethesda).

Results

Single-Marker Transmission Disequilibrium. We genotyped a total of 147 SNPs distributed through the 1.5-Mb region surrounding JA04 in 153 nuclear RD families recruited by the CLDRC. The strongest QTDT peak was with the discriminant score (DISC) phenotype and QTDT scores. Results for the entire SNP panel can be found in Table 1 and Fig. 1 provide the results from a selected subset of the most significant QTDT scores. Results for the entire SNP panel can be found in Table 3, which is published as supporting information on the PNAS web site. Five SNPs yielded a \( P \) value of \( \leq 0.01 \); two of these SNPs were located in DCDC2. Thirty-seven SNPs yielded a \( P \) value of \( \leq 0.05 \); 11 of these SNPs were located in DCDC2. Of the 31 SNPs distributed through DCDC2 (average minor allele frequency = 0.24), 10 were associated with the DISC phenotype (\( P \leq 0.05 \)).

Intermarker Linkage Disequilibrium. We constructed an intermarker linkage disequilibrium map (Fig. 2a) spanning the 1.5 Mb with GOLD and HAPLOVIEW. There was evidence for five linkage disequilibrium blocks (A to E) spanning small clusters of SNPs in DCDC2 (Fig. 2b). There were three blocks (F to H) centromeric of DCDC2 that corresponded to single-marker QTDT peaks.

Haplotype-TDT. All five haplotype blocks in DCDC2 showed significant transmission disequilibrium with reading performance tasks; three of these blocks, A, B, and D, did not contain single-marker QTDT peaks. Fig. 3 is a graphic presentation of the haplotype transmission disequilibrium data, which is also provided in tabular form in Tables 4 and 5, which are published as supporting information on the PNAS web site. A haplotype in each of blocks A, C, D, E, F, and G was associated with compromised performance in several reading tasks in the context of preserved IQ. Haplotype blocks A, C, D, E, and H were located in DCDC2. There were no haplotypes in block H that showed significant association with any of the cognitive phenotypes.

A Previously Uncharacterized Deletion in DCDC2. C.449792, located in intron 2 of DCDC2 (Fig. 1), showed non-Mendelian allele transmission errors in 10 RD families. To ensure that this error was not an artifact of whole genome amplification, we confirmed these initial genotypes by sequencing PCR products derived from unamplified genomic DNA templates for all 10 families. Allele transmission from the two flanking SNPs, 41 and 42, were typically Mendelian and defined initially the outer boundaries of a 17-kb region with loss-of-heterozygosity (LOH). To identify the extent of the deletion, we interrogated for LOH by sequencing SNPs within the 17-kb genomic region in RD trios. Additional flanking SNPs limited the deletion to 3,848 bp. Finally, we amplified and sequenced a 1,200-bp fusion fragment in subjects with LOH, which
assigned the breakpoints to 24,433,346 and 24,435,659 (ENSEMBL, Fig. 2). Primer walking was used to sequence the nondeleted fragment from the same subjects with LOH. These results confined the deletion to 2,445 bp. Overall, the deletion was 60% AT and contained a 168-bp purine-rich (98% AG) region.

**A Compound STR in the Deletion.** Within the 168-bp purine-rich region was a polymorphic compound STR (dbSTS ID no. 808238) comprised of 10 alleles containing variable copy numbers of (GAGGAAGGAAA)$_n$ and (GGAA)$_n$ repeat units (Table 6, Fig. 2). Primer walking was used to sequence the nondeleted fragment from the same subjects with LOH. These results confined the deletion to 2,445 bp. Overall, the deletion was 60% AT and contained a 168-bp purine-rich (98% AG) region.

By combining the deletion and 10 minor alleles, QTDT showed a peak of transmission disequilibrium with homonym choice ($P = 0.00002$, Table 7, which is published as supporting information on the PNAS web site). TESS (25) comparison to the TRANSFAC database identified 131 putative transcription factor binding sites distributed through the 168 bp of the purine-rich region, including four copies each of $PEA3$ (AGGAAA) and $NF-ATp$ (AGGAAA) sites in repeat unit 1 of dbSTS ID no. 808238. Both transcription factors are expressed in mouse brain.

RNAi of DCDC2 Impairs Radial Neuronal Migration. In utero RNAi was used to test for a functional role of $DCDC2$ in neuronal migration. Cotransfection of plasmid vectors encoding shRNA targeted against $DCDC2$ sequence in developing neocortex or control scrambled sequence along with an $EGFP$ expression plasmid was performed at gestational day 14 in the rat. This transfection method initially labeled ~1% of cells at the surface of the ventricles where new neurons undergo their terminal mitoses. Cells migrate from this surface to the pial surface in 4 to 6 days. We assessed the progress in migration 4 days after transfection for the two conditions. As shown in Fig. 5, cells transfected with control plasmids progressed significantly further away from the ventricular surface and toward the pial surface than did cells transfected with a vector targeted against $DCDC2$. The mean migration distance in matched littermate controls was 606 ± 178 µm, and in the $DCDC2$ shRNA transfection group, the mean migration distance was 376 µm + 135 ($n = 4, P < 0.01$).

**Discussion**

After our previous studies that showed transmission disequilibrium to JA04, we systematically interrogated the 6p22 DYS2 locus for a candidate gene that could confer susceptibility for RD. Starting with single-marker QTDT analysis, we found the strongest peak and concentration of transmission disequilibrium with SNPs in $DCDC2$. The extent of intermarker linkage disequilibrium clustered through the 1.5 Mb of genomic sequence suggests adequate marker density in this region and seven haplotype blocks. Blocks spanning $DCDC2$ also show significant transmission disequilibrium with several quantitative reading phenotypes in the context of preserved IQ, suggesting a specific effect on reading performance and not generalized or global effects on brain function.

We report here the results from 147 SNP markers, but we originally queued 152 consecutive markers in our high-throughput genotyping strategy. Four markers failed PCR and were dropped from the analysis. A fifth marker, C.449792, was flagged for non-Mendelian transmission and was set aside. Only after completion of the single-marker QTDT analysis did we confirm LOH with C.449792 in samples not subjected to multiple displacement amplification and discovered the 2,445-bp deletion in intron 2 of $DCDC2$, between the exons encoding the two doublecortin domains (Fig. 1c).
The deletion, including minor alleles of dbSTS ID no. 808238, is in strong linkage disequilibrium with reading performance ($P = 0.00002$; Table 7). Furthermore, dbSTS ID no. 808238 encodes multiple copies of PEA3 and NF-ATp sites that are active in brain. Loss of this entire regulatory region, as would happen with the common large deletion we found in dyslexics, would therefore have profound affects on DCDC2 function. Polymorphisms would disrupt PEA3 and NF-ATp sites, which in future studies may explain dyslexia in subjects without the common deletion or the variation of reading ability due to allelic heterogeneity.

DCDC2 (also called RU2 and KIAA1154, MIM:605755) is located in the D YX2 locus 500 kb from J A04. The function is unknown but it contains two doublecortin peptide domains that were originally described in the doublecortin gene (DCX, MIM: 300121) encoded on the X chromosome (Fig. 1c). DCX encodes a cytoplasmic protein that directs neuronal migration by regulating the organization and stability of microtubules and is mutated in human X-linked lissencephaly (28) and double cortex syndrome. Lissencephaly is a neuronal migration defect that produces profound mental retardation and seizures (29). Double cortex syndrome is caused by arrested migration halfway to the cortex, producing a subcortical neuronal band heterotopia or “double cortex.” For both syndromes, the large majority of point mutations cluster within the conserved doublecortin peptide motifs of DCX, which are also encoded in DCDC2.

Converging imaging data implicate three important regions in the left hemisphere that are important for fluent reading: the anterior system in the inferior frontal region, the dorsal parieto-temporal system involving the angular, supramarginal, and posterior portions of the superior temporal gyri, and the ventral occipito-temporal system involving portions of the middle temporal and middle occipital gyri (3, 30). Imaging studies of dyslexic adults and children show a disruption of posterior reading systems in parieto-temporal and occipito-temporal regions (31). Yet DCDC2 is highly expressed in the same regions activated by fluent and dyslexic readers, suggesting that dysregulation, attributable to polymorphisms of a regulatory region, and not complete disruption of a protein product, could explain the expression patterns.

These findings are consistent with the hypothesis that dyslexia is associated with subtle changes, like the anecdotal microscopic anomalies reported by Galaburda and colleagues (32), in the migration of neurons in developing neocortex. Similarities in structure and cellular function between DCDC2 and DCX, a gene known to be critical to neuronal migration, further supports a hypothesis for impaired neuronal migration. Loss of function of DCX causes severe developmental disruption in neocortex. In contrast, dyslexia is not characterized by large malformations of neocortex. The DCDC2 alleles that associate with dyslexia, however, would not be expected to be nulls, and so even if DCX and DCDC2 had similarly critical roles in neuronal migration, large malformations would not be an expected phenotype for the described alleles. In addition, a comparison of the RNAi results after DCX RNAi (24), with that after DCDC2 RNAi, suggest that DCX may be necessary for neuronal migration, whereas DCDC2 may be more modulatory.

**Fig. 5.** In utero RNAi against DCDC2. (a) Control transfection of a neutral shRNA vector and EGFP shows normal migration after 4 days. Most neurons have migrated well away from the ventricular surface (Vent) toward the pial surface (Pia). (b) Neurons transfected with an shRNA vector directed against DCDC2 migrate abnormally. (c) Cumulative probability plot of the migration distances from the ventricular surface of all transfected EGFP plus cells shown in a and b in the two transfection conditions. (Scale bar: 100 μm.)
Unlike the effects of DCX RNAi treatment (20), DCDC2 RNAi treatment allows cells to migrate further, attain typical migratory bipolar morphologies, and does not induce the formation of large subcortical band heterotopia. Although the RNAi treatment does not exclusively target mRNAs that populate the reading centers, when considered in the context of DCDC2 expression in inferior and medial temporal cortex, it offers a plausible pathophysiological mechanism for RD due to genetic expression heterogeneity. DCDC2 heterogeneity is also consistent with other pathophysiological mechanisms. Imaging studies have shown a functional disruption of a more subtle nature, demonstrable only in composite maps of pooled subjects imaged at 1.5 tesla, in areas where heterotopias have not been described. Accordingly, it may be that DCDC2 heterogeneity sensitizes the dyslexic reader to disruption in the development of “a hierarchy of local combination detectors” in the occipitotemporal system, as postulated most recently by Dehaene and colleagues (33).

Previous attempts at transmission disequilibrium mapping with sparse densities of SNP markers in this region, 31 SNPs over 10 Mb (34) and 57 SNPs over 5.7 Mb (35), proved inconclusive. One of these studies, which found significant linkage disequilibrium with markers around the TTRAP gene (35), did not include markers over DCDC2. A recent study covering VMP1, DCDC2, KIAA0319, TTRAP, and THEM2 identified maximum association with KIAA0319 (36). Given its specificity of expression in brain and the location of JA04 in the 5′ untranslated region (15), KIAA0319 is a reasonable candidate, but the reported paucity of polymorphisms in disequilibrium with reading phenotypes (35), confirmed by sequencing in the CLDRC cohort, made it less attractive (see DNA Sequence Analysis in Supporting Materials and Methods).

Furthermore, in our population, transmission disequilibrium was mostly from short haplotypes confined to DCDC2 (blocks A–E), with minimal support for association from single markers within MRS2L, GPLD1, KIAA0319, TTRAP, and THEM2 (Table 3). Block F, spanning GPLD1 just telomorphic of DCDC2, also has one haplotype in disequilibrium. HAPLOVIEW and GOLD show, however, that the strongest marker in F, SNP 72, shares weak intermarker disequilibrium with SNP 33 (D′ = 0.41 and 0.49, respectively) located in block C, suggesting transmission disequilibrium is due to polymorphisms in DCDC2. No other haplotypes spanning GPLD1 show significant disequilibrium (data not shown). The transmission disequilibrium from block G is unknown and it spans no recognizable coding sequences. Although it is located with 118 kb of a published peak in THEM2, we found no disequilibrium with any HAPLOVIEW block on either side of block G or spanning THEM2 (35). Haplotypes within block H, telomeric to G and also void of recognizable coding sequences, do not show significant disequilibrium with RD phenotypes. Overall, these results suggest that intermarker linkage disequilibrium blocks in this region are relatively short, therefore it is unlikely that transmission disequilibrium from DCDC2 in the CLDRC cohort is due to risk alleles of genes located elsewhere in the DYY2 locus.

The brain is a highly intricate organ that requires a complex orchestra of changes and growth to fully develop in humans. Regardless of the pathophysiological mechanisms, RD is a complex phenotype and several, if not many, genes are involved. Because they are often functionally grouped on chromosomes, it is possible that variations within more than one gene on 6p22, such as DCDC2 and KIAA0319, and/or epistatic effects are responsible for interindividual differences, which may become apparent in further studies of additional populations.

In summary, we saturated the region of the genome around JA04, which led to the identification of an intronic polymorphic deletion of DCDC2. Alleles of dBSTS ID no. 808258 within the region that the deletion spans are in significant disequilibrium with multiple RD traits. RT-PCR data suggest that DCDC2 localizes to the region of the brain where fluent reading occurs, and RNAi studies show that down-regulating DCDC2 leads to alteration in neuronal migration, again within the brain regions of interest. Thus, we propose that DCDC2 is a susceptibility gene for RD.

We thank Jennifer Koch at the Yale Center for the Study of Learning and Attention for subject recruitment and follow-up; Dr. Shrikant Mane (Yale University School of Medicine) for genotyping support and Prof. Allen Bale (Yale University School of Medicine) for critical review and guidance; and Prof. Anthony Monaco (Oxford University, Oxford) for suggesting collaborative functional studies. J.R.G. and H.M. are supported by The Charles H. Hood Foundation (Boston), The Robert Leet and Clara Guthrie Partnership Trust (Hartford, CT), and National Institutes of Health (NIH) Grant R01 NS43530. J.R.G. and J.G. are supported by NIH Grants R01 AA11330, R01 DA12690, and R01 DA12849. M.H. is supported by NIH Grant R01 NS43530. The CLDRC (S.D.S., R.K.O., B.F.P., and J.C.D.) is funded by NIH-National Institute of Child Health and Human Development (NICHD) Grant 5P50HD027802 (to J.C.D.), B.A.S., T.E.S., and K.M. are supported by NICHD Grants PO1 HD21888 and P01 HD25802. Y.W., M.P., and J.J.L. are supported by NIH Grants HD20806 and MH56524.