

Breakthroughs in the search for dyslexia candidate genes

Lauren M. McGrath¹, Shelley D. Smith² and Bruce F. Pennington¹

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

²University of Nebraska Medical Center, 985456 Nebraska Medical Center, Omaha, NE 68198-5456, USA

Four genes have recently been proposed as candidates for dyslexia: dyslexia susceptibility 1 candidate 1 (*DYX1C1*), roundabout *Drosophila* homolog 1 (*ROBO1*), doublecortin domain-containing protein 2 (*DCDC2*) and *KIAA0319*. Each gene is implicated in global brain-development processes such as neural migration and axonal guidance, with the exception of *DYX1C1*, the function of which is still unknown. The most immediate clinical prospect of the discovery of these genes is the possibility of early identification of dyslexia via genetic screening. However, research efforts have yet to identify a functional mutation in any of these genes. When causal variants are identified, they will need to be considered within a multifactorial framework, which is likely to involve gene–gene and gene–environment interactions, to make accurate predictions of diagnostic status.

Introduction

Dyslexia, also known as reading disability (RD), is a complex neurobehavioral disorder that affects ~5–10% of school-aged children [1]. A consensus definition of RD was developed in 2002 by the International Dyslexia Association (www.interdys.org) and adopted by the National Institute of Child Health and Human Development (NICHD): ‘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 candidate genes for RD. In 2003, the first candidate gene for RD was proposed: dyslexia susceptibility 1 candidate 1

(*DYX1C1*, also known as *EKN1*) [6]. In 2005, there was considerable excitement in the field of RD as three new candidate genes were proposed: *KIAA0319* [7], doublecortin domain-containing protein 2 (*DCDC2*) [8] and roundabout *Drosophila* homolog 1 (*ROBO1*) [9]. Identifying genes for complex behavioral disorders has proven difficult in the past [10]. The discovery of these candidate genes for RD has marked the arrival of complex behavioral disorders into the realm of molecular medicine. The success of this research has been based on: (i) careful application of cognitive methods to develop precisely-defined phenotypes [11]; (ii) systematic analysis of linkage sites identified in genome-wide linkage studies [5]; and, in some cases, (iii) the helping hand of serendipity via informative chromosomal translocations. Here, we will discuss the groundwork that sets the stage for the recent identification of the genes involved in RD. We will also discuss the genes that have been proposed as candidates for RD, highlighting their functional relevance in the brain and replication efforts. Finally, we will examine the clinical prospects for this work.

Phenotypic definition of RD

RD has benefited from cognitive science methods to develop phenotypes for use in molecular genetic studies [11]. The current interest of psychiatry 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 [5] (Box 1).

Each of these cognitive components of the RD phenotype and measures of reading proficiency itself have been used successfully as phenotypes in behavioral and molecular genetic studies. The precision of the phenotypic definition initially inspired optimism that specific phenotypes might be associated with specific genomic regions [13]. However, follow-up studies showed that mapping of specific cognitive processes to genes was more complicated than this hypothesis would suggest [11,14–16], especially because cognitive components of the RD phenotype were correlated. Although it has been difficult to find evidence of differential linkage of phenotypes to specific loci, some specificity at the genetic level might exist. Indeed, behavioral genetic results show that some of the cognitive

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

Available online 16 June 2006

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' and 't'. Children must be able to segment the speech stream into phonemes to learn the letter-sound correspondences characteristic of an alphabetic language such as English.
- Phonological coding: the ability to pronounce letter strings that have never been seen before, often measured by the ability to pronounce pseudowords such as 'joop'. This skill demonstrates an understanding of letter-sound correspondences (e.g. gave), which exist even in languages (e.g. 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 such as 'yacht'.
- Rapid serial naming: the ability to retrieve rapidly 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.

components of reading have partially independent genetic influences [17]. Currently, linkage and association studies of RD 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 might hold promise for capitalizing on the precisely defined RD phenotype.

Behavioral genetic and 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) using a large sample of twins showed that >50% of the group deficit could be attributed to genetic influences [20]. Similarly, high-heritability estimates were obtained when the cognitive components of reading – phoneme awareness, phonological decoding and orthographic coding – were analyzed [17].

After establishing RD familiarity and heritability, RD research began to focus on the identification of specific genes involved in this disorder. Genes that affect a quantitatively measured trait such as reading are termed quantitative-trait loci (QTLs) [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 such as Attention-Deficit/Hyperactivity Disorder (ADHD), RD does not have any obvious *a priori* candidate genes. Thus, investigators have relied on genome-wide linkage scans to identify QTLs. Linkage findings have been notoriously difficult to replicate in complex disorders [23], but many targeted and genome-wide linkage studies have generated the following seven replicated linkage regions: 1p36-p34 (susceptibility to dyslexia 8, *DYX8*), 2p16-p15 (susceptibility to dyslexia 3, *DYX3*), 3p12-q13 (susceptibility to dyslexia 5, *DYX5*), 6p22.2 (susceptibility to dyslexia 2, *DYX2*), 15q21 (susceptibility to dyslexia 1, *DYX1*), 18p11.2 (susceptibility to dyslexia 6, *DYX6*) and Xq27.3 (susceptibility to dyslexia 9, *DYX9* [5,24] (Table S1 in the supplementary material online). Subsequent to the linkage findings, 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 the identification of two other candidate genes, *DCDC2* and *KIAA0319* (Table S2 in the supplementary material online) (Box 2).

Candidate genes and association studies

DYX1C1 on chromosome 15q21

DYX1C1 was proposed as a candidate gene following the discovery of a two-generation Finnish family with the translocation t(2;15)(q11;q21), which disrupted a gene in the 15q21 region and co-segregated with RD [25]. Taipale *et al.* [6] observed that *DYX1C1* encodes a protein with three tetratricopeptide repeat domains, which are believed to subserve protein-protein interactions. Apart from 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 *DYX1C1* is unknown.

To determine whether *DYX1C1* might be implicated in RD in individuals without the translocation, Taipale *et al.* [6] screened for single nucleotide polymorphisms (SNPs) in the DNA of 20 RD individuals. These authors tested the identified SNPs in two separate groups, each consisting of RD patients and controls, and found association with two SNPs, -3G→A and 1249G→T. However, a concern with the first group was that some RD patients and controls were related to each other. In the second group, the authors found weaker association with the -3G→A variant and no significant association with the 1249G→T variant. Taipale *et al.* [6] hypothesized that both these SNPs might be causal in the RD phenotype because the -3G→A SNP was located in the translation initiation sequence and in a transcription-factor binding site, and the 1249G→T SNP produced a premature stop codon, although only the last four amino acids were lost.

Six studies attempted to replicate the association between *DYX1C1* and RD, but none of them found support for the specific risk alleles that were proposed by Taipale *et al.* [6], despite the diverse methods and samples used [26–31] (Table S2 in the supplementary material online). Four of these studies did not find any association between *DYX1C1* variants and RD phenotypes. [26,28,30,31]. Two additional studies observed associations of RD with the same SNPs reported by Taipale *et al.* [6], but these associations were in the opposite direction from the original findings [27,29]. Wigg *et al.* [29] showed that the -3G and 1249G alleles, which are more common alleles in the population, were associated with RD. This study also

Box 2. Description of association methods

Case-control association: it refers to the design in which the genotypes of unrelated individuals with a disorder are compared with a population of individuals without the disorder. In the most basic design, a chi-square test compares the allele frequencies between the two groups to test whether a particular allele is significantly more frequent (or less frequent) in the affected population than in controls. 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 powerful and efficient method for detection of allelic association [62], 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 between the two groups are the same. However, small samples might lack the power to detect population differences [63].

Family-based association: it refers to analyses that were developed to eliminate the problem of ethnic stratification.

- In the transmission disequilibrium test (TDT) [64], affected individuals and their parents are genotyped, and the frequencies of the non-transmitted alleles are compared with the frequencies of the transmitted alleles in the affected individuals. The disadvantages are that the parental genotypes must be known (or accurately inferred) and must be segregating for the alleles to be tested. Thus, only a portion of the tested population might be informative.
- Quantitative TDT (QTDT) is an adaptation of the TDT by Abecasis *et al.* [65]. It uses parental alleles and can also take sibling alleles into account to determine whether an allele is transmitted significantly more often to an affected individual than to an unaffected individual. This package offers both regression-based and variance-component-based analyses. Caution is needed with the variance-component framework because power is lost if the

phenotype is not normally distributed (e.g. when the population is selected for occurrence and severity of the disorder).

- Family-based association test (FBAT) [66] examines the covariance of alleles with the severity of the phenotype scores and is less susceptible to deviations from normality in the phenotype than other methods [67]. It also enables the analysis of different models with additive and/or dominance components and for haplotype analysis.
- UNPHASED [68] is a collection of programs that use different approaches for determining multilocus haplotypes depending upon the study design (case-control versus family-based and qualitative versus 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 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 might increase power depending on the nature of the causal alleles and the degree of linkage disequilibrium [69], particularly when the haplotypes are as unambiguous as possible.

Comparison of studies: differences in association results between studies of RD might 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.

identified an intronic variant that was significantly associated with RD [29]. Consistent with the findings by Wigg *et al.* [29], Scerri *et al.* [27] also found a marginally significant association of RD with the more common 1249G SNP. These authors noted that this association only occurred in one of the six RD phenotypes tested, and it was not significant after adjustment for multiple tests. Because it is unlikely that the proposed risk alleles are detrimental in one group and not in the other, 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 summary, the results seem to indicate that the proposed *DYX1C1* risk alleles are not associated with RD. Nevertheless, it is important to point out that *DYX1C1* might be causally implicated in the RD phenotype in a Finnish family without being causally related to more-common forms of RD.

Although the risk alleles proposed by Taipale *et al.* [6] might not be causal factors in RD, the fact that the same SNPs showed association in at least two independent studies [6,29], albeit in opposite directions, is intriguing. A simple explanation, which is consistent with the four studies that did not find any association, is that the results obtained by Taipale *et al.* [6] and Wigg *et al.* [29] might be false positives. However, an alternative explanation for these findings is that the proposed risk alleles might 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 might explain why the opposite association profile was obtained in different populations. This explanation would also be

consistent with the association of RD with a third SNP within *DYX1C1*, which was found by Wigg *et al.* [29]. Thus, future research is needed to determine whether 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 question that needs to be answered concerns the relationship between the 15q21 linkage peak and *DYX1C1*. Linkage peaks in the 15q region are several centimorgans away from the breakpoint described in the study by Taipale *et al.* [6] (Table S1 in the supplementary material online). Linkage studies are known to be imprecise in specifying exact QTL locations; therefore, *DYX1C1* might cause the 15q linkage peaks. Alternatively, there might be another candidate gene for RD residing in this region.

ROBO1 on chromosome 3p12-q13

Similar to *DYX1C1*, a serendipitous discovery of an individual with RD and the translocation t(3;8)(p12;q11) within the 3p12-q13 region led to the proposition that *ROBO1* might be a candidate gene for RD [9]. Before this individual was identified, a study that investigated a large four-generation family with 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 was used to resolve the region to the gene level. Hannula-Jouppi *et al.* [9] have shown that the translocation disrupted *ROBO1*, which is implicated in guidance of axons that cross between brain hemispheres and guidance

of dendritic connections [33–35]. The function of *ROBO1* in axon guidance is consistent with diffusion-tensor imaging results that suggest degradation in the micro-structural integrity of temporal-parietal white matter in poor readers [36,37].

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

To date, no other studies have investigated the association of *ROBO1* with RD and, therefore, this candidate gene for RD has yet to be replicated. A concern raised by Hannula-Jouppi *et al.* [9] is that the individual with the translocation had siblings who were diagnosed with RD but did not have the translocation. Hence, the causal role of the translocation in the diagnosis of RD in this individual is questionable. The fact that a unique SNP set within *ROBO1* segregated with RD in the large multiplex family is encouraging because it suggests that the *ROBO1* gene might have a causal role in RD in the multiplex family despite the ambiguous implications of the results in the individual with the translocation. However, it remains to be determined whether disruptions in *ROBO1* 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 *DYX1C1* and *ROBO1*, which were identified through rare translocations, a candidate gene in this region is likely to contribute to common forms of RD. To resolve the locus to the level of genes, researchers have targeted the 6p22 region with association studies using high-density SNP maps (Table S2 in the supplementary material online). So far, two candidate genes in this region have been proposed, *DCDC2* and *KIAA0319*. It is still unclear whether only one gene or both genes might contribute to the RD phenotype. In both cases, functional variants have yet to be identified, in contrast to studies investigating *DYX1C1* that tested specific variants, which 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 yet [7,8,38–40].

Both candidate genes are likely to have a role in neural migration [8,41], consistent with early autopsy studies by Galaburda *et al.* [42], who reported the presence of ectopias in RD. *DCDC2* is expressed in the inferior temporal cortex and medial temporal cortex, and in

other brain areas [8]. The inferior temporal and medial temporal cortices have been implicated in functional imaging studies of reading [43]. A clue about the function of *DCDC2* is provided by its homology with the double-cortin 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 RNA interference (RNAi) analysis of rats *in utero*. Vectors targeted against *DCDC2* were transfected into cells at the cerebral ventricular zone of living rat embryos. Results showed a significant reduction in mean neural-migration distance for the transfected cells [8].

A similar *in utero* RNAi analysis of *KIAA0319* implicated this gene also in neural migration [41]. Paracchini *et al.* [41] transfected vectors targeted against *KIAA0319* into cells at the cerebral ventricular zone of living rat embryos. Results showed a significant reduction in neural-migration distance for the transfected cells. The vector not only shortened the migration distance but also affected the orientation of some migrating neurons so that they were oriented orthogonally to the radial glial fibers. This abnormal orientation suggests that *KIAA0319* might be important during migration for adhesion of the neurons along the radial glial fibers [41]. Similarly, Paracchini *et al.* [41] observed that the spatio-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, *DCDC2* and *KIAA0319* are indistinguishable because they are only ~500 kb apart. However, there is disagreement among research groups at the level of association analysis. Deffenbacher *et al.* [38] were the first to report a refinement of the 6p22 QTL region. These authors employed linkage analyses using a high-density SNP map, several RD phenotypes and a large sib-pair sample to narrow the peak linkage region at 6p22. They identified ten plausible candidate genes in the peak linkage region. They reported association of RD with five genes in two clusters: (i) vesicular membrane protein (*VMP*)–*DCDC2* and (ii) *KIAA0319*–*TRAF* and *TNF* receptor-associated protein (*TTRAP*)–thioesterase superfamily member 2 (*THEM2*). These authors noted that the associations of RD with *VMP*–*DCDC2* were more robust across different association analyses and phenotypes and, thus, should be prioritized in future studies. Following on these results, Meng *et al.* [8] found association of RD with *DCDC2*, but not *KIAA0319*. They also identified a previously uncharacterized deletion within intron 2 of *DCDC2* in ten RD families. Notably, the region of the deletion contained several transcription-factor binding sites, implying that deletion of this region might 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. It remains an open question whether the intronic deletion is causal in at least a subset of families. Regardless, it is likely that other causal variants that

have yet to be identified exist. A study conducted by Schumacher *et al.* [39] also reported significant association of RD with *DCDC2*, but not *KIAA0319*, and replicated the result of Meng *et al.* [8] in an independent sample. They also conducted a mutation analysis, but they were unable to identify a possible functional allele. Unfortunately, the authors did not investigate the deletion that was identified by Meng *et al.* [8].

In addition to the studies that have shown evidence for *DCDC2* as a candidate gene for RD, there have also been studies that have found significant evidence for *KIAA0319*. Francks *et al.* [40] conducted an association study that investigated a region of interest identified through a preliminary linkage study. This peak linkage region encompassed *DCDC2* and *KIAA0319*. Within this linkage region, the authors chose eight genes for further association studies. *DCDC2* was not one of the candidates chosen, whereas *KIAA0319* was. They observed association of RD with a gene cluster that included *KIAA0319*, and replicated these findings in two independent samples. Cope *et al.* [7] conducted an association study of a region encompassing both *KIAA0319* and *DCDC2*. The authors used a hierarchical study design and several samples. At each level of analysis, *KIAA0319* emerged as the most likely candidate gene. Both Francks *et al.* [40] and Cope *et al.* [7] screened *KIAA0319* for variants with possible functional effects, but neither group identified a plausible functional variant, suggesting that gene regulatory regions were likely to be involved. Consistent with this hypothesis, Paracchini *et al.* [41] examined transcription regulation in cell lines heterozygous for a SNP haplotype of *KIAA0319* that was associated with RD. These authors showed that the risk haplotype was associated with a transcription reduction of *KIAA0319*, but not of its neighboring genes, *TTRAP* or *THEM2*, which had previously been suggested as candidates for RD [38]. These results provided further evidence that *KIAA0319* might be the candidate gene for RD, and suggested a possible biological mechanism for its action.

As reported earlier, there is contrasting evidence that either *DCDC2* or *KIAA0319* might be the candidate gene for RD in the 6p22 region. Each finding has been replicated in at least one independent sample. An explanation for the conflicting results might be that both genes contribute to the RD phenotype. If so, the reason why the 6p22 QTL has been so well-replicated in linkage studies might be explained [5]. The fact that both candidate genes are involved in neural migration might suggest that disruption of this process, regardless of its etiology, conveys susceptibility for RD.

Alternatively, methodological considerations might help to explain the inconsistencies among studies. 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 reached different conclusions [8,38,40]. One methodological consideration that might be important is the sample phenotypic severity or the severity of the reading deficit in the children that are included in the analysis. The phenotypic severity of the sample seems to be particularly important for the *DCDC2*

locus, which tends to show stronger linkage and association results in more severely selected samples [38,39]. This might explain why some studies that did not select for phenotype severity found stronger evidence for *KIAA0319* compared with that for *DCDC2* [7], although there are exceptions to this pattern [8]. In addition, severity selection does not seem to preclude finding association with *KIAA0319* [40]. Another phenotypic methodological variation concerns whether the intelligence quotient (IQ) is covaried out of the phenotype. Francks *et al.* [40] showed that this technique increased the strength of the linkage in the 6p22 region, but most studies do not employ this procedure. Because Francks *et al.* [40] tested the association of RD with *KIAA0319* and not *DCDC2*, it is unclear whether this procedure might 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. There is some discrepancy in the regions that are covered by each study, especially in potential regulatory regions. For example, Meng *et al.* [8] observed a deletion within intron 2 of *DCDC2*, whereas other studies did not cover this region [7]. Finally, issues of power of the study and population heterogeneity might explain inconsistencies among studies [39].

Summary of candidate-gene findings

To date, all the candidate genes for RD are involved in global brain-development processes such as axonal guidance and neural migration, with the exception of *DYX1C1*, the function of which is still unknown. This is not surprising because reading is a recent cultural invention for which specific genes are unlikely to exist [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 that are involved in global brain development can produce relatively specific phenotypes such as RD.

Although the role of these genes in global brain development complicates the understanding of RD-phenotype specificity, the fact that the functions of these genes correspond to previous imaging studies provides a reason for cautious optimism that the gene and brain levels of analysis might reach convergence. Caution is necessary because it has been difficult to find convergence of imaging findings in RD, particularly structural findings [46]. Nonetheless, there is at least a rough correspondence between gene functions and imaging results. For example, the role of *ROBO1* 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 the autopsy studies by Galaburda *et al.* [42] that demonstrated the presence of ectopias in RD. More generally, the neural and/or axonal migration functions of these genes are consistent with imaging studies that have shown a disruption in the reading networks in the left hemisphere associated with RD [46,47].

Finally, it is important to highlight that studies have not yet identified a replicable functional mutation in any of the coding regions for the four candidate genes, despite several attempts. Some authors have suggested that the inability to identify a functional variant might indicate that the mutation influences gene regulation rather than the protein [8,40]. In fact, there is direct evidence for *ROBO1* and *KIAA0319* that their variants are likely to be involved in gene regulation [9,41]. The fact that regulatory regions might be involved is consistent with the fact that RD is a mild phenotype in the grand scheme of brain development. The fact that the brain develops mostly appropriately with only a few aberrations would be consistent with gene-regulation problems rather than complete loss of a protein.

Clinical prospects for the candidate-gene findings

The most immediate clinical prospect, as a result of the discovery of these candidate genes, is the possibility of early identification of RD. Typically, RD is not diagnosed until a child has already shown problems in reading and, therefore, early identification would enable early, possibly preventative, training in pre-literacy skills. Unfortunately, there are several issues that must be resolved before early identification might 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, genetics is in its infancy with respect to understanding the complexity of gene regulatory regions. Sequence variations in these regions will be difficult to interpret, especially if there are many different mutations in the RD population. Thus, until additional knowledge is gained about these regulatory regions, it will be difficult to predict 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 to determine the sensitivity and specificity of the mutations in the general population. To date, studies have focused on individuals and families with RD; therefore, there is little information available regarding the expected rate of polymorphisms in 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 might contribute to early identification of RD, there are several levels of analysis that need to be completed before these predictions can be clinically useful. In the future, we might expect to do a microarray-type analysis of several loci, calculating probabilities based on patterns of risk alleles.

In addition to the difficulty of dealing with many potential QTLs, a complex behavior disorder such as RD is likely to show gene–gene and gene–environment interactions. There are already indications that gene–gene interactions might be operating in RD [48]. So far, gene–environment interactions have been relatively neglected in the reading field, despite evidence that

literacy environments can exert main effects on reading phenotypes [49]. Gene–environment interactions are currently an important area of investigation in behavioral disorders [50]. The recent identification of candidate genes for RD makes the study of gene–environment interactions more tractable in RD than in most other behavioral disorders. Rutter *et al.* [50] noted that satisfactory studies of gene–environment interactions will require molecular genetic measures of genetic risk rather than behavioral genetic measures, which have been customary [51]. RD research is moving closer to using molecular genetic indices of genetic risk to test for gene–environment interactions. These analyses might indicate optimal environments that can serve as protective factors for background genetic risk.

Another clinical application of the identification of candidate genes for RD is the potential to advance the understanding of common comorbidities associated with RD, such as ADHD and speech and/or language disorders. Clinically, these comorbidities can often help in determining the risk for a child to develop RD because the comorbid disorders are usually diagnosed before RD. For example, RD is comorbid with ADHD at a rate of ~25–40% [52]. Similarly, RD is comorbid with speech and/or language disorders at a rate of ~25–30% [53–55]. So far, several overlapping linkage regions have been found for ADHD, speech and/or language disorders and RD [24,56–59]. An explanation of these findings is that these disorders share some QTLs but not others, and the pattern of unique and shared QTLs determines the phenotypic outcome [60]. However, currently there is still a limited understanding of which neural vulnerabilities might 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 QTLs for RD sometimes convey vulnerability for other developmental disorders.

Another clinical application of the identification of candidate genes for RD is the potential to combine knowledge of candidate genes with neuroimaging methods to understand the functional effects of the genes in the brain. This method has previously been implemented with impressive results in studies that investigated variations in the serotonin-transporter gene associated with human amygdala responses [61]. Alternatively, it might be possible to use imaging phenotypes in association screens to identify genes that might contribute to an observed variation in brain activity. These lines of research at the interface between genes and brain 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*. Each gene that has been identified as a candidate for RD (with the exception of *DYX1C1*, the function of which 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

disruptions in global brain development result in such a specific phenotype? The identification of these genes has opened the possibility of early identification of RD via genetic screening, but several unresolved issues currently limit its clinical utility. First and most importantly, causal mutations in these genes have yet to be identified. Indeed, it is possible that the causal variants will reside in regulatory regions where they will be difficult to be detected. In addition, RD is a complex behavioral disorder with a multifactorial etiology that is likely to involve gene–gene and gene–environment interactions. Thus, even when causal mutations are identified, they will need to be considered in the context of a multifactorial framework to make accurate predictions about diagnostic status. The identification of candidate genes for RD can be now used to understand gene–environment interactions and the comorbidities of RD, and to integrate the analysis of genes and brain by using brain imaging technologies. Results of such studies might inform clinical practice by specifying environmental risk and protective factors, increasing the understanding of comorbidity in RD and helping to determine which structures in the neural networks that subserve reading are most important in the pathogenesis of RD.

Although the identification of these genes represents an exciting breakthrough, in many cases it has raised more questions than it has answered (Box 3). To address these questions, studies will require large and diverse samples. There is currently a large European Consortium (Neurodys) (www.neurodys.com) that has been created to put together existing samples across countries and ethnic groups. These samples will give more power to address several of these questions (Box 3). For example, these samples will provide increased power to determine whether genes that were identified through rare translocations are causal in the general population. It will also give increased power to resolve linkage peaks to determine whether an identified candidate gene or a neighboring gene is responsible for the linkage signal. The heterogeneity in genetic background of these samples will also enable studies to test directly 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 that have traditionally been neglected when selecting genetic markers for association analyses. Mutation screening of exons of the candidate genes has yet to find replicable functional mutations in any of these genes. Therefore, studies should begin typing markers in the regulatory regions of genes in addition to exons. The area of gene–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]) to test whether it might serve as a moderator for the association results. Finally, studies should begin to gather phenotypic information about the common comorbidities of RD, such

Box 3. Outstanding questions

DYX1C1

- What is the function of *DYX1C1*?
- Does *DYX1C1* 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 *DYX1C1* or is there another plausible candidate?

ROBO1

- Is the translocation causal in the individual with the translocation and RD [9]?
- Does *ROBO1* account for more-common cases of RD in the population?

DCDC2 and KIAA0319

- Why are there conflicting findings for the candidate genes in the 6p22 region?
- What role does severity selection of the phenotype in the sample have in gene findings?
- What role does covarying IQ from the reading measures have in gene findings?

General questions

- How do genes involved in global brain developmental processes cause relatively specific phenotypes such as RD?
- Which mutations in each gene contribute to the RD phenotype?
- 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–gene and gene–environment interactions have in the development of the phenotype?
- Can the gene identifications help explain common comorbidities of RD?

as ADHD and speech and/or language disorders, for the purpose of conducting bivariate linkage and association studies to determine whether the candidate genes impact the comorbidities of RD, in addition to the RD phenotype itself.

Overall, the identification of these genes has provided glimpses into possible links between genes, brain and behavior, but large gaps in this knowledge still remain. For example, a causal mutation in any of these genes has yet to be identified, leaving the mechanisms by which these 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 the knowledge suggest directions for future research and caution in prematurely applying the findings to clinical issues. Nonetheless, the rapid progress of the field of RD indicates 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).

Supplementary data

Supplementary data associated with this article can be found at [doi:10.1016/j.molmed.2006.05.007](https://doi.org/10.1016/j.molmed.2006.05.007)

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.*, pp. 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, e50
- 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, I.I. 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) Genome-wide 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 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
- 39 Schumacher, J. *et al.* (2006) Strong genetic evidence of *DCDC2* as a susceptibility gene for dyslexia. *Am. J. Hum. Genet.* 78, 52–62
- 40 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
- 41 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.* 15, 1659–1666
- 42 Galaburda, A.M. *et al.* (1985) Developmental dyslexia: four consecutive patients with cortical anomalies. *Ann. Neurol.* 18, 222–233
- 43 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
- 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), pp. 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. *Behav. Genet.* 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 Leffly, D.L. (2001) Early reading development in children at family risk for dyslexia. *Child Dev.* 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. From single to multiple deficit models of developmental disorders. *Cognition* (in press)
- 61 Hariri, A.R. *et al.* (2002) Serotonin transporter genetic variation and the response of the human amygdala. *Science* 297, 400–403
- 62 Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science* 273, 1516–1517
- 63 Newton-Cheh, C. and Hirschhorn, J.N. (2005) Genetic association studies of complex traits: design and analysis issues. *Mutat. Res.* 573, 54–69
- 64 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
- 65 Abecasis, G.R. *et al.* (2001) Association analysis in a variance components framework. *Genet. Epidemiol.* 21(Suppl. 1), S341–S346
- 66 Horvath, S. *et al.* (2001) Family-based association test method: age of onset traits and covariates. *Genet. Epidemiol.* 21(Suppl. 1), S403–S408
- 67 Diao, G. and Lin, D.Y. (2006) Improving the power of association tests for quantitative traits in family studies. *Genet. Epidemiol.* 30, 301–313
- 68 Dudbridge, F. (2003) Pedigree disequilibrium tests for multilocus haplotypes. *Genet. Epidemiol.* 25, 115–121
- 69 Schaid, D.J. (2004) Evaluating associations of haplotypes with traits. *Genet. Epidemiol.* 27, 348–364

Have you contributed to an Elsevier publication?

Did you know that you are entitled to a 30% discount on books?

A 30% discount is available to ALL Elsevier book and journal contributors when ordering books or stand-alone CD-ROMs directly from us.

To take advantage of your discount:

1. Choose your book(s) from www.elsevier.com or www.books.elsevier.com

2. Place your order

Americas:

TEL: +1 800 782 4927 for US customers

TEL: +1 800 460 3110 for Canada, South & Central America customers

FAX: +1 314 453 4898

E-MAIL: author.contributor@elsevier.com

All other countries:

TEL: +44 1865 474 010

FAX: +44 1865 474 011

E-MAIL: directorders@elsevier.com

You'll need to provide the name of the Elsevier book or journal to which you have contributed. Shipping is FREE on pre-paid orders within the US, Canada, and the UK.

If you are faxing your order, please enclose a copy of this page.

3. Make your payment

This discount is only available on prepaid orders. Please note that this offer does not apply to multi-volume reference works or Elsevier Health Sciences products.

For more information, visit www.books.elsevier.com

Supplementary Data

Breakthroughs in the search for dyslexia candidate genes

Lauren M. McGrath¹, Shelley D. Smith² and Bruce F. Pennington¹

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

²University of Nebraska Medical Center, 985456 Nebraska Medical Center, Omaha, NE 68198-5456, USA

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

Table S1. Linkage and association studies for replicated linkage peaks

Linkage Regions	Supportive Results	Negative Results
1p36-p34 (DYX8)	Rabin et al. (1993) [1]	
	Grigorenko et al. (2001) [2]	
	Tzenova et al. (2004) [3]	
2p16-p15 (DYX3)	Fagerheim et al. (1999) [4]	Chapman et al. (2004) [5]
	Fisher et al. (2002) [6]	
	Francks et al. (2002) [7]	
	Petryshen et al. (2002) [8]	
	Kaminen et al. (2003) [9]	
	Peyrard-Janvid et al. (2004) [10]	
3p12-q13 (DYX5)	Nopola-Hemmi et al. (2001) [11]	
	Fisher et al. (2002) [6]	
6p22.2 (DYX2)	Smith et al. (1991) [12]	Field and Kaplan (1998) [13]
	Cardon et al. (1994, 1995) [14,15]	Nöthen et al. (1999) [16]
	Grigorenko et al. (1997) [17]	Petryshen et al. (2000) [18]
	Fisher et al. (1999) [19]	Chapman et al. (2004) [5]
	Gayán et al. (1999) [20]	
	Grigorenko et al. (2000) [21]	
	Fisher et al. (2002) [6]	
	Kaplan et al. (2002) [22]	
	Turic et al. (2003) [23]	
	Marlow et al. (2003) [24]	
	Grigorenko et al. (2003) [25]	
15q21 (DYX1)	Smith et al. (1983) [26]	Rabin et al. (1993) [1]
	Smith et al. (1991) [12]	Bisgaard et al. (1987) [27]
	Fulker et al. (1991) [28]	
	Grigorenko et al. (1997) [17]	
	Nöthen et al. (1999) [16]	
	Nopola-Hemmi et al. (2000) [29]	
	Morris et al. (2000) [30]	
18p11.2 (DYX6)	Chapman et al. (2004) [5]	
	Fisher et al. (2002) [6]	Chapman et al. (2004) [5]
	Marlow et al. (2003) [24]	Schumacher et al. (2006) [31]
Xq27.3 (DYX9)	Fisher et al. (2002) [6]	
	de Kovel et al. (2004) [32]	

Table S2. Summary of candidate gene and association studies^a

Genes	Sample (country)	Type of analysis ^b	Genotyped markers	Summary of primary results	Proposed gene function	Refs.
DYX1C1 (EKN1) Chromosome 15q21	One affected two-generation family with a translocation (Finland)	Translocation t(2;15) (q11;q21)		Translocation cosegregated with dyslexia	Unknown. It contains motifs that are thought to be involved in protein-protein interactions	[33]

	54 dyslexics and 113 controls (Finland)	Case-control association	Eight SNPs identified by screening the cDNA of 20 dyslexics	-3G→A, p = 0.006 1249G →T, p = 0.02 -3A/1249T haplotype, p = 0.02	[33]
	52 dyslexics and 81 controls (Finland)	Case-control association	Eight SNPs	-3G→A, p = 0.02 1249G →T, p = 0.1	[33]
	148 nuclear families (Canada)	TDT and FBAT	Six SNPs (two SNPs from [33] and four SNPs from public databases)	-3A→G, p<.05 -3G/1249G haplotype, p = 0.03 rs11629841, p = 0.02	[34]
	264 nuclear families (UK)	QTDT	Eight SNPs reported by [33]	1249T→ G, p < 0.02, association was not significant after adjustment for multiple tests	[35]
	150 nuclear families (USA)	QTDT	Two significant SNPs from [33]	No significant associations	[36]
	158 nuclear families (Italy)	TDT and FBAT	Three SNPs (including two significant SNPs from [33])	No significant associations	[37]
	57 dyslexics and 96 controls (Italy)	Case-control association	Three SNPs (including two significant SNPs from [33])	No significant associations	[38]
	247 parent-proband trios (UK)	TDT and FBAT	Three SNPs (including two significant SNPs from [33] and one significant SNP from [34])	No significant associations	[39]
ROBO1 Chromosome 3p12-q13	One dyslexic individual with a translocation (Finland)	Translocation t(3;8) (p12;q11)		Individual with the translocation was diagnosed with dyslexia	[40]
	Sixteen members from an affected four-generation family showing linkage to the 3p12-q13 region (Finland)	Polymorphism screening	Seven SNPs identified by polymorphism screening	A specific SNP haplotype segregated with dyslexia in the multiplex family	[40]
DCDC2 and KIAA0319 Chromosome 6p22					
	114 nuclear families (USA)	FBAT and QTDT	31 SNPs spanning 680 kb and ten genes, including KIAA0319 and DCDC2	Significant association with five genes in two clusters: VMP-DCDC2 and KIAA0319-TTRAP-THEM2	[41]

Guidance of axons crossing between brain hemispheres and of dendritic connections.

Both candidates are involved in neural migration

89 nuclear families (UK) (sample 1)	QTD	57 SNPs spanning 225 kb and 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> . Strongest results were obtained in a severity-selected sample	[42]
175 nuclear families (UK) (sample 2)	QTD	20 SNPs that showed association in sample 1	Replication of the associations in sample 1, especially when the sample was severity-selected	[42]
159 nuclear families (USA) (sample 3)	QTD	20 SNPs that showed association in sample 1	Replication of the associations in sample 1, especially when the sample was severity-selected	[42]
240 cases and 312 controls (UK) (sample 1)	Case-control association	Pooled analysis of 137 SNPs in the eight genes identified by [41,42]	17 SNPs showed significant association; 13 of these were with <i>KIAA0319</i>	[43]
223 cases and 273 controls (UK) (sample 2)	Case-control association	Ten SNPs in the region showing association in sample 1	Significant association with <i>KIAA0319</i> , <i>MRS2L</i> and <i>THEM2</i>	[43]
143 parent-proband trios (UK) (sample 3)	UNPHASED	Seven SNPs showing significant association in sample 2	Significant association with <i>KIAA0319</i> , <i>MRS2L</i> and <i>THEM2</i>	[43]
153 nuclear families (USA)	QTD	147 SNPs spanning 1.5 Mb and 18 genes, including <i>KIAA0319</i> and <i>DCDC2</i>	Strongest association peak in <i>DCDC2</i> . A deletion in intron 2 of <i>DCDC2</i> in ten dyslexic families was found	[44]
137 parent-proband trios (Germany) (sample 1)	TDT and QTD	16 STR markers spanning 24Mb and encompassing the <i>VMP-DCDC2-KAAG1</i> gene cluster. Follow-up analysis of the cluster with 25 SNPs and four STR markers	Most significant association with a marker in <i>DCDC2</i> . Results became stronger with severity selection	[45]
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	[45]
376 parent-proband trios (Germany) (sample 3)	TDT	Ten SNPs spanning the <i>KIAA0319-TTRAP-THEM2</i> gene cluster	No significant associations	[45]

*Abbreviations: *DCDC2*, doublecortin domain-containing protein 2; *DYX1C1*, dyslexia susceptibility 1 candidate 1; FBAT, family-based association test; *KAAG1*, kidney-associated antigen 1; QTD, quantitative transmission disequilibrium test; *ROB1*, roundabout *Drosophila* homolog 1; SNP, single-nucleotide polymorphism; STR, short tandem repeat markers; TDT, transmission disequilibrium test; *THEM2*, thioesterase superfamily member 2; *TTRAP*, TRAF and TNF receptor-associated protein; *VMP*, vesicular membrane protein.

*See Box 2 for a full description of the association methods.

References

- 1 Rabin, M. *et al.* (1993) Suggestive linkage of developmental dyslexia to chromosome 1p34–p36. *Lancet* 342, 178
- 2 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
- 3 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
- 4 Fagerheim, T. *et al.* (1999) A new gene (*DYX3*) for dyslexia is located on chromosome 2. *J. Med. Genet.* 36, 664–669
- 5 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
- 6 Fisher, S.E. *et al.* (2002) Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. *Nat. Genet.* 30, 86–91
- 7 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
- 8 Petryshen, T.L. *et al.* (2002) Supportive evidence for the *DYX3* dyslexia susceptibility gene in Canadian families. *J. Med. Genet.* 39, 125–126
- 9 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
- 10 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
- 11 Nopola-Hemmi, J. *et al.* (2001) A dominant gene for developmental dyslexia on chromosome 3. *J. Med. Genet.* 38, 658–664
- 12 Smith, S.D. *et al.* (1991) Screening for multiple genes influencing dyslexia. *Reading and Writing* 3, 285–298
- 13 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
- 14 Cardon, L.R. *et al.* (1994) Quantitative trait locus for reading disability on chromosome 6. *Science* 266, 276–279
- 15 Cardon, L.R. *et al.* (1995) Quantitative trait locus for reading disability: correction. *Science* 268, 1553
- 16 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
- 17 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
- 18 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
- 19 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
- 20 Gayan, J. *et al.* (1999) Quantitative-trait locus for specific language and reading deficits on chromosome 6p. *Am. J. Hum. Genet.* 64, 157–164
- 21 Grigorenko, E.L. *et al.* (2000) Chromosome 6p influences on different dyslexia-related cognitive processes: further confirmation. *Am. J. Hum. Genet.* 66, 715–723
- 22 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
- 23 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
- 24 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
- 25 Grigorenko, E.L. *et al.* (2003) Continuing the search for dyslexia genes on 6p. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 118, 89–98
- 26 Smith, S.D. *et al.* (1983) Specific reading disability: identification of an inherited form through linkage analysis. *Science* 219, 1345–1347
- 27 Bisgaard, M.L. *et al.* (1987) Dyslexia and chromosome 15 heteromorphism: negative lod score in a Danish material. *Clin. Genet.* 32, 118–119
- 28 Fulker, D.W. *et al.* (1991) Multiple regression analysis of sib-pair data on reading to detect quantitative trait loci. *Reading and Writing* 3, 299–313
- 29 Nopola-Hemmi, J. *et al.* (2000) Two translocations of chromosome 15q associated with dyslexia. *J. Med. Genet.* 37, 771–775
- 30 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
- 31 Schumacher, J. *et al.* (2006) Linkage analyses of chromosomal region 18p11–q12 in dyslexia. *J. Neural Transm.* 113, 417–423
- 32 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
- 33 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
- 34 Wigg, K.G. *et al.* (2004) Support for *EKN1* as the susceptibility locus for dyslexia on 15q21. *Mol. Psychiatry* 9, 1111–1121
- 35 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
- 36 Meng, H. *et al.* (2005) TDT-association analysis of *EKN1* and dyslexia in a Colorado twin cohort. *Hum. Genet.* 118, 87–90
- 37 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
- 38 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
- 39 Cope, N.A. *et al.* (2005) No support for association between dyslexia susceptibility 1 candidate 1 and developmental dyslexia. *Mol. Psychiatry* 10, 237–238
- 40 Hannula-Jouppi, K. *et al.* (2005) The axon guidance receptor gene *ROBO1* is a candidate gene for developmental dyslexia. *PLoS Genet.* 1, e50
- 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 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
- 43 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
- 44 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
- 45 Schumacher, J. *et al.* (2006) Strong genetic evidence of *DCDC2* as a susceptibility gene for dyslexia. *Am. J. Hum. Genet.* 78, 52–62