

**McGrath, L. M.,** Smith, S. D., Pennington, B. F. (2006). Breakthroughs in the search for dyslexia candidate genes. *Trends in Molecular Medicine*, 12, 333-341.

## Breakthroughs in the Search for Dyslexia Candidate Genes

Lauren M. McGrath<sup>1</sup>, Shelley D. Smith<sup>2</sup>, and Bruce F. Pennington<sup>1</sup>

<sup>1</sup> University of Denver, Department of Psychology, Frontier Hall, 2155 S. Race St., Denver, CO 80208, USA

<sup>2</sup> University of Nebraska Medical Center, 985456 Nebraska Medical Center, Omaha, NE 68198-5456, USA

Corresponding author: McGrath, L. M. ([lmcgrath@du.edu](mailto:lmcgrath@du.edu))

Keywords: dyslexia, reading disability, candidate genes, linkage, association, *DYX1C1*, *EKN1*, *ROBO1*, *DCDC2*, *KIAA0319*.

## Abstract

Four candidate genes for dyslexia, or reading disability (RD), have recently been proposed: *DYX1C1*, *ROBO1*, *DCDC2*, and *KIAA0319*. Each of the genes is implicated in brain development processes, such as neural migration and axonal guidance, with the exception of *DYX1C1* whose function is still unknown. The most immediate clinical prospect of these gene identifications is the possibility of early identification via genetic screening. However, the field has yet to identify a functional mutation in any of the genes, which currently limits this future prospect. When causal variants are identified, they will need to be considered within a multifactorial framework, likely involving gene x gene and gene x environment interactions, in order to make accurate predictions of diagnostic status.

## Breakthroughs in the Search for Dyslexia Candidate Genes

### *Background*

Dyslexia, or reading disability (RD), is a complex neurobehavioral disorder affecting approximately 5-10% of school-aged children [1]. A consensus definition of RD was developed in 2002 by the International Dyslexia Association and adopted by the National Institute of Child Health and Human Development (NICHD) ([www.interdys.org](http://www.interdys.org)):

“Dyslexia is a specific learning disability that is neurological in origin. It is characterized by difficulties with accurate and/or fluent word recognition and by poor spelling and decoding abilities. These difficulties typically result from a deficit in the phonological component of language that is often unexpected in relation to other cognitive abilities and the provision of effective classroom instruction. Secondary consequences may include problems in reading comprehension and reduced reading experience that can impede growth of vocabulary and background knowledge.”

Since the 1950s, there has been considerable interest in the genetics of RD. Studies have shown that the disorder is familial [2, 3] and heritable [4]. In addition, recent molecular genetic studies have identified and replicated several linkage peaks in the genome [5]. These genetic findings have recently culminated in the identification of several candidate genes for RD. In 2003, the first candidate gene for RD was proposed: *DYX1C1* (or *EKN1*) [6]. In 2005, there was considerable excitement in the field as three new candidate genes were proposed: *KIAA0319* [7], *DCDC2* [8], and *ROBO1* [9]. In the past, identifying genes for complex behavioral disorders has proven difficult [10], so RD is being celebrated as a success story, which marks the arrival of complex behavioral disorders into the realm of molecular medicine. The success of this research has rested on the careful application of cognitive methods to develop precisely-defined phenotypes [11], the systematic analysis of linkage sites identified in genome-wide linkage studies [5], and in some cases, the helping hand of serendipity via informative chromosomal translocations. In the following review, we will discuss the groundwork that set the stage for the

recent gene identifications. Next, we will discuss the candidate genes that have been proposed, highlighting the functional relevance of these genes in the brain and replication efforts (see Table 2 for a summary). Finally, we will examine the clinical prospects for this work.

### *Phenotypic Definition*

As previously discussed, RD has benefited from a sophisticated cognitive science underlying the development of phenotypes for use in molecular genetic studies [11]. Psychiatry's current interest in endophenotypes reflects this same emphasis on developing precisely-defined phenotypes with a biological basis [12]. In the case of RD, the phenotype has been dissected into several main cognitive components that contribute to the acquisition of reading skill: phonological awareness, phonological coding, orthographic coding, and rapid serial naming (see Box 1 for an explanation) [5].

Each of these cognitive components of the RD phenotype, as well as measures of reading proficiency itself, have been utilized successfully as phenotypes in behavioral and molecular genetic studies. The precision of the phenotypic definition initially inspired optimism that specific phenotypes could be associated with specific genomic regions [13]. However, follow-up studies showed that the mapping of specific cognitive processes to genes was more complicated than this parsimonious hypothesis would suggest [11, 14-16], especially since the cognitive components were correlated. Although it has been difficult to find evidence of differential linkage of the phenotypes to specific loci, there is likely to be some specificity at the genetic level because behavioral genetic results show that some of the cognitive components of reading possess partially independent genetic influences [17]. Currently, linkage and association studies of RD typically use several cognitive measures of reading and look for convergence among the measures. Recent studies have investigated the possibility of using multivariate methods, which

capture the covariance of the cognitive components, to conduct molecular genetic studies [18]. These methods may hold the most promise for fully utilizing the sophisticated cognitive science underlying the RD phenotypes.

### *Behavioral Genetic & Linkage Studies*

As noted earlier, the familiarity [2, 19] and heritability [4] of RD has been firmly established. Recent heritability estimates for a composite reading measure (word reading, spelling, and reading comprehension) utilizing a large twin sample showed that more than half of the group deficit could be attributed to genetic influences ( $h^2_g = .58$ ) [20]. Similarly high heritability estimates are obtained when the component cognitive processes of reading: phoneme awareness, phonological decoding, and orthographic coding, were analyzed [17].

Once the familiarity and heritability of RD was established, the field began to focus on identifying the specific genes involved in RD. Genes that affect a quantitatively measured trait, such as reading, are termed quantitative trait loci (QTL) [21, 22]. When the notion of a QTL is applied to complex behavioral traits, such as RD, the QTLs are usually considered part of a multifactorial etiology in which any single QTL is neither necessary nor sufficient to cause the disorder [22]. Unlike other complex behavioral disorders, like Attention-Deficit/Hyperactivity Disorder, RD does not have any obvious theoretically-driven candidate genes. Thus, the field has relied on genome-wide linkage scans to identify QTLs. Linkage findings have been notoriously difficult to replicate in complex disorders [23], but a large body of targeted and genome-wide linkage studies have generated the following 7 replicated linkage regions: 1p36-p34 (DYX8), 2p16-p15 (DYX3), 3p12-q13 (DYX5), 6p22.2 (DYX2), 15q21 (DYX1), 18p11.2 (DYX6), and Xq27.3 (DYX9) (see Table 1 for supporting references) [for a review see 5, 24].

From this point, the identification of candidate genes in the 15q21 (*DYX1C1*) and 3p12-q13 (*ROBO1*) regions was aided by the discovery of rare translocations that segregated with RD. In the 6p22 region, fine-scale association mapping of the region resulted in identification of two candidate genes, *DCDC2* and *KIAA0319* (see Table 2 for a summary of the candidate gene and association studies. Box 2 describes the association methods.)

### *Candidate Gene and Association Studies*

*DYX1C1 (EKN1) on Chromosome 15q21.* The *DYX1C1* candidate gene was proposed following the discovery of a two-generation family with a translocation t(2;15)(q11;q21) disrupting a gene in the 15q21 region and cosegregating with RD [25]. Taipale et al. [6] found that the gene, labeled *DYX1C1* (dyslexia linkage region 1 – candidate 1), encodes a protein with three tetratricopeptide repeat domains, which are believed to subserve protein-protein interactions. Other than these domains, the protein contains no other homology to known proteins. The gene is expressed in several tissues, including the brain, where it is found in a subset of cortical neurons and white matter glial cells. Beyond these details, the function of the protein product is unknown.

To determine if the *DYX1C1* gene could be implicated in RD more generally, Taipale et al. [6] screened for polymorphisms in the DNA of 20 RD individuals. They tested the identified SNPs in two separate samples of cases and controls finding association with two SNPs: -3G→A and 1249G →T. However, one concern with the first sample was that some of the cases and controls were related to each other. In the second sample, the authors found weaker results for the -3G→A variant, and no significant association for the 1249G→T variant. Taipale et al. [6] hypothesized that both of these SNPs could be causal in the RD phenotype because the -3 SNP

was located in the translation initiation sequence and in a transcription factor binding site and the 1249 SNP produced a premature stop codon, although only the last 4 amino acids were lost.

There have been six attempted replications of the *DYX1C1* association, but none of these studies has found support for the specific risk alleles proposed by Taipale et al. [6] despite the diverse methods and samples employed (see Table 2) [26-31]. Four of these studies have been unable to find any association between *DYX1C1* variants and RD phenotypes. [26, 28, 30, 31]. Two additional studies have found associations with the same SNPs reported by Taipale et al. [6] but the associations were in the *opposite* direction, such that the more common allele in the population was associated with poorer reading performance [27, 29]. Wigg et al. [29] found that the more common alleles, -3G, 1249G, were associated with RD. The study also identified an intronic variant that showed significant association [29]. Consistent with Wigg et al. [29], Scerri et al. [27] also found a nominally significant association with the more common 1249G SNP. The authors noted that this association only occurred for 1 of the 6 RD phenotypes tested and it was not significant after adjustment for multiple tests. Since it is unlikely that the proposed risk alleles could be detrimental in one sample and not another, Wigg et al. [29] and Scerri et al. [27] concluded that the specific risk alleles proposed by Taipale et al. [6] were unlikely to be causal variants for RD. In sum, the weight of the evidence seems to indicate that the proposed *DYX1C1* risk alleles are not associated with RD. Nevertheless, it is important to point out that the *DYX1C1* gene could be causally implicated in the RD phenotype in the Finnish family without being causally related to more common forms of RD.

Although the risk alleles proposed by Taipale et al. [6] may not be causal factors in RD, the fact that the same SNPs showed association in two independent studies [6, 29] (and possibly three if the nominally significant Scerri et al. [27] result is considered), albeit in opposite

directions, is intriguing. One simple explanation, which is consistent with the four studies reporting null results, is that both results could be false positives. However, an alternative explanation for these findings could be that the proposed risk alleles may be in linkage disequilibrium with a nearby DNA change in the *DYX1C1* gene or even neighboring genes. If this putative causal variant had different founders, it could explain why the opposite association profile was obtained in different populations. This explanation would also be consistent with the finding of association with a third SNP within *DYX1C1* by Wigg et al. [29]. Thus, future research will be needed to determine if the associations reported by Taipale et al. [6], Wigg et al. [29], and Scerri et al. [27] are true associations and whether these associations point to nearby causal variants. Another remaining question concerns the relationship between the 15q21 linkage peak and the *DYX1C1* candidate gene. The linkage peaks from previous studies of the 15q region (see Table 1) are several centimorgans away from the breakpoint described in the Taipale et al. [6] study. Linkage studies are notoriously imprecise in specifying exact QTL locations, so the *DYX1C1* candidate may indeed be causing the 15q linkage peaks. Alternatively, there may be another gene for RD residing in this region.

*ROBO1 on Chromosome 3p12-q13.* Like *DYX1C1*, a serendipitous discovery of an individual with RD and a translocation t(3;8)(p12;q11) within the 3p12-q13 region led to the advancement of *ROBO1* as a candidate gene for RD [9]. Before this individual was found, a study investigating a large four-generation family showing a dominant inheritance pattern for RD had mapped the susceptibility gene to the 3p12-q13 region, but could not resolve the region further [32]. Thus, the individual with the translocation provided a natural experiment which allowed the region to be resolved down to the gene level. Hannula-Jouppi et al. [9] found that

the translocation disrupted the *ROBO1* gene, which is implicated in guidance of axons crossing between brain hemispheres and guidance of dendritic connections [33-35]. The function of *ROBO1* in axon guidance is consistent with diffusion tensor imaging results suggesting degradation in the microstructural integrity of temporo-parietal white matter in poor readers [36, 37].

Hannula-Jouppi et al. [9] further examined *ROBO1* through polymorphism screening to determine if it was responsible for the linkage signal in the 3p12-q13 region for the large four-generation family [32]. They found a specific SNP haplotype that segregated with RD in the family. To examine the functional implications of the SNP haplotype, the authors conducted an analysis of *ROBO1* gene expression in lymphocytes, hypothesizing that the findings would generalize to brain expression as well. They found that *ROBO1* mRNA expression was attenuated in individuals with RD who were carrying the susceptibility haplotype, compared to controls. Thus, although the specific mutation associated with RD in this family remained elusive, the functional implications of the mutation for gene regulation were clear.

The *ROBO1* candidate gene for RD has yet to be replicated. One concern raised by Hannula-Jouppi et al. [9] is that the individual with the translocation had siblings diagnosed with RD that did not possess the translocation. Hence, the causal role of the translocation in the individual's RD diagnosis is questionable. The fact that a unique SNP set within *ROBO1* segregated with RD in the large multiplex family is encouraging. However, it remains to be seen whether disruptions in the *ROBO1* gene can be implicated in more common forms of RD.

*DCDC2 and KIAA0319 on Chromosome 6p22.* The 6p22 locus is one of the most well-replicated QTLs in linkage studies of RD [5]. Thus, unlike the *DYX1C1* and *ROBO1* candidate

genes that were identified through rare translocations, a candidate gene in this region is more likely to contribute to common forms of RD. In order to further resolve the locus down to the level of genes, researchers have targeted the 6p22 region with association studies using high density SNP maps (see Table 2). So far, two candidate genes in this region have been proposed, *DCDC2* and *KIAA0319*. At this point, it is unclear whether either gene or both might contribute to the RD phenotype. In both cases, functional variants have yet to be identified. This state of the literature contrasts with studies investigating *DYX1C1* which tested specific variants that were proposed by Taipale et al. [6] to have functional consequences. In the case of *DCDC2* and *KIAA0319*, studies have identified and replicated association but they have not converged on a replicable causal variant, to date.

Both candidate genes are believed to play a role in neural migration [8, 38], consistent with the early autopsy studies by Galaburda et al. [39] reporting ectopias in RD. *DCDC2* is expressed in the inferior temporal cortex and medial temporal cortex, as well as in other brain areas [8]. The inferior temporal and medial temporal cortex has been implicated in functional imaging studies of reading [40]. One clue about the function of *DCDC2* is provided by its homology with the doublecortin gene on the X chromosome (*DCX*) which is mutated in human X-linked lissencephaly and double cortex syndrome, both disorders of neural migration [8]. Empirical evidence of the role of *DCDC2* in neural migration was obtained from an RNAi analysis of rats *in utero*, which showed a significant reduction in mean neural migration distance for rats transfected with a vector targeted against *DCDC2* [8].

A similar RNAi analysis of *KIAA0319* also implicated this gene in neural migration [38]. Results showed a significant reduction in neural migration distance for rats transfected with a vector targeted against *KIAA0319*. Not only did the vector shorten the migration distance, but it

also affected the orientation of the migrating neurons in the intermediate zone such that they were oriented orthogonally to the radial glial fibers. This abnormal orientation suggests that *KIAA0319* may be important for adhesion during migration of the neurons along the radial glial fibers [38]. Paracchini et al. [38] also found that the spatial-temporal distribution of *KIAA0319* in the developing mouse and human fetal brain was consistent with its putative role in neural migratory processes.

At the level of linkage analysis, the two candidate genes are indistinguishable because they are only about 500kb apart, but there is disagreement among research groups at the level of association analysis. Deffenbacher et al. [41] were the first to report a refinement of the 6p22 QTL using a high density SNP map. They reported association with 5 genes in 2 clusters: *VMP/DCDC2* and *KIAA0319/TTRAP/THEM2*. They noted that the associations with *VMP/DCDC2* were more robust across analytical strategies and phenotypes and so should be prioritized in future studies. Following on these results, Meng et al. [8] found association in the *DCDC2* gene, but not the *KIAA0319* gene. They also identified a previously uncharacterized deletion within intron 2 of *DCDC2* in 10 RD families. Notably, the deletion region contained several transcription factor binding sites, implying that deletion of this region could affect transcription regulation. Unfortunately, this deletion occurred too infrequently to test for transmission disequilibrium directly. Only when the deletion was pooled with other minor alleles did the association with RD reach significance. As such, it remains an open question whether the intronic deletion is causal in at least a subset of families. It is also likely that there are other causal variants that have yet to be identified. A second study conducted by Schumacher et al. [42] also reported significant association with the *DCDC2*, but not *KIAA0319*, and replicated the result in a second independent sample. They also conducted a mutation

analysis, but they were unable to identify a likely functional allele. Unfortunately, the authors did not test for the deletion identified by Meng et al. [8].

In addition to the studies that have found evidence for *DCDC2* as a candidate gene for RD, there have also been studies that have found significant evidence for a neighboring gene, *KIAA0319*. Francks et al. [43] conducted an association study that investigated a region of interest identified through a preliminary linkage study. This peak linkage region encompassed the *DCDC2* gene and the *KIAA0319* gene. Within this linkage region, the authors chose 8 genes for further association studies. The *DCDC2* gene was not one of the candidates chosen, but *KIAA0319* was included in the study. They found association with a cluster that included *KIAA0319* and replicated these findings in 2 independent samples. Cope et al. [7] conducted an association study with a region encompassing both *KIAA0319* and *DCDC2*. The authors employed a hierarchical study design and several samples. At each level of analysis, *KIAA0319* emerged as the most likely candidate gene. Both Francks et al. [43] and Cope et al. [7] screened the *KIAA0319* gene for variants with possible functional effects, but neither group identified a plausible functional variant, suggesting that gene regulatory regions were likely involved. Consistent with this hypothesis, Paracchini et al. [38] examined transcription regulation in cell lines that were heterozygous for a SNP haplotype of *KIAA0319* that was associated with RD. Results showed that the risk haplotype was associated with a transcription reduction of *KIAA0319*, but not its neighboring genes, *TTRAP* or *THEM2*, which had previously been suggested as candidates [41]. These results provided further evidence for *KIAA0319* as the candidate gene and suggested a possible biological mechanism for its action.

Taken together, there is evidence for both proposed candidate genes and each has been replicated in at least one independent sample. One explanation for the conflicting results could

be that both genes contribute to the RD phenotype. If so, this might explain why the 6p22 QTL has been so well-replicated in linkage studies [5]. The fact that both candidate genes are involved in neural migration may suggest that disruption in this process, regardless of its etiology, conveys susceptibility for RD.

Alternatively, methodological considerations may help to explain the inconsistencies. This possibility is also highlighted by the fact that several studies have used overlapping samples drawn from the U.S. Colorado Learning Disabilities Research Center [CLDRC: 44], but they have obtained different results. Meng et al. [8] and Deffenbacher et al. [41] obtained stronger evidence for *DCDC2* in this sample but Francks et al. [43] found evidence for *KIAA0319* (although the authors did not test for association with *DCDC2*). These somewhat inconsistent results may be partially explained by methodological differences between the studies. For example, the sample's phenotypic severity may be important, particularly for the *DCDC2* locus which tends to show stronger results in more severely selected samples [41, 42]. This might explain why some studies that did not select for phenotype severity found stronger evidence for *KIAA0319* compared to *DCDC2* [e.g. 7], although there are exceptions to this pattern [e.g. 8]. It also does not appear to be the case that severity selection precludes finding association with the *KIAA0319* gene [43]. Another phenotypic methodological variation concerns whether IQ is covaried out of the phenotype. Francks et al. [43] used this technique and found that the linkage peak in the 6p22 region increased and was more refined, but most studies do not employ this procedure. Because Francks et al. [43] tested for association with *KIAA0319* and not *DCDC2*, it is unclear whether this procedure could have differential effects on association tests with both genes. In addition to these phenotypic methodological variations, studies have also chosen different SNPs that span the 6p22 region of interest. As such, there is some discrepancy in the

regions covered by each study, especially in potential regulatory regions. For example, Meng et al. [8] found a deletion within intron 2 of *DCDC2* while other studies did not cover this region [7]. Finally, issues of power and population heterogeneity could explain inconsistencies between studies [42].

### *Key Points*

To date, all of the candidate genes are involved in general brain development processes such as axonal guidance and neural migration, with the exception of *DYX1C1* whose function is still unknown. This is not surprising because reading is a recent cultural invention for which there are unlikely to be specific genes [4, 45]. However, the fact that RD impairs some language and cognitive skills, but not others, creates an interesting puzzle. Future research will need to address the mechanisms by which genes for general brain development can produce relatively specific phenotypes like RD.

Although the role of these genes in global brain development creates a puzzle for understanding the specificity of the RD phenotype, the fact that the gene functions generally correspond with previous imaging studies provides a reason for cautious optimism. Caution is necessary because it has been fairly difficult to find convergence of imaging findings in RD, particularly structural findings [46]. Nonetheless, there is at least a surface correspondence between the gene functions and imaging results. For example, *ROBO1*'s role in axon guidance is consistent with diffusion tensor imaging results suggesting aberrant white matter pathways in RD [36, 37]. Similarly, the role of *DCDC2* and *KIAA0319* in neural migration is consistent with Galaburda et al.'s [39] classic autopsy studies demonstrating ectopias in RD. More generally,

the neural/axonal migration functions of these genes are consistent with imaging studies showing a disruption in reading networks in the left hemisphere associated with RD [46, 47].

Finally, it is important to highlight that studies have not yet identified a functional mutation in any of the coding regions for the 4 candidate genes, despite several attempts. Some authors have commented that an inability to identify a functional variant may suggest that the mutation is likely to influence gene regulation rather than the protein product [8, 43]. In fact, for *ROBO1* and *KIAA0319* there is empirical evidence that the variant is likely to be involved in gene regulation [9, 38]. The fact that regulatory regions are likely involved is consistent with the fact that RD is a fairly mild phenotype in the grand scheme of brain development. The fact that the brain develops mostly appropriately with a few aberrations would be consistent with gene regulation problems rather than the complete loss of a protein product.

### *Clinical Prospects*

The clinical prospect that comes immediately to mind as a result of these candidate gene identifications is the possibility of early identification. Typically, RD is not diagnosed until a child had already fallen behind in reading and so early identification would enable early, possibly preventative, training in pre-literacy skills. Unfortunately, there are several issues that must be resolved before the goal of early identification could be realized. So far, the candidate genes have not shown clear mutations in their coding regions, implying that regulatory regions are likely to be involved. In general, the genetics field is in its infancy with respect to understanding the complexity of gene regulatory regions. Sequence variations in these regions will be more difficult to interpret, especially if there are many different mutations in the RD population. Thus, until more knowledge can be gained about these regulatory regions, it will be

difficult to make predictions about RD status based on specific base-pair changes in these regions. Once candidate polymorphisms in these regions are identified and replicated, the next step towards early identification will require population screening for the polymorphisms in unselected epidemiological samples in order to determine the sensitivity and specificity of the mutations in the general population. To date, studies have emphasized RD selected samples so there is very little information available regarding the expected rate of polymorphisms in the RD candidate genes in the general population. The multifactorial etiology of RD will also make it difficult to predict diagnostic status based on genetic screening until we fully understand the relative contribution of each QTL to the RD phenotype. Thus, although there is optimism that the identified candidate genes may someday contribute to early identification efforts for RD, there are several levels of analysis left to be completed before these predictions will be clinically useful. Eventually, we might expect to do a microarray-type analysis of a number of loci, calculating probabilities based on patterns of risk alleles.

To further complicate matters beyond simply dealing with a myriad of potential QTLs, a complex behavior disorder like RD is likely to show gene x gene and gene x environment interactions. There are already hints that gene x gene interactions may be operating in RD [48]. Thus far, gene x environment interactions have been relatively neglected in the reading field, despite evidence that literacy environments can exert main effects on reading phenotypes [49]. Investigations of gene x environment interactions are currently an important area of investigation in behavioral disorders, [50] and the recent candidate gene identifications in RD make these questions more tractable in RD than in most other behavioral disorders. Rutter et al. [50] noted that satisfactory investigations of gene x environment interactions will require molecular genetic measures of genetic risk, rather than behavioral genetic measures which have been customary

[e.g. 51]. Hence, with the recent candidate gene identifications, the RD field is well-poised to begin using molecular genetic indices of genetic risk to test for gene x environment interactions. These analyses may point to optimal environments that can serve as protective factors in the face of background genetic risk.

Another clinical application is the potential for the gene identifications to advance our understanding of the common comorbidities associated with RD. Comorbidity is the rule rather than the exception with RD. Clinically, these comorbidities can often help in determining a child's risk for RD because the comorbid disorders are usually diagnosed before RD. For example, RD is comorbid with ADHD at a rate estimated at 25-40% [52]. Similarly, RD is comorbid with speech/language disorders at a similar rate of 25-30% [53-55]. Thus far, several overlapping linkage regions have been found for ADHD, speech/language disorders, and RD [24, 56-59]. The most parsimonious explanation of these findings is that the disorders share some QTLs, but not others, and the pattern of unique and shared QTLs determines the phenotypic outcome [60]. However, we currently have a limited understanding of what neural vulnerabilities could place a child at risk for two or more disorders at the same time. Identification of the function of RD genes will help to determine why the QTLs for RD sometimes convey vulnerability for other developmental disorders. This line of research will advance the understanding of the etiology of RD as well as its common comorbid disorders.

Another clinical prospect is the potential to combine knowledge of candidate genes with neuroimaging methods in order to understand the functional effects of the genes in the brain. This method has previously been implemented with impressive results in studies investigating variations in the serotonin transporter gene associated with human amygdala responses [61]. Alternatively, it may be possible to use imaging phenotypes themselves in association screens to

identify genes that may contribute to an observed variation in brain activation. These lines of research at the interface of gene and brain levels of analyses have the potential to make substantial contributions to the understanding of the RD phenotype.

### *Concluding Remarks*

In conclusion, the genetics of RD has advanced to the point that it has approached the realm of molecular medicine with the proposal of four candidate genes: *DYX1C1*, *ROBO1*, *DCDC2*, and *KIAA0319*. One theme that emerges from these candidate gene identifications is that each gene (with the exception of *DYX1C1* whose function is unknown) has been implicated in global brain developmental processes, such as neural migration and axonal guidance. This finding creates a puzzle for future research to unravel: how can a disruption in global brain development result in such a specific phenotype? The most immediate clinical prospect of these gene identifications is the possibility of early identification of RD via genetic screening, but several unresolved issues currently limit the clinical utility of the gene identifications for this purpose. First and foremost, the field has yet to identify a causal mutation and, in fact, it seems likely that the causal variants will reside in regulatory regions where they will be difficult to detect. In addition, RD is a complex behavioral disorder with a multifactorial etiology that likely involves gene x gene and gene x environment interactions. Thus, even when causal mutations are identified, they will need to be considered in the context of a multifactorial framework in order to make accurate predictions about diagnostic status. Although genetic screening for RD may be further in the future than many would hope, these gene identifications can be used now to better understand gene x environment interactions and the comorbidities of RD, and to integrate the genetic and brain levels of analysis by using brain imaging technologies. Results of such

studies could inform clinical practice by further specifying environmental risk and protective factors, sharpening our understanding of comorbidity in RD, and helping to determine which structures in the neural network that mediate reading are most important in the pathogenesis of RD.

Although these gene identifications represent an exciting breakthrough, in many cases they have raised more questions than they have answered (see Box 3). In order to address several of the questions raised in Box 3, studies will require large, diverse samples. There is currently a large European Consortium (Neurodys: [www.neurodys.com](http://www.neurodys.com)) that has been formed for the purpose of pooling existing samples across countries and ethnic groups. This sample will provide more power to address several of the outstanding questions listed in Box 3. For example, the Neurodys sample will provide increased power to determine whether genes that were identified through rare translocations are causal in the general population. It will also provide increased power to further resolve linkage peaks in order to determine whether the identified candidate gene or a neighboring gene is responsible for the linkage signal. The heterogeneity in genetic background in this sample will also enable studies to directly test whether genetic background variability is a plausible explanation for conflicting results (e.g. *DYX1C1*).

Several methodological additions to association studies would also help to address some of the remaining questions (Box 3). Future studies should pay special attention to the regulatory regions of the candidate genes, which have traditionally been neglected when selecting genetic markers for association analyses. Mutation screening of the exons of the candidate genes has yet to turn up replicable functional mutations in any of the genes, suggesting that studies should begin typing markers in the regulatory regions of genes in addition to the exons. The area of

gene x environment interactions is currently an exciting topic in behavioral disorders.

Association studies should begin to gather information about the environment (such as the home literacy environment [49]) in order to test whether it serves as a moderator for the association results. Finally, studies should begin to gather phenotypic information about the common comorbidities of RD, such as ADHD and speech/language disorders, for the purpose of conducting bivariate linkage and association studies to determine if the candidate genes impact the comorbidities of RD as well as the RD phenotype itself.

Overall, these gene identifications have provided glimpses into possible links from genes to brain to behavior, but large gaps in our knowledge still remain. The fact that we have progressed in our understanding at each level of analysis should not trick us into believing that we have identified the links in the chain. For example, we have yet to identify a causal mutation in any of the genes, leaving the mechanisms by which the genes cause altered brain development unknown. Furthermore, the causal links between brain and behavior have been difficult to disentangle because the most promising method, functional neuroimaging, is a correlational method. Taken together, these gaps in our knowledge suggest directions for future research as well as caution in prematurely applying the findings to clinical issues. Nonetheless, the rapid progress of the field suggests that future breakthroughs are likely to be on the horizon.

#### Acknowledgements

This work was supported by two grants from the U.S. National Institutes of Health (HD049027-22 and HD027802-15).

## References

- 1 Shaywitz, S.E., *et al.* (1990) Prevalence of reading disability in boys and girls. Results of the Connecticut Longitudinal Study. *Jama* 264, 998-1002
- 2 Hallgren, B. (1950) Specific dyslexia (congenital word-blindness); a clinical and genetic study. *Acta Neurol Scand Suppl* 65, 1-287
- 3 DeFries, J.C., *et al.* (1978) Familial nature of reading disability. *Br J Psychiatry* 132, 361-367
- 4 Pennington, B.F., *et al.* (2005) Genetics of Dyslexia. In *The science of reading: A handbook.*, 453-472, Blackwell Publishing
- 5 Fisher, S.E., and DeFries, J.C. (2002) Developmental dyslexia: genetic dissection of a complex cognitive trait. *Nat Rev Neurosci* 3, 767-780
- 6 Taipale, M., *et al.* (2003) A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proc Natl Acad Sci U S A* 100, 11553-11558
- 7 Cope, N., *et al.* (2005) Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *Am J Hum Genet* 76, 581-591
- 8 Meng, H., *et al.* (2005) DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proc Natl Acad Sci U S A* 102, 17053-17058
- 9 Hannula-Jouppi, K., *et al.* (2005) The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS Genet* 1, 1-8
- 10 Plomin, R. (2005) Finding genes in child psychology and psychiatry: when are we going to be there? *J Child Psychol Psychiatry* 46, 1030-1038
- 11 Pennington, B.F. (1997) Using genetics to dissect cognition. *Am J Hum Genet* 60, 13-16
- 12 Gottesman, II, and Gould, T.D. (2003) The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry* 160, 636-645
- 13 Grigorenko, E.L., *et al.* (1997) Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. *Am J Hum Genet* 60, 27-39
- 14 Fisher, S.E., *et al.* (1999) A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. *Am J Hum Genet* 64, 146-156
- 15 Gayan, J., *et al.* (1999) Quantitative-trait locus for specific language and reading deficits on chromosome 6p. *Am J Hum Genet* 64, 157-164
- 16 Grigorenko, E.L., *et al.* (2000) Chromosome 6p influences on different dyslexia-related cognitive processes: further confirmation. *Am J Hum Genet* 66, 715-723
- 17 Gayan, J., and Olson, R.K. (2001) Genetic and environmental influences on orthographic and phonological skills in children with reading disabilities. *Dev Neuropsychol* 20, 483-507
- 18 Marlow, A.J., *et al.* (2003) Use of multivariate linkage analysis for dissection of a complex cognitive trait. *Am J Hum Genet* 72, 561-570
- 19 DeFries, J.C., *et al.* (1987) Evidence for a genetic aetiology in reading disability of twins. *Nature* 329, 537-539
- 20 Wadsworth, S.J., *et al.* (2000) Differential genetic etiology of reading disability as a function of IQ. *J Learn Disabil* 33, 192-199
- 21 Eley, T.C., and Craig, I.W. (2005) Introductory guide to the language of molecular genetics. *J Child Psychol Psychiatry* 46, 1039-1041
- 22 Plomin, R., *et al.* (1997) *Behavioral Genetics*. W. H. Freeman & Co.

- 23 Altmuller, J., *et al.* (2001) Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet* 69, 936-950
- 24 Gayan, J., *et al.* (2005) Bivariate linkage scan for reading disability and attention-deficit/hyperactivity disorder localizes pleiotropic loci. *J Child Psychol Psychiatry* 46, 1045-1056
- 25 Nopola-Hemmi, J., *et al.* (2000) Two translocations of chromosome 15q associated with dyslexia. *J Med Genet* 37, 771-775
- 26 Meng, H., *et al.* (2005) TDT-association analysis of EKN1 and dyslexia in a Colorado twin cohort. *Hum Genet* 118, 87-90
- 27 Scerri, T.S., *et al.* (2004) Putative functional alleles of DYX1C1 are not associated with dyslexia susceptibility in a large sample of sibling pairs from the UK. *J Med Genet* 41, 853-857
- 28 Marino, C., *et al.* (2005) A family-based association study does not support DYX1C1 on 15q21.3 as a candidate gene in developmental dyslexia. *Eur J Hum Genet* 13, 491-499
- 29 Wigg, K.G., *et al.* (2004) Support for EKN1 as the susceptibility locus for dyslexia on 15q21. *Mol Psychiatry* 9, 1111-1121
- 30 Cope, N.A., *et al.* (2005) No support for association between dyslexia susceptibility 1 candidate 1 and developmental dyslexia. *Mol Psychiatry* 10, 237-238
- 31 Bellini, G., *et al.* (2005) No evidence for association between dyslexia and DYX1C1 functional variants in a group of children and adolescents from Southern Italy. *J Mol Neurosci* 27, 311-314
- 32 Nopola-Hemmi, J., *et al.* (2001) A dominant gene for developmental dyslexia on chromosome 3. *J Med Genet* 38, 658-664
- 33 Kidd, T., *et al.* (1998) Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors. *Cell* 92, 205-215
- 34 Kidd, T., *et al.* (1999) Slit is the midline repellent for the robo receptor in Drosophila. *Cell* 96, 785-794
- 35 Seeger, M., *et al.* (1993) Mutations affecting growth cone guidance in Drosophila: genes necessary for guidance toward or away from the midline. *Neuron* 10, 409-426
- 36 Klingberg, T., *et al.* (2000) Microstructure of temporo-parietal white matter as a basis for reading ability: evidence from diffusion tensor magnetic resonance imaging. *Neuron* 25, 493-500
- 37 Deutsch, G.K., *et al.* (2005) Children's reading performance is correlated with white matter structure measured by diffusion tensor imaging. *Cortex* 41, 354-363
- 38 Paracchini, S., *et al.* (2006) The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. *Hum Mol Genet*, [epub ahead of print]
- 39 Galaburda, A.M., *et al.* (1985) Developmental dyslexia: four consecutive patients with cortical anomalies. *Ann Neurol* 18, 222-233
- 40 Shaywitz, S.E., *et al.* (1998) Functional disruption in the organization of the brain for reading in dyslexia. *Proc Natl Acad Sci U S A* 95, 2636-2641
- 41 Deffenbacher, K.E., *et al.* (2004) Refinement of the 6p21.3 quantitative trait locus influencing dyslexia: linkage and association analyses. *Hum Genet* 115, 128-138
- 42 Schumacher, J., *et al.* (2006) Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. *Am J Hum Genet* 78, 52-62

- 43 Francks, C., *et al.* (2004) A 77-kilobase region of chromosome 6p22.2 is associated with  
dyslexia in families from the United Kingdom and from the United States. *Am J Hum*  
*Genet* 75, 1046-1058
- 44 DeFries, J.C., *et al.* (1997) Colorado Learning Disabilities Research Center. *Learning*  
*Disability Quarterly* 8, 7-19
- 45 Plomin, R., and Kovas, Y. (2005) Generalist genes and learning disabilities. *Psychol Bull*  
131, 592-617
- 46 Eckert, M. (2004) Neuroanatomical markers for dyslexia: a review of dyslexia structural  
imaging studies. *Neuroscientist* 10, 362-371
- 47 Demonet, J.F., *et al.* (2004) Developmental dyslexia. *Lancet* 363, 1451-1460
- 48 Grigorenko, E.L., *et al.* (2001) Linkage studies suggest a possible locus for  
developmental dyslexia on chromosome 1p. *Am J Med Genet* 105, 120-129
- 49 Phillips, B.M., and Lonigan, C.J. (2005) Social correlates of emergent literacy. In *The*  
*Science of Reading: A Handbook* (Snowling, M.J., and Hulme, C., eds), 173-187,  
Blackwell Publishing
- 50 Rutter, M., *et al.* (2006) Gene-environment interplay and psychopathology: multiple  
varieties but real effects. *J Child Psychol Psychiatry* 47, 226-261
- 51 Kremen, W.S., *et al.* (2005) Heritability of Word Recognition in Middle-Aged Men  
Varies as a Function of Parental Education. *Behavior Genetics* 35, 417-433
- 52 Willcutt, E.G., and Pennington, B.F. (2000) Comorbidity of reading disability and  
attention-deficit/hyperactivity disorder: differences by gender and subtype. *J Learn*  
*Disabil* 33, 179-191
- 53 Gallagher, A., *et al.* (2000) Precursors of literacy delay among children at genetic risk of  
dyslexia. *J Child Psychol Psychiatry* 41, 203-213
- 54 Pennington, B.F., and Lefly, D.L. (2001) Early reading development in children at family  
risk for dyslexia. *Child Development* 72, 816-833
- 55 Scarborough, H.S. (1990) Very early language deficits in dyslexic children. *Child Dev*  
61, 1728-1743
- 56 Willcutt, E.G., *et al.* (2002) Quantitative trait locus for reading disability on chromosome  
6p is pleiotropic for attention-deficit/hyperactivity disorder. *Am J Med Genet* 114, 260-  
268
- 57 Smith, S.D., *et al.* (2005) Linkage of speech sound disorder to reading disability loci. *J*  
*Child Psychol Psychiatry* 46, 1057-1066
- 58 Stein, C.M., *et al.* (2004) Pleiotropic effects of a chromosome 3 locus on speech-sound  
disorder and reading. *Am J Hum Genet* 74, 283-297
- 59 Loo, S.K., *et al.* (2004) Genome-wide scan of reading ability in affected sibling pairs  
with attention-deficit/hyperactivity disorder: unique and shared genetic effects. *Mol*  
*Psychiatry* 9, 485-493
- 60 Pennington, B.F. (in press) From Single to Multiple Deficit Models of Developmental  
Disorders. *Cognition*
- 61 Hariri, A.R., *et al.* (2002) Serotonin transporter genetic variation and the response of the  
human amygdala. *Science* 297, 400-403
- 62 Rabin, M., *et al.* (1993) Suggestive linkage of developmental dyslexia to chromosome  
1p34-p36. *Lancet* 342, 178

- 63 Tzenova, J., *et al.* (2004) Confirmation of a dyslexia susceptibility locus on chromosome  
1p34-p36 in a set of 100 Canadian families. *Am J Med Genet B Neuropsychiatr Genet*  
127, 117-124
- 64 Fagerheim, T., *et al.* (1999) A new gene (DYX3) for dyslexia is located on chromosome  
2. *J Med Genet* 36, 664-669
- 65 Chapman, N.H., *et al.* (2004) Linkage analyses of four regions previously implicated in  
dyslexia: confirmation of a locus on chromosome 15q. *Am J Med Genet B*  
*Neuropsychiatr Genet* 131, 67-75
- 66 Fisher, S.E., *et al.* (2002) Independent genome-wide scans identify a chromosome 18  
quantitative-trait locus influencing dyslexia. *Nat Genet* 30, 86-91
- 67 Francks, C., *et al.* (2002) Fine mapping of the chromosome 2p12-16 dyslexia  
susceptibility locus: quantitative association analysis and positional candidate genes  
SEMA4F and OTX1. *Psychiatr Genet* 12, 35-41
- 68 Petryshen, T.L., *et al.* (2002) Supportive evidence for the DYX3 dyslexia susceptibility  
gene in Canadian families. *J Med Genet* 39, 125-126
- 69 Kaminen, N., *et al.* (2003) A genome scan for developmental dyslexia confirms linkage  
to chromosome 2p11 and suggests a new locus on 7q32. *J Med Genet* 40, 340-345
- 70 Peyrard-Janvid, M., *et al.* (2004) Fine mapping of the 2p11 dyslexia locus and exclusion  
of TACR1 as a candidate gene. *Hum Genet* 114, 510-516
- 71 Smith, S.D., *et al.* (1991) Screening for multiple genes influencing dyslexia. *Reading and*  
*Writing* 3, 285-298
- 72 Field, L.L., and Kaplan, B.J. (1998) Absence of linkage of phonological coding dyslexia  
to chromosome 6p23-p21.3 in a large family data set. *Am J Hum Genet* 63, 1448-1456
- 73 Cardon, L.R., *et al.* (1994) Quantitative trait locus for reading disability on chromosome  
6. *Science* 266, 276-279
- 74 Cardon, L.R., *et al.* (1995) Quantitative trait locus for reading disability: correction.  
*Science* 268, 1553
- 75 Nothen, M.M., *et al.* (1999) Genetic linkage analysis with dyslexia: evidence for linkage  
of spelling disability to chromosome 15. *Eur Child Adolesc Psychiatry* 8 Suppl 3, 56-59
- 76 Petryshen, T.L., *et al.* (2000) Absence of significant linkage between phonological  
coding dyslexia and chromosome 6p23-21.3, as determined by use of quantitative-trait  
methods: confirmation of qualitative analyses. *Am J Hum Genet* 66, 708-714
- 77 Kaplan, D.E., *et al.* (2002) Evidence for linkage and association with reading disability  
on 6p21.3-22. *Am J Hum Genet* 70, 1287-1298
- 78 Turic, D., *et al.* (2003) Linkage disequilibrium mapping provides further evidence of a  
gene for reading disability on chromosome 6p21.3-22. *Mol Psychiatry* 8, 176-185
- 79 Grigorenko, E.L., *et al.* (2003) Continuing the search for dyslexia genes on 6p. *Am J Med*  
*Genet B Neuropsychiatr Genet* 118, 89-98
- 80 Smith, S.D., *et al.* (1983) Specific reading disability: identification of an inherited form  
through linkage analysis. *Science* 219, 1345-1347
- 81 Bisgaard, M.L., *et al.* (1987) Dyslexia and chromosome 15 heteromorphism: negative lod  
score in a Danish material. *Clin Genet* 32, 118-119
- 82 Fulker, D.W., *et al.* (1991) Multiple regression analysis of sib-pair data on reading to  
detect quantitative trait loci. *Reading & Writing* 3, 299-313
- 83 Morris, D.W., *et al.* (2000) Family-based association mapping provides evidence for a  
gene for reading disability on chromosome 15q. *Hum Mol Genet* 9, 843-848

- 84 Schumacher, J., *et al.* (2006) Linkage analyses of chromosomal region 18p11-q12 in  
dyslexia. *J Neural Transm* 113, 417-423
- 85 de Kovel, C.G., *et al.* (2004) Genomewide scan identifies susceptibility locus for dyslexia  
on Xq27 in an extended Dutch family. *J Med Genet* 41, 652-657
- 86 Risch, N., and Merikangas, K. (1996) The future of genetic studies of complex human  
diseases. *Science* 273, 1516-1517
- 87 Newton-Cheh, C., and Hirschhorn, J.N. (2005) Genetic association studies of complex  
traits: design and analysis issues. *Mutat Res* 573, 54-69
- 88 Spielman, R.S., *et al.* (1993) Transmission test for linkage disequilibrium: the insulin  
gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52, 506-  
516
- 89 Abecasis, G.R., *et al.* (2001) Association analysis in a variance components framework.  
*Genet Epidemiol* 21 Suppl 1, S341-346
- 90 Horvath, S., *et al.* (2001) Family-based association test method: age of onset traits and  
covariates. *Genet Epidemiol* 21 Suppl 1, S403-408
- 91 Diao, G., and Lin, D.Y. (2006) Improving the power of association tests for quantitative  
traits in family studies. *Genet Epidemiol*, [epub ahead of print]
- 92 Dudbridge, F. (2003) Pedigree disequilibrium tests for multilocus haplotypes. *Genet  
Epidemiol* 25, 115-121
- 93 Schaid, D.J. (2004) Evaluating associations of haplotypes with traits. *Genet Epidemiol*  
27, 348-364

Box 1. Cognitive components of word reading.

**Phonological Awareness:** an oral language skill characterized by the ability to dissect a spoken word into smaller sound units, the smallest of which are phonemes. For example, the word “cat” is composed of three phonemes, /k/, /a/, /t/. Children must be able to segment the speech stream into phonemes in order to learn the letter-sound correspondences characteristic of an alphabetic language like English.

**Phonological Coding:** the ability to pronounce letter strings that have never been seen before, often measured by the ability to pronounce pseudowords like “joop.” This skill demonstrates an understanding of letter-sound correspondences (such as in “gave”) which exist even in languages (like English) that have many words with irregular spellings (e.g. “have).

**Orthographic Coding:** the ability to encode the specific spelling pattern of a word, including words that are pronounced the same but spelled differently (e.g. “gate” and “gait”). Orthographic coding also aids in the identification and spelling of exception words like “yacht.”

**Rapid Serial Naming:** the ability to rapidly retrieve names for items presented in a series, often measured by the time taken to name an array of letters, numbers, colors, or objects. This skill is often associated with reading fluency.

Table 1. Linkage and association studies for replicated linkage peaks.

<i>Linkage Regions</i>	<i>Supportive Results</i>	<i>Negative Results</i>
<i>1p36-p34 (DYX8)</i>	[62] Rabin et al. (1993) [48] Grigorenko et al. (2001) [63] Tzenova et al. (2004)	
<i>2p16-p15 (DYX3)</i>	[64] Fagerheim et al. (1999) [66] Fisher et al. (2002) [67] Francks et al. (2002) [68] Petryshen et al. (2002) [69] Kaminen et al. (2003) [70] Peyrard-Janvid et al. (2004)	[65] Chapman et al. (2004)
<i>3p12-q13 (DYX5)</i>	[32] Nopola-Hemmi et al. (2001) [66] Fisher et al. (2002)	
<i>6p22.2 (DYX2)</i>	[71] Smith et al. (1991) [73, 74] Cardon et al. (1994, 1995) [13] Grigorenko et al. (1997) [14] Fisher et al. (1999)	[72] Field & Kaplan (1998) [75] Nöthen et al. (1999) [76] Petryshen et al. (2000) [65] Chapman et al. (2004)

	[15] Gayán et al. (1999)	
	[16] Grigorenko et al. (2000)	
	[66] Fisher et al. (2002)	
	[77] Kaplan et al. (2002)	
	[78] Turic et al. (2003)	
	[18] Marlow et al. (2003)	
	[79] Grigorenko et al. (2003)	
<b><i>15q21 (DYX1)</i></b>	[80] Smith et al. (1983)	[62] Rabin et al. (1993)
	[71] Smith et al. (1991)	[81] Bisgaard et al. (1987)
	[82] Fulker et al. (1991)	
	[13] Grigorenko et al. (1997)	
	[75] Nöthen et al. (1999)	
	[25] Nopola-Hemmi et al. (2000)	
	[83] Morris et al. (2000)	
	[65] Chapman et al. (2004)	
<b><i>18p11.2 (DYX6)</i></b>	[66] Fisher et al. (2002)	[65] Chapman et al. (2004)
	[18] Marlow et al. (2003)	[84] Schumacher et al. (2006)
<b><i>Xq27.3 (DYX9)</i></b>	[66] Fisher et al. (2002)	
	[85] de Kovel et al. (2004)	

Table 2. Summary of candidate gene and association studies.

<i>Reference</i>	<i>Sample (country)</i>	<i>Type of analysis<sup>a</sup></i>	<i>Genotyped markers</i>	<i>Summary of primary results</i>	<i>Proposed gene function</i>
<b><i>EKN1 (DYX1C1)</i></b> <b><i>Chromosome</i></b> <b><i>15q21</i></b>					Unknown – contains motifs thought to be involved in protein-protein interactions.
[6] Taipale et al. (2003)	1 affected two-generation family with a translocation (Finland)	Translocation t(2;15)(q11;q21)		Translocation cosegregated with dyslexia.	
	54 dyslexics/113 controls (Finland)	Case-control Association	8 SNPs identified by screening the cDNA of 20 dyslexics	-3G→A, $p=.006$ 1249G →T, $p=.02$ -3A/1249T haplotype, $p=.02$	
	52 dyslexics/81 controls (Finland)	Case-control Association	8 SNPs (same as above)	-3G→A, $p=.02$ 1249G →T, $p=.1$	
[29] Wigg et al. (2004)	148 nuclear families (Canada)	TDT & FBAT <sup>c</sup>	6 SNPs (2 sig. Taipale SNPs and 4 SNPs from public databases)	-3A→G, $p<.05$ -3G/1249G haplotype, $p =.03$ rs11629841, $p =.02$	
[27] Scerri et al. (2005)	264 nuclear families (U.K)	QTDT	8 SNPs reported by Taipale	1249T→G, $p<.02$ , association was not significant after adjustment for multiple tests.	
[26] Meng et al. (2005)	150 nuclear families (United States)	QTDT <sup>d</sup>	2 SNPs (the 2 sig. Taipale SNPs)	No significant associations.	
[28] Marino et al. (2005)	158 nuclear families (Italy)	TDT & FBAT	3 SNPs (including the 2 sig. Taipale SNPs)	No significant associations.	
[31] Bellini et al. (2005)	57 dyslexics/96 controls (Italy)	Case-control Association	3 SNPs (including the 2 sig. Taipale SNPs)	No significant associations.	
[30] Cope et al. (2005)	247 parent-proband trios (UK)	TDT & FBAT	3 SNPs (the 2 sig. Taipale SNPs and the 1 sig. Wigg SNP)	No significant associations.	

<b>ROBO1</b>					Guidance of axons crossing between brain hemispheres and guidance of dendritic connections.
<b>Chromosome 3p12-q13</b>					
[9] Hannula-Jouppi et al. (2005)	1 dyslexic individual with a translocation (Finland)	Translocation t(3;8)(p12;q11)		Individual with the translocation diagnosed with dyslexia.	
	16 members from an affected four-generation family showing linkage to the 3p12-q13 region (Finland)	Polymorphism screening	7 SNPs identified by polymorphism screening	A specific SNP haplotype segregated with dyslexia in the multiplex family.	
<b>DCDC2 &amp; KIAA0319</b>					Both candidates are involved in neural migration.
<b>Chromosome 6p22</b>					
[41] Deffenbacher et al. (2004)	114 nuclear families (US)	FBAT & QTDT	31 SNPs spanning 680 kb & 10 genes including <i>KIAA0319</i> and <i>DCDC2</i>	Significant association with 5 genes in 2 clusters: <i>VMP/DCDC2</i> & <i>KIAA0319/TTRAP/THEM2</i> .	
[43] Francks et al. (2004)	89 nuclear families (UK) (sample 1)	QTDT	57 SNPs spanning 225kb & 8 brain-expressed genes, including <i>KIAA0319</i> , but not <i>DCDC2</i>	Significant association with a 77 kb region that spans <i>TTRAP/KIAA0319</i> . Results strongest in a severity-selected sample.	
	175 nuclear families (UK) (sample 2)	QTDT	20 SNPs that showed association in sample 1 above	Replication of the associations in sample 1, especially when the sample was severity-selected.	
	159 nuclear families (US)	QTDT	20 SNPs that showed association in sample 1	Replication of the associations in sample 1, especially when the	

	(sample 3)		above	sample was severity-selected.
[7] Cope et al. (2005)	240 cases/312 controls (UK) (sample 1)	Case-control association	Pooled analysis of 137 SNPs in the 8 genes identified by Deffenbacher and Francks	17 SNPs showed significant association, 13 of these were with <i>KIAA0319</i> .
	223 cases/273 controls (UK) (sample 2)	Case-control association	10 SNPs in the region showing association in sample 1 above	Significant association with <i>KIAA0319</i> , <i>MRS2L</i> , <i>THEM2</i> .
	143 parent-proband trios (UK) (sample 3)	UNPHASED	7 SNPs showing significant association in sample 2 above	Significant association with <i>KIAA0319</i> , <i>MRS2L</i> , <i>THEM2</i> .
[8] Meng et al. (2005)	153 nuclear families (US)	QTDT	147 SNPs spanning 1.5 Mb & 18 genes including <i>KIAA0319</i> & <i>DCDC2</i>	Strongest association peak in <i>DCDC2</i> ; discovered a deletion in intron 2 of <i>DCDC2</i> in 10 dyslexic families.
[42] Schumacher et al. (2006)	137 parent-proband trios (Germany) (sample 1)	TDT & QTDT	16 STR <sup>e</sup> markers spanning 24Mb and encompassing the <i>VMP/DCDC2/KAAG1</i> gene cluster; follow-up analysis of the cluster with 25 SNPs and 4 STR markers	Most significant association with a marker in <i>DCDC2</i> . Results became stronger with severity selection.
	239 parent-proband trios (Germany) (sample 2)	TDT	Three significant markers from sample 1	Replicated associations with <i>DCDC2</i> when the sample was severity selected.
	376 parent-proband trios (Germany) (sample 3)	TDT	10 SNPs spanning the <i>KIAA0319/TTRAP/THEM2</i> gene cluster	No significant associations.

<sup>a</sup> see Box 2 for a full description of the association methods.

<sup>b</sup> Transmission Disequilibrium Test (TDT)

<sup>c</sup> Family-based Association Test (FBAT)

<sup>d</sup> Quantitative Transmission Disequilibrium Test (QTDT)

<sup>e</sup> Short tandem repeat markers (STR)

Box 2. Description of Association Methods.

**Case-control association** refers to the design in which the genotypes of unrelated individuals with a disorder are compared to a population of individuals without the disorder. In the most basic design, a chi-square test can compare the allele frequencies between the two groups to test if a particular allele is significantly more frequent (or less frequent) in the affected population. For quantitative traits, generalized linear modeling methods can be used to detect the effect of genotype on the phenotypic variation. Case-control association can be a very powerful and efficient method for detecting allelic association [86] as long as ethnic stratification is not present (i.e., both affected and unaffected subjects must be drawn from the same population). Some techniques are now available to control for this problem, such as typing an established set of alleles to ensure that the frequencies are the same between the two groups; however, small samples may lack the power to detect population differences [87].

**Family-based association** analyses were developed in an effort to eliminate the problem of ethnic stratification.

**Transmission Disequilibrium Test** [TDT: 88]. In the TDT test, affected individuals and their parents are genotyped and the frequencies of the non-transmitted alleles are compared to the frequencies of the transmitted alleles in the affected individuals. The disadvantages are that the parental genotypes must be known (or accurately inferred) and they must be segregating for the alleles to be tested; thus, only a portion of the tested population may be informative.

**QTDT** (quantitative TDT) is an adaptation of the TDT by Abecasis et al. [89]. It uses parental alleles and can also take sibling alleles into account to determine if an allele is transmitted significantly more often to an affected individual. This package offers both regression-based and variance component-based analyses. Caution is needed with a variance component framework since power is lost if the phenotype is not normally distributed (as would be the case when the population is selected for occurrence and severity of the disorder).

**FBAT** [90] examines the covariance of alleles with the severity of the phenotype scores and is somewhat less susceptible to deviations from normality in the phenotype [91]. It also allows for the analysis of different models with additive and/or dominance components and for haplotype analysis.

**UNPHASED** [92] is a collection of programs that use different approaches for determining multilocus haplotypes depending upon the study design (case-control or family-based, qualitative or quantitative) and the information on the haplotype phase that is available from the population

**Power.** For dichotomous variables, case control studies are the most efficient and the most powerful. For quantitative traits or in cases where population stratification cannot be rigorously eliminated, variance-component tests such as the QTDT are considered to be more powerful as long as the trait distribution does not deviate significantly from normality. Otherwise, tests such as the FBAT are preferred. Multilocus tests may increase power depending on the nature of the causal alleles and the degree of linkage disequilibrium [93] particularly when the haplotypes are as unambiguous as possible.

***Comparison of Studies.*** Differences in association results between studies of RD may be due to several factors, including different population selection and analysis techniques, small sample size, differences in allele frequencies between populations, differences in the test SNPs, and genetic heterogeneity. These variations can result in lack of replication, in replication of an association but with a different allele of a SNP, or in different peaks of association within the same region.

Box 3. Remaining questions.

***DYX1C1***

What is the function of the gene?

Does the gene account for more common cases of RD in the population?

Can the conflicting association results be accounted for by genetic background differences?

Can the 15q21 linkage peaks be attributed to this gene or might there be another plausible candidate?

***ROBO1***

Is the translocation causal in the individual with the translocation and RD?

Does the gene account for more common cases of RD in the population?

***DCDC2/KIAA0319***

Why are there conflicting findings for the candidate gene in the 6p22 region?

What role does severity selection of the phenotype in the sample play in the gene findings?

What role does covarying IQ from the reading measures play in the gene findings?

***General Questions***

How do genes for general developmental processes cause relatively specific phenotypes like RD?

Which mutations in each gene contribute to the RD phenotypes?

Does the fact that functional mutations have yet to be discovered imply that the variants are likely to be in regulatory regions?

What role do gene x gene and gene x environment interactions play in the development of the phenotype?

Can the gene identifications help explain common comorbidities of RD?