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

Haiying Meng[a], Shelley D. Smith[b], Karl Hager[c], Matthew Held[a], Jonathan Liu[d], Richard K. Olson[e], Bruce F. Pennington[f], John C. DeFries[g], Joel Gelernter[h,i], Thomas O'Reilly-Pol[a], Stefan Somlo[j], Pawel Skudlarski[a], Sally E. Shaywitz[a], Bennett A. Shaywitz[a], Karen Marchione[a], Yu Wang[k], Murugan Paramasivam[k], Joseph J. LoTurco[k], Grier P. Page[l], and Jeffrey R. Gruen[a,m]

[a]Department of Pediatrics, Yale Child Health Research Center, Yale University School of Medicine, New Haven, CT 06520; [b]Munroe Meyer Institute, University of Nebraska Medical Center, Omaha, NE 68198; [c]W. M. Keck Foundation Biotechnology Resource Laboratory, Yale University School of Medicine, New Haven, CT 06511; [d]SoftGenetics, State College, PA 16803; [e]Department of Psychology and [g]Institute for Behavioral Genetics, University of Colorado, Boulder, CO 80309; [f]Department of Psychology, University of Denver, Denver, CO 80210; [h]Department of Psychiatry, Veterans Affairs Medical Center, West Haven, CT 06516; Departments of [i]Psychiatry and [j]Internal Medicine, Yale University School of Medicine, New Haven, CT 06519; [k]Department of Physiology and Neurobiology, University of Connecticut, Storrs, CT 06269; and [l]Department of Biostatistics, 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.

dyslexia | DYX2 | disequilibrium | haplotype | doublecortin

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). Subsequently, loci were described on chromosomes 1, 2p15-16, 3p13, 6p (refs. 7–14), 6q, 7q32, 11, 15q21, and 18p11.2.

Of the reported susceptibility loci, the most widely reproduced is DYX2. Reported linkage intervals range widely: 13.4 cm (16.9 Mb) spanning D6S422 (pter) through D6S291 (14), 5 cm (4.8 Mb) spanning D6S464 through D6S258 (13), and 1.8 cm (7.9 Mb) spanning D6S299 through D6S273 (12) (physical distances were previously described in ref. 11). Subsequently, we identified a peak of association with a short tandem repeat (STR) marker, JA04 [NCBI identification (ID) no. G72384], located in the 5′ untranslated region of *KIAA0319*, an uncharacterized gene that is expressed in the brain (7, 8). However, there are at least 19 genes and two pseudogenes encoded within 1.5 Mb of JA04, most of which are expressed in brain (15).

Our previous association study of quantitative and discrete traits used 29 informative STR markers spanning the 10 Mb from D6S1950 through D6S478 (7, 8). We identified a peak of total association at JA04 (Fig. 1*a*) with orthographic choice ($P = 0.0007$), a reading performance task that requires the rapid recognition of a target word versus a phonologically identical background foil that is not a word (i.e., rain, rane; sammon, salmon; see ref. 16). The goal of the present study was to identify the DYX2 gene and the corresponding alleles that create susceptibility for developing RD. To confine an association interval to the smallest possible number of candidate genes, we assembled a high-density marker panel of 147 single nucleotide polymorphisms (SNPs) covering the 1.5 Mb surrounding JA04. We used this panel to assess single-marker and haplotype transmission disequilibrium with quantitative reading performance assessments in RD families. We also correlated quantitative expression studies of eight genes included in the panel with 18 regions of human brain corresponding to the primary functional reading centers.

Materials and Methods

Colorado Learning Disabilities Research Center (CLDRC) RD Family Samples. The 536 samples (parents and siblings) consisted of 153 nuclear families collected by the CLDRC (17). Subjects with an intelligence quotient (IQ) < 80 or for whom English was a second language were not included in the initial sample. Further descriptions of subjects and RD phenotypes are found in *Supporting Methods and Materials*, which is published as supporting information on the PNAS web site. Genomic DNA samples were amplified by multiple displacement 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.

Freely available online through the PNAS open access option.

Abbreviations: CLDRC, Colorado Learning Disabilities Research Center; DISC, discriminant score; ID, identification; IQ, intelligence quotient; LOH, loss-of-heterozygosity; RD, reading disability; RNAi, RNA interference; shRNA, small hairpin RNA; SNP, single nucleotide polymorphism; STR, short tandem repeat.

Data deposition: The sequences reported in this paper have been deposited in the dbSTS database, www.ncbi.nlm.nih.gov/projects/SNP (dbSTS ID no. 808238).

[m]To whom correspondence should be addressed. E-mail: jeffrey.gruen@yale.edu.

GENETICS

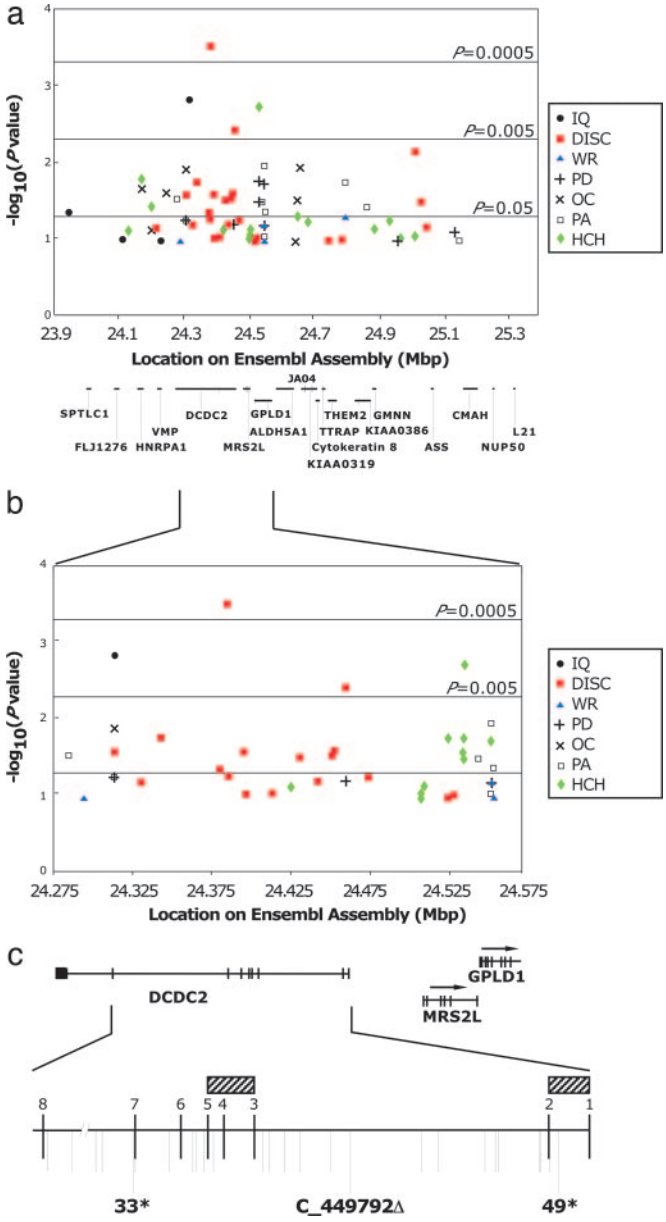

Fig. 1. High density SNP QTDT analysis. (*a*) Evidence for transmission disequilibrium for 147 SNPs as $-\log_{10} P$ value, and plotted against position in the ENSEMBL human genomic reference sequence. The locations of 18 genes encoded in this region are provided. The vertical lines on the genes are coding SNPs (cSNPs). The location of marker JA04 is shown above the gene map. The longest distance between SNPs was 332 kb, located at the centromeric end of the region. The shortest distance was 14 bp in exon 1 of *MRS2L*. There were 20 cSNPs within exons of nine genes and 12 nonsynonymous cSNPs in five genes (*DCDC2*, *MRS2L*, *GPLD1*, *KIAA0319*, and *TTRAP*). The average minor allele frequency was 0.28 in the RD probands, not including the five previously uncharacterized private SNPs in *MRS2L*. (*b*) $-\log_{10} P$ value for 33 SNPs ($P < 0.1$) located within *DCDC2*, *MRS2L*, and part of *GPLD1*. (*c*) Further expansion of a 110-kb region within *DCDC2*. SNPs labeled with an asterisk are associated with RD phenotypes with $P < 0.005$. C_449792 is located within the deleted 2,445 bp in intron 2 of *DCDC2* and designated by a triangle. The heavy vertical black lines represent exons in *DCDC2*. The hatched rectangles above exons 1 and 2, and above exons 3–5 highlight the coding regions for the *DCX* doublecortin peptide domains.

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, Uppsala). 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 (AGCCTGCCTACCACAGAGAA), reverse primer for nondeleted chromosomes (GGAACAACCTCACAGAAATGG), and reverse primer for deleted chromosomes (TGAAACCCCGTCTCTACTGAA). Reaction products were resolved on 1.5% agarose gels. The deletion fusion fragment was 225 bp, and the nondeleted fragment was 550bp.

dbSTS ID no. 808238 genotype. The compound STR, dbSTS ID no. 808238, was genotyped by sequencing PCR products generated with a forward primer (TGTTGAATCCCAGACCACAA) and reverse primer (ATCCCGATGAAATGAAAAGG). 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*, *PPIA*, 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/~ayoung/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. Haplotype-TDT 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′-cccaccaagcaattccagacaa(aca)ttgtctggaattgcttggtggg-3′ and the control sequence was 5′-cccagtcaaggcattgaattaaa(aca)tttaattcaatgccttgactggg-3′. The sequence was selected by its asymmetry and for the absence of any matches to rat genomic sequence in

Table 1. Single-marker QTDT analysis for markers with $P \leq 0.01$

SNP ID	Gene	DISC	PTP	HCH	ENSEMBL location	Celera location
33	*DCDC2* Int 6	0.0003			24386848	25512242
49	*DCDC2* Int 1	0.0035			24463129	25588523
72	*GPLD1* Int 24			0.0018	24539037	25664490
117	Intergene		0.0077		24872844	25998238
130	Intergene	0.0067	0.055	0.0811	25022795	26142106

PTP, phenome transposition; HCH, homonyn choice.

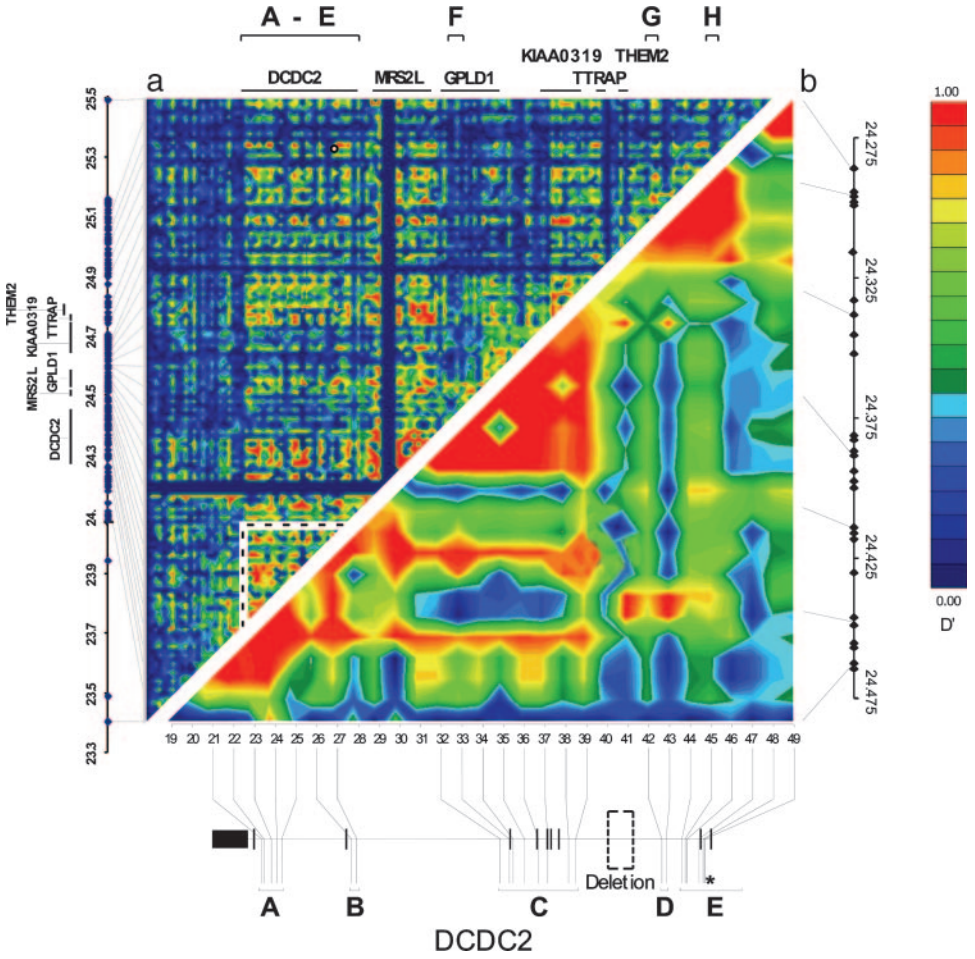

Fig. 2. Linkage disequilibrium between pairs of SNPs. Color-coded *D′* values for pairs of SNPs are plotted with the GOLD program. (*a*) Linkage disequilibrium between pairs of SNPs in the 1.5-Mb region. The location of the 147 SNPs in this region are provided in Table 3. Gene and haplotype block depictions on the top are relative to marker number and not actual physical distances. Gene and marker locations on the left are proportional to physical distances. (*b*) Triangular excerpt in lower left corner of *a* with higher resolution of SNPs 19–49 covering 180 kb and haplotype blocks A–E in *DCDC2*. *, SNPs with $P < 0.005$. Block A spanned five SNPs (SNP ID nos. 21–25) and 6.5 kb in intron 8. Block B spanned two SNPs (SNP ID nos. 26 and 27) and 23 kb in intron 7, including the single marker peak at SNP 26 with IQ. Block C spanned eight SNPs (SNP ID nos. 32–39) and 34.2 kb from introns 2 to 7, including the highest single marker peak at SNPs 33 with DISC. Block D spanned five SNPs (SNP ID nos. 42–46) and 11.5kb in intron 2. Block E spanned three SNPs (SNP ID nos. 47, 49, and 50) and 16 kb in from introns 1 to 2 and the 5′ untranslated region, including the single marker peak at SNP 49 with DISC. Block F spanned five SNPs (SNP ID nos. 68–72) and 5.4kb from *MRS2L* to *GPLD1*, including the nonsynonymous cSNP in *MRS2L*, SNP 69. Block G spanned three SNPs (SNP ID nos. 117–119) and 34.4 kb, including the single marker peak at SNP 117 with phoneme transposition. Block H spanned three SNPs (SNP ID nos. 128–130) and 13.5 kb, including the single marker peak at SNP 130 with DISC.

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 SNP 33 located in intron 6 of *DCDC2* ($P = 0.0003$). 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 ≤0.01; two of these SNPs were located in *DCDC2*. Thirty-seven SNPs yielded a *P* value of ≤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. 2*a*) 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. 2*b*). 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, and E 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

GENETICS

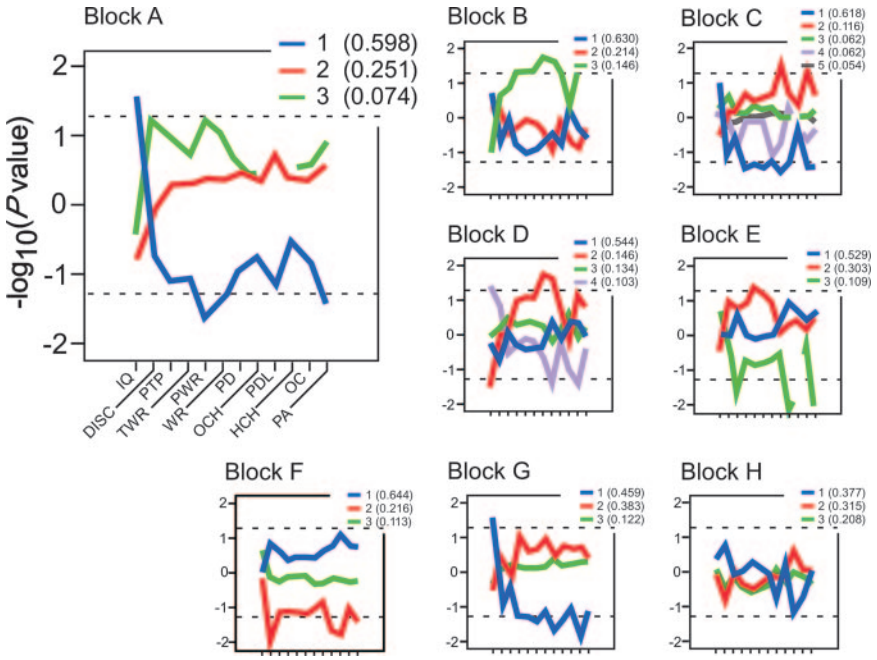

Fig. 3. Haplotype-TDT analyses. FBAT results for 12 cognitive phenotypes at haplotype blocks A–H. The locations of the haplotype blocks are presented in Fig. 2. The markers comprising each haplotype block are described in Fig. 2 legend and Tables 4 and 5. Evidence for transmission disequilibrium is plotted as $-\log_{10}P$ along the *y* axis, for each phenotype represented by tick marks along the *x* axis from left to right as: IQ, DISC, phoneme transposition (PTP), timed Peabody Individual Achievement Test word recognition (TWR), untimed Peabody Individual Achievement Test word recognition (PWR), word recognition (WR), phonological decoding (PD), orthographic choice (OCH), phoneme deletion (PDL), homonym choice (HCH), orthographic coding (OC), and phoneme awareness (PA). Positive or negative values for $-\log_{10}P$ value reflect the direction of the *z* score derived by FBAT, so that *z* scores below the population mean are plotted as $-\log_{10}P$ value <0 and visa versa. Dashed lines represent $P < 0.05$. Haplotypes within each block are numbered 1–5 and are represented by different colors. The alleles that define each haplotype are presented in Table 4. Frequencies of each haplotype in the CLDRC cohort are presented in the legend. Blocks A–E span *DCDC2*.

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 $(GAGAGGAAGGAAA)_n$ and $(GGAA)_n$ repeat units (Table 6, which is published as supporting information on the PNAS web site). In the CLDRC cohort, some alleles were present only in the parents (five) and others, including the deletion, occurred too infrequently in probands to compute transmission disequilibrium. 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* (AGGAAAG) sites in repeat unit 1 of dbSTS ID no. 808238. Both transcription factors are expressed in mouse brain. *PEA3* is associated with sexual function and peripheral motor neuron arborization (26). *NF-ATp* mediates rapid embryonic axon extension necessary for forming neuronal connections (27), which would complement the putative function of the doublecortin peptide domains in *DCDC2*.

Quantitative Real-Time RT-PCR. Fig. 4 shows the expression levels of eight genes in 17 regions of human brain normalized to thalamus by quantitative real-time RT-PCR. The most variably expressed genes were *KIAA0319*, *MRS2L*, and *DCDC2*. *KIAA0319* was most highly expressed in the superior parietal cortex, primary visual cortex, and occipital cortex. *MRS2L* was most highly expressed in the superior temporal cortex, hypothalamus, and amygdala. *DCDC2* was most highly expressed in the entorhinal cortex, inferior temporal cortex, medial temporal cortex, hypothalamus, amygdala, and hippocampus. Expression of *TTRAP*, *THEM2*, *GMNN*, and *ALDH5A* in the 17 regions of the brain did not differ significantly from thalamus.

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 scrabbled 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 367 μm + 135 ($n = 4$, $P < 0.01$).

Discussion

After our previous studies that showed transmission disequilibrium to JA04, we systematically interrogated the 6p22 DYX2 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. 1*c*).

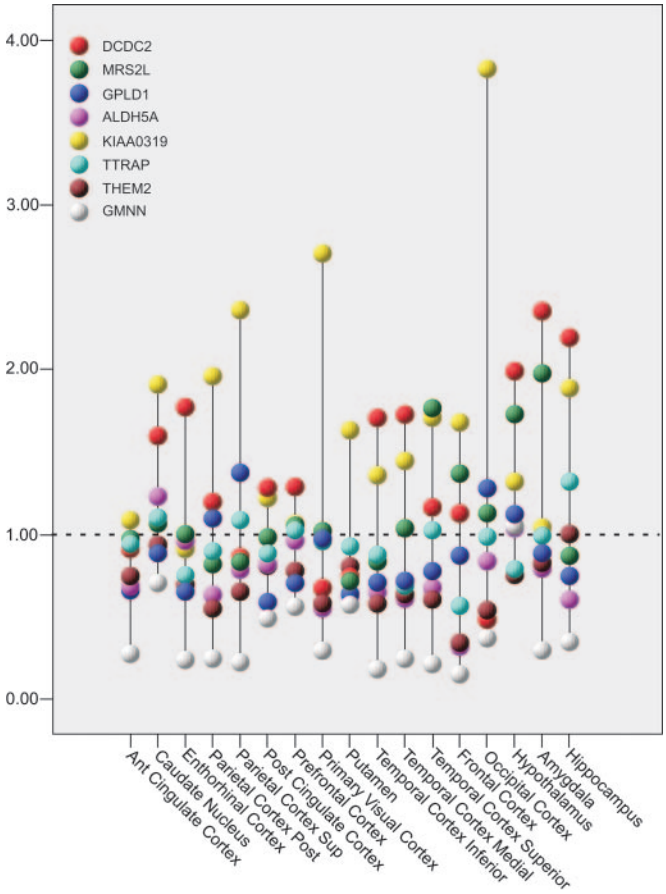

Fig. 4. RT-PCR results for *DCDC2*, *MRS2L*, *GPLD1*, *ALDH5A*, *KIAA0319*, *TTRAP*, *THEM2*, and *GMNN* in 17 areas of anonymous donor human brain regions normalized to thalamus (=1.00).

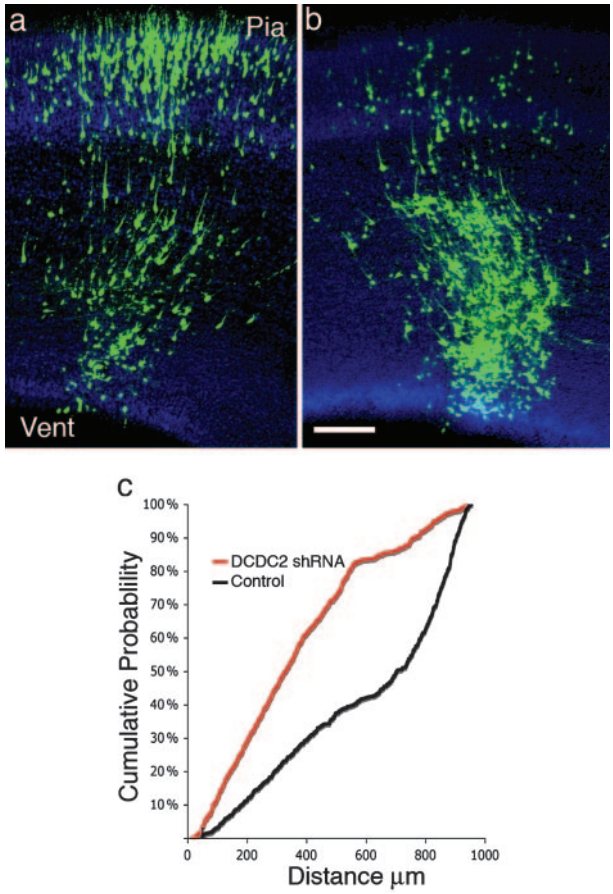

Fig. 5. *In utero* RNAi against *DCDC2*. (*a*) Control transfection of a neutral shRNA vector and *EGFP* shows normal migration after 4 days. Most neurons have migrated well away from the ventricular surface (Vent) toward the pial surface (Pia). (*b*) Neurons transfected with an shRNA vector directed 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.)

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 DYX2 locus 500 kb from JA04. 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. 1*c*). *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 occipitotemporal 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 occipitotemporal 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.

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 neurons that populate reading centers, when considered in the context of *DCDC2* expression in inferior and medial temporal cortex, it offers a plausible pathophysiologic mechanism for RD due to genetic expression heterogeneity. *DCDC2* heterogeneity is also consistent with other pathophysiologic 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 *VMP*, *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 telomeric 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 origin of the transmission disequilibrium from block G is unknown and it spans no recognizable coding sequences. Although it is located within 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 then, conservative estimates of 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 DYX2 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 pathophysiologic 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 inter-individual 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. 808238 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 Patterson 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., S.E.S., and K.M. are supported by NICHD Grants PO1 HD 21888 and P50 HD25802. Y.W., M.P., and J.J.L. are supported by NIH Grants HD20806 and MH56524.

1. Shaywitz, S. E. & Shaywitz, B. A. (2003) *Pediatr. Rev.* **24,** 147–153.
2. Shaywitz, S. E., Shaywitz, B. A., Pugh, K. R., Fulbright, R. K., Constable, R. T., Mencl, W. E., Shankweiler, D. P., Liberman, A. M., Skudlarski, P., Fletcher, J. M., *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95,** 2636–2641.
3. Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Mencl, W. E., Fulbright, R. K., Skudlarski, P., Constable, R. T., Marchione, K. E., Fletcher, J. M., Lyon, G. R., *et al.* (2002) *Biol. Psychiatry* **52,** 101–110.
4. Finucci, J. M., Guthrie, J. T., Childs, A. L., Abbey, H. & Childs, B. (1976) *Ann. Hum. Genet.* **40,** 1–23.
5. DeFries, J. C., Fulker, D. W. & LaBuda, M. C. (1987) *Nature* **329,** 537–539.
6. Smith, S. D., Kimberling, W. J., Pennington, B. F. & Lubs, H. A. (1983) *Science* **219,** 1345–1347.
7. Turic, D., Robinson, L., Duke, M., Morris, D. W., Webb, V., Hamshere, M., Milham, C., Hopkin, E., Pound, K., Fernando, S., *et al.* (2003) *Mol. Psychiatry* **8,** 176–185.
8. Kaplan, D. E., Gayan, J., Ahn, J., Won, T. W., Pauls, D., Olson, R. K., DeFries, J. C., Wood, F., Pennington, B. F., Page, G. P., *et al.* (2002) *Am. J. Hum. Genet.* **70,** 1287–1298.
9. Fisher, S. E., Francks, C., Marlow, A. J., MacPhie, I. L., Newbury, D. F., Cardon, L. R., Ishikawa-Brush, Y., Richardson, A. J., Talcott, J. B., Gayan, J., *et al.* (2002) *Nat. Genet.* **30,** 86–91.
10. Barr, C. L., Shulman, R., Wigg, K., Schachar, R., Tannock, R., Roberts, W., Malone, M. & Kennedy, J. L. (2001) *Am. J. Med. Genet.* **105,** 250–254.
11. Ahn, J., Won, T. W., Zia, A., Reutter, H., Kaplan, D. E., Sparks, R. & Gruen, J. R. (2001) *Genomics* **78,** 19–29.
12. Grigorenko, E. L., Wood, F. B., Meyer, M. S. & Pauls, D. L. (2000) *Am. J. Hum. Genet.* **66,** 715–723.
13. Gayán, J., Smith, S. D., Cherny, S. S., Cardon, L. R., Fulker, D. W., Brower, A. M., Olson, R. K., Pennington, B. F. & DeFries, J. C. (1999) *Am. J. Hum. Genet.* **64,** 157–164.
14. Fisher, S. E., Marlow, A. J., Lamb, J., Maestrini, E., Williams, D. F., Richardson, A. J., Weeks, D. E., Stein, J. F. & Monaco, A. P. (1999) *Am. J. Hum. Genet.* **64,** 146–156.
15. Londin, E. R., Meng, H. & Gruen, J. R. (2003) *BMC Genomics* **4,** 25.
16. Olson, R., Wise, B., Conners, F., Rack, J. & Fulker, D. (1989) *J. Learn. Disabil.* **22,** 339–348.
17. DeFries, J. C., Filipek, P. A., Fulker, D. W., Olson, R. K., Pennington, B. F., Smith, S. D. & Wise, B. W. (1997) *Learn. Disabil.* **8,** 7–19.
18. Dean, F. B., Nelson, J. R., Giesler, T. L. & Lasken, R. S. (2001) *Genome Res.* **11,** 1095–1099.
19. Sobel, E. & Lange, K. (1996) *Am J. Hum. Genet.* **58,** 1323–1337.
20. Abecasis, G. R., Cookson, W. O. & Cardon, L. R. (2000) *Eur. J. Hum. Genet.* **8,** 545–551.
21. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. (2005) *Bioinformatics* **21,** 263–265.
22. Abecasis, G. R. & Cookson, W. O. (2000) *Bioinformatics* **16,** 182–183.
23. Horvath, S., Xu, X., Lake, S. L., Silverman, E. K., Weiss, S. T. & Laird, N. M. (2004) *Genet. Epidemiol.* **26,** 61–69.
24. Bai, J., Ramos, R. L., Ackman, J. B., Thomas, A. M., Lee, R. V. & LoTurco, J. J. (2003) *Nat. Neurosci.* **6,** 1277–1283.
25. Schug, J. (2003) in *Current Protocols in Bioinformatics*, eds. Baxevanis, A., Davison, D., Page, R., Petsko, G., Stein, L. & Stormo, G. (Wiley & Sons, New York), Unit 2.6.
26. Laing, M. A., Coonrod, S., Hinton, B. T., Downie, J. W., Tozer, R., Rudnicki, M. A. & Hassell, J. A. (2000) *Mol. Cell. Biol.* **20,** 9337–9345.
27. Graef, I. A., Wang, F., Charron, F., Chen, L., Neilson, J., Tessier-Lavigne, M. & Crabtree, G. R. (2003) *Cell* **113,** 657–670.
28. Dobyns, W. B., Truwit, C. L., Ross, M. E., Matsumoto, N., Pilz, D. T., Ledbetter, D. H., Gleeson, J. G., Walsh, C. A. & Barkovich, A. J. (1999) *Neurology* **53,** 270–277.
29. Dobyns, W. B. & Truwit, C. L. (1995) *Neuropediatrics* **26,** 132–147.
30. Horwitz, B., Rumsey, J. M. & Donohue, B. C. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 8939–8944.
31. Shaywitz, S. E. & Shaywitz, B. A. (2005) *Biol. Psychiatry* **57,** 1301–1309.
32. Galaburda, A. M., Sherman, G. F., Rosen, G. D., Aboitiz, F. & Geschwind, N. (1985) *Ann. Neurol.* **18,** 222–233.
33. Dehaene, S., Cohen, L., Sigman, M. & Vinckier, F. (2005) *Trends Cogn. Sci.* **9,** 335–341.
34. Deffenbacher, K. E., Kenyon, J. B., Hoover, D. M., Olson, R. K., Pennington, B. F., DeFries, J. C. & Smith, S. D. (2004) *Hum. Genet.* **115,** 128–138.
35. Francks, C., Paracchini, S., Smith, S. D., Richardson, A. J., Scerri, T. S., Cardon, L. R., Marlow, A. J., Macphie, I. L., Walter, J., Pennington, B. F., *et al.* (2004) *Am. J. Hum. Genet.* **75,** 1046–1058.
36. Cope, N., Harold, D., Hill, G., Moskvina, V., Stevenson, J., Holmans, P., Owen, M. J., O'Donovan, M. C. & Williams, J. (2005) *Am. J. Hum. Genet.* **76,** 581–591.