Published in final edited form as:

Behav Genet. 2011 January ; 41(1): 58–66. doi:10.1007/s10519-010-9408-3.

A Dyslexia-Associated Variant in *DCDC2* Changes Gene Expression

Haiying Meng,

Department of Pediatrics, Yale Child Health Research Center, Yale University School of Medicine, 464 Congress Avenue, New Haven, CT 06520-8081, USA

Natalie R. Powers,

Department of Pediatrics, Yale Child Health Research Center, Yale University School of Medicine, 464 Congress Avenue, New Haven, CT 06520-8081, USA

Department of Genetics, Yale University School of Medicine, 464 Congress Avenue, New Haven, CT 06520-8081, USA

Ling Tang,

Department of Pediatrics, Yale Child Health Research Center, Yale University School of Medicine, 464 Congress Avenue, New Haven, CT 06520-8081, USA

Natalie A. Cope,

Department of Pediatrics, Yale Child Health Research Center, Yale University School of Medicine, 464 Congress Avenue, New Haven, CT 06520-8081, USA

Ping-Xia Zhang,

Department of Pediatrics, Yale Child Health Research Center, Yale University School of Medicine, 464 Congress Avenue, New Haven, CT 06520-8081, USA

Ramsay Fuleihan,

Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

Christopher Gibson,

Department of Pediatrics, Yale Child Health Research Center, Yale University School of Medicine, 464 Congress Avenue, New Haven, CT 06520-8081, USA

Grier P. Page, and

Genomics and Statistical Genetics Research Unit, Research Triangle Institute International, Atlanta, GA, USA

Jeffrey R. Gruen

Department of Pediatrics, Yale Child Health Research Center, Yale University School of Medicine, 464 Congress Avenue, New Haven, CT 06520-8081, USA

© Springer Science+Business Media, LLC 2010

Correspondence to: Jeffrey R. Gruen, jeffrey.gruen@yale.edu.

Edited by Elena Grigorenko and Brett Miller.

H. Meng and N. R. Powers contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10519-010-9408-3) contains supplementary material, which is available to authorized users.

Financial Disclosures This manuscript describes the characterization of an enhancer element in *DCDC2*. Yale University has applied for a patent covering this element; authors Jeffrey Gruen and Haiying Meng are inventors on this patent. Furthermore, the patent rights have been licensed to a start-up company founded by Dr. Gruen.

Department of Genetics, Yale University School of Medicine, 464 Congress Avenue, New Haven, CT 06520-8081, USA

Investigative Medicine Program, Yale University School of Medicine, 464 Congress Avenue, New Haven, CT 06520-8081, USA

Jeffrey R. Gruen: jeffrey.gruen@yale.edu

Abstract

Reading disability (RD) or dyslexia is a common neurogenetic disorder. Two genes, *KIAA0319* and *DCDC2*, have been identified by association studies of the DYX2 locus on 6p21.3. We previously identified a 2445 bp deletion, and a compound STR within the deleted region (BV677278), in intron 2 of *DCDC2*. The deletion and several alleles of the STR are strongly associated with RD ($P = 0.00002$). In this study we investigated whether BV677278 is a regulatory region for *DCDC2* by electrophoretic mobility shift and luciferase reporter assays. We show that oligonucleotide probes from the STR bind nuclear protein from human brain, and that alleles of the STR have a range of *DCDC2*-specific enhancer activities. Five alleles displayed strong enhancer activity and increased gene expression, while allele 1 showed no enhancer activity. These studies suggest that the association of BV677278 with RD reflects a role as a modifier of *DCDC2* expression.

Keywords

Dyslexia; Reading disability; *DCDC2*; Regulatory region; Association

Introduction

Reading disability (RD), also known as dyslexia, is characterized by unexpected difficulty learning to read despite intelligence, motivation, and sociocultural opportunity adequate for literacy. Estimates of RD prevalence vary from 5 to 17% depending on the population studied, and on the diagnostic criteria employed (Katusic et al. 2001; Shaywitz et al. 1990). RD has been known for decades to be highly familial, and heritability estimates suggest that 44–75% of RD risk is explained by genetic factors (DeFries et al. 1987). Starting in 1983, several genetic linkage and association studies were undertaken to localize susceptibility genes for RD. Nine putative RD risk loci (DYX1–DYX9) have been reported, on chromosomes 1p (DYX8), 2p (DYX3), 3 (DYX5), 6p (DYX2), 6q (DYX4), 11p (DYX7), 15 (DYX1), 18p (DYX6), and Xq (DYX9). DYX2, the RD locus on 6p22, was identified by linkage studies of three independent cohorts from the US and the UK (Cardon et al. 1994; Fisher et al. 1999; Gayán et al. 1999; Grigorenko et al. 2000). Subsequent association studies identified two risk genes within the DYX2 locus: *KIAA0319* and *DCDC2* (Cope et al. 2005; Deffenbacher et al. 2004; Francks et al. 2004; Meng et al. 2005; Paracchini et al. 2006; Schumacher et al. 2006). *DCDC2* stands for ‘doublecortin domain containing 2,’ a doublecortin-family gene, while *KIAA0319* encodes a transmembrane protein of unknown function. The *doublecortin* or ‘*DCX*’ gene encodes a microtubule-binding protein, and resides on the X chromosome in humans (Taylor et al. 2000). Mutations in *doublecortin* cause one of two neuronal migration disorders—subcortical laminar heterotopia or X-linked lissencephaly—depending on dosage of the mutant allele (Online Mendelian Inheritance in Man, <http://www.ncbi.nlm.nih.gov/omim>). Genes in the doublecortin family all share the microtubule-binding doublecortin domain. It is interesting to note that several of the other eight RD loci contain a gene from this family or have one nearby: *DCDC2B* in DYX8, *DCDC1* and *DCDC5* in DYX7, and *doublecortin* near DYX9. Association of RD with *KIAA0319* was found in cohorts from Wales, England, and the US (Cope et al. 2005;

Francks et al. 2004), while association with *DCDC2* was found in cohorts from the US, Germany, and Australia (Meng et al. 2005; Schumacher et al. 2006; Wilcke et al. 2009).

We previously reported strong transmission disequilibrium (in 153 RD families from Colorado) for a highly polymorphic purine-rich compound short tandem repeat (STR, GenBank ID BV677278), which is located in intron 2 of *DCDC2*. Ten BV677278 alleles were present in the Colorado cohort, which varied in length from 289 to 314 bp depending on the number of several different repeat units (Table 1) (Meng et al. 2005). In addition, we reported a 2445 bp deletion, present in 8.5% of parents of probands, encompassing BV677278. Chromosomes with this deletion are therefore missing the STR. Most of the BV677278 alleles were of insufficient frequency for single marker analysis. Since we could not combine alleles based on any biological information available at the time, we grouped BV677278 alleles by frequency, so as to include all eleven in the analysis. By combining minor alleles (the deletion and alleles 2, 5, 6, 7, 8, 9, and 10), we showed association of the STR with homonym choice ($P = 0.00002$), which had been previously shown to be heritable and to correlate with component measures of reading and language (Gayan and Olson 2001). Furthermore, 11 SNPs located within *DCDC2*, but outside the purine-rich region, also showed association with RD in the same study ($P < 0.05$). Two SNPs, rs1087266 (ABI TaqMan/Celera marker C_7454790) in intron 1 ($P = 0.0035$) and rs807724 (C_7454704) in intron 6 ($P = 0.0003$), showed association with a weighted composite (DISC phenotype) of the reading recognition, reading comprehension, and spelling subtests of the Peabody Individual Achievement Test. Four haplotypes, together spanning most of *DCDC2*, were associated with poor performance on several reading measures among those individuals with $IQ > 80$.

Independent replication of the *DCDC2*-RD association came from Schumacher et al. (2006), who showed RD association with a two-marker *DCDC2* haplotype, rs793862–rs807701, in a German-speaking sample. Interestingly, the association was strengthened when subjects with more severe RD were selected and analyzed as a subgroup (1SD discrepancy between spelling score and IQ, GRR = 4.11; 2SD, GRR = 4.81; 2.5SD, GRR = 11.31) (Schumacher et al. 2006). Recently, Lind et al. (2010) reported a study supporting association of two SNPs in *DCDC2* with regular word reading and spelling (rs1419228, $P = 0.002$) and irregular word reading (rs1091047, $P = 0.003$) in 522 families from Australia that were unselected for reading impairment (Lind et al. 2010). Weak evidence for association between the intron 2 deletion and quantitative RD component phenotypes has also recently been published by Brkanac et al. (2007). Wilcke et al. (2009) recently described an RD association with three SNPs and a haplotype from *DCDC2* ($P < 0.05$), as well as the BV677278 deletion ($P < 0.01$), in a study from Germany (Wilcke et al. 2009). This group also performed a clinical subgroup analysis, and showed that the strongest association with the BV677278 deletion allele occurred in nondysphonetic (i.e. ‘surface RD:’ thought to be caused by an impairment in visual perception of words) ($P < 0.005$) and nonsevere ($P < 0.01$) forms of RD. However, Ludwig et al. (2008) also attempted to replicate the association of BV677278 with RD in a German-speaking sample, but it failed to replicate in that study.

Having shown association of BV677278 with RD, we became interested in the mechanism of this association. According to the TRANSFAC database, the purine-rich STR encodes over 100 transcription factor (TF) recognition sites (www.transfac.com). Additionally, this sequence appears to be conserved between human, mouse and rat homologs of *DCDC2* (see Results). This led us to hypothesize that BV677278 might be a regulatory element, and we set out to determine whether this sequence specifically binds protein in human brain nuclear lysate, and whether the sequence itself could specifically change *DCDC2* expression. We used electrophoretic mobility shift assays (EMSAs) to interrogate short sequences extracted

from the BV677278 STR for sequence-specific binding to protein in human brain nuclear lysate. Next, we interrogated the BV677278 alleles that were common in our RD sample for their ability to enhance transcription, using luciferase reporter constructs. These in vitro experiments show that, in our system, various BV677278 alleles associated with RD change the expression level of *DCDC2*, and that BV677278 may function as an enhancer of *DCDC2*.

Methods and materials

Cell culture

P19 cells were purchased from ATCC (Manassas, VA), and grown in alpha minimum essential medium with 7.5% bovine calf serum, 2.5% fetal bovine serum (FBS), and penicillin/streptomycin. Raji cells were purchased from ATCC (Manassas, VA), and grown in RPMI 1640 medium with 10% fetal bovine serum (FBS).

Electrophoretic mobility shift assays (EMSAs)

Nuclear extract of whole human brain was purchased from Active Motif (Carlsbad, CA) and from Novus Biologicals (Littleton, CO). For the first EMSA (Fig. 1b), four pairs of complementary 20-mer oligonucleotides were synthesized as probes spanning the purine-rich region of BV677278 (Table 2). The location of each probe within BV677278 is presented in Fig. 1a and Table 3. The oligonucleotides were digoxigenin (DIG) end-labeled by terminal transferase and digoxigenin-11-ddUTP. Double-stranded probes were made by annealing complementary oligonucleotides—by heating to 95°C for 15 min and cooling gradually to room temperature. Binding reactions were performed by preincubating each labeled probe with human brain nuclear lysate, polydI/dC, and poly L-lysine to block non-specific binding. For competition assays, a 100-fold molar excess of unlabeled double-stranded probe was incubated with the reaction mixture prior to the addition of labeled probe. Reaction products were then fractionated on a 6% polyacrylamide native gel in 0.5× tris–borate–EDTA buffer (TBE), pH 8.3 at 70 V for 2 h.

For the subsequent EMSAs (Fig. 2b, c), the Lightshift Chemiluminescent EMSA kit from Thermo Scientific (Pierce) was used. This kit uses biotin-labeled probes instead of DIG-labeled, so to construct the labeled EMSA-3 probe, a pair of 20-mer complimentary oligos were synthesized—one of them with a 5'-biotin tag. These two oligos were annealed to make the labeled EMSA3 dsP-robe. The unlabeled probes (EMSA3, EMSA4, EMSA3-Scram1, EMSA3-Scram2, and EBNA1) were made by synthesizing and annealing complementary oligos without a biotin label. Binding reactions were performed by preincubating each labeled probe with human brain nuclear lysate, Raji nuclear lysate, or P19 nuclear lysate. Salmon sperm DNA (1 µg/reaction) was used to block non-specific binding. For competition assays, a 200-fold molar excess of unlabeled double-stranded probe was incubated with the reaction mixture for 10 min prior to the addition of labeled probe. After addition of the labeled probe, binding reactions were allowed to proceed for an hour. Reaction products were then fractionated on a 6% polyacrylamide native gel in 0.5× tris–borate–EDTA buffer (TBE), pH 8.3 at 100 V for 1.5 h. All other steps, including transfer to a nylon membrane and probing with streptavidin-conjugated horseradish peroxidase, were done as per the manufacturer's instructions. Raji and P19 nuclear lysates were prepared according to the protocol of Wu (2006).

Bv677278 genotyping

The compound STR, BV677278, was genotyped by sequencing PCR products amplified from genomic DNA with forward primer (TGTTGAATCCCAGACCACAA) and reverse primer (ATCCCGATGAAATGAAAAGG). Conditions for amplification have been

previously described (Meng et al. 2005). Chromatograms were analyzed and alleles assigned with Mutation Surveyor version 3.1 (SoftGenetics, State College), by comparing samples to reference traces after alignment.

Luciferase reporter assay

To assess enhancer activity of BV677278 alleles, we used the pGL3 Basic vector (Promega, Carlsbad, CA). The 600 bp 5-prime of the *DCDC2* ATG start codon (*DCDC2* promoter region) was cloned into the Sac I site in the promoter position, upstream of the firefly luciferase reporter gene in the pGL3 Basic vector (Fig. 3a). Individual BV677278 alleles (1, 3, 4, 5, 6, or 10), amplified from genomic DNA with the primers described above, were inserted into the *Bam*HI site in the enhancer position, downstream of the luciferase gene. Unmodified pGL3-Basic, lacking promoter and enhancer elements, served as a negative control. The pGL3 promoter and enhancer vectors were also used as positive and negative controls.

For transfection, P19 cells were seeded at 1×10^5 cells/well in 96-well plates and grown to 80% confluency at 37°C in a 5% CO₂ incubator. P19 is a multipotent murine embryonal carcinoma cell line. Constructs were added into each well together with Transfast™ liposome transfection reagent (Promega), as per supplier's instructions. Renilla luciferase reporter vector pRL-SV40 (25 ng) was co-transfected in all experiments as an internal standard for transfection efficiency. One hour after transfection the cells were overlaid with 1 ml of prewarmed complete medium (distributed evenly among the wells and added gently from the side of each well to avoid disturbing the monolayer), and incubated overnight at 37°C in 5% CO₂. Cells were harvested after 24 h or 48 h and lysed. A dual luciferase reporter assay system (Promega) and luminometer (Turner Designs, Sunnyvale, CA) were used to assess luciferase activity. For each construct, experiments were performed in triplicate and the luciferase activity measured twice for each of the three replicates.

Statistical analysis

For the luciferase reporter gene assay, Student's *t* tests were used to compare relative luciferase expression levels between pairs of alleles, in all possible combinations (SPSS13) (Supplemental Table 1). Results were corrected for multiple tests using false discovery rate (FDR), $\alpha = 0.01$. (Benjamini and Hochberg 1995) All *P* values reported are two-tailed.

Sequence analysis

DCDC2 genomic sequences were compared in Megalign (DNASTAR, Inc., Madison, WI), using the Wilbur-Lipman, Clustal V, and Clustal W algorithms (standard method parameters). Human (NCBI 36 human assembly, November 2005), mouse (NCBI m37 mouse assembly, April 2007), and rat (Ensembl build RGSC3.4, November 2004 update) sequences were downloaded from the Ensembl Genome Browser (Birney et al. 2004).

Results

BV677278 is conserved

Alignment of the purine-rich sequence in human BV677278 (Ensembl chr6:24,326,300 > 24,326,498) by the Wilbur-Lipman method shows 71% identity to the orthologous purine-rich sequences in intron 2 of mouse *Dcdc2a* (Ensembl chr13:25,181,319 > 25,181,511), and 60% to intron 1 of rat *Dcdc2* (Ensembl chr17:47,089,537 > 47,089,699) (Supplemental Fig. 1). However, in terms of repeat unit structure, the STRs in mouse and rat are considerably more similar to each other than either is to the human STR. Based on standard Clustal W multiple alignment parameters, rat and mouse sequences show 75.3% identity and a divergence value of 32.2, while human and mouse sequences show 60.4% identity and a

divergence of 65.0, and human and rat sequences show 55.8% identity and a divergence of 91 (Supplemental Fig. 2). Human *DCDC2* cDNA shows 79.6% identity to the orthologous mouse *Dcdc2a* cDNA and 84.5% to rat *Dcdc2* cDNA (Supplemental Fig. 3).

BV677278 specifically binds a nuclear protein expressed in human brain

To test whether nuclear proteins expressed in human brain bind the purine-rich sequence in BV677278, EMSA was performed using nuclear extract from human brain. Two probes from BV677278 were bound by nuclear proteins from human brain tissue (Fig. 1b, lanes 5 and 8). An identical but unlabeled “cold” probe was used in each experiment to show specificity; Fig. 1b, lanes 6 and 9 show that unlabeled probe competes with labeled probe for binding with the protein. This result suggests that some nuclear protein expressed in human brain binds with specificity to the purine-rich region of BV677278.

The protein–BV677278 interaction is present and specific in human brain, Raji, and P19 nuclear lysates

To further test the specificity of the EMSA3 shift we observed in Fig. 1b (lanes 4–6), we performed competition assays using probes with the EMSA3 sequence scrambled two different ways (EMSA3-Scram1 and EMSA3-Scram2, Table 2), as well as with a 30-mer probe of unrelated sequence, which contains a binding site for the Epstein-Barr virus protein EBNA1 (Table 2). In human brain nuclear lysate (Fig. 2b), these competition assays show that the observed EMSA3–protein interaction is unequivocally specific. The major shift (arrow) is competed away completely by a 200-fold molar excess of unlabeled EMSA3 probe (lane 3). However, if the unlabeled probe used is one of the scrambled EMSA3 probes, the EBNA1 probe, or the EMSA4 probe, the opposite effect is seen—these probes appear to actually increase the strength of binding to the labeled EMSA3 probe (lanes 4–7). Interestingly, this effect is very dramatic when unlabeled EMSA4 probe is used (lane 7).

We were interested to see whether this shift was present in undifferentiated P19 nuclear lysate, as this would confirm the presence of the shift-causing DNA-binding protein in these cells, and validate the use of this cell line in our luciferase assays. We were also interested to see whether this shift was present in a human tissue other than brain, so we chose Raji, a genomically stable, karyotypically normal line of human Burkitt lymphoma cells. Figure 2c shows that the observed EMSA3 shift (arrow) is present and unequivocally specific in both cell lines, and behaves similarly to its counterpart in human brain. These experiments were also performed with labeled EMSA4 probe (and corresponding unlabeled probes including two with the EMSA4 sequence scrambled). The results suggest that the originally observed EMSA4 shift (Fig. 1b, lanes 7–9) is a nonspecific interaction (data not shown).

BV677278 has enhancer activity

To test whether BV677278 could be a regulatory element, we cloned the two most common alleles, 1 and 4 (Table 1), in both orientations into the promoter site of the luciferase reporter vector, pGL3-enhancer (with SV40 enhancer), and also into the enhancer site of the reporter vector pGL3-promoter (with SV40 promoter). In P19 cells, BV677278 showed no promoter activity. However, in the presence of the SV40 promoter, allele 4, cloned into the enhancer site in either orientation, increased luciferase activity threefold. This result suggests that BV677278 has the capacity for enhancer activity.

Modulation of luciferase reporter expression by BV677278 is allele-dependent

To examine each of the BV677278 polymorphisms associated with RD for enhancer activity specifically related to *DCDC2*, we cloned the 600 bp 5-prime of the *DCDC2* ATG start codon (with the assumption that this sequence contains at least some *DCDC2* promoter

elements) into the promoter site of pGL3-Basic, upstream of the firefly luciferase reporter gene. In P19 cells these 600 bp upstream of *DCDC2* alone (Basic/600) increased luciferase gene expression four to fivefold over background (Fig. 3b, column 4). We therefore used this region as the *DCDC2* promoter in subsequent experiments. We then tested the six most common BV677278 alleles 1, 3, 4, 5, 6, and 10, by cloning them one at a time into the enhancer position in both orientations, and transfecting the resulting constructs into P19 cells. Allele-specific expression varied one to threefold, depending on the allele (Fig. 3b, columns 5–16). Alleles 3, 4, 5, 6, and 10 showed strong enhancer activity. Allele 1 and the construct without a BV677278 allele cloned into the enhancer site showed no enhancer activity (Fig. 3b, column 4). *P*-values for differences in enhancer function among pairs of alleles in all possible combinations (Student's *t* test) are presented in Supplemental Table 1.

Discussion

In this study we attempt to link functional regulation of *DCDC2* expression to BV677278, a highly polymorphic STR sequence in intron 2, which is strongly associated with RD. The purine-rich sequence in BV677278 has conserved orthologs in mouse intron 2 and rat intron 1 of *DCDC2*, even though exhaustive RT-PCR studies show it is not transcribed (data not shown). The extent to which this non-coding region is conserved is characteristic of TF binding sites and other classes of regulatory regions (Gibbs et al. 2004). We garnered further support for our hypothesis that BV677278 is a functional element by searching the sequence in the TRANSFAC database and finding that it encoded many known regulatory motifs.

There are at least ten sequence variations of the BV677278 STR (Table 1), with significant differences among them in the number of putative transcription factor binding sites (TRANSFAC). In addition, we previously reported a microdeletion variant in RD families, wherein a 2445 bp sequence encompassing the STR was deleted. In a recent study of 72 RD cases and 184 controls from Germany, Wilcke et al. (2009) reported that deletion heterozygotes (del/non-del) were more commonly found in RD cases ($P < 0.01$), and that deletion homozygotes (del/del) were only found among RD cases ($P < 0.005$).

This paper is the first to attempt to elucidate the mechanism by which BV677278 exerts its effect on RD. The difference in STR allele frequency observed in RD cases compared to fluent controls could be explained by variation in regulatory function among the BV677278 STR alleles. If *DCDC2* does indeed modulate brain development and neuroarchitecture through its effects on neuronal migration (Gabel et al. 2010), then differences in *DCDC2* expression could have a direct effect on specialized reading centers in the brain, or the neural circuits that connect structures or modalities important for reading fluency.

This putative role of *DCDC2* in neuronal migration is the reason we chose P19 cells for the luciferase study. This cytogenetically normal cell line is multipotent, and will readily differentiate into neurons and neuroglia in the presence of retinoic acid. (Jones-Villeneuve et al. 1982) Though these cells are not strictly neural progenitor cells (NPCs), the fact that they have this potential makes them more suitable for these experiments than other cell lines. We chose to use them in their undifferentiated state because NPCs, and not terminally differentiated neurons, migrate during development. In spite of these advantages, P19 is a murine cell line, and there may be some question regarding its biological relevance to a reporter assay with a human putative enhancer sequence. However, inter-specific luciferase reporter assays of this type are commonly described in the literature (see Lu et al. 2002; Jin et al. 2006). Recently, Hirunsatit et al. (2009) used P19 cells, in addition to the human cell lines HEK-293 and SK-N-BE(2), to show enhancer activity of a 21 bp insertion in the human *SLC6A1* gene. Furthermore, Fig. 2c shows that the same EMSA3 shift we observed in human brain nuclear lysate is also present in Raji nuclear lysate, and more importantly, in

undifferentiated P19 nuclear lysate. This indicates that the EMSA3-binding protein of interest is present in P19 cells, supporting their validity for use in our luciferase assay.

Figure 2 shows that the interaction between the EMSA3 probe and its cognate binding protein is specific in all tissues/cell lines tested, including P19. In human brain nuclear lysate, the scrambled EMSA3 and EBNA1 unlabeled probes (Fig. 2b, lanes 4–6) seem to increase the amount of protein-labeled probe interaction, probably through additional blocking of nonspecific or weakly specific proteins that would otherwise compete. However, this effect is very dramatic when unlabeled EMSA4 probe is used. Though subsequent experiments suggested that the original shift we saw with the EMSA4 probe (Fig. 1b, lanes 7–9) is a nonspecific interaction, the presence of the EMSA4 sequence in the reaction seems to facilitate the EMSA3 interaction, though it is possible that this is an artifact, or simply the result of additional blocking of nonspecific interactions. It should also be noted that the protein-binding EMSA3 sequence is not localized exclusively to intron 2 of *DCDC2*, but is present in multiple sites in the genome. It is possible that EMSA3 alone is sufficient to confer regulatory activity, and that it is therefore active at these other sites. It is also possible, however, that BV677278 acts more like an enhanceosome, and other DNA sequences in addition to EMSA3, and possibly additional proteins, are necessary for optimal activity. In any case, genes that are spatially or temporally co-expressed are often modulated by common regulatory elements; that EMSA3 and its surrounding sequences could control a set of developmentally regulated genes represents an intriguing possibility.

It is interesting to note that while the presence of a purine-rich STR in rat and mouse *DCDC2* is conserved, the repeat units in rat and mouse show significant structural differences from those in human, and significant structural similarity to each other (Supplemental Fig. 2), while the *DCDC2* coding region itself is much more conserved among all three species (Supplemental Fig. 3). Curiously, the most common allele (allele 1) and the Basic/600 construct had a similarly small effect on the magnitude of *DCDC2* expression in our reporter system. However, the Basic/600 construct is by no means a complete model of the deletion allele, and it is possible that the large number of putative regulatory binding sites missing from the deletion allele could conceivably have a larger effect on spatial and temporal regulation of *DCDC2* expression that could not be captured in the reporter assay. This could explain the strength of the association between the deletion allele and RD. Intriguingly, it could also explain the disproportionate frequency of the deletion allele present in a functional subgroup of RD, nondysphonetic dyslexics ($P < 0.005$), in the Wilcke et al. (2009) study. Though our luciferase results suggest a *cis* effect of BV677278 on the *DCDC2* promoter, it remains possible that BV677278 could modulate the expression of other *DYX2* genes (such as *KIAA0319*) in *cis*, or even of distant genes in *trans* via long-range interaction with their promoters. Identification of the EMSA3-binding protein(s) will allow for further elucidation of the function of BV677278.

These data suggest, for the first time, that BV677278 alleles function as regulatory elements for *DCDC2* expression, which may link to neuronal migration in the development of the central nervous system. To date, not a single null mutation within any RD candidate gene has been described, indicating instead that polymorphisms within regulatory elements may be the genetic factors critical to reading performance. In the case of BV677278, there could be a wide range of allelic and even combinatorial genotypic effects, given the large number of alleles in samples from the US and Europe. The hypothesized regulatory effect of BV677278, for which evidence is presented in this study, appears intuitively compatible with RD, a highly heritable syndrome which nonetheless shows an intriguing lack of pleiotropy for a brain disorder. A change in regulation of a gene, rather than its gain or loss of function, could parsimoniously explain the relative subtlety and specificity of the RD phenotype and the reason it lies on a continuum. Moreover, regulatory elements are

modifiable by epigenetic, environmental, and stochastic factors, and by genetic alterations to their cognate binding proteins, which could account for some of the non-genetic variation attributed to the disorder, and perhaps even the non-Mendelian pattern of inheritance that is commonly observed in RD. Assignment of allele-specific or genotype-specific risks, yet to be determined through clinical studies, will likely identify deleterious as well as protective alleles of various strengths. In turn, these will inform additional functional studies, especially with regard to identification and further molecular studies of the nuclear protein shown to bind the EMSA probes. Beyond *DCDC2*, clinical studies that determine the RD risk contributed by other genes will richly inform functional studies of gene–gene and gene–environment interactions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by the International Dyslexia Association (R07420 to H.M.), and National Institutes of Health/National Institute of Neurological Disorders and Stroke (R01 NS43530 to J.R.G.). The authors thank Dr. Satish Ghatpande for kindly providing the cell lines, Dr. Patrick G. Gallagher for scientific suggestions, and Dr. Seiyu Hosono, Dr. Zhi-jia Ye, Dr. Queenie Tan, and Dr. Rong Cong for technical assistance. We also thank Susan Chan for editing the manuscript.

References

- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Methodol)* 1995;57(1):289–300.
- Birney E, Clamp M, Durbin R. GeneWise and Genomewise. *Genome Res* 2004;14(5):988–995. [PubMed: 15123596]
- Brkanac Z, Chapman NH, Matsushita MM, Chun L, Nielsen K, Cochrane E, et al. Evaluation of candidate genes for DYX1 and DYX2 in families with dyslexia. *Am J Med Genet B Neuropsychiatr Genet* 2007;144B(4):556–560. [PubMed: 17450541]
- Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC. Quantitative trait locus for reading disability on chromosome 6. *Science* 1994;266(5183):276–279. [PubMed: 7939663]
- Cope N, Harold D, Hill G, Moskvina V, Stevenson J, Holmans P, et al. Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *Am J Hum Genet* 2005;76(4):581–591. [PubMed: 15717286]
- Deffenbacher KE, Kenyon JB, Hoover DM, Olson RK, Pennington BF, DeFries JC, et al. Refinement of the 6p21.3 quantitative trait locus influencing dyslexia: linkage and association analyses. *Hum Genet* 2004;115:128–138. [PubMed: 15138886]
- DeFries JC, Fulker DW, LaBuda MC. Evidence for a genetic aetiology in reading disability of twins. *Nature* 1987;329(6139):537–539. [PubMed: 3657975]
- Fisher SE, Marlow AJ, Lamb J, Maestrini E, Williams DF, Richardson AJ, et al. A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. *Am J Hum Genet* 1999;64(1):146–156. [PubMed: 9915953]
- Francks C, Paracchini S, Smith SD, Richardson AJ, Scerri TS, Cardon LR, et al. 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* 2004;75(6):1046–1058. [PubMed: 15514892]
- Gabel LA, Gibson CJ, Gruen JR, Loturco JJ. Progress towards a cellular neurobiology of reading disability. *Neurobiol Dis* 2010;38(2):173–180. [PubMed: 19616627]
- Gayan J, Olson RK. Genetic and environmental influences on orthographic and phonological skills in children with reading disabilities. *Dev Neuropsychol* 2001;20(2):483–507. [PubMed: 11892949]

- Gayán J, Smith SD, Cherny SS, Cardon LR, Fulker DW, Brower AM, et al. Quantitative-trait locus for specific language and reading deficits on chromosome 6p. *Am J Hum Genet* 1999;64(1):157–164. [PubMed: 9915954]
- Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, et al. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 2004;428(6982):493–521. [PubMed: 15057822]
- Grigorenko EL, Wood FB, Meyer MS, Pauls DL. Chromosome 6p influences on different dyslexia-related cognitive processes: further confirmation. *Am J Hum Genet* 2000;66(2):715–723. [PubMed: 10677331]
- Hirunsatit R, George ED, Lipska BK, Elwafi HM, Sander L, Yrigollen CM, Gelernter J, Grigorenko EL, Lappalainen J, Mane S, Nairn AC, Kleinman JE, Simen AA. Twenty-one-base-pair insertion polymorphism creates an enhancer element and potentiates *SLC6A1* GABA transporter promoter activity. *Pharmacogenet Genomics* 2009;19(1):53–65. [PubMed: 19077666]
- Jin ZG, Liu L, Zhong H, Zhang KJ, Chen YF, Bian W, Cheng LP, Jing NH. Second intron of mouse nestin gene directs its expression in pluripotent embryonic carcinoma cells through POU factor binding site. *Acta Biochim Biophys Sin* 2006;38(3):207–212. [PubMed: 16518546]
- Jones-Villeneuve EM, McBurney MW, Rogers KA, Kalnins VI. Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glial cells. *J Cell Biol* 1982;94(2):253–262. [PubMed: 7107698]
- Katusic SK, Colligan RC, Barbaresi WJ, Schaid DJ, Jacobsen SJ. Incidence of reading disability in a population-based birth cohort, 1976–1982, Rochester, Minn. *Mayo Clin Proc* 2001;76(11):1081–1092. [PubMed: 11702896]
- Lind PA, Luciano M, Wright MJ, Montgomery GW, Martin NG, Bates TC. Dyslexia and DCDC2: normal variation in reading and spelling is associated with DCDC2 polymorphisms in an Australian population sample. *Eur J Hum Genet*. 2010 [Epub ahead of print].
- Lu C, Li Y, Zhao Y, Xing G, Tang F, Wang Q, Sun Y, Wei H, Yang X, Wu C, Chen J, Guan KL, Zhang C, Chen H, He F. Intracrine hepatopoietin potentiates AP-1 activity through JAB1 independent of MAPK pathway. *FASEB J* 2002;16(1):90–92. [PubMed: 11709497]
- Ludwig KU, Schumacher J, Schulte-Körne G, König IR, Warnke A, Plume E, et al. Investigation of the DCDC2 intron 2 deletion/compound short tandem repeat polymorphism in a large German dyslexia sample. *Psychiatr Genet* 2008;18(6):310–312. [PubMed: 19018237]
- Meng H, Smith SD, Hager K, Held M, Liu J, Olson RK, et al. DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proc Natl Acad Sci USA* 2005;102(47):17053–17058. [PubMed: 16278297]
- Paracchini S, Thomas A, Castro S, Lai C, Paramasivam M, Wang Y, et al. The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. *Hum Mol Genet* 2006;15(10):1659–1666. [PubMed: 16600991]
- Schumacher J, Anthoni H, Dahdouh F, König IR, Hillmer AM, Kluck N, et al. Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. *Am J Hum Genet* 2006;78(1):52–62. [PubMed: 16385449]
- Shaywitz SE, Shaywitz BA, Fletcher JM, Escobar MD. Prevalence of reading disability in boys and girls. Results of the Connecticut Longitudinal Study. *JAMA* 1990;264(8):998–1002. [PubMed: 2376893]
- Taylor KR, Holzer AK, Bazan JF, Walsh CA, Gleeson JG. Patient mutations in *doublecortin* define a repeated tubulin-binding domain. *J Biol Chem* 2000;275:34442–34450. [PubMed: 10946000]
- Wilcke A, Weissfuss J, Kirsten H, Wolfram G, Boltze J, Ahnert P. The role of gene DCDC2 in German dyslexics. *Ann Dyslexia* 2009;59(1):1–11. [PubMed: 19238550]
- Wu, KK. Analysis of protein-DNA binding by streptavidin-agarose pulldown. In: Bina, M., editor. *Methods in molecular biology*, vol 338: gene mapping, discovery, and expression: methods and protocols. Totowa, NJ: Humana Press, Inc; 2006. p. 281–290.

Fig. 1. EMSA detects oligonucleotide probes from BV677278 that bind protein in human brain nuclear lysate. **a** Sequence of EMSA probes relative to the BV677278 repeat sequence. **b** EMSA detects oligonucleotide probes from BV677278 that bind protein in human brain nuclear lysate. Lanes 1–3 are the *Oct2A* positive control, which contains a specific, well-characterized recognition sequence present in the probe. EMSA3 (lanes 4–6) and EMSA4 (lanes 7–9) are non-overlapping double-stranded, 20-mer oligonucleotide probes with sequence extracted from BV677278 (**a**). Lanes 1, 4, and 7 are baseline mobilities of labeled probes without human brain nuclear lysate. Lanes 2, 5, and 8 show probes bound to a protein in human brain nuclear lysate and the consequent electrophoretic mobility shifts. Lanes 3, 6, and 9 show that binding can be competed with identical “cold” unlabeled probe

Fig. 2.

The EMSA3 interaction is specific. **a** Summary of the labeled and unlabeled probes used in **(b)** and **(c)**. Labeled EMSA3 probe was used in all lanes, in both **(b)** and **(c)**. Competition assays were done with a 200-fold molar excess of unlabeled EMSA3 probe in lane 3, of unlabeled EMSA3-Scram1 probe in lane 4, of unlabeled EMSA3-Scram2 probe in lane 5, of unlabeled EBNA1 probe in lane 6, and of unlabeled EMSA4 probe in lane 7 (**b** only). **b** The EMSA3 interaction is specific in human brain nuclear lysate. The shift observed in lane 2 (*arrow*) vanishes when a 200-fold molar excess of unlabeled EMSA3 probe is added to the binding reaction (lane 3). However, a 200-fold molar excess of either one of the two unlabeled scrambled EMSA3 probes (lanes 4 and 5) does not compete with the labeled probe, nor does a 200-fold molar excess of unlabeled EBNA1 probe (lane 6) or unlabeled EMSA4 probe (lane 7). **c** The EMSA3 interaction is present and specific in the Raji cell line (human Burkitt lymphoma) and in the P19 cell line (mouse embryonal carcinoma, undifferentiated). In both of these cell lines, the behavior of the EMSA3 shift (*arrow*) is similar to that observed in human brain nuclear lysate (**b**)

Fig. 3.

BV277278 enhancer activity with *DCDC2*-specific promoter. **a** Schematic of pGL3-Basic vector constructs. LUC+ is the luciferase gene. BV677278(+) means the orientation of the sequence is the same as that of the *DCDC2* transcript. BV677278(-) means the orientation of the sequence is reversed with respect to the *DCDC2* transcript. “600” in “B/600” is the 600 bp 5-prime of the ATG start codon of *DCDC2*, serving as the *DCDC2*-specific promoter. “B” in “B/600” stands for Basic vector from Promega. “A” in “B/600/A” stands for alleles of BV677278. **b** BV677278 enhancer activity with *DCDC2*-specific promoter. The X-axis shows the names of the constructs. The Y-axis shows signal intensity, normalized to co-transfected *Renilla* luciferase, and averaged among three replicates. B/EA1(-) and B/EA1(+): Basic vector with BV677278 allele 1 in the enhancer position, in the + or - orientation, respectively. Basic/600: Basic vector with *DCDC2*-specific promoter at the 5' position, upstream of luciferase reporter gene (conventional promoter position). B/600/A1(-) and B/600/A1(+): Basic/600 construct with allele 1 of BV677278 3-prime of the luciferase gene (conventional enhancer position). This nomenclature continues for alleles 3, 4, 5, 6, and 10. SV40 Promoter: Promoter vector from Promega with heterologous SV40 promoter. (-) and (+) indicate the orientation of the BV677278 alleles relative to *DCDC2*'s orientation along 6p22; (+) denotes the same orientation as the *DCDC2* transcript. Bars represent standard deviation for triplicate experiments with each construct. Asterisks represent a statistically significant difference in luciferase signal intensity relative to the Basic/600 construct ($P < 0.01$), after correction for multiple testing

DCDC2 intron 2 deletion/compound STR polymorphism alleles (GenBank accession No. BV677278)

Table 1

Allele	Repeat unit 1	Repeat unit 2	SNP1	Repeat unit 3	Repeat unit 4	Repeat unit 5	Length (BP)	Freq ^a
1	(GAGAGGAAAGGAAA)2	(GGAA)7		(GGAA)2	(GGAA)4	(GGGA)2	305	0.624
2	(GAGAGGAAAGGAAA)1	(GGAA)9	DelGAAA	(GGAA)0	(GGAA)4	(GGGA)2	288	0.003
3	(GAGAGGAAAGGAAA)1	(GGAA)6		(GGAA)2	(GGAA)4	(GGGA)2	288	0.060
4	(GAGAGGAAAGGAAA)2	(GGAA)6		(GGAA)2	(GGAA)4	(GGGA)2	301	0.106
5	(GAGAGGAAAGGAAA)2	(GGAA)8		(GGAA)2	(GGAA)4	(GGGA)2	309	0.028
6	(GAGAGGAAAGGAAA)2	(GGAA)8		(GGAA)2	(GGAA)3	(GGGA)2	305	0.039
7	(GAGAGGAAAGGAAA)2	(GGAA)8		(GGAA)1	(GGAA)4	(GGGA)2	305	0.003
8	(GAGAGGAAAGGAAA)2	(GGAA)7	DelGAAA	(GGAA)0	(GGAA)4	(GGGA)2	293	0.003
9	(GAGAGGAAAGGAAA)1	(GGAA)7		(GGAA)2	(GGAA)4	(GGGA)2	292	0.005
10	(GAGAGGAAAGGAAA)2	(GGAA)4		(GGAA)2	(GGAA)4	(GGGA)2	293	0.044
Deletion	-	-	-	-	-	-	0	0.085

^aFrequency among parents, from Meng et al. (2005)

Table 2

BV677278 probes for EMSA

	Probe	Complementary sequence
EMSA1	5'-TAAAAAGAAGGAAAGAGAGG-3'	5'-CCTCTCTTTCCTTCTTTTA-3'
EMSA2	5'-GAGAGGAAGGAAAGAGAGGA-3'	5'-TCCTCTCTTTCCTTCCTCTC-3'
EMSA3	5'-GAGAGGAAGGAAAGGAAGGA-3'	5'-TCCTTCCTTTCCTTCCTCTC-3'
EMSA3- Scram1	5'-GGGGGGGGGAAAAAAAAA-3'	5'-TTTTTTTTTCCCCCCCCC-3'
EMSA3-Scram2	5'-GAGAGAGAGAGAGAGAGAGA-3'	5'-TCTCTCTCTCTCTCTCTC-3'
EMSA4	5'-AAGGAAGGAAGGAAAGAATG-3'	5'-CATTCCTTCCTTCCTTCCTT-3'
Oct2a ^a	5'-GTACGGAGTATCCAGCTCCGTAGCAT GC AAATCCTCTGG-3'	5'-CCTCATAGGTCGAGGCATCG TACGTTT AGGAGACCAGCT-3'
EBNA1	5'-ATTAGGATAGCATATGCTACCCAGATATAG-3'	5'-CTATATCTGGGTAGCATATGCTATCCTAAT-3'

Bolded residues represent the known binding site for OCT2A protein

^aPositive control oligos

Table 3

Representation of BV677278 alleles in the EMSA probes

Allele	EMSA1	EMSA2	EMSA3	EMSA4
1	+	+	+	+
2	+	+	+	+
3	+	–	+	+
4	+	+	+	+
5	+	+	+	+
6	+	+	+	+
7	+	+	+	+
8	+	+	+	+
9	+	–	+	+
10	+	+	+	+
Deletion	–	–	–	–