9-14-2009

The regulation of DCDC2, a candidate gene for dyslexia

Christopher J. Gibson
Yale University

The Regulation of *DCDC2*, a Candidate Gene for Dyslexia

A Thesis Submitted to the

Yale University School of Medicine

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Medicine

by

Christopher J. Gibson

2009

THE REGULATION OF *DCDC2*, A CANDIDATE GENE FOR DYSLEXIA

Christopher J Gibson and Jeffrey R. Gruen. Departments of Pediatrics, Genetics, and Investigative Medicine, Yale University School of Medicine, New Haven, CT.

Within the human genome, genetic mapping studies have identified ten regions of different chromosomes, known as DYX loci, in genetic linkage with dyslexia. The gene *DCDC2*, located within the DYX2 region on chromosome 6p22, has been shown to have genetic association with dyslexia in several independent studies. Functional assays of *DCDC2* indicate that it may help guide the migration of neurons during early brain development. *DCDC2* polymorphisms that display the strongest association with dyslexia are located in a highly GC-rich region in intron 2 known as *BV677278*. These polymorphisms contain several transcription factor binding sites, including the canonical 8-base recognition site for PEA3, a transcription factor known to modulate neuronal migration in mice. We hypothesized that 1) *BV677278* is an enhancer element for *DCDC2* that regulates its expression level, location, or timing, and that 2) PEA3 regulates *DCDC2* expression by binding *BV677278*. To test these hypotheses we showed that PEA3 binds to regions within *BV677278*, and that siRNA knockdown of PEA3 appears to delay the expression of *DCDC2* during neuronal differentiation of mouse cells. We concluded that PEA3 was a viable candidate transcription factor for *DCDC2*, with the ability to bind *BV677278*. Taken together, these data suggest a possible mechanism by which *BV677278* polymorphisms alter PEA3 binding and DCDC2 expression, which in turn may modulate neuronal migration and affect the risk of dyslexia.

Acknowledgements

This work would not have been possible without the generous help of many people. I would like to thank the Office of Student Research and the Leon Rosenberg Genetics Fellowship, which funded my research year. I had fabulous support within the Gruen lab: Haiying Meng taught me the principles of molecular genetics, and Natalie Powers helped me refine and troubleshoot my ideas. In the spirit of scientific collaboration, I also had help from members of several outside labs. At Yale, my friend David Merrick, from the Caplan lab in the Department of Physiology, taught me the basics of cell culture and western blotting. Yu Wang and Joe LoTurco, at the University of Connecticut, and Hunter Cai and Laura Grabel, at Wesleyan University, taught me additional techniques of cell culture and provided me with rare cell lines that helped further my research.

I would like to extend particular thanks to Dr. Jeffrey Gruen, associate professor of Pediatrics, Genetics, and Investigative Medicine. He was the ideal mentor: demanding but understanding, challenging but forgiving, and above all, incredibly generous with his time and enthusiasm. I have friends who feel fortunate if they see their mentor once a month; I saw mine every day, and his door was always open to me.

I have always thought, and sometimes announced, that I was fortunate enough to grow up in the best family on the planet. My parents, Sandra and Jim Gibson, have supported me with love through every chapter of my life, and I cannot thank them enough. My sister, Laura Gibson, is one of my best friends and constantly inspires me to look for joy and excitement in every corner of life.

Finally, my life has been incomparably enriched for the last four years by the love and support of my fiancée, Laura Dichtel, to whom I dedicate this work. I love you and I could not imagine life without you.

Table of Contents

Introduction

Reading Disability (RD, also known as developmental dyslexia), is the most common learning disability among school-age children, with an incidence between 5% and 17% in countries where it has been studied. It is defined as difficulty learning to read despite normal intelligence and adequate instruction. The idea that differences in language and reading abilities are partially attributable to genetics is not new; as early as the 19th century, for instance, educators and physicians described families in which more than one member had difficulty learning to read.[1] With the evolution of more sophisticated techniques of genetic analysis, our understanding of the biologic basis of these language disorders continues to grow. This introduction summarizes some of the current understanding of the genes and proteins that are thought to affect RD. The remainder of the paper describes experiments conducted to elucidate the regulation of *DCDC2*, a particular candidate gene for the disease.

I. The Process of Identifying Language Disorder Genes

To understand how the genes for language disorders are identified, one must first understand the types of studies geneticists utilize. These can be visualized as a multi-step process of increasingly narrow scope, starting with heritability studies, proceeding to genetic linkage analysis and high-resolution genetic association studies (a process termed "positional cloning"[2]), and ending with functional assays of candidate genes. In specific cases, other techniques, such as classical karyotyping, may also be used.

The simple observation that a trait runs in families is not enough to establish that it is genetic, since family members often share the same environment, and some of the

same genes. The most common method used to confirm that a familial trait is at least partly heritable is a twin study, in which the concordance of the trait in monozygotic twins, who are genetically identical, is compared to its concordance in dizygotic twins (or any pair of non-twin siblings), who are not.[3] A significantly higher concordance in monozygotic twins than in dizygotic twins implies that the trait has a genetic component. Statistical analysis can estimate the degree of phenotypic variation that is due solely to additive genetic factors, which is known as the trait's *heritability*.[4] For RD, twin studies have shown a heritability of 44% to 77%,[5] implying that at least half of the entire range of disease can be attributed to genetic, rather than environmental, factors.

Sometimes, the next level of genetic study may be karyotype analysis, which has been used since the 1950s. It involves light microscopic analysis of peripheral white blood cell chromosomes arrested in metaphase, and stained with giemsa to distinguish characteristic banding patterns for each chromosome. Classical karyotype analysis can detect chromosomal deletions or duplications, as well as the exchange of large chromosomal segments, called translocations, on the order of 10 million bases (Mb). With higher resolution labeling, such as that used in Fluorescence In Situ Hybridization (FISH), smaller chromosomal abnormalities on the order of 100 kb can be identified.

Although karyotype analysis is sometimes used in clinical genetics, its use is often limited, and it is not commonly used in large research studies. This is due in part to its inability to analyze sequence and DNA microstructure. The presence of regulatory elements, epigenetic modification, and diverse gene-gene interactions, which affect gene penetrance and the heterogeneity of expression, often mean that siblings who possess identical chromosomal macrostructure may nevertheless display different phenotypes.

Nevertheless, we mention karyotyping here because two of the four RD genes, *DYX1C1* and *ROBO1*, were identified through the serendipitous discovery of chromosomal translocations by karyotype analysis. We will detail its application when we discuss the specific genes below.

Traditionally, once a trait or disease has been shown to have a genetic component, the next step has been genetic linkage analysis, which determines the chromosomal regions, usually 1-20 Mb in size, that contribute to the development of the trait.[6] In performing these studies, researchers compare genotypes from multiple members across several generations of families affected by the trait in question and consider markers distributed across all chromosomes. Although older studies utilized microsatellites or short tandem repeats (STRs), the markers most commonly used in modern genetic studies are Single Nucleotide Polymorphisms (SNPs), variations in single bases that occur on the order of one per 100 bases of DNA.[7] Traditionally, by tracing the lineage of marker alleles across generations and comparing it with the lineage of the trait in the family, researchers have identified markers within a specific chromosomal region that are inherited in the same pattern as the trait. These markers are said to be "in genetic linkage" with the trait because the marker and trait appear together more often than would be expected by chance alone (that is, they are physically "linked" by their proximity on a chromosome and cosegregate during recombination events in meiosis); the chromosomal region in which they are encoded is termed a "susceptibility locus."

Classical linkage analysis using large pedigrees is best suited to Mendelian diseases. More recently, researchers have developed models that determine linkage by comparing the rate of allele sharing in sibling pairs (or other relational groupings) to the

rate of sharing predicted by chance alone. This method eliminates the need to obtain DNA from multiple generations, which is often difficult to accomplish and has traditionally been a barrier to efficient linkage studies.[8] Perhaps more importantly, it allows for analysis of complex diseases because it de-emphasizes the mode of disease inheritance, which for complex diseases is never well defined.

Numerous linkage studies have been performed for RD and have identified about ten susceptibility loci, which are collectively designated by the label *DYX*. The evidence supporting the involvement of some of these loci in RD is relatively weak, but four have been consistently replicated: DYX1 on 15q[9, 10], DYX2 on 6p[11], DYX5 on 3p[12], and DYX6 on 18p[13]. These regions may encode genes that affect heritable susceptibility to RD across the general population, while the other six loci likely encode genes with RD polymorphisms unique to small populations. It is also likely that other loci affecting reading and RD exist in the genome but have not yet been discovered.

After a linkage study has implicated a chromosomal region in the development of a disease, the next step has traditionally been to identify candidate genes using genetic association studies, which compare the frequency of marker alleles in affected subjects (cases) to the frequency in matched unaffected controls. The premise of these studies is that, due to historical recombination events over many generations, marker alleles found significantly more often in affected individuals must be in close physical proximity to the disease-causing mutation. Genetic linkage analysis are powerful and can sensitively determine the location of a disease-related gene to within 5 to 10 million base pairs, but their poor resolution precludes them from more finely localizing disease-related genes. Association analysis, on the other hand, can be much more precise. Its major weakness is

that it is vulnerable to false-positive association due to hidden population stratification (that is, occult allele sharing between cases and controls as a result of ancient admixture) when unrelated cases and controls are matched. Thus most widely-used association tests employ a family-based structure in which the cases and controls are not individuals but genotypes; the parental genotype found in the affected individual serves as the case, and the untransmitted parental genotype is the control. Using an array of finely spaced markers in a family-based association study, researchers can pinpoint a peak of association within a locus previously identified by a linkage study with both precision and relative certainty.

The need to perform a linkage study before testing association can be eliminated if the entire genome, as opposed to a single susceptibility locus, is interrogated by genetic association for disease-associated marker alleles. The advent of gene-chip technology, in which more than one million markers spanning the genome can be tested for association with a trait or disease, has made such genome-wide association (GWA) studies possible. Several well-publicized GWA studies have already been performed for common complex disorders (due to a combination of multiple genes and environmental factors), such as coronary artery disease,[14] breast cancer,[15] Type 2 Diabetes,[16] and multiple sclerosis.[17] The major drawback to GWA studies is the large number of markers needed to cover the genome, which greatly increases the likelihood of false-positive associations due to multiple testing. For complex diseases, in which the effect size may be small, the need to control for multiple testing requires the recruitment of several thousand case subjects and at least as many controls, making GWA studies expensive and complicated, and effectively impossible for rare diseases.[18]

II. Candidate Genes for Reading Disability: Discovery and Functional Assays

The methods described above have resulted in the identification of four candidate genes for dyslexia: *DYX1C1* in DYX1,[19] *DCDC2*[20] and *KIAA0319*[21] in DYX2, and *ROBO1* in DYX5.[22] Three of these are discussed here; *DCDC2*, the subject of this study, is discussed in section III.

Once candidate genes have been identified by genetic association studies, a common step in elucidating the pathway through which they contribute to the disease is the determination of their physiologic function. The disease variants of the gene, whether full translocations or smaller sequence variations, can then be evaluated in the context of the gene's function to understand how they confer disease susceptibility. In the case of RD, three genes (*DCDC2*, *KIAA0319*, and *DYX1C1*) appear to influence the migration of developing neurons during early embryogenesis, while *ROBO1* appears to affect the extension of axons from neuron cell bodies. As above, discussion of *DCDC2* is reserved for section III.

KIAA0319

In 2005, Cope et al identified *KIAA0319*, located on 6p21, by interrogating 5.3 million bases (Mb) spanning the region with 57 SNPs, in a sample of 143 parent-proband RD trios in the United Kingdom.[21] The study found association not within the gene itself, but in a region immediately adjacent to the gene's transcription start site. These regions, called 5' untranslated regions (5' UTRs), often regulate the timing and degree of transcription. In this case, the SNPs associated with the development of RD appeared to decrease the transcription of *KIAA0319*. The results of the initial association study for

KIAA0319 have been replicated in two further studies, one in a cohort from Wales[23] and one in a cohort from the United States.[24] However, a more recent study in a large population failed to show the same degree of association.[25]

KIAA0319 mRNA transcripts are present at very high levels within the human brain, particularly in the visual and parietal cortices. RNAi knockdown studies of *KIAA0319* have resulted in aberrant migration of affected neurons, though in a pattern different from that of *DCDC2* knockdown.[26] The molecular mechanism by which *KIAA0319* exerts its neurostructural effects has not yet been elucidated, though recent studies have shown that the full protein product of the gene resides in the plasma membrane of neurons and may mediate interactions between these and the supporting cells of the brain, known as glial cells. [27]

DYX1C1

As previously mentioned, *DYX1C1* (originally called *EKN1*) was initially discovered by karyotype analysis, which showed a translocation segregating with RD in a family in Finland.[12] In this case, the q11 portion of chromosome 15, which had already been identified by linkage analysis and designated DYX1, was translocated onto chromosome 2, and the q21 portion of chromosome 2 was translocated onto chromosome 15 (notated t(2;15)(q11;q21)); analysis of the exact position of the chromosome 15 breakpoint showed that it lay within a gene called *EKN1*. A small genetic study of families in Finland showed association with *EKN1*/*DYX1C1*,[19] as did a study in Canada,[10] but studies in the United States[28] and Italy[29] failed to show an association. This implies that the

importance of *DYX1C1* as a susceptibility gene for RD may be limited to specific populations.

Like *KIAA0319*, *DYX1C1* has been shown to be highly expressed in brain.[19] Furthermore, RNAi knockdown of *DYX1C1* significantly impaired neuronal migration, implying that it likely plays a role in early brain development.[30]

ROBO1

Similar to *DYX1C1*, *ROBO1* was first identified by finding a translocation in an RD family[31] through karyotype analysis. In this case, the translocation was between the p12 region of chromosome 3 and the q11 region of chromosome 8 (t(3;8)(p12;q11)). Since 3p12 had already been identified as DYX5 in a previous linkage study, an association study using SNPs was conducted on the original linkage cohort, which showed an association between certain SNPs within *ROBO1* and decreased expression of the gene.[22] However, there has not yet been independent validation of this gene in a separate RD cohort.

As opposed to the other three RD susceptibility genes, *ROBO1* does not appear to affect neuronal migration. Rather, it has been shown to encode an axonal guidance receptor, that is, a protein involved in receiving cellular signals to help direct the projection of axons from the neuron cell body.[22] In the absence of strong evidence for genetic association, understanding the molecular role of the *ROBO1* protein product lends some credibility to assertions of its role in reading.

III. The identification, function, and regulation of *DCDC2*

Meng et al first identified *DCDC2* as a candidate gene for dyslexia in 2005 in a genetic association study of 220 RD families from Colorado,[20] which was confirmed by Schumacher et al in a different German cohort shortly thereafter.[32] In the initial study, Meng et al examined 149 SNPs over 1.2 Mb of 6p22 and found a peak of association in six non-coding SNPs within an intron of *DCDC2*. Since introns are spliced out of the RNA transcript and do not affect the sequence of the ultimate protein product, these results imply that the susceptibility polymorphisms for *DCDC2* occur in a regulatory region of the gene. In other words, the known polymorphisms do not affect the amino acid sequence of the *DCDC2* protein product, but rather affect where (e.g., specific brain region or type of neuron), when (e.g., during a specific period of brain development), or how much of the protein is produced.

Meng et al surveyed levels of *DCDC2* messenger RNA (mRNA) levels and showed that it is expressed in the brain, especially in the temporal cortex and cingular gyrus, regions known to be involved in reading. They also performed functional studies of the gene, which implied that it may be necessary for the proper migration of neurons from the region around the brain ventricles, where they originate during early embryogenesis, to the outermost layer of the cerebral cortex, where they reside in maturity. In these studies, called RNAi knockdown assays, *DCDC2* mRNA in embryonic rat brain was specifically prevented from being translated into protein in early neural progenitor cells. Cells with decreased *DCDC2* levels stopped migrating at intermediate areas of the cerebrum, whereas wildtype cells expressing normal amounts of *DCDC2* migrated to the outer layers of the cortex.

At a molecular level, *DCDC2*'s effect on neuronal migration may occur via interactions between the *DCDC2* protein product and the cellular scaffolding known as the cytoskeleton. One of the functional domains of the *DCDC2* protein product, known as the doublecortin domain, has been shown in other studies to stabilize the assembly of important cytoskeletal components called microtubules.[33] If this proposed pathway is correct, decreased levels of the *DCDC2* protein product caused by dyslexia-associated polymorphisms may confer disease susceptibility by destabilizing microtubule structures and impairing the ability of neurons to migrate through the developing brain.

It is difficult, however, to reconcile the results of these functional studies with the observation that the polymorphisms identified within *DCDC2* all lie within the intronic region *BV677278*. Since this region is not translated into protein, it is unlikely that RD-related polymorphisms of *DCDC2* directly affect the gene's protein product. Instead, it is most likely that the RD-related polymorphisms alter a critical sequence necessary for the binding of an important transcription factor (TF). Indeed, reporter assays of *BV677278* have suggested that it may have just such a function (Meng et al 2006, unpublished data). When the region was cloned into an expression vector downstream (3') of firefly luciferase gene, the light emitted by human Jurkat cells transfected with the vector increased compared to control (vectors containing firefly luciferase but without *BV677278*). When *BV677278* was inserted upstream (5') of the luciferase gene, however, there was no significant difference in light production compared to control cells. This suggests that *BV677278* may act as a distal regulatory element, or "enhancer," rather than as a proximal regulatory element, or "promoter." Whereas promoters are regions of DNA at which transcription initiation complexes bind (and are thus essential to

the successful transcription of genes), enhancers serve as binding sites for transcription factors that may finely modulate, rather than switch on or off, gene transcription.

These results thus suggest that *BV677278* may act as an enhancer at which a critical transcription factor binds *DCDC2.* This TF may modulate the production of the *DCDC2* protein product, or it may affect the product of another nearby gene, such as *KIAA0319,* which, as previously noted, lies in the same 500 Kb locus as *DCDC2.* The existence of such a relationship, in fact, could reconcile the apparent coincidence that two genes showing independent association within RD lie in such close proximity to each other. Figure 1, shown on the following page, details possible mechanisms by which such a regulatory element could ultimately influence the development of neural architecture and a complex cognitive phenotype such as reading.

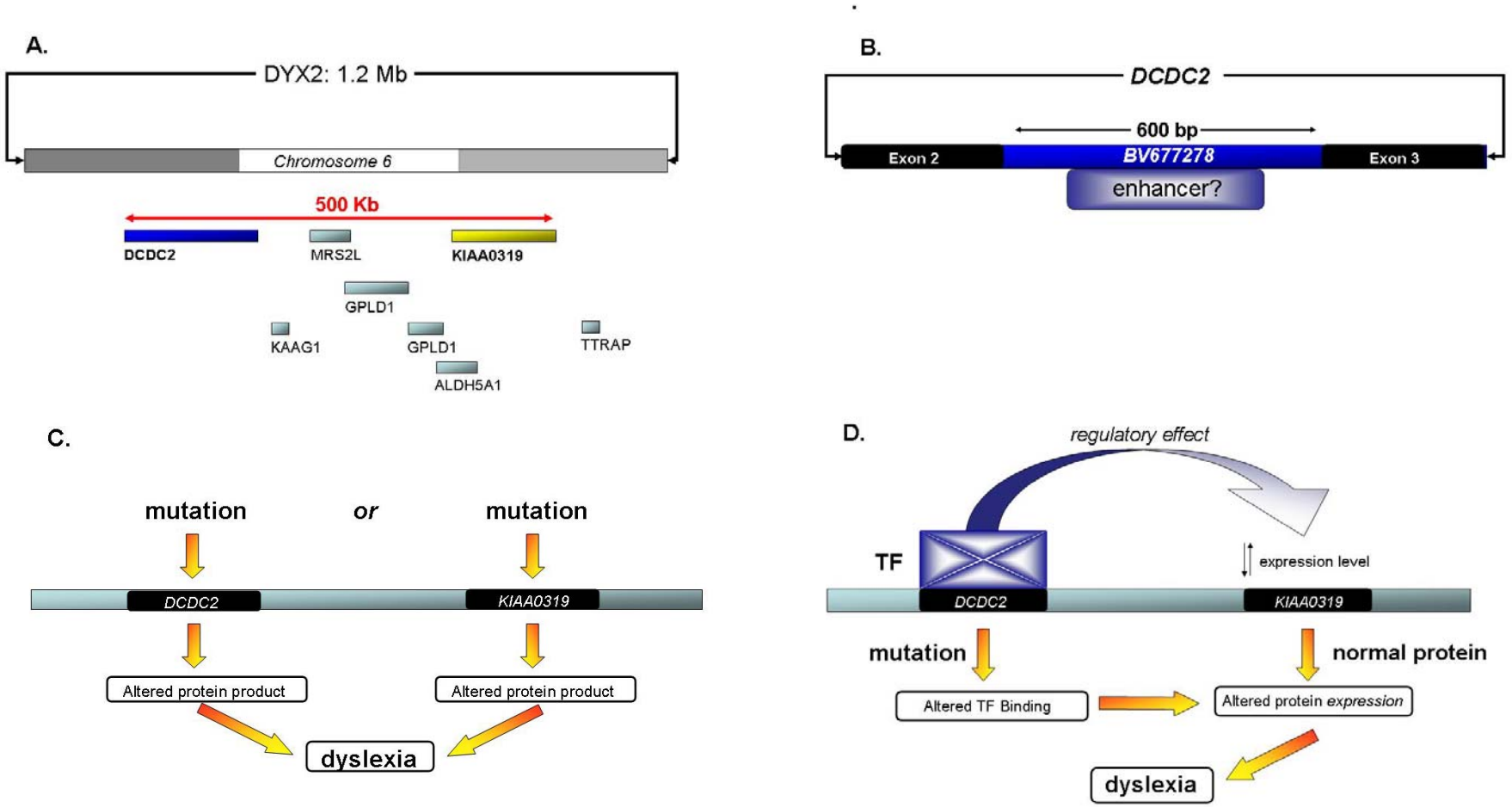

Figure 1. A: The DYX2 locus on 6p22 contains two candidate dyslexia genes, *DCDC2* and *KIAA0319*, within a 500 Kb span. **B:** *BV677278*, an approximately 600-bp region in intron 2 of *DCDC2*, has been shown by luciferase assay to have enchancer activity. **C.** One possible scenario by which both *DCDC2* and *KIAA0319* might contribute to dyslexia: if protein products of both genes are involved in reading, then risk polymorphisms in either could directly influence disease development. **D.** An alternative explanation might be that a transcription factor binds at an enhancer within one gene (e.g. *BV677278* in *DCDC2*) and regulates the expression of the other. The altered expression of the latter protein product is therefore responsible for the development of disease.

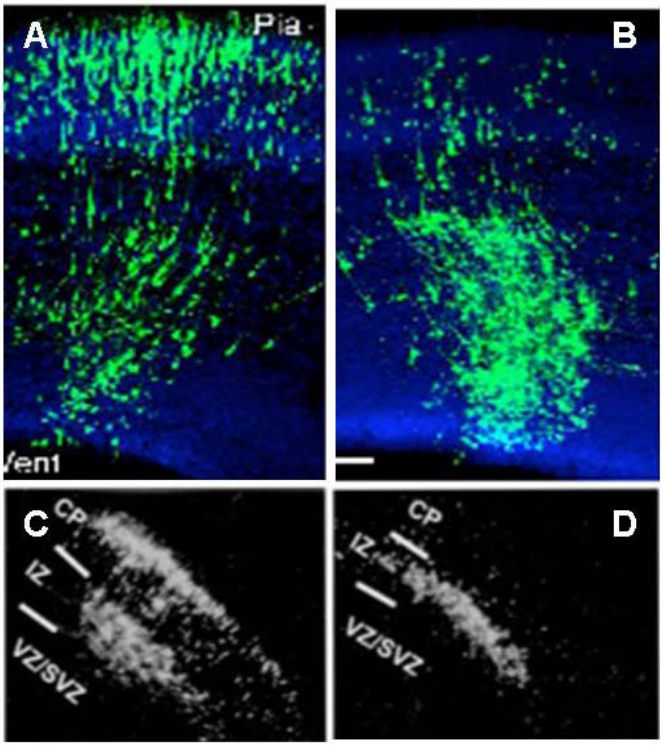

Figure 2. Comparison of neural migration defect conferred by knockdown of *DCDC2* or dominant negative of PEA3. In wildtype rats, Meng et al (2005) showed that the majority of normal neural progenitors, labeled green, migrate to the outer layers of the cerebral cortex (A). *DCDC2* knockdown cells (B) arrested predominantly in the intermediate zone. Hasegawa et al (2004) showed a similar population of neural progenitors reaching the outer cortex (CP) in wildtype mice (C). Neurons transfected with a PEA3 dominant negative construct (D) cluster overwhelmingly in the intermediate zone.

If there is a TF that binds *BV677278,* which one is it and how exactly does it exert its effects? One particularly interesting candidate is PEA3 (whose identification is described in Methods, below), a member of the ETS family of transcription factors expressed in human neocortex. In 2005, Hasegawa et al showed that PEA3 expression is dependent on the activation of fibroblast growth factor receptors (FGFRs), and that neurons with dominant negative mutations of PEA3 display an aberrant migration pattern when implanted in developing mouse brains in a pattern of intermediate-zone arrest identical to that of the *DCDC2* knockdowns described by Meng in 2005.[34] Figure 2, above, compares the migration deficit shown by these two papers.

Two questions must be answered to say with confidence that PEA3 is in fact a critical transcription factor for either *DCDC2* or *KIAA0319*. First, can PEA3 bind to *BV677278*? Second, does changing the expression of PEA3 (in this case, knocking it down completely) affect the expression of either *DCDC2* or *KIAA0319?* The answer to these questions is the subject of the experiments described here.

Hypothesis and Specific Aims

Based on the background presented here, we hypothesized that the intronic region *BV677278* within *DCDC2* is a critical regulatory region for *DCDC2* or *KIAA0319*. To test this hypothesis, we proposed:

Specific Aim 1: Using a bioinformatics approach, identify viable candidate transcription factor(s) based on binding properties, tissue localization, and known function;

Specific Aim 2: Test the candidate transcription factor(s) for binding to *BV677278* using electrophoretic mobility shift assays; and

Specific Aim 3: Characterize the regulatory effect of the candidate transcription factor(s) on *DCDC2* and *KIAA0319*, two proposed dyslexia genes, using siRNA knockdown of PEA3 and quantitative rt-PCR assays of gene expression.

Methods

Statement of Medical Student Contribution

The bioinformatics approach to transcription factor discovery, and the initial supershift assays, were conceived by Haiying Meng in the Gruen lab. The author refined the bioinformatics search and conceptualized the final design of the supershift assays. All other project design was conceived by the author. The author solely performed all experiments and analysis, wrote all text, and designed all figures, except for Figure 2 in the Introduction, which as noted presents previously published data from papers by Meng et al 2005 and Hasegawa et al 2004.

Identification of Candidate Transcription Factors

Four oligonucleotide sequences, each 22 nucleotides in length, were identified within *BV677278*. Each of these was shown to be altered or obliterated by the insertion or deletion of the short tandem repeats associated with RD. These four probes (TAAAAAGAAGGAAAGAGAGG, GAGAGGAAGGAAAGAGAGGA, GAGAGGAAGGAAAGGAAGGA, AAGGAAGGAAGGAAAGAATG), were each queried in the Transcription Element Search Software database (TESS, University of Pennsylvania) for all known transcription factor binding sites lying within them.

Cell Culture and Neural Differentiation

P19 (mouse teratocarcinoma) cells were obtained from ATCC (Manassas, Virginia). These were cultured in α-MEM supplemented with 7.5% fetal bovine serum, 2.5% calf serum, 5 ml 1x glutamine and 5 ml penicillin-streptomycin and split 1/8 every other day

with TrypLE (proprietary trypsin formulation, Invitrogen, Carlsbad, CA). Mouse embryonic stem cells containing GFP in the Sox1 promoter (referred to as "Sox1GFP cells") were originally engineered by the Smith lab, Cambridge, UK, and were obtained as a generous gift from Professor Laura Grabel, Wesleyan University. These were cultured in Glasgow-MEM supplemented with 10% FBS (lot-tested), 1% 1x glutamine, 1% penicillin-streptomycin, 1% non-essential amino acids, 1% sodium pyruvate, 3. 5 μl B-mercaptoethanol, and 0.2% leukocyte inhibitory factor (LIF). Passaging was the same as for P19s.

The differentiation of P19s into neural-like cells has previously been described in detail.[35] Briefly, after trypsinization, cells were plated into 6-well trays at a density of 1 to 3 x 10^5 cells/well in P19 growth medium supplemented with 5 x 10^{-7} M all-*trans*-retinoic acid (atRA). This medium was changed daily until differentiation day 4, when it was replaced by normal P19 growth medium to minimize atRA toxicity. Cells were trypsinized and replated when they reached 100% confluency, with complete differentiation lasting six to eight days.

The differentiation of mouse embryonic stem cells into neuronal precursors has also been described in detail elsewhere.[36] Briefly, Sox1GFP cells were plated 1 x 10^5 cells/well in serum-free medium supplemented with B27 (Invitrogen; lot-tested) and N2 (progesterone, insulin-transferrin-selenium, and putrescine). This medium was changed daily. Full differentiation took twelve days, and cells were generally split and replated at day 7-8.

RNA Extraction

RNA extraction was performed using RNEasy Plus© kits from Qiagen. Briefly, cells were trypsinized, then homogenized in QiaShredder homogenizer columns. The rest of the extraction was performed according to the RNEasy protocol. RNA concentration was assayed by assessing absorbance at 260 and 280 nm.

RT-PCR

Prior to PCR, 1 μg RNA from each sample was reverse transcribed into cDNA using an archive kit based on random primers (ABI). Quantitative RT-PCR was performed with the Taqman® system from ABI using 48-well plates. Reactions were all performed either in duplicate or triplicate. Fluorescent primers used (all mouse) were *dcdc2a* (the full-length transcript of mouse *dcdc2), d130043k22rik* (mouse ortholog of *KIAA0319;* hereafter referred to as *kiaa0319*), *pea3*, *sox1*, and *mtap2*. Mouse *gadph* was used as an internal control.

Electrophoretic Mobility Shift Assays (EMSAs)

These were performed using the Lightshift® kit (ThermoFisher Scientific, Rockford IL). The four DNA oligonucleotides (hereafter Probes 1-4) from within *BV677278* were annealed and labeled with biotin. They were then incubated with human brain nuclear lysate under varying reaction conditions. For each experiment, at least three reaction mixtures were incubated: one containing labeled probe only (the baseline mixture), one containing labeled probe plus brain lysate (the experimental mixture), and one containing

labeled probe, brain lysate, and an excess of unlabeled probe, to test the specificity of binding between probe and lysate.

After incubation, the reaction mixtures were loaded into a native (non-denaturing) 4% polyacrylamide gel and run at 80 V for 2 hours at 4 degrees Celsius. The protein was then electrophoretically transferred to a positively charged nylon membrane at 400 mA for 30 minutes, and crosslinked to the membrane using ultraviolet radiation in a UV Spectralinker (Stratagene). The blocking, hybridization, and washing steps were performed using provided Lightshift products; the hybridization was to a strepavidin-horseradish peroxidase conjugate. Membranes were developed using ECL and exposed to x-ray film for two minutes.

Supershift Assays

These were performed similarly to EMSAs, as described above, except that during incubation, mouse monoclonal or rabbit polyclonal anti-PEA3 (Santa Cruz, sc-113 or H120) was added to the reaction mixture. For each experiment, one additional reaction mixture substituting mouse anti-IgG, goat anti-DCDC2, or mouse anti-Oct2A was incubated as a control.

Results

Identification of Candidate Transcription Factors

The four oligonucleotides identified as the critical regions of *BV677278* for dyslexia risk were each entered into the TESS database. The queries showed that each probe contained binding sites for about twenty transcription factors, but only two, PEA3 (consensus sequence: AGGAAR, where R = purine) and NF-1 (Nuclear Factor-1, consensus sequence: TGGCA) are expressed in human brain. The expression of NF-1 is ubiquitous, but little is known about its function in humans, and it has mostly been characterized in adenovirus, where it was discovered.[37] PEA3 is expressed strongly in human brain.[38] The full results of the search are shown in Appendix 1.

Supershift Assays

To determine whether the transcription factor PEA3 can bind *in vitro* to *BV677278*, supershift assays were performed using labeled DNA probes and human brain nuclear lysate. The results are displayed in Figure 3, below. In lane 1, labeled probe alone migrated to the bottom of a native polyacrylamide gel, marked by arrow B. When incubated only with brain lysate (lane 2, marked with upward pointing arrow), the original band was greatly reduced and a new band appeared at the top of the gel, marked by arrow A, indicating a binding reaction between the probe and some component of the lysate mixture. When labeled probe was incubated with both lysate and non-specific antibody (anti-Oct2A, lane 3), the appearance of the lane was similar to that of lane 2 (no band at arrow B).

Lanes 4 through 7 show the effects of adding increasing amounts of monoclonal anti-PEA3 antibody. A band in the region of unbound probe (arrow B) becomes increasingly prominent as the concentration of antibody is incrementally increased, indicating that the antibody and probe compete for binding to the unknown component of brain lysate. A similar band with the addition of specific, unlabeled probe is seen in lane 8. Adding both monoclonal anti-PEA3 and an excess of unlabeled specific probe produces a band more intense than the addition of either species alone (lane 9). Finally, incubation of labeled probe and lysate with a non-specific competitor DNA (Epstein Barr Nuclear Antigen DNA) failed to produce a competitive band in the region of arrow B.

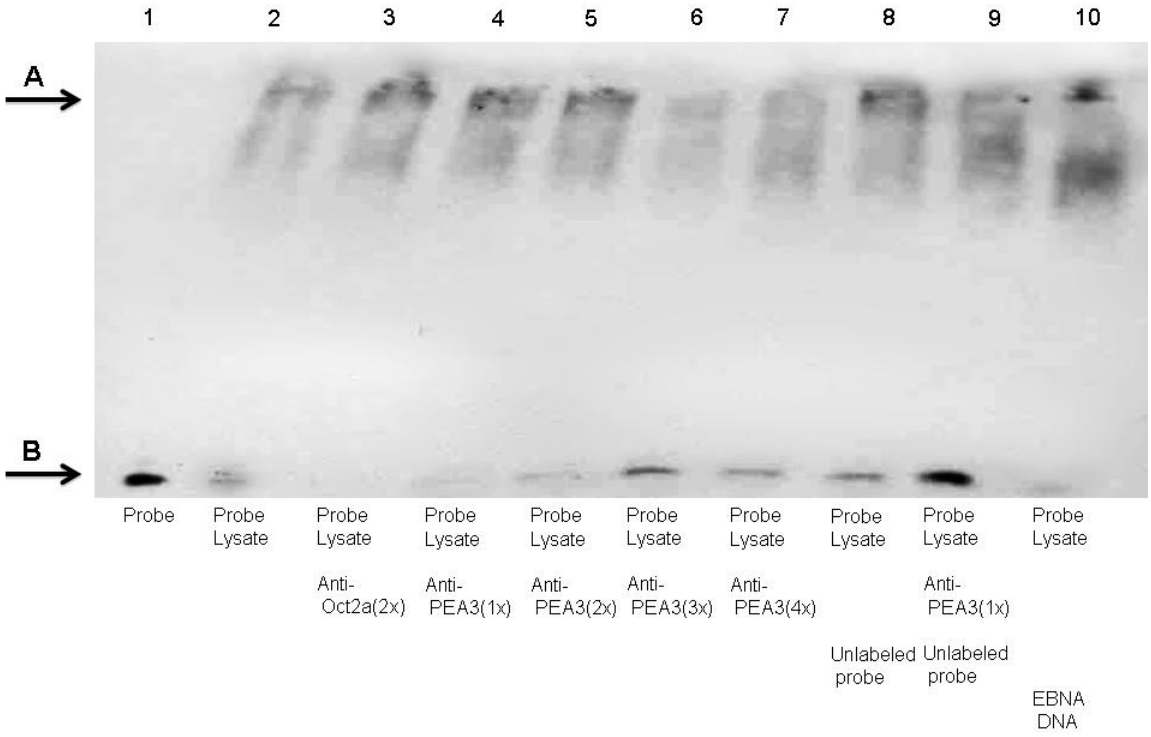

Figure 3. Electrophoretic Mobility Shift Assay testing interaction between probes from *BV677278* and PEA3. In lane 1, labeled probe alone migrated to the bottom of the gel (arrow B). When probe was incubated with human brain lysate, lane 2, a band appeared at the top of the gel (arrow A), indicating binding between probe and lysate (known as "shift"). Incubation of labeled probe and lysate with PEA3-specific antibody produced a competitive band (arrow B) that increased in intensity as the concentration of antibody increased (lanes 4-7). Incubation of labeled probe and lysate with either non-specific antibody (anti-Oct2a, lane 3) or non-specific probe (EBNA DNA, lane 10) failed to produce a similar band.

Quantitative RT-PCR

To avoid confusion, the following conventions of terminology apply: "wildtype" refers to cells that were not transfected with anti-*pea3* siRNA, whereas "naïve" refers to cells that were not treated with atRA (and thus not differentiated into neural-like cells).

P19 cells were treated with a 6-day course of 5 x 10^{-6} M all-*trans*-retinoic acid (atRA) to induce neural differentiation. To determine whether cells had in fact committed to a neural genetic profile, qRT-PCR was used to determine the relative levels

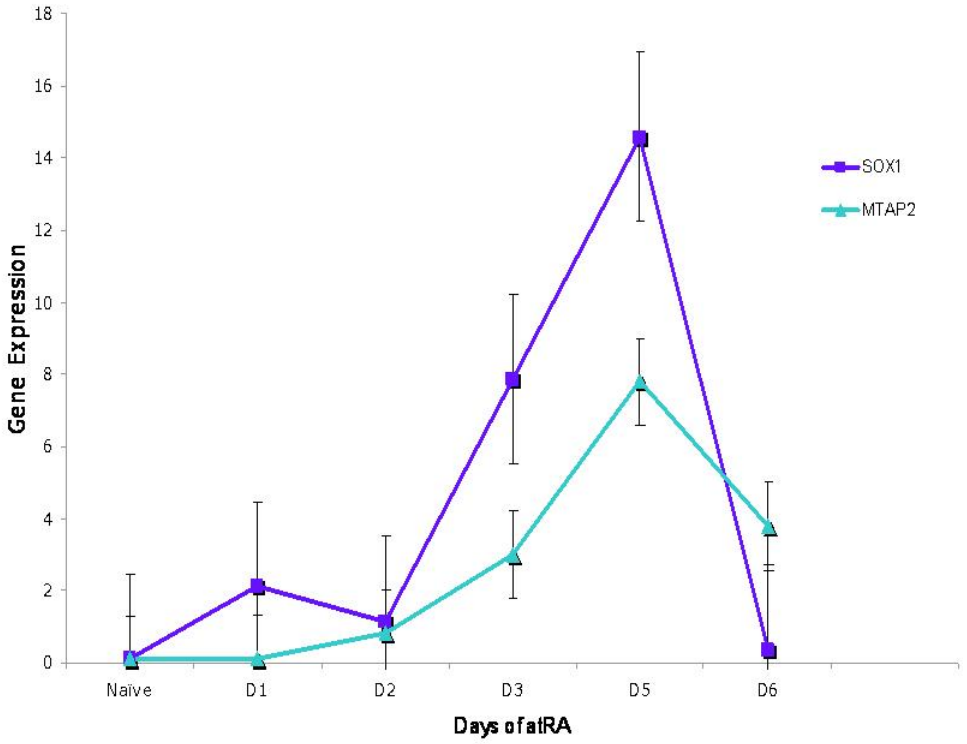

Figure 4. Expression of two neural markers, *sox1* and *mtap2*, in P19 cells during exposure to all-*trans*-retinoic acid, ±SEM. The expression of both genes was initially low, but rose with increasing time of exposure.

of *sox1*, a marker of early neural development, and *mtap2*, a marker of later neural development. Results are shown in Figure 4, above. The expression of *sox1* began to rise at atRA day 3 and peaked at day 5, with a marked decrease by day 6. The expression

of *mtap2* followed essentially the same profile, though levels declined less between days 5 and 6 than for *SOX1*.

qRT-PCR was then used to determine the relative expression of *dcdc2a*, *kiaa0319,* and *pea3* in the same P19 cells. The relative expression of each gene product was plotted against time treated with atRA to give a rough timeline of changes in

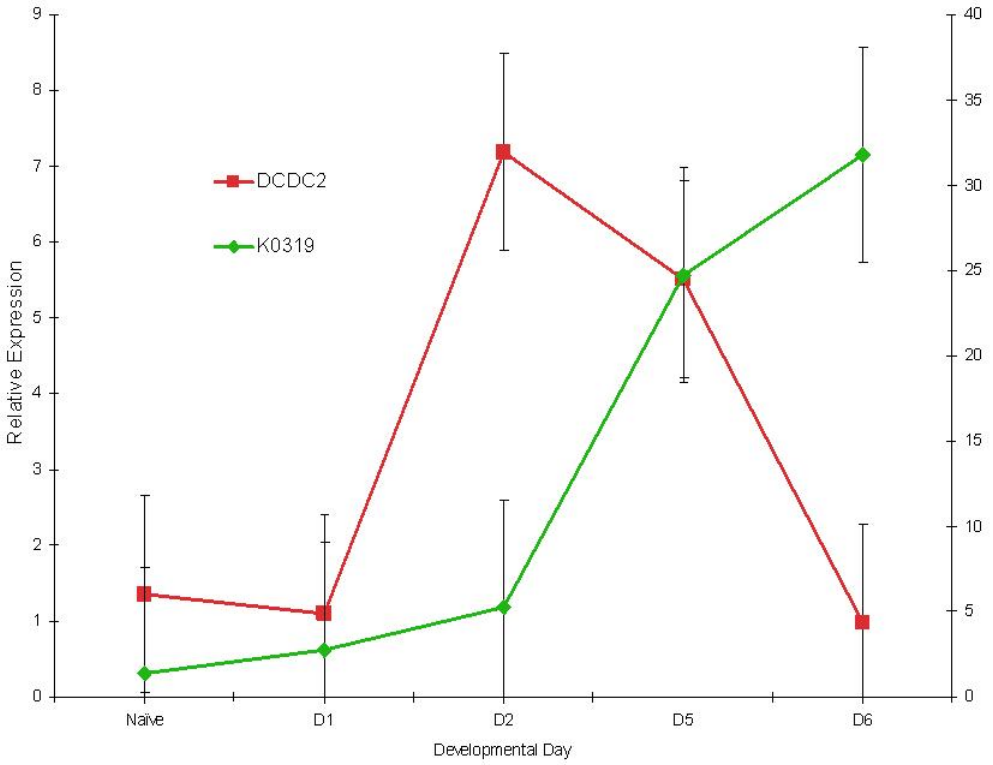

Figure 5. Expression of *dcdc2a* and *kiaa0319* during exposure to all-*trans*-retinoic-acid in P19 cells, ±SEM. *Dcdc2a* expression is initially low, then rises abruptly at day 2. *kiaa0319* expression is relatively stable until late in neural development. Left y-axis: *dcdc2a*; right y-axis: *kiaa0319.*

expression during neural differentiation. The results of these experiments are shown in Figure 5, above. In wildtype cells, the expression of *dcdc2a* was relatively low before the initiation of treatment with atRA; expression peaked at day 2 and returned to baseline levels by day 6. Expression of *kiaa0319* was also initially low in wildtype cells before the initiation of atRA, and then rose steadily, peaking at day 6. The expression *kiaa0319*

was much higher than that of *dcdc2*; for purposes of graphical representation, different scales are represented on two y-axes.

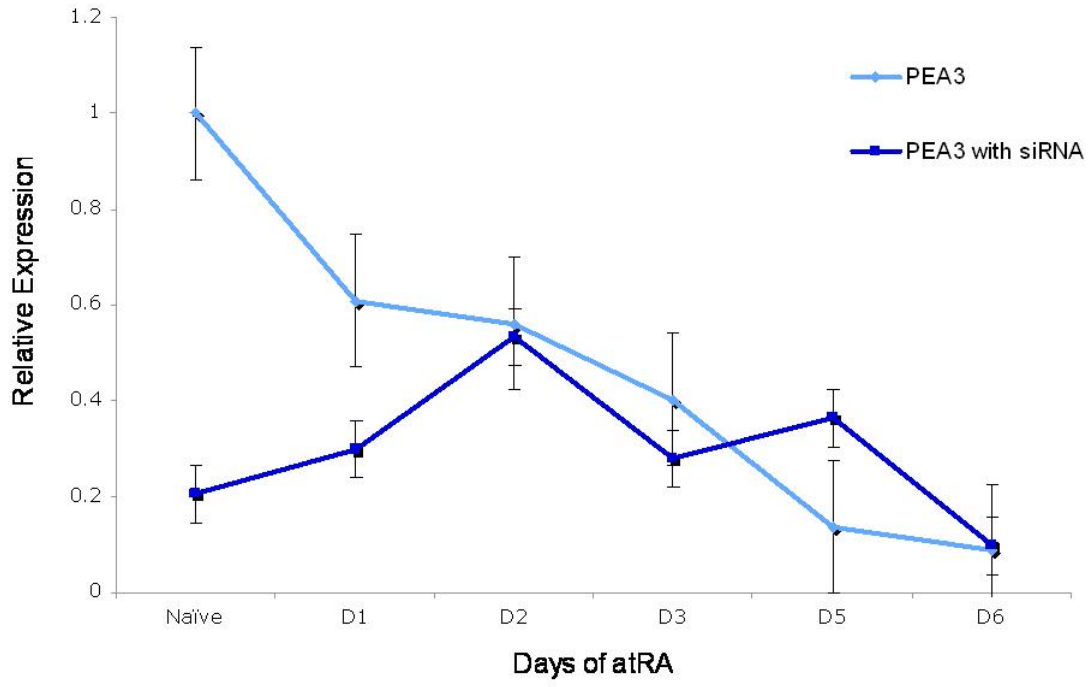

Figure 6. Expression of *pea3* in P19 cells during exposure to all-*trans*-retinoic acid, ±SEM. The light blue line represents *pea3* expression in wildtype P19s, whereas the dark blue line represents expression in cells exposed to *pea3* siRNA. Knockdown of *pea3* ceased to be effective at day 2.

To determine the success of *pea3* knockdown, relative *pea3* mRNA expression was plotted versus time for wildtype (non-transfected) and experimental (transfected with *pea3* siRNA) P19 cells. The results are shown in Figure 6, above. In wildtype cells, *pea3* expression was highest in the naïve state (not exposed to atRA), and then steadily declined with increasing exposure to atRA, consistent with assays of *pea3* expression during mouse cell differentiation described by Xin et al.[38] Knockdown of *pea3* was successful in naïve cells and after one day of atRA treatment, but with longer exposure to

atRA (two days and beyond), knockdown ceased to be effective. This result was consistent across several repetitions of the experiment.

The expression of each gene in wildtype cells was then compared to the expression in cells that had been treated with anti-*pea3* siRNA. The results of these experiments are shown in Figures 7 (*dcdc2a*) and 8 (*kiaa0319*), below. In naïve cells, expression of *dcdc2a* was significantly higher in *pea3* knockdown cells than in wildtype cells.

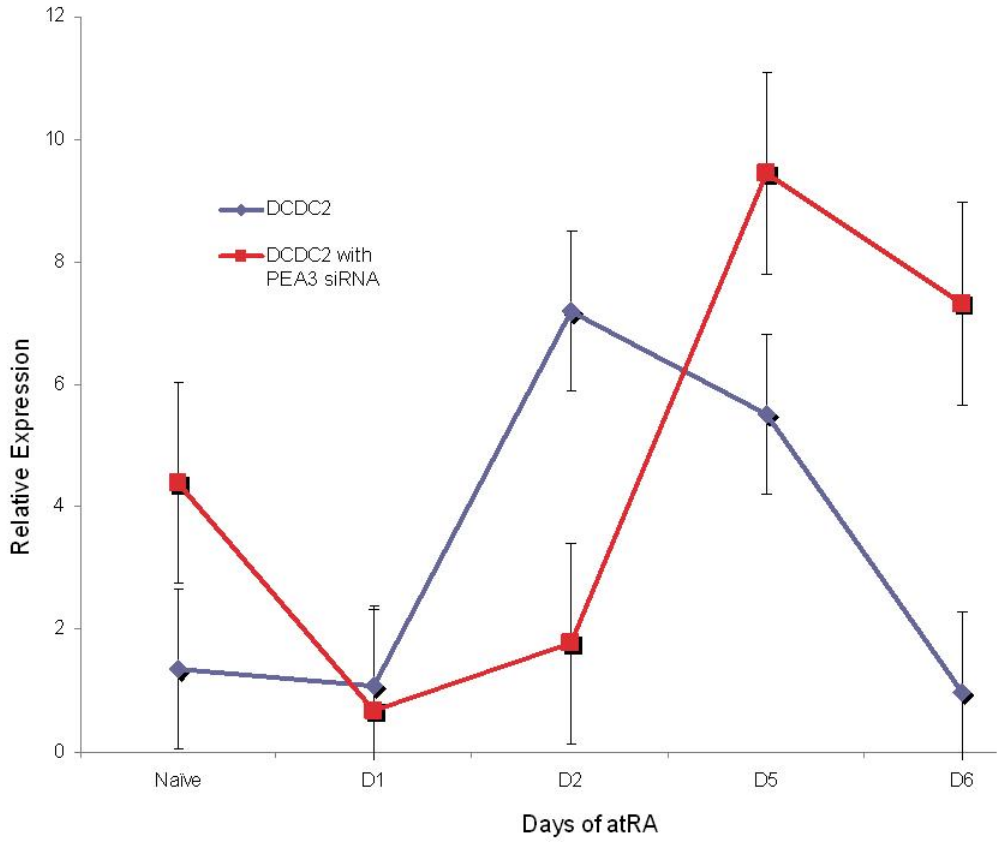

Figure 7. Change in *dcdc2a* expression with *pea3* knockdown, ±SEM. In wildtype P19 cells, *dcdc2a* surges after two days of treatment with all-*trans*-retinoic acid, then returns to baseline. This surge is delayed until day 5 in P19 cells that have been transfected with anti-*pea3* siRNA.

On day 1 of neural differentiation, however, there was no significant difference in *dcdc2a* expression between wildtype and knockdown cells, even though the knockdown was successful. On day 2 of differentiation, the expression of *dcdc2a* was significantly

different in the two populations of cells, but the expression of *pea3* did not differ significantly (i.e. the knockdown was not successful). The expression of *kiaa0319* was not significantly different between the two groups on any of these early days of differentiation.

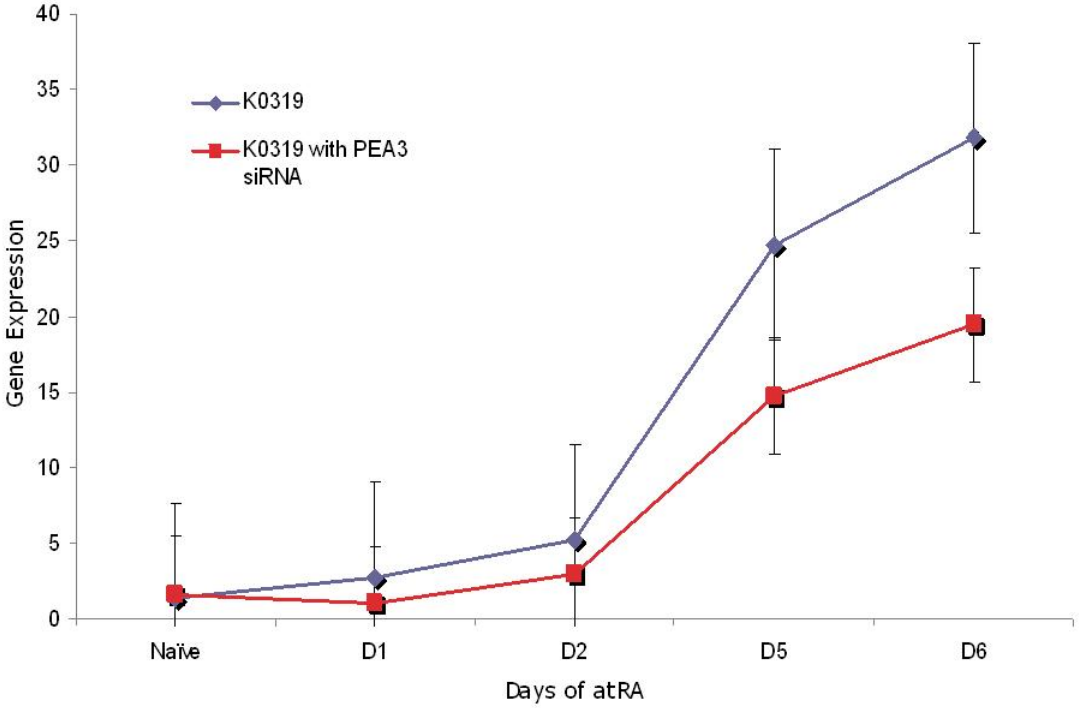

Figure 8. Change in *kiaa0319* expression with *pea3* knockdown, ±SEM. In wildtype P19 cells, *kiaa0319* remains at low baseline levels until after 2 days of treatment with all-*trans*-retinoic acid, when expression begins to increase. There is no significant difference in expression when *pea3* is knocked down.

Discussion

Our understanding of the genetics underlying reading disability has improved greatly over the last century. We have moved from a general observation that reading difficulties often appear to segregate in families to the identification of specific candidate genes postulated to exert effects on brain development. Yet despite these advances, many questions remain. Exactly what is the function of these genes at the molecular level? More broadly, how do these effects work together at the levels of brain structure, function, and development to influence such a complex phenotype as reading ability?

Crucial to answering these questions is an understanding of how the genes implicated in reading ability are regulated. As our comprehension of genetics has improved, we have come to realize that many diseases are not simply the results of mutations affecting the structure of a gene product *per se*, but are rather due to subtle changes in where, when, or how much of the gene product is made. This is certainly implied by polymorphisms identified in studies of RD. In *DCDC2* and *KIAA0319*, the two genes for which the most credible evidence of association with RD exists, candidate polymorphisms have only been identified in non-coding areas. Sequence changes in these regions could then change the binding affinity of transcription factors, either increasing or diminishing their ability to affect transcription of a target gene.

We were particularly interested in whether *BV677278*, the highly polymorphic region in intron 2 of *DCDC2* that is associated with RD, might harbor such a transcription factor binding site. This study sought to identify which TF(s) might bind in this region, and to describe the effects any candidate TFs might have on the transcription of downstream gene targets, namely *DCDC2* and *KIAA0319*.

The rough pathway for discovery and verification of a transcription factor is shown in Figure 9, below. Speaking generally, after identification of candidate TFs, each candidate must be tested for binding to the DNA sequence under investigation (the "probe"). After a candidate has been shown to bind the target sequence, it must then be shown to affect the expression of the gene in question. These were the criteria we decided any viable candidate TF must meet.

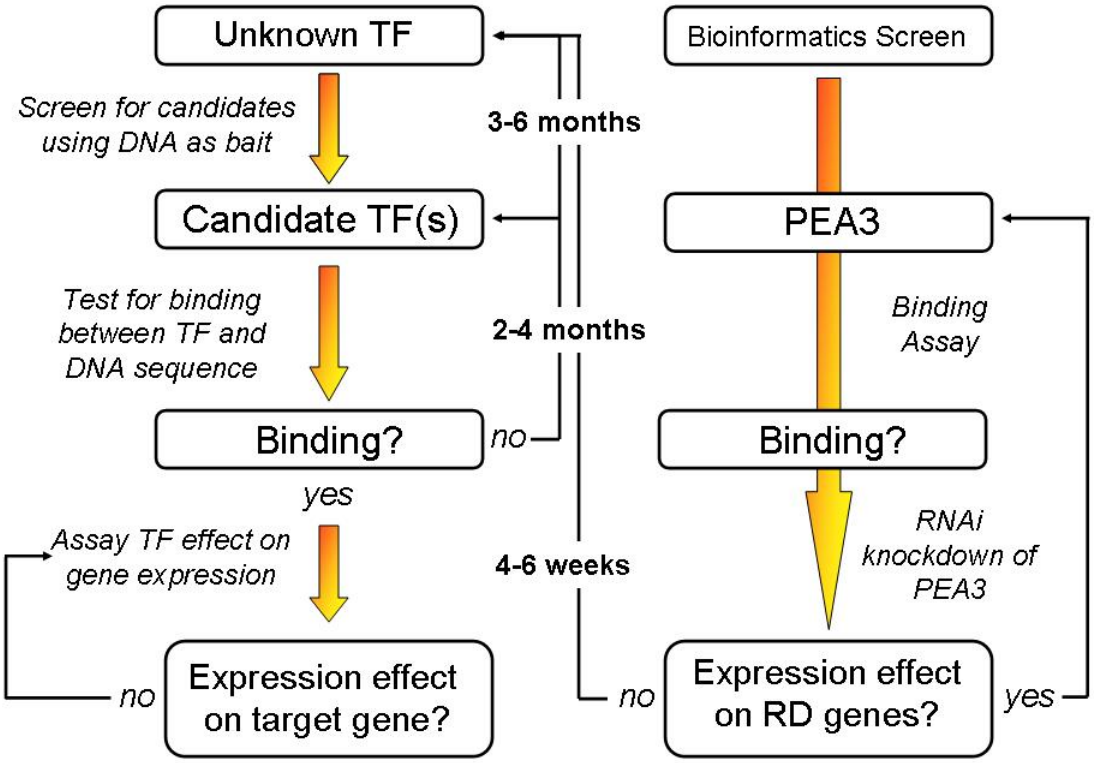

Figure 9. Hypothetical (left column) and actual (right column) algorithms for evaluating candidate transcription factors. After an initial screen to find initial pool of candidates, each must be tested for its ability to bind the specific DNA sequence. Once this has been shown, the TF must also be shown to affect the expression of target genes. If either the EMSA or the siRNA studies of PEA3 had been unsuccessful, the appropriate next step would be to return to broad screen, such as a yeast one-hybrid.

There are thousands of known transcription factors, so our first problem was finding a way to eliminate most of these and focus our investigation on a manageable number of candidates. The classic method of identifying candidate transcription factors

is through an assay known as a yeast one-hybrid. In this assay, a pre-determined library of transcription factors is screened for the ability to bind a particular DNA sequence. Only yeast clones in which a particular candidate TF stably binds this sequence will grow on selective media. Unfortunately, the yeast one-hybrid often returns TFs that appear to bind the target region in the assay but do not display regulatory activity *in vivo.*[39, 40] Furthermore, it is a labor-intensive assay that often yields over ten candidates, all of which must be tested for appropriate *in vivo* behavior.

A less exhaustive but quicker method is a bioinformatics approach.[41, 42] The Transcription Element Search Software (TESS) is an online database sponsored by the University of Pennsylvania that allows the input of DNA sequences and returns all known transcription factors with binding sites in the queried sequence. *BV677278* is over 2 Kb long, and interrogating the entire sequence would likely yield an unacceptably high number of possible TFs. Since it is comprised of compound short tandem repeats, however, we realized that there were only four small regions whose sequences were altered by the expansion or contraction of these repeats. We reasoned that if *BV677278* indeed affects reading ability through the binding of transcription factors, those factors most likely bind in the areas that are disrupted by dyslexia-associated polymorphisms, so we felt confident using these four 22-nucleotide sequences as input for the TESS search.

The search returned a number of hits, but only two really viable possibilities. One, NF-1 (Nuclear Factor-1), is ubiquitously expressed in humans, but its role has so far mainly only been shown to be important in adenoviruses,[37] so it seemed unlikely to be contributing to reading disability. The other, PEA3, was intriguing. Most studies of

PEA3 so far have been focused on its role in cancer; it has been noted, for example, to be overexpressed in breast tumors.[43]

There is evidence, however, that PEA3 likely plays a role in neural development as well. First, it is expressed at high levels in human brain, particularly in the telencephalon,[44] where activities of higher cognition take place. Second, there is evidence that it is specifically important in guiding neural precursor cells from their origin in the periventricular region of brain to their ultimate destination in the outer layers of the cerebral cortex. In 2004, a Japanese group created a dominant negative mutation of PEA3 in mouse neuronal precursors by transfecting them with a vector that overexpressed only the DNA-binding portion of the protein, thereby blocking functional, wildtype PEA3.[34] When labeled with a fluorescent marker and transplanted into the brains of developing mice *in utero*, the PEA3-mutant neuronal precursors did not migrate to the outer layers of the cortex, as they normally should have, but instead halted in the intermediate zone of the brain. Their axonal projections displayed aberrant architecture as well.

This morphology is strikingly reminiscent of the aberrant migration pattern of *DCDC2*-knockdown cells when implanted into developing rat brain. These experiments were conducted in cooperation between our lab and the LoTurco lab at the University of Connecticut and reported by Meng et al in 2005. Although only a circumstantial observation, the similar laminar patterning in *DCDC2* knockdowns and *PEA3* dominant negatives raises the possibility that PEA3 may bind *DCDC2* (or, perhaps, *KIAA0319*, which is nearby) and upregulate its expression. When PEA3 function is ablated, as in the Japanese experiments, the resultant decrease in *DCDC2* or *KIAA0319* expression might

be similar to the decrease that results when the genes are knocked down directly, as in our experiments. In either case, whatever effect the downstream gene product (i.e. protein) has on directing neuronal migration is abolished, resulting in an abnormal architectural phenotype.

Our hypothesis, prior to the experiments described here, was that first, PEA3 could bind to *BV677278*, and second, that it affects the expression of *DCDC2* or *KIAA0319*. The results of the experiments appear to support both claims.

The first question—whether PEA3 can bind to *BV677278*—was approached by supershift assay, a two-step process. The first step was simply to show that the 22-nucleotide probes taken from *BV677278* bind human brain nuclear lysate, which would by inference imply that the probes bind to some transcription factor found in human brain. This result, known as "shift," is evident by comparing lanes 1 and 2 of figure 2. The naked probe, labeled with horseradish peroxidase, quickly migrated to the bottom of the native polyacrylamide gel during electrophoresis. When nuclear lysate was added, the probe's mobility was greatly retarded, and the dominant band appeared near the top of the gel.[45, 46] Our results consistently showed strong shift for all four 22-nt probes, indicating that these sections of *BV677278* indeed bind a transcription factor found in human brain.

The second step of the assay—the "supershift" portion—was designed to identify the specific transcription factor in the soup of lysate that caused the change in mobility of the labeled probe during the shift assay. In this step, an antibody to PEA3 was incubated in the same reaction mixture as the labeled probe and the lysate. The idea here was that if PEA3 is the actual transcription factor causing the original shift, the interaction of the

antibody and PEA3 will further change the mobility of the entire complex, though the direction of migration depends on the nature of the exact interaction. If the binding of the antibody and the probe to the TF is "cooperative"—that is, if the two bind to different sites on the TF—the entire complex would be larger than the shift complex and would migrate relatively higher on the gel. If the binding of the antibody and the probe is "competitive," i.e. the two attempt to bind at the same site on the TF, the addition of the antibody would result in more unbound labeled probe than in the shift step, and a strong band would reappear at the bottom of the gel. The details of this assay are shown in Figure 10, next page.

Our results identified a competitive interaction between anti-PEA3 and probe 1 from *BV677278*. Adding an increasing amount of anti-PEA3 to the mixture of probe and lysate produced an increasingly intense band at the bottom of the gel (seen in lanes 4-7 of figure 2). Reactions using either a control probe (DNA encoding the Epstein-Barr Nuclear Antigen, EBNA) or control antibody (human anti-Oct2a) showed the same pattern as that seen with probe and nuclear lysate only, implying that the interaction between anti-PEA3 and the target antigen was specific. The fact that only the combination of the experimental probe, nuclear lysate, and anti-PEA3 created supershift implies that the target antigen for the probe is, in fact, PEA3.

One problem with the supershift assay is its artificiality. Manipulating the parameters of the binding reaction—changing the pH, adding monovalent and divalent ions, raising or lowering the reaction temperature—affect the binding avidity of the components.[47] In this case, adding glycerol and potassium and incubating the antibody with the lysate for two hours before adding the labeled probe produced supershift, while

other sets of reaction conditions did not. Given the delicacy of this interaction, it is unclear how applicable these results are to understanding how PEA3 and *BV677278* actually interact in a physiologic system.

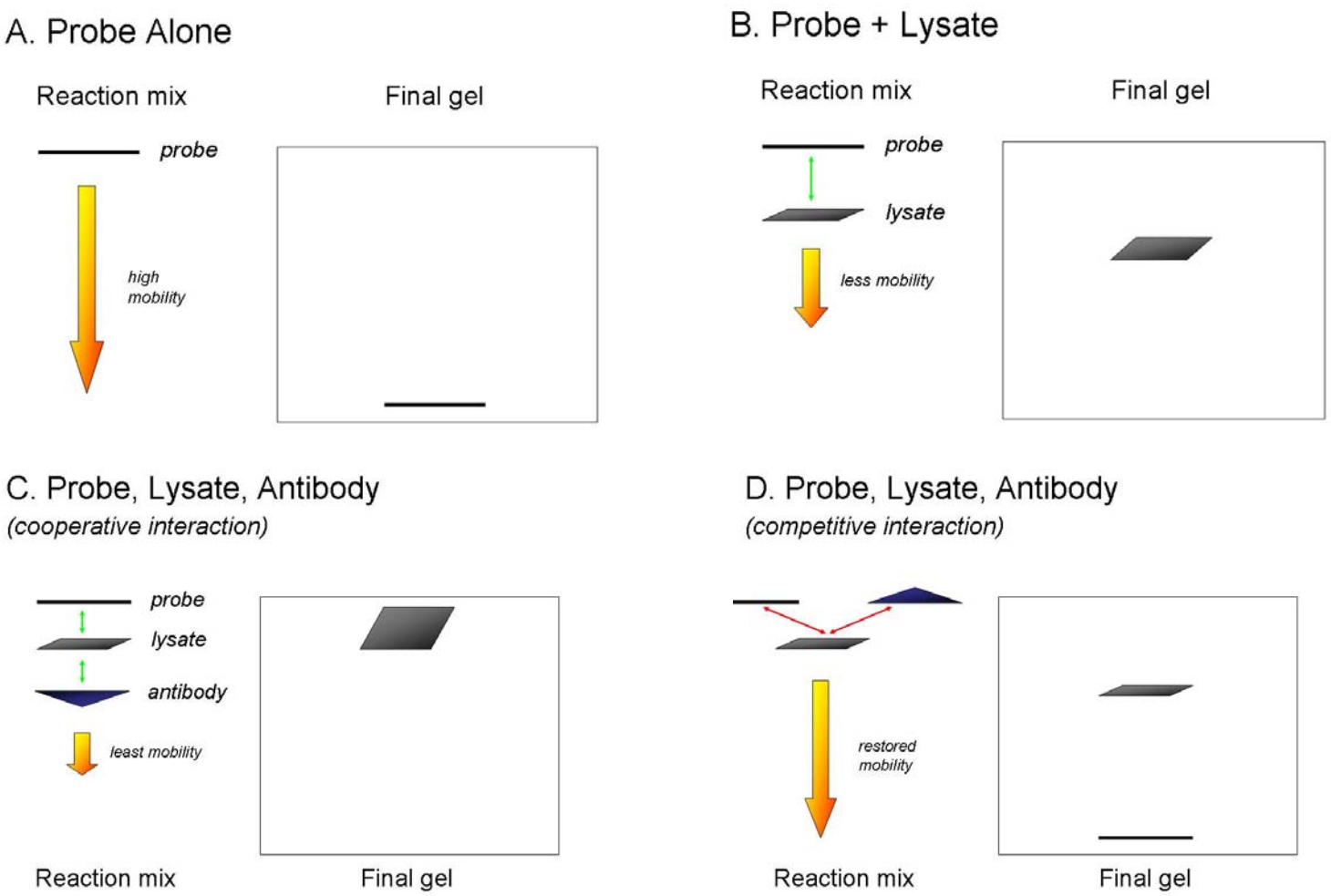

Figure 10. Details of EMSA supershift assays. **A.** When labeled probe is incubated alone, its high mobility leads to a band at the bottom of the gel. **B.** When probe interacts with lysate (*green arrow*), the mobility of the complex is much lower than probe alone, and a large band appears higher on the gel. **C.** If antibody specific to the correct transcription factor is added, and if this antibody binds the factor at different site than the probe (*green arrows*), the entire complex will have greatly reduced mobility and a very large band will appear at the top of the gel. **D.** If the antibody and probe bind at the same site (*red arrows*), the competition for binding will lead to an increase in the amount of unbound labeled probe, restoring the small band at the bottom of the gel. Because some probe will also bind the lysate, a larger band of high molecular weight will usually remain.

In addition, an ideal experiment would have shown supershift using all four probes, since all four contain the canonical binding site for PEA3. Unfortunately, human brain nuclear lysate is relatively difficult to come by, and our supplier quit stocking it shortly after we had completed the assay with probe 1. Although there are other suppliers selling similar product, we would ideally use lysate from the same lot for all experiments, to control for variation in lysate contents due to genetic heterogeneity of the donors.

More physiologic assays than the EMSA exist. One would be a Chromatin ImmunoPrecipitation (ChIP), which, unfortunately, limited time did not allow us to perform. In a ChIP, the chromatin of a cell population is cross-linked, usually with formaldehyde, to preserve all bound transcription factors, and the cells are lysed. The chromatin-TF mixture is then applied to a bead matrix on which has been fixed an antibody to the target antigen (e.g. PEA3). After the pull-down, the crosslinking is reversed and the recovered DNA is subjected to PCR with the appropriate primers (e.g., to *BV677278* or fragments thereof).[48] A successful ChIP using this strategy would show not only that *BV677278* and PEA3 *can* bind to each other under artificial conditions, but that they actually *are* bound to each other in live cells. Completing this assay would require careful selection of an appropriate cell line (e.g. neurons or neuronal precursors), since the interaction is likely tissue-specific. Furthermore, since the relationship may be temporally specific, it may be necessary to study differentiating cells at distinct time points, or to use primary cells harvested from animals of different ages.

After addressing the question of binding, we turned to the second question—whether PEA3 expression affects that of either *DCDC2* or *KIAA0319*—by asking the inverse: if PEA3 expression is abolished, does the expression of either target gene

change? We identified at least two ways to abolish PEA3 expression. The first, described by Hasegawa et al in 2005, would be to engineer an expression vector that produces only the DNA-binding portion of PEA3, but not the functional domain. The product of such a vector would thus compete directly with wildtype PEA3 for binding to target sites, but would be essentially non-functional. If produced in saturating quantities (for instance, under a constitutive promoter), the non-functional PEA3 would block almost all wildtype protein from binding, effectively abolishing its function.

The other strategy would be to use small inhibitor RNAs (siRNA, also known as RNAi knockdown), to prevent *PEA3* RNA from being translated into protein. In this technique, an siRNA containing a sequence complementary to a specific region of the target RNA is transfected into cells and, after binding its target, signals the cells to degrade the entire complex.[49, 50] One potential drawback to using siRNA to specifically block the production of transcription factors is that some TFs are "pre-manufactured" in large quantities and stored within the nucleus until needed.[51] If this were the case with our target TF, blocking the translation of RNA to protein would not affect the molecules of TF that have already been produced. On the other hand, the technique is relatively easy, as commercially produced anti-*PEA3* siRNA is readily available, whereas the production of a dominant negative vector like that described above would take several months of molecular genetics. Thus, we elected to try the siRNA approach first.

We also needed to find an appropriate biologic model, since we cannot easily analyze the molecular interactions of the developing human brain. Because *PEA3*, *DCDC2*, and *KIAA0319* are all highly conserved, we decided it would be acceptable to begin our studies in cells derived from lower mammals, e.g. mice or rats. However, since

we were interested in neural-specific interactions, we reasoned that we would need to study neural cell lines. Furthermore, these interactions presumably change during the course of brain development, so we would need to use a cell line that can differentiate from immature precursors into mature neurons. We chose cell lines for this purpose: mouse embryonic stem cells (mESCs)[52] and P19 cells, a line derived from a mouse embryonal carcinoma.[35, 53] The mESCs unfortunately yielded very small amounts for RNA and protein and were thus not well suited for our purposes. The data presented here thus come from P19s.

Our results from the siRNA knockdown of *PEA3* mRNA in P19s imply a relationship between PEA3 and *DCDC2* in early neural development. A similar relationship was not seen between PEA3 and *KIAA0319*. Before discussing these results in detail, it is important to note that the protocol for differentiating P19s into neural-like cells utilizes all-trans-retinoic acid (atRA), which is somewhat toxic to the cells and causes them to apoptose after several days of exposure. This may explain why the siRNA was only effective at knocking down PEA3 early in the differentiation timeline (i.e. days 0 and 1), but not at days 5 and 6.

Wildtype P19 cells initially express *DCDC2* at low levels. Upon exposure to atRA and the initiation of neural differentiation, *DCDC2* levels surge, reaching a peak at day 2 and returning to baseline by day 6 (a timepoint roughly analogous to late neural development in an animal model). The introduction of anti-PEA3 siRNA appears to delay this surge and decline until later in neural development. On the other hand, there is no such effect on *KIAA0319* expression, which in both wildtype and PEA3 knockdown cells remains low during early development and rises around day 5. One plausible

explanation for the lack of relationship between PEA3 and *KIAA0319* is that the mouse *KIAA0319* ortholog is located on a different chromosome than mouse *DCDC2*. This is quite different than the location of their human counterparts, which lie in very close proximity on chromosome 6 within the same dyslexia locus. If a later study were to show a regulatory relationship between PEA3 and *KIAA0319* in humans, this finding might imply the migration of *KIAA0319* into the vicinity of a new regulatory complex, which could arguably have contributed to the development of higher cognitive functions like reading.

The exact nature of the relationship between *DCDC2* and PEA3 is not clear from these experiments. Knocking down PEA3 in P19 cells appears to delay the surge in *DCDC2* expression by about two to three days in the atRA differentiation timeline. Furthermore, to achieve this change in *DCDC2* expression, it is sufficient to knock down PEA3 only in the first two days of differentiation. This implies that PEA3 normally plays some role in *DCDC2*'s upregulation and that, since the effect of knocking down *PEA3* mRNA is not immediate, the interaction between the two must either be indirect or take some time to manifest even under normal circumstances.

These results must be interpreted with caution and will require more complex physiologic studies to support them. In this study, the baseline gene expression of wildtype P19 cells is used as a negative control. To say with certainty that the effect on *DCDC2* expression is truly due to the knockdown of PEA3, however, one would like to see that cells transfected with a "dummy" siRNA, containing scrambled sequences that are not complimentary to any known RNA sequence, express *DCDC2* similarly to wildtype cells. We began our experiments with such a control, but the cells transfected

with that construct died early and yielded insufficient RNA for RT-PCR. Ideally, this entire experiment should be repeated, with the use of scrambled siRNA, rather than wildtype cells, as a negative control.

Although interrogation of interactions between a transcription factor and target genes by siRNA knockdown, termed "pathway studies," has precedent in the current literature, the specifics of our approach appear to be novel; we were unable to identify a study that examined the effect of TF knockdown on a handful of candidate genes during the course of tissue development. One recently published study, for instance, examined the effect of knocking down *NPAS2*, a TF that has been identified as a potential tumor suppressor, on the expression of genes known to repair DNA damage.[54] The study used techniques of siRNA transfection and qRT-PCR analysis similar to ours, but the authors also utilized gene chip technology to observe changes in the expression of thousands of genes, rather than the two targets our study identified. Other recent examples of pathway studies are similarly broad in scope.[55]

Examples of pathway studies with a narrower focus exist in the current literature as well, but many treat the interaction of transcription factor and targeted gene as a binary, on-off proposition, isolated in space and time. One recent study, for example, explored the relationship of the transcription factor PU.1 on the expression of a pathogen receptor gene, *LSECtin*, in human Kupffer cells.[56] The authors found that knockdown of PU.1 in Kupffer cells decreased *LSECtin* expression, which they felt showed a regulatory relationship between the two elements. Unlike our paradigm, this relationship was assumed to be immediate and was not assessed in the context of developing tissue. Several other examples of such studies can be found in the literature as well.[57, 58]

Finally, perhaps the broadest group of pathway studies examines the effect of transcription factor knockdown not on the expression of target genes, but on phenotypes. A recent study in MLL-fusion leukemia cells, for instance, showed that knockdown of the homeobox gene *Meis1* inhibited tumor growth.[59] Another study of neurogenesis in mice showed that knockdown of the transcription factors COUP-TFI and II prolonged the ability of embryos to generate new neurons.[60] The observation of a relationship between a transcription factor and a phenotype is valuable, but these studies are generally only the first step towards understanding the underlying mechanism behind this relationship, and further studies to identify intermediate effector genes will always be necessary to elucidate the entire pathway.

The value of our approach is twofold. First, it provides evidence for a pathway connecting PEA3 and an observed phenotype, aberrant neuronal migration, by examining the regulatory effect of PEA3 on a downstream gene, *DCDC2*. Although the details of the pathway are not yet fully understood (mainly, how does *DCDC2* affect neuronal migration at a molecular level?), these data begin to outline a rough framework on which the rest of the pathway may be based. Second, our approach appears to be novel in that it examines the effect of a transcription factor over the course of cell development and differentiation, not at a static point in space and time. Despite the limitations of our system, such an approach may be more physiologic than those employed by other studies, given its recognition that interactions within genetic regulatory networks are dynamic, and that perturbations in the networks may not have an immediately observable effect. Overall, our results imply that both normal expression PEA3, and a normal early surge in *DCDC2*, are necessary for normal neuronal migration, and that decreased expression of

either element is sufficient to distort the process (Figure 11, below). In particular, we emphasize our hypothesis that decreased PEA3 expression leads to aberrant migration through a delay in the *DCDC2* surge.

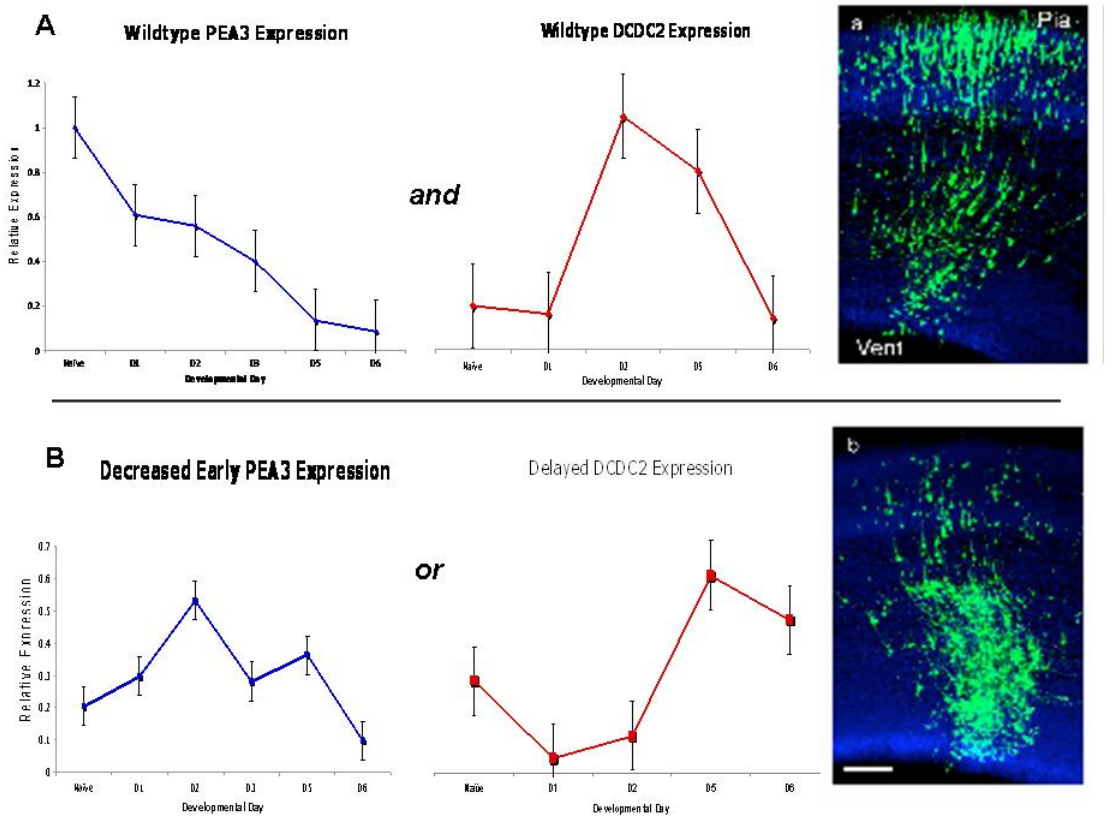

Figure 11. Proposed pathway linking PEA3, *DCDC2*, and neuronal migration. **A:** Both high early expression of PEA3 *and* a normal early surge of *DCDC2* expression, as seen in wildtype P19s, are necessary for normal neural migration to the outer layers of cortex. **B:** Either decreased early levels of PEA3 *or* delay of *DCDC2* surge is sufficient to distort neural migration and cause progenitors to arrest in the intermediate zone of the brain.

The results of this study imply the potential for more complex studies in an animal model. Such a study might begin by replicating the experiments of Hasegawa et al to show a migration defect in neural progenitors that have been transfected with a dominant-negative mutation for PEA3. The experiment would continue on to determine whether the same migration deficit can be replicated by knocking out the *BV677278*

region from implanted cells so that, presumably, PEA3 would be unable to bind *DCDC2* at this location. Finally, one would have to determine whether the migration deficit observed in PEA3-negative cells can be mitigated by the overexpression of *DCDC2* through the transfection of an expression vector under a strong promoter. This would be a complex and involved experiment that would require expertise in both molecular genetics (for the development of the PEA3-negative vector and the overexpression vectors) and animal experiments (for the implantation of transfected cells into the brains of fetal mice or rats).

Overall, these experiments imply that our hypothesis was correct. *BV677278* appears to be a critical regulatory element for *DCDC2*. This effect appears to be modulated by PEA3, which binds *BV677278* and increases the expression of *DCDC2*, with little evidence to suggest that, at least in mouse cells, it affects the expression of *KIAA0319*. Further experiments are necessary to confirm this and to prove, by extension, that gene regulation governed by PEA3 is in part responsible for aspects of brain development that regulate reading ability. One day, confirming these types of relationships between genotype and phenotype may enable us to identify children at risk of becoming poor readers and to institute early educational interventions that can improve their performance before they fall behind their normal-reading peers. Before this can occur, however, we need a more well-founded and complete understanding of these interactions, which will further our comprehension of the underlying genetic structure of reading.

References

1. Hinshelwood J. Congenital Word-Blindness. Chicago: Medical Book Co.; 1917.
2. Collins FS. Positional cloning: let's not call it reverse anymore. Nature genetics 1992;1(1):3-6.
3. Farrer L. Collection of Clinical and Epidemiological Data for Linkage Studies. Current Protocols in Human Genetics 2004;S40:1.. - ..17.
4. Rice TK, Borecki IB. 4 Familial resemblance and heritability. In: Advances in Genetics: Academic Press; 2001:35-44.
5. DeFries JC, Fulker DW, LaBuda MC. Evidence for a genetic aetiology in reading disability of twins. Nature 1987;329(6139):537-9.
6. Pericak-Vance MA. Analysis of Genetic Linkage Data for Mendelian Traits. Current Protocols in Human Genetics 1996;S9:1.4.1 - .4.21.
7. Gregory S, and John Gilbert. Strategies for Genotype Generation. Current Protocols in Human Genetics 2005;S47:1.3.1 - .3.16.
8. Borecki IB, Suarez BK. Linkage and Association: Basic Concepts. Advances in Genetics 2001;42:45-66.
9. Grigorenko EL, Wood FB, Meyer MS, et al. Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. American Journal of Human Genetics 1997;60(1):27-39.
10. Wigg KG, Couto JM, Feng Y, et al. Support for EKN1 as the susceptibility locus for dyslexia on 15q21. Mol Psychiatry 2004;13:13.
11. Gayán J, Smith SD, Cherny SS, et al. Quantitative-trait locus for specific language and reading deficits on chromosome 6p. American Journal of Human Genetics 1999;64(1):157-64.
12. Nopola-Hemmi J, Taipale M, Haltia T, Lehesjoki AE, Voutilainen A, Kere J. Two translocations of chromosome 15q associated with dyslexia. J Med Genet 2000;37(10):771-5.
13. Fisher SE, Francks C, Marlow AJ, et al. Independent genome-wide scans identify a chromosome 18 quantitative- trait locus influencing dyslexia. Nature genetics 2002;30(1):86-91.
14. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. The New England journal of medicine 2007;357(5):443-53.
15. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447(7148):1087-93.
16. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 2007;316(5829):1341-5.
17. Hafler DA, Compston A, Sawcer S, et al. Risk alleles for multiple sclerosis identified by a genomewide study. The New England journal of medicine 2007;357(9):851-62.
18. Zondervan KT, Cardon LR. Designing candidate gene and genome-wide case-control association studies. Nature protocols 2007;2(10):2492-501.
19. Taipale M, Kaminen N, Nopola-Hemmi J, et al. A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. Proc Natl Acad Sci U S A 2003;100(20):11553-8.
20. Meng H, Smith SD, Hager K, et al. DCDC2 is associated with reading disability and modulates neuronal development in the brain. Proc Natl Acad Sci U S A 2005;102(47):17053-8.

21. Cope N, Harold D, Hill G, et al. Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. Am J Hum Genet 2005;76(4):581-91.
22. Hannula-Jouppi K, Kaminen-Ahola N, Taipale M, et al. The axon guidance receptor gene ROBO1 Is a candidate gene for developmental dyslexia. PLoS Genet 2005;1(4):e50.
23. Harold D, Paracchini S, Scerri T, et al. Further evidence that the KIAA0319 gene confers susceptibility to developmental dyslexia. Mol Psychiatry 2006.
24. Bates TC, Luciano M, Castles A, Coltheart M, Wright MJ, Martin NG. Replication of reported linkages for dyslexia and spelling and suggestive evidence for novel regions on chromosomes 4 and 17. Eur J Hum Genet 2007;15(2):194-203.
25. Paracchini S, Steer CD, Buckingham LL, et al. Association of the KIAA0319 dyslexia susceptibility gene with reading skills in the general population. The American journal of psychiatry 2008;165(12):1576-84.
26. Paracchini S, Thomas A, Castro S, 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):1659-66.
27. Velayos-Baeza A, Toma, C., Paracchini, S., Monaco, A. P. The Dyslexia-Associated Gene Kiaa0319 Encodes Highly N- and O-Glycosylated Plasma Membrane and Secreted Isoforms. Hum Mol Genet 2007.
28. Meng H, Hager K, Held M, et al. TDT-association analysis of EKN1 and dyslexia in a Colorado twin cohort. Human genetics 2005;118(1):87-90.
29. Marino C, Giorda R, Lorusso ML, et al. A Family-Based Association Study Does Not Support DYX1C1 on 15q21.3 as a Candidate Gene in Developmental Dyslexia. European Journal of Human Genetics 2005:1-9.
30. Wang Y, Paramasivam M, Thomas A, et al. DYX1C1 functions in neuronal migration in developing neocortex. Neuroscience 2006;143(2):515-22.
31. Nopola-Hemmi J, Myllyluoma B, Haltia T, et al. A dominant gene for developmental dyslexia on chromosome 3. J Med Genet 2001;38(10):658-64.
32. Schumacher J, Anthoni H, Dahdouh F, et al. Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. Am J Hum Genet 2006;78(1):52-62.
33. Francis F, Koulakoff A, Boucher D, et al. Doublecortin is a developmentally regulated, microtubule-associated protein expressed in migrating and differentiating neurons. Neuron 1999;23(2):247-56.
34. Hasegawa H, Ashigaki S, Takamatsu M, et al. Laminar patterning in the developing neocortex by temporally coordinated fibroblast growth factor signaling. J Neurosci 2004;24(40):8711-9.
35. Jones-Villeneuve EM, McBurney MW, Rogers KA, Kalnins VI. Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glial cells. The Journal of cell biology 1982;94(2):253-62.
36. Cai C, Thorne J, Grabel L. Hedgehog serves as a mitogen and survival factor during embryonic stem cell neurogenesis. Stem cells (Dayton, Ohio) 2008;26(5):1097-108.
37. Gronostajski RM, Nagata K, Hurwitz J. Isolation of human DNA sequences that bind to nuclear factor I, a host protein involved in adenovirus DNA replication. Proc Natl Acad Sci U S A 1984;81(13):4013-7.

38. Xin JH, Cowie A, Lachance P, Hassell JA. Molecular cloning and characterization of PEA3, a new member of the Ets oncogene family that is differentially expressed in mouse embryonic cells. Genes & development 1992;6(3):481-96.
39. Sieweke M. Detection of transcription factor partners with a yeast one hybrid screen. Methods in molecular biology (Clifton, NJ 2000;130:59-77.
40. Vidal M, Legrain P. Yeast forward and reverse 'n'-hybrid systems. Nucleic acids research 1999;27(4):919-29.
41. Taverner NV, Smith JC, Wardle FC. Identifying transcriptional targets. Genome biology 2004;5(3):210.
42. Wang JC. Finding primary targets of transcriptional regulators. Cell cycle (Georgetown, Tex 2005;4(3):356-8.
43. Myers E, Hill AD, Kelly G, McDermott EW, O'Higgins NJ, Young LS. A positive role for PEA3 in HER2-mediated breast tumour progression. British journal of cancer 2006;95(10):1404-9.
44. Chotteau-Lelievre A, Dolle P, Peronne V, Coutte L, de Launoit Y, Desbiens X. Expression patterns of the Ets transcription factors from the PEA3 group during early stages of mouse development. Mechanisms of development 2001;108(1-2):191-5.
45. Hellman LM, Fried MG. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nature protocols 2007;2(8):1849-61.
46. Lin JJ, Grosskurth SE, Harlan SM, Gustafson-Wagner EA, Wang Q. Characterization of cis-regulatory elements and transcription factor binding: gel mobility shift assay. Methods in molecular biology (Clifton, NJ 2007;366:183-201.
47. Siu FK, Lee LT, Chow BK. Southwestern blotting in investigating transcriptional regulation. Nature protocols 2008;3(1):51-8.
48. Mukhopadhyay A, Deplancke B, Walhout AJ, Tissenbaum HA. Chromatin immunoprecipitation (ChIP) coupled to detection by quantitative real-time PCR to study transcription factor binding to DNA in Caenorhabditis elegans. Nature protocols 2008;3(4):698-709.
49. Kim D, Rossi J. RNAi mechanisms and applications. BioTechniques 2008;44(5):613-6.
50. Gomase VS, Tagore S. RNAi--a tool for target finding in new drug development. Current drug metabolism 2008;9(3):241-4.
51. Zaidi SK, Young DW, Javed A, et al. Nuclear microenvironments in biological control and cancer. Nature reviews 2007;7(6):454-63.
52. Aubert J, Stavridis MP, Tweedie S, et al. Screening for mammalian neural genes via fluorescence-activated cell sorter purification of neural precursors from Sox1-gfp knock-in mice. Proc Natl Acad Sci U S A 2003;100 Suppl 1:11836-41.
53. McBurney MW, Jones-Villeneuve EM, Edwards MK, Anderson PJ. Control of muscle and neuronal differentiation in a cultured embryonal carcinoma cell line. Nature 1982;299(5879):165-7.
54. Hoffman AE, Zheng T, Ba Y, Zhu Y. The circadian gene NPAS2, a putative tumor suppressor, is involved in DNA damage response. Mol Cancer Res 2008;6(9):1461-8.
55. Miura H, Tomaru Y, Nakanishi M, Kondo S, Hayashizaki Y, Suzuki M. Identification of DNA regions and a set of transcriptional regulatory factors involved in transcriptional regulation of several human liver-enriched transcription factor genes. Nucleic acids research 2008.

56. Dominguez-Soto A, Aragoneses-Fenoll L, Gomez-Aguado F, et al. The pathogen receptor liver and lymph node sinusoidal endotelial cell C-type lectin is expressed in human Kupffer cells and regulated by PU.1. Hepatology (Baltimore, Md 2009;49(1):287-96.
57. Xu B, Yang WH, Gerin I, Hu CD, Hammer GD, Koenig RJ. DAX-1 and Steroid Receptor RNA Activator (SRA) Function as Transcriptional Coactivators for Steroidogenic Factor-1 in Steroidogenesis. Molecular and cellular biology 2009.
58. Wang X, Fukuda T, Li W, et al. Requirement of Fut8 for the expression of vascular endothelial growth factor receptor-2: a new mechanism for the emphysema-like changes observed in Fut8-deficient mice. Journal of biochemistry 2009.
59. Kumar AR, Li Q, Hudson WA, et al. A role for MEIS1 in MLL-fusion gene leukemia. Blood 2008.
60. Naka H, Nakamura S, Shimazaki T, Okano H. Requirement for COUP-TFI and II in the temporal specification of neural stem cells in CNS development. Nature neuroscience 2008;11(9):1014-23.

Appendix 1. Transcription Factors returned by TESS Search

(Transcription Element Search Software: http:// www.cbil.upenn.edu/cgi-bin/tess/tess?RQ=WELCOME)

Nucleic Acid Codes: A, Adenine; G, Guanine; T, Thymine; C, Cytosine; Y, Pyrimidine (T/C); R, Purine (A/G); M, A/C; S, C/G; W, A/T; N, any.

Factor	Consensus Sequence	Known Species	Known Tissue	Molecular Function
PEA3	AGGAAR	Human, mouse, rat, pig	Brain, epidermis, lung	Interacts with oncogenes
NF-1	TGGCA	Human; Adenovirus	Ubiquitous	Enhances viral replication
TCF-1	MAMAG	Human	T cells, intestinal and mammary epithelia	Multiple downstream targets in T-cells
LEF-1	MAMAG	Human, many	T-cells	Member of Wnt cascade
c-ETS-1	SMGGAWGY	Human, mouse, chick, xenopus	T-cells	Unknown; levels highest in resting T-cells

Factor	Consensus Sequence	Known Species	Known Tissue	Molecular Function
YY1	SSGCCATCTTSNCTS	Human, mouse	Skeletal Muscle, other?	Catecholamine synthesis?
Elf-1	GAAGAGGAAAAA	Humans, Mouse	Lymphoid	Unknown; expressed in activated T-cells
PU.1	AGAGGAACT	Human, Mouse, Chick, Yeast	Lymphoid, Liver, Testes	Interferes with erythroblast maturation
NP-TCII	ANANTTTCC	Human, Mouse	Lymphoid	Unknown
GT-IIA	AGCTGGTTCTTTCC	Human	HeLa, Molt-4, BJA-B	Unknown
GAL4	CGGAGGACAGTACTCCG	Yeast	Yeast	Mediates galactose response
Dof2	AAAG	Maize	Leaves	Promoter-specific actions

Factor	Consensus Sequence	Known Species	Known Tissue	Molecular Function
MNB1a	AAAG	Maize	Leaves, stems, roots	Promoter-specific actions
Delta Factor	SSGCCATCTTSNCTS	Mouse	Brain	Activates ribosomal genes

Appendix 2

Notes on Conventions and List of Abbreviations

By convention, gene names and mRNA transcripts are listed in italics. Protein names, even if the same as gene names, are not italicized. Human genes and proteins are capitalized, whereas mouse genes and proteins are in lower case. In some cases, human genes and their mouse orthologs have different official names; in these cases, this difference is noted, but the mouse ortholog is referred to by the human name to avoid confusion.

Genes, Proteins, and Regions

1. *DCDC2*: Candidate gene for dyslexia, on 6p22 in the DYX2 locus.
2. *KIAA0319*: Another candidate gene for dyslexia, on 6p22 in the DYX2 locus.
3. *BV677278:* an approximately 600-bp region in intron 2 of *DCDC2*, hypothesized to have regulatory activity related to the development of dyslexia.
4. PEA3: a member of the ETS family of transcription factors. Previously shown to be involved in oncogenesis and neuronal migration.
5. *DYX1C1:* A candidate gene for dyslexia, on 15q11 in the DYX1 locus.
6. *ROBO1:* A candidate gene for dyslexia on 3p12 in the DYX5 locus.
7. *DYX* loci: regions shown to be in linkage with dyslexia (10 total).

Abbreviations and Terms

1. atRA: all-*trans*-retinoic acid, used to differentiate P19s into neuron-like cells.

2. Knockdown: inhibition of mRNA transcripts using siRNA (blocks RNA to protein translation).
3. Knockout: inhibition of gene through targeted deletion (blocks gene to RNA transcription).
4. P19: mouse embryonal carcinoma cell line.
5. RD: Reading Disability, also known as developmental dyslexia.
6. siRNA: small inhibitory RNA. Used to knock down mRNA transcripts
7. SNP: single nucleotide polymorphism, the most common type of marker used in modern genetic linkage and association studies.
8. TF: transcription factor