UNIVERSITY OF CALGARY

Localization of susceptibility genes involved in phonological coding dyslexia by family linkage and linkage disequilibrium studies

by

Tracey Lynn Petryshen

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MEDICAL SCIENCE

CALGARY, ALBERTA
JULY, 2001

Abstract

Dyslexia is a specific disability in learning to read that is independent of normal intelligence and education. The root of this disorder is difficulty processing phonemes, the basic sounds of language. Neurobiological studies demonstrate that dyslexic individuals have a number of anatomical and functional brain anomalies, indicating that dyslexia is a neurodevelopmental disorder. Dyslexia has a significant genetic basis, and is a "multifactorial" or "complex trait" in that multiple genes of small to moderate effect confer compounding risk of the disorder through interactions with each other and environmental risk factors. Identification of susceptibility genes would greatly increase understanding of the biological basis of this condition, and potentially lead to better treatments and earlier diagnosis of children at risk of dyslexia.

Dr. Leigh Field, a geneticist at the University of Calgary, and Dr. Bonnie Kaplan, a behavioural psychologist at the Alberta Children's Hospital and University of Calgary, head a study to identify dyslexia susceptibility genes in 100 families with dyslexic members. The primary goal of this Ph.D. thesis was to investigate a region on chromosome 6 where preliminary studies indicated a dyslexia locus may exist. Using linkage and linkage disequilibrium methods, strong evidence was found for a dyslexia locus (named DYX4) on 6q11.2-q12, and two candidate regions were identified where the *DYX4* gene is most likely located. Two dyslexia loci identified by other researchers were also studied, and while quantitative-trait locus linkage analysis did not replicate the DYX2 locus on chromosome 6p21.3, in agreement with a previous qualitative linkage report by Dr. Field and Dr. Kaplan, the DYX3 locus on chromosome 2p15-p16 was confirmed and the gene was localized to a small interval. In addition, candidate neurotransmitter receptor and transporter genes were investigated and evidence was found for the involvement of dopamine receptor genes *DRD4* and *DRD5*, and gamma-aminobutyric receptor genes *GABRB2* and *GABRG2* in dyslexia susceptibility. In conclusion, this thesis work has contributed significantly to the field of dyslexia genetics, and has provided valuable information for further studies to identify the *DYX3* and *DYX4* genes and to clarify the roles of the neurotransmitter receptor genes in dyslexia.

Acknowledgements

There are a number of people that I would like to acknowledge and thank for their influence on me, particularly while I was carrying out my Ph.D. First, my parents, Ted and Noreen Petryshen, for their unending love and support, and for encouraging me in everything that I do but never expecting more of me than I am capable. Also, my big sister, Heather Chen, for her love and support, and for being my role model and my friend through good times and bad, and my zany brother-in-law, Alex Chen, for being..... ummm.... zany ☺ I would also like to acknowledge Dr. Charlie Boone, who played a key role in my decision to do my Ph.D. (without him, I'd be an optometrist!). A big "thank you" to my supervisor, Dr. Leigh Field, for her excellent training and support, and for allowing my studies to take me all over the world, and my supervisory committee, Dr. Johanna Rommens, Dr. Peter Bridge, and Dr. Robert Bell, for their guidance and encouragement during my thesis project. Many other people were essential for the success of my research, particularly Dr. Bonnie Kaplan, who was always keen to answer my questions and discuss the behavioural psychology side of the project, and Susan Crawford, who was invaluable in providing much of the data used in my research. In addition, the following past and present members of Dr. Field's laboratory: Rose Tobias, Martha Hughes, Elzbieta "Dear" Swiergala, Norma Schmill de French, Sherry Olmstead, Michelle Engel, Christine Binns, Ming Fu Liu, Jordana Tzenova, Malgorzata Zapala, Zhao Zhang, Tam Lu, and Zhengweng Wang. I would also like to acknowledge Francis Ouellette for hosting me in his lab during a "side project" and providing great ideas and advice, as well as a few lattés ☺ And I want to thank my Calgary friends with whom I shared a lot of great times, and who have given me many special memories that I will always treasure. I would especially like to thank my friends Elspeth Twist and Sophie Chabot, who were there for me when I needed help through a difficult time. And last but not least, I want to acknowledge my dog Kona, the most beautiful yellow lab in the world, for many relaxing walks and runs (and swims) at the Bow River.

Table of Contents

List of Tables

List of Figures

ADHD	attention-deficit/hyperactivity disorder
AFBAC	affected family-based controls
ASP	affected sibpair
BAC	bacterial artificial chromosome
CEPH	Centre d'Etude du Polymorphisme Humain
CHLC	Cooperative Human Linkage Center
cM	centiMorgan
DNA	deoxyribonucleic acid
ddH_2O	double-distilled water
df	degrees of freedom
fMRI	functional magnetic resonance imaging
GABA	gamma-aminobutyric acid
GDB	Genome Database
HBPPL	haplotype-based posterior probability of linkage
HLA	human leukocyte antigen
hlod	heterogeneity log of the odds ratio
IBD	identical by descent, identity by descent
IQ	intelligence quotient
kb	kilobase
LD	linkage disequilibrium
LDB	Genetic Location Database
lod	log of the odds ratio (also LOD)
Mb	megabase
MGN	medial geniculate nucleus
MLE	maximum likelihood estimate
MRC	Medical Research Council
NPL	nonparametric linkage
PAC	P1 phage artificial chromosome

PCD	phonological coding dyslexia
PCR	polymerase chain reaction
QTL	quantitative-trait locus
RAN	rapid automatized naming
RFLP	restriction fragment length polymorphism
RH	radiation hybrid
SNP	single nucleotide polymorphism
STS	sequence-tagged site
UV	ultraviolet
WICGR	Whitehead Institute/MIT Center for Genome Research
YAC	yeast artificial chromosome

Chapter One: Introduction

1.1. Dyslexia

Dyslexia is traditionally defined as severe difficulty in learning to read despite normal intelligence, visual acuity, motivation, and instruction (Critchley 1970). Dyslexia affects 3-10% of school-age children (Lerner 1989) and persists into adulthood (Felton et al. 1990). As a result, dyslexia has major negative social, educational, emotional, and economic repercussions (Spreen 1988). Early studies suggest that the ratio of males to females with reading disability is between 2:1 and 5:1 (Critchley 1970; Finucci and Childs 1981). However, more recent studies indicate that this skewed ratio is probably due to biases in subject ascertainment, and that the prevalence rates of reading disability in males and females are nearly 1:1 (Shaywitz et al. 1990; Willcutt and Pennington 2000). The predominant approach to treating dyslexic children is highly structured phonic teaching, in which children are taught to read and pronounce words by learning the basic sounds (phonemes) of letters and letter combinations. This method of treatment is based on substantial evidence that the majority of dyslexic individuals have a specific deficit in the phonological domain of reading, to be discussed below.

Dyslexia co-occurs with a number of emotional and behavioural problems, specifically attention-deficit/hyperactivity disorder (ADHD), anxiety, and depression, more frequently than expected by chance (Gilger et al. 1992; Shaywitz et al. 1995; Willcutt and Pennington 2000). With regard to ADHD, between 15% and 26% of individuals with dyslexia also have ADHD, and while some studies have indicated that the two disorders are genetically independent (Gilger et al. 1992; Faraone et al. 1993), a recent study has found evidence for a genetic overlap between dyslexia and an inattentive subtype of ADHD, but not a hyperactive/impulsive ADHD subtype (Willcutt et al. 2000). There also appears to be a link between dyslexia and autism, since parents of autistic children have an increased frequency of reading problems (Folstein et al. 1999). The co-occurrence of dyslexia with other behavioural disorders suggests that there may be genes with pleiotropic effects (i.e., one gene leads to many different phenotypes) involved in the manifestation of these conditions.

1.2. Cognitive studies of reading and dyslexia

Reading is a multifaceted process that depends upon a number of components, many of which are phenotypically correlated with one another and overlap in a continuum of basic through complex reading skills. Although this overlap often makes it difficult to distinguish one component from another, most reading experts agree that the most basic reading component is phonological awareness. This process is the ability to perceive and manipulate phonemes, which are used in different combinations to make up all words. Phonological awareness develops early in childhood and is used during simple tasks such as rhyming. A related but slightly higher-level reading component, phonological coding (also referred to as phonological decoding), is the ability to sound out written words by using the appropriate grapheme-phoneme (letter-sound) correspondence rules. This skill is often assessed by non-word reading, in which the subject reads nonsense words, such as "pid" or "golup", which follow regular grapheme-phoneme rules but do not exist in the English language, thus preventing reading by use of visual memory. A somewhat distinct component of reading is orthographic coding, which is the ability to recognize words' specific grapheme patterns. This skill is very important in English, since the same word sounds can be represented by different grapheme patterns (e.g. pair, pare, pear). Skilled reading also requires the rapid and automatic identification of words, which calls upon a process referred to as rapid automatized naming (RAN). RAN is thought to be important for the orthographic aspects of word identification, but only weakly involved in the phonological aspects of this task (Bowers and Swanson 1991). At the highest levels of the reading hierarchy are components involved with semantics (vocabulary or word meaning), syntax (grammatical structure), and discourse (connected sentences).

Research on the mechanisms involved in reading development has led to the formulation of several models of reading. In the dual route model, beginning readers recognize printed words using a slow phonological procedure (termed "indirect access"), in which words are sounded out using grapheme-phoneme correspondence rules. Skilled readers, however, recognize familiar words by a rapid lookup procedure (termed "direct access") in which orthographic cues are used to automatically and effortlessly retrieve

the words' meanings and pronunciations. Only when a skilled reader encounters an unfamiliar word will the slower, "indirect" route be utilized. However, closer examination of normal reading has found that recognition of printed words always requires at least some degree of phonological processing and that phonological processing is often rapid (Van Orden et al. 1990; Gough and Walsh 1991; Frost 1998), thus questioning the dual route model. To resolve this issue, some researchers have modified the dual route model such that both routes are activated and may interact during word recognition, and the route that predominates in performance will be the fastest route for that particular situation. Other researchers, however, have turned to a connectionist model of reading (Seidenberg and McClelland 1989). In this model, word recognition occurs by a three-layer neural network, in which the input layer corresponds to graphemes and the output layer corresponds to phonemes, with these layers encoding the regular grapheme-phoneme correspondence rules, and the middle layer allowing for learning of word-specific patterns. However, like the dual route model of reading, the connectionist model is also debated, and further research on reading development, skilled reading, and reading failure is needed to clarify the mechanisms involved in reading.

As mentioned previously, dyslexia is a specific disability in reading that is independent of general intelligence. Under such a broad definition, dyslexia is not a single disorder, but rather a "mixed bag" of heterogeneous reading disorders with a variety of cognitive defects and a wide range of severity. The most notable deficits observed in dyslexic individuals are difficulties in phonological processing, spelling, verbal memory, auditory perception, and rapid naming. It is a common fallacy that letter reversal is the hallmark of dyslexia; however, many beginning normal readers make this error (as do some skilled readers). There is now a consensus among researchers that the key problem in most dyslexics (approximately 80-90%) is a specific difficulty in phonological processing, with some of the other observed deficits resulting from this phonological difficulty. Originally posited by Stanovich (1988), and recently restated by Frith (1998), the phonological deficit hypothesis of dyslexia states that a biological abnormality in the brain causes a specific phonological deficit, which is manifested in

poor phonological awareness, poor reading acquisition, and impaired verbal memory, since phonological codes are primarily used for the retention of verbal information (Johnston et al. 1987; Holligan and Johnston 1988). While the phonological deficit does not extend into non-overlapping reading or cognitive domains, higher-level reading components such as vocabulary and comprehension may not be accessible due to the block at the phonological level. Some individuals with phonological deficits are able to overcome their reading disability, however, using visual memory skills (i.e., memorizing the letter pattern of words) and other compensatory strategies that do not require phonological processing. Cognitive studies have also demonstrated that many dyslexic individuals have deficits in RAN of letters, colours, objects, and numbers. These deficits are most pronounced in speed of naming, and less so in naming accuracy (Denckla and Rudel 1974,1976). In terms of the dual route model of reading discussed above, deficits in phonological processing would cause a block in the "indirect access" route to word identification, while RAN deficits would block the automatic "direct access" route. In support of the dual route model, Wolf (1999) has identified dyslexic individuals who have a deficit in either phonological processing or rapid naming. However, a large group of more severely affected dyslexic individuals was also identified that have deficits in both phonological and RAN, a so-called "double deficit". Under the dual route model, these individuals would have separate alterations during development in each route, which seems unlikely. A more probable explanation is that the two routes interact somehow during development, thus lending support to interacting dual route models and to the connectionist model of reading.

1.3. Neurological studies demonstrate key differences in dyslexic brains

Neuroanatomical studies demonstrate that the brains of dyslexic individuals show a number of anatomical and functional differences in comparison to the brains of normal readers. One of the earliest findings is that dyslexic brains have an increased symmetry or reversed asymmetry in the language area of the left temporal-parietal cerebral lobe (a region called the planum temporale), compared to that seen in normal populations (see review by Shapleske et al. 1999). More recent functional neuroimaging studies

demonstrate that the cortical regions surrounding the left temporal-parietal junction exhibit altered patterns of activation during reading tasks in dyslexics (Rumsey et al. 1997; Shaywitz et al. 1998; Georgiewa et al. 1999; Simos et al. 2000). Taken together, these studies implicate the temporal-parietal region of the brain, particularly the left hemisphere language area, as the principal region of dysfunction in dyslexia. However, other brain anomalies exist in dyslexic patients, suggesting that the reading problems may be caused by functional deficits in areas other than the temporal-parietal region or, alternatively, by deficits in the temporal-parietal region as well as deficits in other regions.

There is a wide body of evidence showing that dyslexics have a specific deficiency in the magnocellular (transient) visual pathway, which processes rapidly moving, gross detail stimuli (Merigan and Maunsell 1993). In contrast, the parvocellular (sustained) pathway, which processes mostly colour and fine detail stimuli (Merigan and Maunsell 1993), appears to be normal in dyslexics. A study of five autopsied dyslexic brains found smaller and more disorganized magnocellular cells in the lateral geniculate nucleus of the thalamus, which receives inputs from the retinal ganglion cells and in turn sends axons to the visual areas of the occipital and parietal cortices, compared to control brains (Livingstone 1991). Furthermore, psychophysical testing, in which subjects respond to visual images presented on a monitor, found that dyslexic individuals have decreased sensitivity for stimulatory conditions where the magnocellular pathway is utilized (Lovegrove 1980; Cornelissen et al. 1995; Demb et al. 1998; Witton et al. 1998). Also, dyslexics have an abnormal pattern of brain visual evoked potentials in response to stimuli specific for the magnocellular pathway (Livingstone 1991; Lehmkuhle et al. 1993; Kubova et al. 1996). Functional magnetic resonance imaging (fMRI) studies, which determine brain activation patterns by measuring blood oxygenation contrast signals, have shown that dyslexics have decreased activation of visual cortical areas while observing magnocellular pathway-specific stimuli (see review by Eden and Zeffiro 1998). In light of the evidence for visual processing deficits in dyslexics, Breitmeyer (1980) proposed that the visual transient system causes reading disability in that an impaired magnocellular system fails to inhibit each reading fixation during saccades

(rapid eye movements) that occur while reading a line of text, leading to superimposition of images and visual confusion. Similarly, Stein and Walsh (1997) suggested that the visual magnocellular system helps to control eye movements and, thus, deficits in this system destabilize binocular fixation so letters appear to move around during reading. However, as will be discussed below, other conflicting theories of the cause of reading difficulties in dyslexics have been proposed.

Deficits in auditory processing have also been demonstrated in dyslexic individuals. Although separate magnocellular and parvocellular pathways have not been clearly distinguished in the auditory system, a number of relay nuclei, including the medial geniculate nucleus (MGN) of the thalamus, contain large neurons responsible for analyzing acoustic stimulus changes. A study of autopsied dyslexic brains determined that the left hemisphere MGN has more small neural cells and fewer large cells compared to control brains (Galaburda et al. 1994). Psychophysical testing has demonstrated that dyslexic individuals have significant impairments in processing rapid sound sequences (Tallal 1980; Hari and Kiesila 1996) and in detecting small changes in frequency of tones (Witton et al. 1998; McAnally and Stein 1996). Dyslexics also have abnormal brain potentials in response to frequency changes or to bursts of tone or speech (McAnally and Stein 1996; Shulte-Korne et al. 1998a). Thus, dyslexics have deficits in processing auditory stimuli, likely through an aberrant "magnocellular-like" subsystem similar to that seen the visual pathway. Tallal et al. (1993) proposed that rapid auditory processing deficits render dyslexics unable to discriminate phonemes, which leads to poor phonological skills, and thus poor reading acquisition.

Studies also demonstrate that dyslexic individuals have difficulties with coordination (Wolff 1993), balance (Nicolson and Fawcett 1990; Yap et al. 1994), postural stability, and muscle tone (Fawcett and Nicolson 1999). Dyslexic adults have also been found to have abnormal activation of the cerebellum during a task consisting of a sequence of finger movements (Nicolson et al. 1999). Because the cerebellum has been regarded as a motor control area, its role in dyslexia has generally been discounted. However, there is emerging but controversial evidence that the cerebellum has connections to the language area and is activated during language-related activities (see

consecutive papers by Leiner et al. 1993, Ito 1993, Glickstein 1993, and Bloedel 1993 for a debate of this issue). It has been proposed, therefore, that cerebellar dysfunction leads to difficulties in acquisition of articulatory and auditory skills (and hence to difficulties in phonological processing), as well as visual skills, and that these difficulties in turn lead to problems in learning to read (Fawcett et al. 1996).

1.4. The temporal processing hypothesis of dyslexia

Interestingly, recent studies have found that the visual and auditory deficits observed in dyslexics are directly correlated to the degree of phonological deficit (Demb et al. 1998; Witton et al. 1998). While these correlations suggest a causal relationship, it is also possible that the deficits arise from a common underlying biological mechanism. One hypothesis that might reconcile the visual, auditory, and motor control deficits postulates that the different impairments stem from a common deficit in the processing of temporal features of various kinds of stimuli. In other words, dyslexic individuals are unable to process rapidly changing stimuli in all sensory modalities, including visual, auditory, and motor domains. The inability to perceive rapid elements of speech leads to phonological deficits (i.e., dyslexics cannot "hear" distinct phonemes), which in turn lead to reading difficulties. Stein and Walsh (1997) suggest that the biological defect underlying the deficits observed in dyslexia may lie in a particular magnocellular neuronal cell line that plays a major role in temporal processing, noting that the defect may be due to a genetic mechanism.

1.5. Genetic studies demonstrate a genetic basis for dyslexia

1.5.1. Twin and family studies

Familial aggregation of dyslexia has been well documented for almost 50 years (Hallgren 1950; Zahalkova et al. 1972; Gilger et al. 1991), and numerous twin and family studies have shown that there is a significant genetic contribution to dyslexia. A study by DeFries et al. (1987) found significant evidence for a genetic aetiology in reading disability in twins. Further investigation found a dyslexia concordance rate of 68% in monozygotic twins and 38% in dizygotic twins (after correcting for

ascertainment bias), suggesting both genetic and environmental components to dyslexia (DeFries and Alarcón 1996). High heritability estimates for several reading components also support a genetic basis of reading disability, with heritabilities of 0.51 for reading comprehension and 0.73 for spelling (Stevenson et al. 1987), and 0.93 for the phonological coding component of word recognition, whereas the orthographic coding component of word recognition was not heritable (Olson et al. 1989). Despite the data indicating a genetic basis for reading disability, the mode of inheritance is not clear. There is evidence for autosomal dominant, autosomal recessive, and polygenic inheritance of dyslexia, and, as seen with most common disorders, genetic heterogeneity is apparent (Hallgren 1950; Finucci et al.1976; Lewitter et al. 1980; Pennington et al. 1991).

1.5.2. Linkage and linkage disequilibrium studies

Linkage analysis tests for cosegregation of a chromosomal marker and disease within a pedigree (or set of pedigrees) to determine whether the marker and a disease-predisposing gene are physically linked (i.e., in close proximity) to each other. Association analysis, however, compares marker frequencies in patients and control individuals and tests for the co-occurrence of a particular marker variant (allele) and the disease at the population level. This association may be the result of co-transmission of the disease gene and marker allele during meiosis (non-independent assortment) due to their close proximity on the same chromosome. In other words, the disease gene and marker allele are in linkage disequilibrium with each other. However, an association can also result from ethnic differences, and thus marker frequency differences, between patients and controls, referred to as population stratification. Thus, family-based association methods are often employed in which marker alleles that are not transmitted from parents to children within families form the control population in the analysis, thus preventing associations due to population stratification.

The first documented linkage study of dyslexia reported linkage between reading disability and chromosome 15 centromeric heteromorphisms (Smith et al. 1983), which could not be replicated by an independent study (Bisgard et al. 1987). Subsequent analyses

by Smith et al. (1991) using additional families found weakly significant results for markers on chromosome 15q distal to the region previously identified. More recently, Grigorenko et al. (1997) found weak linkage between single-word reading (a global measure of reading ability) and chromosome 15q21-q22 (referred to as *DYX1*), located approximately 40cM from the centromere and 25cM from the 15q region identified by Smith et al. (1991). Supportive evidence for linkage was also found between a spelling component of reading disability and *DYX1* (Schulte-Körne et al. 1998b; Nöthen et al. 1999), but linkage was not found in a large kindred with multiple affected members (Sawyer et al. 1998). Significant linkage disequilibrium has also been detected between reading disability and marker haplotypes spanning ~1cM on the centromeric end of this region (Morris et al. 2000a,b). Finally, two families have been identified with balanced translocations ~6Mb apart in the 15q21-q22 region, further supporting the presence of this locus. However, the fact that the translocation breakpoints are far apart suggests that one (or both) of the translocations may occur by coincidence and may not be involved in reading disability, or that two separate dyslexia genes exist in this region, or that the breakpoint of one family disrupts the dyslexia gene while the breakpoint of the other family disrupts a distant regulatory region of this gene. Further studies are therefore necessary to determine whether one or more than one susceptibility locus exists on 15q21-q22.

The second reported dyslexia-predisposing region was identified by Rabin et al. (1993), who found suggestive evidence for linkage between a reading and spelling phenotype and chromosome 1p34-p36, near the Rh locus. Grigorenko et al. (1998) have also reported moderate evidence for linkage in this region; however, other studies have not been able to replicate this linkage (Smith et al. 1998; Sawyer et al. 1998).

The most convincing evidence for a dyslexia susceptibility locus has been found on chromosome 6p21.3-p22 (referred to as *DYX2*), reported by five independent studies. After an initial report of evidence for linkage to the human leukocyte antigen (HLA) region of chromosome 6 (Smith et al. 1989), Cardon et al. (1994) investigated the same kindreds and an independent sample and reported linkage between a composite measure of reading disability and a 2 cM region within HLA. Grigorenko et al. (1997, 2000) have since found evidence for linkage between several reading phenotypes and 6p21.3-p22, a region

overlapping and telomeric to that suggested by Cardon et al. (1994). Furthermore, two studies employing quantitative trait methods reported linkage of phonological and orthographic measures of reading to 6p21.3-p22 (Fisher et al. 1999; Gayán et al. 1999). The linkage to the HLA region is interesting because of evidence of increased immune deficits in dyslexia families (Behan and Geschwind 1985; Pennington et al. 1987; Crawford et al. 1994). However, Turic et al. (2000) recently reported significant linkage disequilibrium between dyslexia and marker haplotypes spanning 1.5Mb on the telomeric end of this region (6p22), outside of HLA, suggesting that *DYX2* is more probably not an HLA gene and thus may not be involved in immune function. In contrast to the above linkage reports, Field and Kaplan (1998) investigated a large sample of families and were unable to find significant evidence for linkage to *DYX2* using a qualitative phonological coding dyslexia phenotype, and Sawyer et al. (1998) were also unable to replicate this linkage. Thus, while several studies have provided convincing evidence for the existence of the DYX2 dyslexia susceptibility locus, it appears that this gene may be responsible for only a proportion of dyslexia cases.

Most recently, a dyslexia locus on chromosome 2p15-p16 (*DYX3*) has been identified in a large Norwegian family with autosomal dominant inheritance of dyslexia (Fagerheim et al. 1999). While statistically significant evidence for this locus was reported, replication in an independent sample would be helpful in confirming and localizing the locus. However, there have not been any reports to date of either replication or nonreplication of *DYX3*.

1.6. Genetic investigation of complex traits

In the studies discussed above, some groups were able to replicate and thereby confirm previous linkage findings, while other groups were unable to do so. This is a familiar scenario for common genetic disorders, and is the result of the complex genetic and environmental components underlying these disorders (Lander and Schork 1994). With respect to the genetic complexity, the general view is that multiple genes of small to moderate effect confer compounding disease risk through interactions with each other and with non-genetic risk factors. Thus, the mode of inheritance of complex traits is generally

considered to be oligogenic (few genes work together to confer risk) or polygenic (many genes together confer risk). The same genes may be commonly involved in conferring disease risk across populations, or they may vary in number and strength between populations, potentially to the degree that, in certain populations, some genes are "major" susceptibility genes with a significant genetic effect. These "major" genes may display complete penetrance or, more often, incomplete penetrance that is dependent on age, sex, other genes, and/or the environment. In addition to this locus heterogeneity, in which different genes confer variable risk in different populations, complex disorders may also display allelic heterogeneity, in which different mutations in a single gene confer risk of the disorder. To highlight the effect of genetic heterogeneity in dyslexia studies, in a recent follow-up study of the 6p21.3 dyslexia locus, Grigorenko and colleagues reported that after increasing the number of dyslexia families, the reading components showing linkage to this locus were different than in the original study (Grigorenko et al. 1997, 2000). Thus, the detection of linkage (and linkage disequilibrium) appears to depend on the sample being studied, due to varying proportions of families linked to different regions in different samples (i.e., locus heterogeneity).

Another factor that hampers the identification of complex trait genes, and is especially problematic in studies of dyslexia, is phenotypic uncertainty. Because many dyslexic individuals have deficits in reading components in addition to phonological deficits, it is difficult to precisely define the dyslexia phenotype. As a result, many genetic studies discussed above utilized various reading components, usually including phonological coding, to assess reading disability. Several reading skills were found linked to the same chromosomal region, suggesting that one gene may play a role in multiple aspects of reading, which is not surprising given the substantial correlations among reading skills. In contrast, separate reading skills also showed linkage to *different* chromosomal regions, implying that independent genes may be involved in distinct reading skills. However, as Pennington (1997) points out, "the variance and reliability of the behavioral phenotypes and differences in informativeness of (genetic) markers may lead to seemingly different linkage results for two correlated phenotypes". As a result, the reading phenotype may affect the ability to detect linkage and/or linkage

disequilibrium to a susceptibility locus, and the use of different phenotypes in different studies may be one reason that identified loci have not been confirmed in all studies.

Because of the difficulties encountered in genetic studies of complex traits, a number of strategies have been employed to try to define a more genetically homogeneous trait from a complex trait, thereby increasing the chances of identifying susceptibility genes. One strategy is to define a specific phenotype, alluded to above, or to study severe cases of the disorder. Another method is to focus on early-onset cases, which has proven successful in studies of Alzheimer's disease, in which the presenilin 1 (*PS1*), presenilin 2 (*PS2*), and beta amyloid protein (*beta APP*) susceptibility genes were identified, and hereditary breast cancer, in which the breast cancer gene 1 (*BRCA1*) is involved. Focusing on "high-risk" families with multiple affected members also allowed the identification of the *BRCA2* breast cancer susceptibility gene. Another way to improve the prospects of identifying complex trait genes is to focus on specific ethnic groups, in which there will be greater genetic and allelic homogeneity than in a mixed (outbred) population. Having said this, recent linkage disequilibrium studies of outbred populations *have* been successful in implicating particular genes in susceptibility to complex traits, such as the peroxisome proliferator-activated receptor-γ (*PPARγ*) and the calpain-10 (*CAPN10*) genes in type II diabetes (Altshuler et al. 2000; Horikawa et al. 2000). However, in these cases, the disease associations are not as definitive as for the examples discussed above and, ultimately, functional and mutational studies will be required to prove that these genes are indeed involved in type II diabetes. Because of the greater genetic heterogeneity in mixed populations, there is increasing consensus in the literature that extremely large sample sizes will be needed to identify complex trait genes (Altshuler et al. 2000, Rao 2001). However, cost and time constraints may make it more feasible for investigators to combine samples for analysis, as long as the samples are similar with respect to ascertainment scheme and phenotypic measurements, among other considerations. Alternatively, meta-analysis of results from multiple independent linkage studies may be successful in identifying complex trait loci.

1.7. Objective

To identify loci involved in susceptibility to dyslexia, Dr. Leigh Field and Dr. Bonnie Kaplan have been conducting a linkage study of Canadian families with dyslexic members. The primary objective of my Ph.D. thesis was to investigate a region on chromosome 6q where preliminary analyses found evidence for linkage to dyslexia. I also followed up previous studies in the laboratory on the chromosome 6p dyslexia region (*DYX2*) by performing quantitative-trait locus (QTL) linkage analyses of this region. Furthermore, I investigated the chromosome 2p15-p16 region (*DYX3*) for evidence of linkage and linkage disequilibrium in our family dataset. And finally, I investigated markers from loci of candidate gamma-aminobutyric acid (GABA) receptor genes, dopamine receptor genes, and the dopamine transporter gene for linkage and linkage disequilibrium in our family sample, to determine whether any genetic evidence could be established to suggest a role for these genes in dyslexia.

Chapter Two: Subjects and Methods

2.1. Study design

As discussed in the Introduction, the majority of dyslexic individuals have deficits in phonological coding, and the high heritability of phonological coding indicates that this skill has a large genetic basis. In 1991, Dr. Field and Dr. Kaplan began a study to identify the genetic loci involved in susceptibility to phonological coding dyslexia (PCD). In brief, families with reading disabled members were primarily ascertained from local learning disability schools, although local media attention also attracted families to the study. All family members underwent a battery of psychometric reading tests to assess four reading components; adults were also assessed for reading history. This information was used to determine a PCD diagnosis for all members, thus the prior reading disability diagnoses that initially brought the families into the study were confirmed by independent methods. All subjects supplied a blood sample from which DNA was extracted, and microsatellite marker genotyping was performed. To find PCD susceptibility loci, parametric (model-based) and nonparametric (model-free) linkage analyses and linkage disequilibrium analyses of the PCD phenotype were performed. In addition, because analysis of a quantitative trait is often more powerful to detect linkage than analysis of a qualitative trait (Wijsman and Amos 1997), and because quantitative linkage analysis methods for use in human studies have undergone significant development in the past few years, I performed quantitative-trait locus (QTL) linkage analyses of the four reading components.

2.2. Subjects

A total of 100 families with at least two PCD siblings (by our diagnosis, see Section 2.5) were ascertained from Calgary-area schools for learning disabled children and through the local media. Exceptions to the "two PCD siblings" criterion were three families that had only one affected member, and three families that had two or more affected members who were not siblings. These families were included in the study since they had already undergone extensive psychometric testing and had DNA sampled when it

was determined that they did not contain two affected siblings, and they had the potential to contribute information to the linkage analyses, particularly QTL linkage analyses. All subjects were >8 years of age and gave informed consent (for children, one parent gave informed consent) in accordance with the University of Calgary Ethics Review Board. The initial study design was to collect nuclear families consisting of both parents, two or more affected siblings, and unaffected siblings, if available. As outlined in Table 2.1, of the 100 families, 50 were nuclear families (although note that seven of these families had only one parent participate). However, many families that entered the study also had affected extended family members who were invited to participate in the study, thus 50 of the 100 families were extended kindreds (a nuclear family and additional branches). Although detailed ethnic information was not obtained during ascertainment, family surnames and Dr. Kaplan's familiarity with the families were used to determine ethnic backgrounds. Ninety-five of the families were considered to be of European ancestry, and one extended pedigree was of European ancestry except for a married-in African-American father and his two children. Four nuclear families were of non-European ancestry (one nuclear family each consisted of a Chinese father, a Japanese father, a First Nations mother, and a Middle Eastern father). Of the total 919 participants, 554 were adults and 365 were children <18 years of age. The number of subjects that underwent psychometric testing and had DNA sampled was 901, and an additional 18 subjects had DNA sampled but did not undergo psychometric testing, but instead were diagnosed based on reading history only. To create complete pedigrees for linkage analyses, family members who were either deceased or declined participation were included in some pedigrees, resulting in a total of 1092 individuals used in the linkage analyses. Many families appeared to exhibit autosomal dominant transmission, even though ascertainment was without regard to the affection status of parents. Approximately one-third of the families were bilineal pedigrees, in which both parents of a nuclear pedigree, or two parents in an extended pedigree, had a personal or family history of reading problems, suggesting assortative mating (choosing a mate based on phenotype).

Because ascertainment of families spanned several years, genetic analysis began before ascertainment was completed. Thus, the earliest investigation, discussed in

Table 2.1

Description of family samples

Sample	Number of families			participants (adult, child[a])	% PCD affected, unaffected, uncertain	male:female ratios
	nuclear	extended	European/Non-European			
79 families	45	34	76 / 3	615 (374, 241)	53%,33%,14% *adults:* 40%,44%,16% *children:* 75%,16%,9%	1.6:1 affected 1:1.7 unaffected 1.2:1 uncertain
83 families	43	40	83 / 0	805 (485, 320)	50%,34%,16% *adults:* 37%,44%,19% *children:* 71%,18%,11%	1.7:1 affected 1:1.7 unaffected 1:1 uncertain
96 families	46	50	96 / 0	902 (548, 354)	52%,33%,15% *adults:* 39%,42%,18% *children:* 72%,18%,10%	1.8:1 affected 1:1.7 unaffected 1:1 uncertain
100 families	50	50	96 / 4	919 (554, 365)	52%,32%,16% *adults:* 39%,42%,19% *children:* 73%,17%,10%	1.7:1 affected 1:1.6 unaffected 1.1:1 uncertain

[a] Less than 18 years of age.

Chapter Three (Absence of Significant Linkage Between Phonological Coding Dyslexia and Chromosome 6p23-21.3 (*DYX2*) Using Quantitative-Trait Methods), was conducted using a sample of 79 families. As outlined in Table 2.1, 45 of the 79 families were nuclear families (note that three families had only one parent participate), and 34 were extended pedigrees. Note that this sample of 79 families included three of the non-European families and the African-American married-in father and his two children. There were a total of 711 individuals used in linkage analyses, with 615 subjects (86%) participating in the study (i.e., underwent PCD diagnosis and had DNA sampled), of whom 374 were adults and 241 were children.

The investigations discussed in Chapter Four (Evidence for a Dyslexia Susceptibility Locus (DYX4) on Chromosome 6q11.2-q12) and Chapter Five (Confirmation of the *DYX3* Dyslexia Susceptibility Gene on Chromosome 2p15-p16) were conducted after ascertainment was completed. These studies used a sample of 96 European families, which included 76 of the 79 families described above (non-Europeans were excluded), additional members of these families, and new families. As outlined in Table 2.1, 46 families were nuclear pedigrees (note that five families had only one parent participate) and 50 families were extended kindreds. This sample comprised 1071 individuals for linkage analyses, with 902 subjects (84%) participating in the study, of whom 548 were adults and 354 were children.

The investigation discussed in Chapter Six (Involvement of Neurotransmitter Receptor Genes in Susceptibility to Dyslexia) was conducted after ascertainment was completed. This study used a sample of 83 European families, which was the same as the sample of 96 families described above except that it did not include 13 bilineal families. These 13 families were not expected to contribute greatly to the linkage analyses, so they were excluded from the sample to preserve resources. As outlined in Table 2.1, 43 of the 83 families were nuclear pedigrees (note that five families had only one parent participate) and 40 families were extended kindreds. This sample comprised 938 individuals for linkage analyses, with 805 subjects (86%) participating in the study, of which 485 were adults and 320 were children.

2.3. Reading phenotypes

All participants completed a battery of psychometric tests that assessed four components of reading: phonological awareness, phonological coding, spelling, and rapid automatized naming (RAN). Adult subjects also participated in an eight-item structured interview that assessed reading history. The first three reading components were assessed since they utilize some degree of phonological skills, which was our primary interest because of the deficits in these skills in dyslexic individuals. Thus, these three reading measures were used to determine each subject's PCD diagnosis, which was used in qualitative linkage analyses and linkage disequilibrium analyses. In addition, each of these three measures was used separately as a continuous measure in QTL linkage analyses. The fourth reading component (RAN) was assessed because of reports of deficits in rapid naming of pictured objects, colors, numbers, and letters in dyslexic people (Denckla and Rudel 1976; Wolf et al. 1986; Felton et al. 1990; Bowers and Swanson 1991). Since it appears that the RAN deficits and phonological deficits may be independent from each other (Wolf 1999), the RAN measure was not used for PCD diagnosis. RAN speed was used for QTL linkage analyses, however. Full-scale intelligence quotient (IQ) was estimated using a short form of the Wechsler Intelligence Scale for Children for subjects 8-16 years of age (Wechsler 1974), and a short form of the Wechsler Adult Intelligence Scale for adult subjects (Wechsler 1981). Since these IQ tests included reading components, the scores of dyslexic individuals were expected to be depressed. For this reason and because the IQ test only *estimated* the IQ, low IQ was not used as an exclusionary criterion and IQ was not used for diagnosis of PCD or for QTL linkage analyses.

2.4. Quantitative reading measures

2.4.1. Phonological awareness

Phonological awareness, the ability to recognize and manipulate phonemes, was assessed using the Auditory Analysis Test (Rosner and Simon 1971). For this test, the subject was asked to repeat a word, for example the word "cat", and then asked to say the word that results from removing a particular phoneme, for example the phoneme

"kuh". The correct response in this case would be "at". The number of correct responses to 23 questions was recorded. Statistical analyses (see Section 2.6) showed that the distribution of scores was negatively skewed towards the maximum raw score of 23 (15% of children and adults scored 23). This observation, in which the test results plateau at upper scores, is referred to as ceiling effects, and may be due to an inappropriate test design for measurement of higher skill levels, such as in adults. Alternatively, the skill being assessed may be learned at an early age and not improve past a certain age, thus a sample containing a large proportion of older subjects will have a score distribution that is skewed towards high values. Unfortunately, retesting of subjects with a more sensitive test battery without ceiling effects was not possible. Since statistical analyses also showed that 7% of the variation in the Auditory Analysis Test in children was due to variation in age, raw test scores were age-adjusted for subjects less than 18 years of age. To try to ameliorate the ceiling effects, raw scores of 23 were excluded prior to age adjustment, since subjects may have scored higher if the test had allowed, thus scores of 23 may be unreliable. Adjustment was performed on the remaining scores using the formula: raw score + [(18-age) x 0.716], with the age-adjustment factor (0.716) determined by statistical analyses of all children in the dyslexia families and 112 control children (see Section 2.7). However, this adjustment method resulted in a smaller data set with reduced variance, which was expected to have less power to detect linkage by QTL methods. Thus, adjustment was also performed using all scores (i.e., including scores of 23) by the same formula as above. This data set had higher variance and was expected to have more power to detect linkage, thus further analyses were only performed on data adjusted using the latter method.

2.4.2. Phonological coding

Phonological coding is the ability to apply grapheme-phoneme correspondence rules to the pronunciation of nonwords, and was tested by the word attack subtests of the Woodcock Johnson Psychoeducational Battery - Revised (Woodcock and Johnson 1989) and the Woodcock Reading Mastery Test (Woodcock 1987). For these tests, the subject was presented with a written "nonsense word" (a word that does not exist in the English

language), and asked to pronounce this word. The number of correct responses to 30 trials was recorded for both tests. For the former word attack subtest, raw scores were converted into age-adjusted standard scores using norms provided in the test protocol. For the latter word attack subtest, raw scores were converted into age-equivalent scores (maximum 18 years of age) using norms provided in the test protocol. Since these age-equivalent scores are not as informative as standard scores, the latter word attack subtest was not used for QTL linkage analyses, and the descriptive statistics of this measure were not determined. Normative data provided with both phonological coding subtests show ceiling effects in adults.

2.4.3. Spelling

Spelling ability draws upon both phonological skill and orthographic skill (recognition of the symbols that represent sounds), with novice readers primarily relying on phonological cues and experienced readers generally using orthographic cues to spell (Ehri 1995). Poor spelling in dyslexic individuals is resistant to remediation and persists throughout adulthood, although some adults with dyslexia have normal spelling ability, presumably due to superior visual memory (Moats 1995). Spelling skill was measured using the Level 1 (ages 5-11 years) and Level 2 (ages 12-75 years) Spelling subtests of the Wide-Range Achievement Test – Revised (Jastak and Wilkinson 1984). The subject was asked to write a recited word, and the number of correct responses to 45 trials (Level 1 test) or 46 trials (Level 2 test) was recorded. Raw scores were transformed into age-adjusted standard scores using norms provided in the test protocol. Normative data provided with this test show ceiling effects in adults.

2.4.4. Rapid automatized naming speed

RAN, the ability to quickly recall and verbalize the name of a presented object, was assessed using the orthographically based Rapid Automatized Naming of Numbers Test (Denckla and Rudel 1974; 1976) in which subjects quickly recite a list of 50 digits. The number of correct responses (accuracy) and the time required to complete the list (regardless of whether the response was correct or incorrect), which was then converted

into digits recited per second (speed), were recorded. Strong ceiling effects in accuracy were observed across all ages (most subjects obtained no errors), thus the accuracy measure exhibited extremely low variance and therefore was not used in QTL linkage analyses (note that RAN speed *was* used for QTL analyses). Statistical analysis showed that 18% of the variation in RAN speed scores in children was due to variation in age, thus age adjustment was performed for subjects less than 18 years old using the formula: raw score + [(18-age) x 0.148]. The age-adjustment factor (0.148) was determined by statistical analyses of all children in the dyslexia families and 112 control children (see Section 2.7).

2.5. Qualitative phonological coding dyslexia phenotype

Reading is a complex skill drawing on multiple components. Many reading experts consider the best indicator of reading disability to be a clinical diagnosis based on multiple criteria, and they disagree with the partitioning of reading disability into separate traits, which has been done in several dyslexia linkage studies. For this reason, we employed a qualitative diagnosis (affected, unaffected, uncertain) of PCD for use in linkage and linkage disequilibrium analyses.

Each subject was classified as PCD "affected", "unaffected", or "uncertain" by Dr. Kaplan and another psychologist using a consensual coding scheme. Children were diagnosed as affected if there was ≥ 2 year difference between chronological age and reading test performance, primarily using the scores on the two word attack (phonological coding) tests, with the spelling and phonological awareness tests used to refine the diagnostic certainty (RAN was not used for PCD diagnosis). For adults, the phonological awareness, phonological coding, and spelling test scores were all considered in the diagnosis, and particular weight was given to the reading history (for further details, see Field and Kaplan 1998). Particularly for adults, cutoff scores could not be used rigidly for the phenotype definition due to the importance of considering the clinical history: the fact that one of the word attack subtests has published norms only through 18 years of age, and the presence of ceiling effects in adults in some of the tests may render the test results inaccurate. Subjects with ambiguous test scores and untested

pedigree members were classified as uncertain, with the exception of 18 subjects who were diagnosed based on a clear reading history. The use of the uncertain diagnosis allowed subjects to remain neutral with respect to PCD phenotype, but to provide marker genotype information in linkage analyses and linkage disequilibrium analyses. As outlined in Table 2.1, in the sample of 79 families used in the investigation discussed in Chapter Three, 53% of subjects were diagnosed as affected, 33% as unaffected, and 14% as uncertain, with a male:female sex ratio in affected individuals of 1.6:1. In the sample of 83 families used in the investigation discussed in Chapter Six, 50% of subjects were diagnosed as affected, 34% as unaffected, and 16% as uncertain, with the affected male:female sex ratio of 1.7:1. And finally, in the sample of 96 families used in the investigations discussed in Chapters Four and Five, 52% of subjects were diagnosed as affected, 33% as unaffected, and 15% as uncertain, with the affected male:female sex ratio of 1.8:1. The large proportion of affected individuals in the samples was a reflection of the dominant-like transmission observed in many pedigrees.

2.6. Statistical analyses

Statistical analyses were performed using the SAS package (SAS Institute 1990) by Dr. Ming Fu Liu from Dr. Field's laboratory. Specifically, age-adjustment analysis was performed for the phonological awareness and RAN speed raw scores, since the corresponding psychometric tests do not include norms for standardization of scores. Age adjustment factors were determined from analysis of all of the dyslexia family children and 112 control children (Section 2.7). After conversion of the raw scores of the phonological awareness, phonological coding, spelling, and RAN speed tests to either age-adjusted or standardized scores, the descriptive statistics (mean, standard deviation, minimum, maximum, skewness, and kurtosis) were calculated. Finally, Pearson correlation analyses were performed by analyzing the PCD phenotype using three categories (category 1=affected; category 2=uncertain; and category 3=unaffected), and by analyzing the reading measures as continuous variables in which a higher test score corresponded to increased skill.

2.7. Control subjects

Children with normal reading ability (N=112) ascertained for a separate dyslexia study were used as controls in statistical analyses of the quantitative reading measures used in this study. Control ages ranged from 8-16, and the male:female sex ratio was 2.7:1. Each child underwent psychometric testing as described in Section 2.4. Phonological awareness and RAN speed raw test scores of the control children were used to calculate age-adjustment factors (see Section 2.6). Descriptive statistics were determined (see Section 2.6) for the phonological awareness, phonological coding, spelling, and RAN speed measures (after raw scores were age adjusted or converted to standard scores) and the estimated IQ (which was not used in QTL linkage analyses), as shown in Table 2.2. The statistics for the phonological awareness and RAN speed measures were particularly informative for comparison to these measures in the dyslexia family samples, since the corresponding psychometric tests do not include norms.

Table 2.2

Descriptive statistics of phonological awareness, phonological coding, spelling, RAN speed, and estimated IQ in the sample of control children

Trait	Mean	SD	Min	Max	Skewness	Kurtosis
Phonological awareness[a]	23.48	4.00	9.90	28.35	-1.20	1.05
Phonological coding[b]	112.72	13.23	82	138	-0.25	-0.65
Spelling[b]	105.33	12.00	74	133	0.08	0.16
RAN speed[a]	3.52	0.54	2.19	4.80	0.08	-0.48
Estimated IQ[b]	110.74	12.75	80	141	-0.06	-0.45

NOTE — For all traits, lower scores indicate greater deficit.

[a] Data are age-adjusted scores.

[b] Data are standard scores according to test norms (normal population mean = 100, standard deviation =15).

2.8. Microsatellite marker genotyping

Genomic DNA was extracted from 14ml whole blood using a salting out procedure modified from Miller et al. (1988). Blood cells were separated from plasma by centrifugation (1600rpm for 20 minutes) and incubated in 50ml red blood cell lysis buffer (0.144M NH_4Cl, 1mM $NaHCO_3$) for 15 minutes at room temperature. White blood cells were collected by centrifugation (1500rpm for 15 minutes) and incubated in 2ml nuclei lysis buffer (10mM Tris, 0.4M NaCl, 2mM EDTA, pH 8.2) with 0.1% SDS and 50µl 20mg/ml protease K at 37°C for 3 hours to overnight. DNA was separated from proteins by addition of 2ml 6M saturated NaCl and 2ml ddH_2O, vigorous shaking, centrifugation (3000rpm for 45 minutes), and addition of 2.5x volume 95% EtOH to the supernatant. Precipitated DNA was isolated using a glass pipette, rinsed in 70% EtOH, air dried, resuspended in 1ml low TE (10mM Tris pH 8.0, 0.1mM EDTA pH 8.0), and stored at –80°C. DNA concentrations were determined by UV spectrophotometry. 5ng/µl DNA stock samples were prepared and stored at 4°C. The amount of each DNA sample required for genotyping (either 20ng or 25ng, see below) was aliquoted into a 96-well Thermowell PCR plate (Costar) and dried at room temperature, and plates were either used immediately or stored at 4°C until use.

Microsatellite marker genotyping was performed with technical assistance by manual or automated methods. For manual genotyping, the concentrations of reagents in each PCR amplification reaction were as follows: 25ng dried DNA, 0.6µM forward and reverse primers, 1x buffer N with 1.5mM $MgCl_2$ (50mM KCl, 10mM Tris pH 8.3, 170µg/ml bovine serum albumin, 0.05% Tween 20, 0.05% Nonidet P-40, 1.5mM $MgCl_2$), 200µM each of dATP, dGTP, and dTTP, 2.5µM dCTP, 0.4MBq α-^{32}P-dCTP, and 0.5units Taq DNA polymerase (Gibco BRL) in a final volume of 10µl, and each well was overlaid with mineral oil. PCR amplification was performed with either an Ericomp EasyCycler Series thermocycler or a Perkin Elmer GeneAmp PCR System 9600 thermocycler. The PCR program was as follows: 6 minutes denaturation at 94°C, 28 amplification cycles of 1 minute at 94°C, 2 minutes at 55°C, and 1 minute at 72°C, followed by 6 minutes extension at 72°C, and soak at room temperature. In cases of poor amplification, the annealing temperature was altered depending on primer GC contents.

Following PCR amplification, 5µl of loading dye (20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol, 92% formamide) were added to each 10µl sample, of which 2µl were loaded onto a 6% acrylamide vertical gel (6% acrylamide, 420mg/ml urea, 0.1% TEMED, 0.25% ammonium persulfate, 1xTBE), with 2µl bacteriophage M13mp18 DNA sequencing ladder loaded at regular intervals, and electrophoresed in 1xTBE buffer (10.8g/L Tris, 5.5g/L boric acid, 2mM EDTA pH 8.0) at 1000V. Following electrophoresis, gels were exposed to Kodak autoradiography film overnight at room temperature, and processed using an automated developer. Two laboratory members independently called marker alleles, and allele sizes were manually entered twice into the database.

For automated genotyping, the concentrations of reagents in each PCR amplification reaction were as follows: 20ng dried DNA, 0.02µM M13-tailed forward primer, 0.02µM reverse primer, 1x buffer N with 2mM $MgCl_2$ (50mM KCl, 10mM Tris pH 8.3, 170µg/ml bovine serum albumin, 0.05% Tween 20, 0.05% Nonidet P-40, 2mM $MgCl_2$), 200µM each of dATP, dGTP, dTTP, and dCTP, 0.02µM IRDye-700 or -800 (Licor), and 0.25units Taq DNA polymerase in a final volume of 5µl, and each well was overlaid with mineral oil. PCR amplification was performed with an Ericomp EasyCycler Series thermocycler. The PCR program was as follows: 5 minutes denaturation at 94°C, 30 amplification cycles of 30 seconds at 94°C, 75 seconds at 55°C, and 15 seconds at 72°C, followed by 6 minutes extension at 72°C, and soak at room temperature. In cases of poor amplification, the annealing temperature was altered depending on primer GC contents. Following PCR amplification, 2µl of IR^2 Stop Solution (Licor) were added to each 5µl sample, of which 1µl was loaded onto a 5% acrylamide 18cM vertical gel (1x FMC Bioproducts Long Ranger Gel Solution, 350mg/ml urea, 0.05% TEMED, 0.05% ammonium persulfate, 1xTBE), with 1µl IRD700 or IRD800 50-350bp Sizing Standard (Licor) loaded at regular intervals, and electrophoresed in 1xTBE buffer at 1200V on a LI-COR 4200S-2 Gene ReadIR DNA Analyzer. Marker alleles were called by GeneImagR software and confirmed by a laboratory member, and allele sizes were automatically entered into the database. Two

control individuals were genotyped and electrophoresed on every gel to assist in consistency in allele calling.

2.9. Determination of marker allele frequencies

Marker allele frequencies required for linkage analyses were determined by counting alleles in the parents of one nuclear family selected from each pedigree. Nuclear families were selected on the basis of: 1) having at least two affected child siblings (<18 years of age), or if that was not possible, then 2) having one affected child, or if that was not possible, then 3) having two or more affected adult siblings. Marker allele frequencies were determined from sample parents because linkage analyses use allele frequencies to infer missing genotypes in the pedigrees, which thereby has an effect on the calculation of the likelihood of linkage. Thus, it is advisable to perform linkage analyses using allele frequencies estimated from the sample under study, rather than using published allele frequencies that are determined from another sample. However, there were relatively few missing genotypes in the dyslexia pedigrees (e.g. 84% - 86% of individuals were genotyped, depending on the sample; see Section 2.2), thus the linkage analyses were not heavily dependent on marker allele frequencies.

2.10. Sequence-tagged site content mapping

Human physical maps consist of contigs of overlapping clones of human sequence. The earliest physical maps are based on yeast artificial chromosomes (YACs), which consist of yeast centromeres, telomeres, and autonomous-replication sequences (similar to replication origins), and human DNA segments approximately one Megabase (Mb) in size. However, YACs are limited by problems of chimerism (containing fragments from more than one genomic region) and instability in some regions, thus later maps utilize the more stable bacterial artificial chromosomes (BACs) and P1 phage artificial chromosomes (PACs). Sequence-tagged sites (STSs), which are simply unique chromosomal markers in the genome, have been localized on the physical maps by PCR amplification. While physical maps allow localization of markers on a finer scale than genetic maps, they suffer from problems of inappropriate overlaps due to chimerism (for

YAC-based maps) and incorrect contig orientation (thus incorrect marker order) due to gaps between contigs. Thus, physical map marker locations may be inaccurate.

YAC clones created by the Centre d'Etude du Polymorphisme Humain (CEPH) (Dausset et al. 1992) were selected from the Whitehead Institute/MIT Center for Genome Research (WICGR) STS map (Hudson et al. 1995) and were obtained from the MRC Genome Resource Facility in Toronto, Ontario. Each clone was thawed from -80°C, streaked onto selective media (SD-URA-TRP+Amp), and incubated at 30°C for 48 hours. DNA from each clone was prepared with technical assistance in the following manner: a single colony from the selective plate was inoculated into 10ml YEPD+ADE and grown for 36 hours in a 30°C shaking water bath. Cells were isolated by centrifugation (2000rpm for 5 minutes), washed with ddH_2O, and sheared by vortexing for 3 minutes following resuspension in 200µl GDIS (2% Triton X-100, 0.1% SDS, 2% 5N NaCl, 1mM EDTA, 10mM Tris pH 8.0), 200µl phenol:chloroform, and addition of 0.35g acid-washed glass beads. Following centrifugation (12000rpm for 4 minutes) and collection of the aqueous layer, single-stranded RNA was degraded by addition of 12µl 10mg/ml RNAseA and incubation for 20 minutes at room temperature. DNA was isolated by addition of 8µl 7.5M NH_4OAc and 1ml 100% EtOH, centrifugation (12000rpm for 10 minutes), removal of the aqueous layer, and air-drying the DNA pellet. DNA was resuspended in 50µl ddH_2O. DNA concentration was estimated in the following manner: restriction enzyme digestion of 5µl resuspended DNA with 15µl of 1 unit *Eco*RI restriction endonuclease (New England Biolabs) in 1xNEBuffer *Eco*RI, incubation at 37°C for 2 hours, followed by electrophoresis of 10µl of the digested product and 1µl loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol in ddH_2O) on a 1% agarose gel (UltraPure; Gibco BRL) in 1xTAE buffer (4.84g/L Tris, 1.14ml/L glacial acetic acid, 50mM EDTA pH 8.0), with 10µl of 0.1µg/µl 1kb DNA ladder (Gibco BRL) loaded at regular intervals, staining in ~1µg/ml ethidium bromide in 1xTAE, destaining in ddH_2O, photography under UV illumination, and comparison of band intensities to DNA ladder intensity.

The locations of markers on the WICGR STS map (Hudson et al. 1995) were confirmed with technical assistance by PCR amplification of each marker on YAC clone DNA. Amplification was performed with only the YAC clone(s) to which the marker mapped and adjacent YAC clones. The concentrations of reagents in each PCR amplification reaction were as follows: 100ng-400ng YAC clone DNA (or 120ng human genomic DNA for positive control, or no DNA for negative control), 0.5μM forward and reverse marker primers, 1x buffer N with 1.5mM $MgCl_2$, 200μM each of dATP, dCTP, dGTP, dTTP, and 1 unit Taq DNA polymerase in a final volume of 50μl. Each well was overlaid with mineral oil. PCR conditions were as follows: 6 minutes denaturation at 94°C, 25 amplification cycles of 1 minute at 94°C, 2 minutes at 55°C, and 1 minute at 72°C, followed by 10 minutes extension at 72°C, and soak at room temperature. In cases of poor amplification, the annealing temperature was altered depending on primer GC content. Immediately prior to electrophoresis, 10μl of each amplified sample were combined with 5μl 0.2mg/ml RNAse A and 2μl loading buffer, which were then loaded on a 1% agarose gel in 1xTAE buffer, with 10μl of 0.1μg/μl 1kb DNA ladder loaded at regular intervals, and electrophoresed at 100V. Gels were stained with ethidium bromide in 1xTAE, destained in ddH_2O, and photographed under UV illumination. The presence of an amplified product of the same size as expected based on sequence information indicated the presence of the particular marker on that YAC clone.

2.11. Radiation hybrid mapping

Radiation hybrid (RH) mapping is a somatic cell method based on the fusion of lethally irradiated donor cells to non-irradiated recipient cells of another species, which stably retain fragments of the donor chromosomes. RH mapping uses the co-retention frequencies of two markers in the hybrids as a statistical measure of the distance between the markers. Distances are expressed in $centiRay_{(N\ rad)}$ units, where 1 $cR_{(N\ rad)}$ corresponds to a 1% frequency of breakage between two markers after exposure to "N" rad of X-rays. Since RH maps are based on breaks induced by radiation, instead of meiotic recombination as for genetic maps, these two mapping methods are complementary.

The MRC Genome Resource Facility in Toronto, Ontario performed radiation hybrid mapping of selected markers on the WICGR RH framework map (Hudson et al. 1995), a lod 2.5 framework map in which markers are ordered with an odds ratio of ~300:1. Marker primers were PCR amplified on the GeneBridge 4 panel of 93 donor human/recipient hamster RH cell lines (Gyapay et al. 1996). This panel was constructed using 3000rads of X-ray irradiation, with a marker retention frequency of ~29% (i.e., each marker was retained in ~29% of the cell lines), and thus has low resolution (1cR = ~270kb) relative to other panels. The results of the PCR assays (positive, negative, or discrepant amplification between duplicate assays) were emailed to the WICGR Mapping Service (www@genome.wi.mit.edu), where RHMAPPER software (http://www.genome.wi.mit.edu/ftp/pub/software/rhmapper) was used to map each marker using a maximum likelihood model. The mapping results consisted of marker placements on the framework map with distances to framework markers, and the odds ratio of each placement relative to the next most likely placement.

2.12. Genetic mapping

Marker order from published genetic maps was confirmed and intermarker genetic distances were determined from analysis of the dyslexia families by the MultiMap genetic map building program (version 2.0) (Lander and Green 1987; Matise et al. 1994). This program utilizes a novel algorithm that rapidly and sequentially builds a multipoint map from genotype data and validates the map at each stage in the analysis by determining whether the marker order has a higher likelihood than other orders. Specifically, as each marker is placed on the map, the order of each pair of adjacent markers is reversed (the markers are "flipped"). The likelihood of this flipped order will presumably be lower, given the genotype data, otherwise the marker is placed elsewhere on the map until a better map is found. A low-resolution framework map of markers uniquely placed with odds of 1000:1 or greater is first created, and then a comprehensive map containing the remaining markers in their 1000:1 or greater odds locations is built onto the framework map. The marker loci are added to the map in a manner determined by the user, which in the current study was according to heterozygosity (i.e., the most

polymorphic markers were placed first). Markers that cannot be located with odds of 1000:1 or greater are not placed on the map, however their most likely locations with the corresponding odds ratios are given, allowing the user to place the marker on the map, if desired.

2.13. Candidate gene restriction fragment length polymorphism genotyping

Candidate gene restriction fragment length polymorphisms (RFLPs), which are sequence variants that change a restriction endonuclease recognition site, were selected from the literature and genotyped on the family sample with technical assistance for the purpose of linkage disequilibrium analysis. The selected RFLPs were serotonin receptor *HTR1β* G-511T, *HTR1β* T-261G, *HTR1β* del-179/-178, and *HTR1β* T371G (Phe to Cys nonconserved variant) (Nöthen et al. 1994), *HTR1β* G861C (silent variant) (Lappalainen et al. 1995), and *HTR1E* C531T (silent variant) (Shimron-Abarbanell et al. 1995). Unless otherwise noted, the concentrations of reagents in each PCR amplification reaction were as follows: 300ng sample DNA (or no DNA for a negative control), 0.5μM forward and reverse primers, 1x buffer N in 2mM $MgCl_2$, 250μM each of dATP, dCTP, dGTP, and dTTP, and 0.5units Taq DNA polymerase in a total of 25μl, and each sample was overlaid with mineral oil. PCR amplification was performed with either an Ericomp EasyCycler Series thermocycler or a Perkin Elmer GeneAmp PCR System 9600 thermocycler. For most of the RFLPs, 100ng of a plasmid containing the RFLP's restriction endonuclease site was also amplified under the same conditions so that, following PCR amplification, successful restriction endonuclease digestion of the plasmid in PCR reagents would serve as a proper positive control for digestion, as opposed to standard digestion of the plasmid in restriction endonuclease buffer only.

For *HTR1β* G-511T, the PCR conditions were: 6 minutes denaturation at 94°C, 30 amplification cycles of 1 minute at 90°C, 2 minutes at 60°C, and 3 minutes at 72°C, followed by 6 minutes extension at 72°C, and soak at 4°C. The amplified samples and pBR322 plasmid control were each digested at 65°C for 1.5 hours with 0.5 units *Bsm*I restriction endonuclease (New England Biolabs) and 1x NEBuffer2 in a final volume of 5μl, to which 2μl loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol, 30%

glycerol in ddH_2O) were added, of which 10µl were loaded onto a 4:1 Nusieve GTG (Gibco BRL):agarose gel in 0.5xTBE, with 10µl of 0.1µg/µl 1kb DNA ladder loaded at regular intervals, electrophoresed at 100V, stained with ethidium bromide in 1xTAE, destained in ddH_2O, and photographed under UV illumination. RFLP genotypes were determined by presence of two fragments at 144bp and 91bp (nt –511: G), or one fragment at 235bp (nt –511: T), and alleles were manually entered into the database.

For *HTR1β* T-261G, PCR amplification was performed as described above except that 100ng of DNA was amplified and 1x PCR enhancer (Gibco BRL) was also added to the reaction. The PCR conditions were: 6 minutes denaturation at 94°C, 35 amplification cycles of 40 seconds at 94°C, 30 seconds at 61°C, and 30 seconds at 72°C, followed by 6 minutes extension at 72°C, and soak at room temperature. The amplified samples and pBR322 plasmid control were each digested at 55°C for 1.5 hours with 1 unit *Bsm*A1 restriction endonuclease (New England Biolabs) and 1x NEBuffer3 in a final volume of 5µl, to which 2µl loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol in ddH_2O) were added, of which 10µl were loaded onto a 2:1 Nusieve GTG:agarose gel in 0.5xTBE, with 10µl of 0.1µg/µl 1kb DNA ladder loaded at regular intervals, electrophoresed at 100V, stained with ethidium bromide in 1xTAE, destained in ddH_2O, and photographed under UV illumination. RFLP genotypes were determined by presence of two fragments at 146bp and 16bp (nt –261: T) (note that the 16bp fragment could not be seen on this concentration of gel), or one fragment at 162bp (nt –261: G), and alleles were manually entered into the database.

For *HTR1β* del-179/-178, PCR amplification was performed as described above except that 2.5µM dCTP was used, 0.4MBq α-^{32}P-dCTP was added to the reaction, and the total reaction volume was 15µl. The PCR conditions were: 6 minutes denaturation at 94°C, 30 amplification cycles of 40 seconds at 94°C, 30 seconds at 63°C, and 30 seconds at 72°C, followed by 6 minutes extension at 72°C, and soak at room temperature. Following PCR amplification, 5µl of loading buffer (20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol, 92% formamide) were added to each sample, of which 3µl were loaded onto a 6% acrylamide vertical gel, with 2µl bacteriophage

M13mp18 DNA sequencing ladder loaded at regular intervals, and electrophoresed in 1xTBE buffer at 1000V. Following electrophoresis, gels were exposed to Kodak autoradiography film overnight at room temperature, and processed using an automated developer. Genotypes were determined by presence of a 251bp fragment (no del-179/-178) or a 249bp fragment (del-179/-178), and alleles were manually entered into the database.

For *HTR1β* T371G, PCR amplification was performed as described above, and the PCR conditions were: 5 minutes denaturation at 94°C, 35 amplification cycles of 1 minute at 94°C, 2 minutes at 57°C, and 1 minute at 72°C, followed by 10 minutes extension at 72°C, and soak at 4°C. The amplified samples and pBK-RSV plasmid control were each digested at 37°C for 1.5 hours with 0.5 units *Nhe*I restriction endonuclease (New England Biolabs) and 1x NEBuffer2 in a final volume of 5µl, to which 2µl loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol in ddH_2O) were added, of which 10µl was loaded onto a 2:1 Nusieve GTG:agarose gel in 0.5xTBE, with 10µl of 0.1µg/µl 1kb DNA ladder loaded at regular intervals, electrophoresed at 100V, stained with ethidium bromide in 1xTAE, destained in ddH_2O, and photographed under UV illumination. RFLP genotypes were determined by presence of one fragment at 258bp (nt 371: T), or two fragments at 238bp and 20bp (nt 371: G) (note that the 20bp fragment could not be seen on this concentration of gel), and alleles were manually entered into the database.

For *HTR1β* G861C, PCR amplification was performed as described above, and the PCR conditions were: 6 minutes denaturation at 94°C, 30 amplification cycles of 1 minute at 90°C, 2 minutes at 57°C, and 3 minutes at 72°C, followed by 6 minutes extension at 72°C, and soak at 4°C. 10µl of the amplified samples (no plasmid control) were digested at 37°C for 2 hours with 1.25 units *Hinc*II restriction endonuclease (Gibco BRL) and 1x React4 buffer in a final volume of 10µl, to which 2µl loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol in ddH_2O) were added, of which 10µl were loaded onto a 1% agarose gel in 1xTAE, with 10µl of 0.1µg/µl 1kb DNA ladder loaded at regular intervals, electrophoresed at 100V, stained with ethidium bromide in 1xTAE, destained in ddH_2O, and photographed under UV illumination.

RFLP genotypes were determined by presence of two fragments at 260bp and 200bp (nt 861: G) or one fragment at 460bp (nt 861: C), and alleles were manually entered into the database.

For *HTR1E* C531T, PCR amplification was performed as described above, and the PCR conditions were: 6 minutes denaturation at 94°C, 28 amplification cycles of 1 minute at 94°C, 2 minutes at 65°C, and 1 minute at 72°C, followed by 6 minutes extension at 72°C, and soak at 4°C. The amplified samples and pUC18 plasmid control were each digested at 37°C for 1.5 hours with 0.5 units *Nla*III restriction endonuclease (New England Biolabs) and 1x NEBuffer 4 in a final volume of 5µl, to which 2µl loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol in ddH_2O) were added, of which 10µl were loaded onto a 4:1 Nusieve:agarose gel in 0.5xTBE, with 10µl of 0.1µg/µl 100bp DNA ladder (Gibco BRL) loaded at regular intervals, electrophoresed at 100V, stained with ethidium bromide in 1xTAE, destained in ddH_2O, and photographed under UV illumination. RFLP genotypes were determined by presence of one fragment at 73bp (nt 531: C) or two fragments at 67bp and 6bp (nt 531: T) (note that the 6bp fragment could not be seen), and alleles were manually entered into the database.

2.14. Qualitative linkage analyses of PCD

2.14.1. Parametric linkage analysis

Parametric linkage analysis, also referred to as lod score analysis, looks for cosegregation of a polymorphic chromosomal marker and a disease of interest within families to determine whether the marker and a disease-predisposing gene are linked to each other on a chromosome. The analysis is performed under a model that specifies certain parameters, in particular the disease allele population frequency and disease locus genotype penetrances. A likelihood ratio test is performed that assesses the likelihood (L) of the marker and disease being linked at a particular recombination fraction (theta, or θ) versus the likelihood of the marker and disease not being linked ($\theta = 0.5$), which is written as

$$\text{Likelihood ratio} = L^*(\theta) = L(\theta)/L(0.5)$$

Many values of θ are tested to determine the maximum likelihood estimate (MLE) of θ that maximizes L^*. To allow summation of $L^*(\theta)$ across individual families in a sample, the $\log_{10}$ of $L^*(\theta)$, referred to as the "lod score" (or the "LOD score") and denoted $Z(\theta)$, is determined. The maximum lod score (denoted Z_{max}), which is obtained by the MLE of θ, is a measure of the evidence in favor of the hypothesis of linkage.

For Mendelian autosomal (single gene) traits, a lod score > 3 is considered significant evidence for linkage, while a lod score < -2 is considered significant evidence against linkage. Lod scores between –2 and 3 are considered insufficient evidence for or against linkage. The critical lod score of 3, proposed by Morton (1955), corresponds to a 1000:1 odds for linkage, meaning that the observed data is 1000 times more likely to occur under the hypothesis of linkage than under the null hypothesis. The theory of likelihood ratio testing predicts that asymptotically, under the null hypothesis, 4.6 x Z_{max} follows a chi-square distribution with 1 degree of freedom (df). Thus, for $Z_{max} = 3$, the chi-square value is 13.8 (4.6 x 3), which corresponds to a significance level of 0.0002 for a two-sided test. However, because the test is only declared significant when $\theta < 0.5$, but not when $\theta > 0.5$, the test is one-sided, and the significance level for $Z_{max} = 3$ is actually 0.0001 (i.e., significant evidence for linkage is expected to occur at a frequency of 0.01% when there is no linkage). This value is much smaller than the traditional significance levels of 5% or 1% used in statistical tests. The reason for the use of this lower level in linkage tests is that the *a priori* chance that a pair of loci will be within 50cM of each other (the limit of linkage detection) is estimated to be 2%, so if a significance level of 5% is used to claim significant evidence for linkage, the test will detect many false linkages. To solve this dilemma, smaller significance levels are employed, and in fact, a lod score of 3.0 corresponds to a posterior probability of linkage of 95% (i.e., 5% significance level).

For complex traits, where a genome-wide approach is usually used in the linkage study, there is controversy as to the appropriate lod score for declaring significant evidence for linkage. As the number of markers being tested increases, the chance of a false positive result increases, thus it is felt that the critical lod score of 3 should be raised to account for the testing of multiple markers. The accepted view at the moment is

to use a lod score of 3.3 as the value to claim significant evidence for linkage to a complex trait. This value, which corresponds to $P = 5 \times 10^{-5}$, is equivalent to a genome-wide significance level of 5% using a marker density of <0.1cM (Lander and Kruglyak 1995). However, for less dense marker coverage, the recommended critical lod score is somewhere below 3.3, depending on the marker density. Also, the recommended critical lod score depends on the method of linkage analysis. Thus, it is apparent that determining whether a linkage study has found "significant" evidence for linkage using these recommended significance levels is not a straightforward matter. As a result, an alternative method that is gaining support is to determine the probability of a false positive result (the type I error rate) for the particular linkage study at hand using computer simulations, and to assign a critical lod score that surpasses this type I error rate (Ott 1999). Thus, the selected significance level is specific for the particular linkage study, which logically seems more appropriate than using "predetermined" significance levels. However, software to perform the necessary simulations is not yet publicly available.

2.14.1.1. Two-point parametric linkage analysis

Two-point parametric linkage analyses of the PCD phenotype were performed using FASTLINK (version 4.1P) from the LINKAGE package of programs (Lathrop and Lalouel 1984; Lathrop et al. 1984,1986; Cottingham et al. 1993; Schaffer et al. 1994). FASTLINK incorporates a "speed-up" function to allow faster computations than the traditional MLINK program in the LINKAGE package.

2.14.1.2. Multipoint parametric linkage analysis

Multipoint linkage analysis considers the genetic information from all markers in the region of interest when determining the likelihood of linkage to a disease locus. An accurate genetic map with respect to both marker order and intermarker distances is required, since an incorrect map may reduce the power to detect linkage (Halpern and Whittemore 1998).

Multipoint parametric linkage analysis of the PCD phenotype was performed using the GENEHUNTER program (version 2.0) (Kruglyak et al. 1996). Lod scores under linkage heterogeneity (hlod scores) are reported since they provided more significant evidence for linkage than lod scores under homogeneity, suggesting genetic heterogeneity in the sample. Due to algorithm constraints, GENEHUNTER can only accommodate pedigrees of moderate size (the number of pedigree non-founders minus two times the number of founders cannot exceed 18). This necessitated subdivision of large pedigrees into subpedigrees, which was expected to reduce the power to detect linkage. Thus, the linkage results are probably lower than those that would have been obtained had it been possible to analyze the complete pedigrees.

2.14.2. Sibpair linkage analysis

Sibpair linkage analysis determines whether each pair of siblings in a family shares 0, 1, or 2 alleles identical-by-descent (that is, the same parental marker allele) at a marker locus of interest. If the alleles are inherited randomly, the proportion of 0, 1, or 2 alleles shared identical-by-descent (IBD) is 1:2:1. However, if the marker is close to a disease gene, then one parent (for dominant inheritance) or both parents (for recessive inheritance) presumably have a disease allele in coupling with one of the marker alleles, which is then likely to be passed to affected offspring. Thus, with linkage of the marker and a disease locus, a deviation towards higher numbers of alleles shared IBD is expected. Sibpair linkage analysis was originally developed to analyze affected sibpairs for >50% IBD allele sharing, thus the term "ASP method" is often used. However, the method has since been generalized to included unaffected siblings in the analysis, thus three sibpair types are studied: concordant unaffected, discordant, and concordant affected sibpairs.

Sibpair linkage analysis does not require prior assumptions about parameters such as penetrance and disease allele frequency, thus it is often referred to as "nonparametric" analysis. It should be pointed out, however, that the statistical definition of a nonparametric method is that it be parameter-free *and* make no assumptions about normality of the data. Some methods of sibpair analysis, however,

assume a normal distribution of phenotypic noise. Therefore, these sibpair analysis methods are not truly nonparametric methods, and are more correctly referred to as "genetic model-free" methods. Methods that do not require specification of an inheritance model are thought to be more appropriate for the analysis of a complex trait where the mode of inheritance is unclear. However, genetic model-free analysis is generally not as powerful as parametric analysis, when the parametric analysis is performed under an appropriate inheritance model.

Sibpair linkage analysis utilizes nuclear families only, where a nuclear family consists of two parents and their children. Only the phenotypic information of the siblings is used in the analysis, although the genetic information of the siblings and the parents (if available) is used to determined IBD status of alleles. Because only nuclear families are analyzed, a smaller sample was used for the sibpair analyses reported here, thus it was expected that sibpair analyses would have reduced power to detect linkage compared to analyses that used the complete sample. As previously described in Section 2.9, one nuclear family was selected from each pedigree on the basis of: 1) having two or more affected child siblings (<18 years of age), or if that was not possible, then 2) having one affected child, or if that was not possible, then 3) having two or more affected adult siblings. This selection scheme was used to minimize the number of adult sibpairs in the sample, since the psychometric reading tests used for PCD diagnosis (and used for QTL linkage analyses, discussed below) exhibit adult ceiling effects and may not be as accurate for adults as for children. As a result, it was possible that there was an increased rate of misdiagnosis in adults compared to children, although reading history was also used for adult diagnosis to increase reliability. Thus, analyses were performed on two samples of sibpairs: all sibpairs in the nuclear families ("all ages" sample) and sibpairs under 18 years of age ("<18 years" sample). As outlined in Table 2.3, in the data set of nuclear families derived from the 79 pedigrees (used in Chapter Three), there were a total of 241 "all ages" sibpairs (144 independent sibpairs, explained in the next paragraph), and the subsample of "<18 years" sibpairs contained only 68 families with 165 sibpairs (112 independent sibpairs). In the data set of nuclear families derived from the 83 families (used in Chapter Six), there were a total of 305 "all-ages" sibpairs (163

independent sibpairs), and the subsample consisted of 71 nuclear families with 209 "<18 years" sibpairs (131 independent sibpairs). And finally, in the data set of nuclear families derived from the 96 families (Chapters Four and Five), there were a total of 336 "all-ages" sibpairs (188 independent sibpairs), and the subsample consisted of 81 nuclear families with 227 "<18 years" sibpairs (145 independent sibpairs).

Table 2.3

Description of samples used in sibpair linkage analyses

Nuclear family sample	"all ages" sibpairs total (independent)	"<18 years" sibpairs total (independent)
79 families	241 (144)	165 (112)
83 families	305 (163)	209 (131)
96 families	336 (188)	227 (145)

In a sibship containing more than two siblings (i.e., a multiple sibship), the sibpairs are not completely independent of one another. Instead, only the first N-1 sibpairs (where N = the number of siblings) are completely independent, and the remaining sibpairs are at least partially dependent on the first sibpairs. For example, in a sibship containing three siblings, there are three possible sibpairs consisting of the first and second siblings, the second and third siblings, and the first and third siblings. The genetic information of two of these sibpairs is completely independent (for example, the first-second siblings and the second-third siblings). However, the genetic information of the third sibpair (i.e., the first-third siblings) is partially dependent on the other sibpairs, thus contributes less information in linkage analyses than independent sibpairs. Therefore, in analyses using all three of these sibpairs, the information from the third sibpair must be weighted to account for this dependence, thereby preventing inflation of a linkage signal.

2.14.2.1. Two-point sibpair linkage analysis

Two-point sibpair linkage analysis of the PCD phenotype was performed using the SIBPAL (version 3.1) program in the S.A.G.E. package (S.A.G.E. 1997). This program tests for genetic linkage by performing traditional Haseman-Elston linear regression of the squared sibpair trait difference on the estimated proportion of alleles shared IBD by the sibpair for each marker locus. In other words, the disease status of each sibling is given a numerical value (e.g. 0=unaffected, 1=affected), and the squared difference between the values for the two siblings in a sibpair is regressed on the proportion of IBD alleles that the sibpair shares. This is done for all sibpairs in the sample, and the slope of the resulting regression line indicates whether there is significant evidence for linkage between the marker and a disease locus. The program also indicates the calculated proportions of alleles shared IBD for concordant unaffected, discordant, and concordant affected sibpairs in the sample, thus one can determine whether there is skewed allele sharing in a particular type of sibpair (for example, >50% IBD allele sharing in concordant unaffected sibpairs is suggestive of a nearby locus with a protective effect). Multiple sibships within each nuclear family are accommodated in SIBPAL by the use of a modified *t*-test with reduced degrees of freedom based on the effective sample size (the number of independent sibpairs), which may be overconservative since it omits the partially dependent (but partially independent) sibpairs in each family. It should be pointed out that while the Haseman-Elston regression method is "genetic model-free", it is not a nonparametric method since it assumes normally distributed phenotypic noise.

2.14.3. Multipoint nonparametric linkage analysis

Multipoint nonparametric linkage (NPL) analysis of the PCD phenotype was performed using the GENEHUNTER program (version 2.0) (Kruglyak et al. 1996). As previously discussed, multipoint analysis utilizes the genetic information from all markers in the region of interest, which is more informative than analysis of single markers. This program is an affected-pedigree-member method that investigates all affected members within a family for IBD allele sharing, thus this method uses a larger sample size and was expected to be more powerful for detecting linkage than sibpair linkage analysis. The

analysis was performed by simultaneously examining all affected individuals in each family for IBD sharing ("all" option), which is more powerful than analyzing each pair of affected individuals in the family ("pairs" option). For each affected individual that shares the same allele IBD, a sharply increasing NPL_{all} score is assigned to the pedigree. The NPL_{all} scores from the pedigrees in the sample are summed and compared to the score under the null hypothesis (no linkage) to determine whether there is significant evidence for linkage in the region under investigation. As mentioned previously, it was necessary to subdivide large pedigrees for GENEHUNTER analysis, probably resulting in loss of power to detect linkage.

2.14.4. Genetic heterogeneity testing

Genetic heterogeneity testing was performed using the HOMOG program (Ott 1991), which tests for mixture in a group of families under the alternative hypothesis of two family types, one group of families with linkage between the trait locus and a marker ($\theta < 0.5$), and the other group of families without linkage ($\theta = 0.5$). This method is based on two-point parametric linkage analysis, thus requires specification of an inheritance model. When there is no prior evidence for linkage (lod score <3), significant evidence for linkage with heterogeneity is obtained when the likelihood ratio of the alternative hypothesis (linkage with heterogeneity) versus the null hypothesis (no linkage) is >2000, corresponding to $P < 0.0001$ under 2 degrees of freedom (due to two independent variables: the proportion of linked families, α, and the recombination fraction, θ) (Ott 1999, p220). Because genetic heterogeneity testing requires assumptions that cannot be verified until the disease gene(s) is identified and characterized (e.g. all mutations of all genes are equally penetrant), estimation of α is probably inaccurate (Whittemore and Halpern 2001).

2.15. Quantitative-trait locus linkage analyses of reading measures

Analysis of a quantitative (continuous) trait appears to be more powerful for detection of linkage than analysis of a qualitative (discrete) trait (Wijsman and Amos 1997). This is because more information is available in quantitative traits due to the finer, continuous scale of the trait, and quantitative-trait locus (QTL) linkage methods

use information from all individuals in a pedigree, whereas some qualitative linkage methods (such as affected-pedigree-member methods) only use information from affected individuals. Because of the potential increase in power to detect linkage, QTL linkage analyses of the phonological awareness, phonological coding, spelling, and RAN speed measures were performed. Several QTL methods with different properties and statistical assumptions were employed. The first method was two-point QTL sibpair linkage analysis using Haseman-Elston linear regression, which has the drawbacks of analyzing nuclear families only (which have less power for detection of linkage compared to analysis of extended pedigrees) and an underlying assumption of normal phenotypic noise (thus, this method is not truly nonparametric). However, this method does not rely on an accurate marker map (as do multipoint methods), and thus may be more appropriate when the marker map is uncertain. A second QTL method that was used was multipoint nonparametric sibpair linkage analysis, which does not assume normality of the trait but only analyzes nuclear families. The third method employed was variance-component linkage analysis, which has the advantage of analyzing extended pedigrees, and therefore probably has more power to detect linkage than sibpair methods. However, this method has an underlying assumption of a normally distributed trait, which is often violated in genetic studies where there is selection for affected individuals, which may lead to an increase in the type I error rate (Allison et al. 1999). In addition, the variance-component method used in the present studies is a multipoint method, therefore it relies on an accurate marker map.

2.15.1. Quantitative-trait locus sibpair linkage analysis

As discussed above, sibpair linkage analysis determines whether each pair of siblings in a family shares 0, 1, or 2 alleles IBD at a marker locus of interest. If a marker is close to a disease gene, then, for a quantitative trait, sibpairs should show a correlation between the magnitude of their trait difference and the number of marker alleles shared IBD. In other words, siblings that are similar for a trait (e.g. both siblings have low scores on a measure of spelling skill) should share more IBD alleles at a marker that is linked to a locus influencing this trait than if they were different for the trait.

2.15.1.1. Two-point quantitative-trait locus sibpair linkage analysis

Two-point QTL sibpair linkage analyses were performed using the SIBPAL program (version 3.1) in the S.A.G.E. package (S.A.G.E. 1997). Similar to analysis of qualitative phenotypes as discussed above, this program tests for genetic linkage of QTLs by performing traditional Haseman-Elston linear regression of the squared sibpair trait difference on the estimated proportion of alleles shared IBD by the sibpair for each marker locus. The slope of the regression line indicates whether there is significant evidence for linkage between the marker and a QTL involved in the trait. As mentioned previously, multiple sibpairs within each nuclear family are accommodated in SIBPAL by the use of a modified *t*-test with reduced degrees of freedom based on the effective sample size. Analyses were performed using a sample of sibpairs of all ages and, to elucidate whether the psychometric reading test adult ceiling effects reduced the variation in the quantitative reading measures and thereby reduced the power to detect linkage, separate analyses were performed on a subsample of sibpairs <18 years of age.

2.15.1.2. Multipoint quantitative-trait locus sibpair linkage analysis

Multipoint QTL sibpair linkage analyses were performed using the MAPMAKER/SIBS program (version 2.0) (Kruglyak and Lander 1995). This program infers the IBD distribution across the marker region for each sibpair, after which QTL mapping can be performed using Haseman-Elston regression, maximum-likelihood variance estimation, and nonparametric methods. The first two methods have an underlying assumption of phenotypic normality, which, as discussed in Section 2.15, may be violated in genetic studies, and lead to increased type I error. Thus, these methods were not used in the current studies. The nonparametric method, however, is a Wilcoxon rank-sum test that does not assume a normal distribution of the trait. Results are reported as a Z score that is asymptotically normally distributed, thereby allowing determination of significance levels (note that this "Z score" is different from the "Z" used to denote parametric lod scores). Analyses were performed with the option of using all possible pairs of sibs in the sample ("all pairs" option), and multiple sibships were accommodated by factoring in a weight of 2/(number of sibs) that, like the weighting

scheme used by SIBPAL, is probably overconservative. Thus, results are shown for analyses with either no weighting or weighting of sibpairs from multiple sibships, and the "true" result probably lies in between. As for two-point QTL sibpair analyses, a sample of sibpairs of all ages and a subsample of sibpairs <18 years of age were investigated.

2.15.2. Variance-component linkage analysis

Variance-component linkage analysis is based on the theoretical foundations laid by Fisher (1918), one of the major contributors to quantitative genetics. Consider the phenotype of an individual (z) to be the sum of the effects of all genetic loci on the trait (G) and environmental deviation, or residual error, (E),

$$z = G + E$$

Fisher showed that the proportion of a phenotype in a population that is due to genes, G, can be partitioned into its "expected" value based on additive genetic effects ($G\hat{}$), and deviations from this expectation due to dominance genetic effects (δ). To explain this concept in more detail, additive effects are when the effect of two combined alleles is equal to the sum of their individual effects. For example, in the context of a quantitative trait, if allele "a" of a gene contributes a value of 1 to the phenotype, and allele "b" of the gene contributes a value of 2 to the phenotype, then genotype "aa" has a resultant phenotype of 2, "ab" has a phenotype of 3, and "bb" has a phenotype of 4. However, dominance effects cause deviations in these phenotype values, such that they are higher or lower than the values from additive genetic effects. Thus,

$$G = G\hat{} + \delta$$

On the population level, then, the variance that is observed in the phenotype can be partitioned into genetic variance (σ^2_G) from all loci influencing the phenotype, which can further be partitioned into additive variance (σ^2_A) and dominance variance (σ^2_D), and variance due to environment, or residual variance, (σ^2_E). This biometric model is written as

$$\sigma^2 = \sigma^2_{QA} + \sigma^2_{QD} + \sigma^2_{PA} + \sigma^2_{PD} + \sigma^2_E$$

where *Q* refers to a QTL in the region under investigation and *P* refers to other genes. The variance-component method essentially analyzes the different types of relatives in each pedigree for genetic and trait information, and uses this information to determine whether a significant amount of the genetic variance (σ^2_G) of the trait can be attributed to a QTL located in the region under investigation. While variance-component analysis does not utilize an inheritance model, the fact that it is based on a biometric model makes it a semi-parametric method of linkage analysis.

Variance-component linkage analysis was performed using the GENEHUNTER program (version 2.0) (Kruglyak et al. 1996; Pratt et al. 2000), once again necessitating the subdivision of large pedigrees for analysis. This program calculates maximum likelihood estimates of variance components for major QTL, unlinked polygenic, and environmental effects at each region of the marker map. The significance of the QTL effects was tested by comparing this maximum-likelihood model with a model in which the QTL variance components were constrained to equal zero (no linkage). The maximum likelihood method assumed a normal distribution of the trait that, if violated, may lead to a higher number of false linkages, as mentioned above. Four models that all included QTL additive variance, polygenic additive variance, and environmental variance, with dominance variance at neither, both, or either the QTL or polygenes, could be tested using this program. Models that included QTL additive and dominance variance are under two degrees of freedom (df), thus one must account for the extra df when considering significance. To convert a variance-component lod score under 2df to a lod score under 1df, the lod_{2df} was multiplied by 4.6 to obtain the approximate chi-squared value under 2df. The *P* value corresponding to this chi-squared value was determined using chi-squared tables, followed by determination of the chi-squared value under 1df that corresponds to this *P* value. Finally, this chi-squared value was divided by 4.6 to obtain the estimated lod score under 1df.

2.16. Haplotype analysis

Determination of marker haplotypes for each individual was performed using the SimWalk2 program (version 2.60) (Sobel and Lange 1996) or the GENEHUNTER

program (version 2.0) (Kruglyak et al. 1996). Haplotype information was used to identify those families in which all PCD affected individuals within each family share a common haplotype of markers in the chromosomal region under investigation. Families in which affected members share one haplotype have a higher likelihood of carrying a putative dyslexia susceptibility locus located in this region. Unaffected or uncertain family members were also permitted to share the affected haplotype to allow for incomplete disease penetrance. These families were designated as "linked", and their affected haplotypes were investigated for recombination breakpoints to attempt to narrow the candidate susceptibility region.

2.17. Linkage disequilibrium analysis

Linkage disequilibrium (LD) analysis, also referred to as association analysis, tests for significantly different frequencies of specific marker alleles in affected individuals compared to control individuals. This method is based on the phenomenon that when a new disease mutation occurred on a founder chromosome, every allele for every marker on the chromosome was associated with the disease mutation. The chromosome with the disease mutation was then transmitted to the descendents of the founder individual, and transmission over successive generations resulted in recombination between the disease mutation and marker loci alleles. Alleles at loci located further away from the founder disease mutation would have undergone recombination more than markers located closer to the disease mutation. Thus, marker alleles located very close to the founder disease mutation will be found at a higher frequency in affected individuals than unaffected individuals. Linkage disequilibrium between two loci is thus formally defined as an inequality between the product of the individual allele frequencies at two loci and their haplotype frequencies. In other words, if the two loci are in equilibrium, the frequency of haplotype "ab" at locus A and locus B will equal the product of the frequency of allele "a" at locus A and the frequency of allele "b" at locus B.

LD extends over very short chromosomal regions. For outbred populations of north-European descent, the extent of detectable LD is in the range of 100kb, or

approximately 0.1cM (Collins et al. 1999; Abecasis et al. 2001; Reich et al. 2001), and is much less for older (e.g. African) populations (Reich et al. 2001). LD mapping can therefore localize disease genes to a very small region, allowing fine mapping of the disease gene.

One important issue to point out is that for LD studies of a complex, genetically heterogeneous disease in a mixed population, the affected individuals in the population might carry different combinations of founder mutations in the susceptibility genes influencing the disease, and multiple founder mutations might be found in *each* of the genes influencing the disease, greatly complicating the detection of LD to a single marker allele in a mixed population. It should also be pointed out that LD not only results from tight linkage between a marker and disease gene, but may also be the result of interaction (epistasis) between an unlinked marker allele and the disease mutation, or population stratification (i.e., the existence of multiple population subtypes with different marker allele frequencies in what is assumed to be a relatively homogeneous population). Matching appropriate controls to the patient sample compensates for stratification, whereas incorrect control matching is unable to compensate for stratification and may lead to spurious LD results.

2.17.1. Single marker linkage disequilibrium analysis

Single-marker LD analyses and candidate gene RFLP LD analyses of the PCD phenotype were performed using the Affected Family-Based Controls (AFBAC) program (Thomson 1995). This program analyzes genotypic data from nuclear families (two parents and one or more affected children) for association with the trait. Thus, one nuclear family was selected from each pedigree, as described in Section 2.9. The AFBAC method compares transmission and non-transmission of parental alleles to affected children, where the alleles not transmitted to affected children form the control population of alleles, thus avoiding the problem of association due to population stratification that may confound case-control studies. Analyses were performed using either the first affected sibling (simplex analysis option) or the first two affected siblings with weighted transmissions (multiplex analysis option) in each nuclear family. Statistical significance was determined

by a chi-squared test using each marker's 2 x *n* contingency table, where *n* was the number of alleles. Analyses were initially performed without grouping rare alleles into one category. However, the chi-squared test statistic is based on a large sample approximation, thus the method is invalid if a considerable number of table cells have expected values <5 (i.e., less than 5 expected occurrences in the sample). The guidelines used to determine how many cells may have an expected value <5 for the chi-squared method to be valid, attributed to W.G. Cochran, are that 80% of the cells should have expected values >5, and all cells should have expected values >1. However, for samples such as those used in our linkage disequilibrium analyses, few alleles will be expected to occur less than 5 times in the sample, but rare alleles should still be accommodated to avoid violation of the method. Thus, linkage disequilibrium analyses were also performed with grouping of rare alleles with <5% observed frequency in both the transmitted and non-transmitted categories. Note that this method of grouping alleles based on expected *frequencies* is more conservative than the original guideline of grouping alleles based on expected *occurrences* in the sample. Only the results from AFBAC analyses performed with grouping of rare alleles are reported.

2.17.2. Marker haplotype linkage disequilibrium analysis

Multiple-marker haplotype LD analysis yields more power to detect LD than single markers (Martin et al. 2000a; Akey et al. 2001). This is because several markers within small regions may be in strong LD with both each other and the disease locus, and analysis of these markers in haplotypes will extract all of the available LD information, whereas analysis of the single markers may not. Haplotype LD analysis was performed using the TRIMHAP program (MacLean et al. 2000), which supports analysis of multiplex pedigrees by conditioning on linkage (i.e., the linkage disequilibrium signal is not inflated by the presence of linkage). The analysis can be performed in two ways. The first method is similar to family-based LD methods in that the test sample consists of founder haplotypes transmitted to PCD affected individuals and the control sample consists of non-transmitted founder haplotypes. The user defines the number of transmissions for a haplotype to be selected for the test or control samples. In the present

analyses, the criteria used were a minimum of two transmissions in the family for selection to the test sample, and no transmissions for selection to the control sample (similar to AFBAC multiplex analysis). The test haplotypes are given equal weight in the analysis, regardless of the number of transmissions to affected individuals in the pedigree. For the second method, all transmissions of a founder haplotype to affected pedigree members are considered jointly, and a haplotype-based posterior probability of linkage (HBPPL) statistic that weights transmitted haplotypes within the pedigree is calculated. All haplotypes are selected for both the test and control samples, with a weight of HBPPL in the test sample and a weight of (1-HBPPL) in the control sample. For both of the analysis methods, a likelihood ratio of the probability of the observed haplotypes in the test sample (under the alternative hypothesis of LD) versus the probability of the haplotypes in the control sample (under the null hypothesis of no LD) is calculated. TRIMHAP calculates the significance level for the disease locus being located in the interval between two markers (regardless of the size of the haplotype being analyzed), and thereby allows for recombinants in haplotypes. Because of the potentially large number of haplotypes tested, a normal sampling distribution cannot be assumed, thus standard chi-squared testing is not appropriate. Instead, empirical significance levels are calculated using random-permutation replications or "bootstrapping", which make no assumptions regarding the sampling distribution (i.e., correction for multiple tests is not required).

Because the TRIMHAP program is new and has not been tested by independent investigators, an alternative haplotype LD analysis program, HAPMAX (http://www.uwcm.ac.uk/uwcm/mg/download), was used to support TRIMHAP results. HAPMAX analyzes parent-affected child trios for transmission and non-transmission of haplotypes, where the non-transmitted haplotypes form the control population of alleles, thus it is similar to simplex AFBAC analysis. Families with recombinations in transmitted haplotypes are excluded from analysis. Since only trios are analyzed, the power to detect LD is probably decreased compared to analysis of extended pedigrees, as in the TRIMHAP program. In addition, chi-squared testing is used to determine significance levels, which are not accurate under high degrees of freedom (i.e., a large

number of tested haplotypes). For these latter two reasons, HAPMAX *P* values were assumed to be estimates only.

2.18. Statistical correction for multiple testing

As recommended by Elston (1997,1998), precise significance levels are reported for all statistical analyses, rather than values adjusted for multiple comparisons. The rationale is that the only accepted method of correction for multiple testing, Bonferoni correction, assumes that the statistical tests are independent, which is not appropriate for genetic studies. In the present investigations, the reading disability phenotypes (PCD, phonological awareness, phonological coding, spelling, and RAN speed) are correlated with each other, the genetic markers that were investigated are located near each other, and several of the methods of linkage analysis and linkage disequilibrium analysis are similar to one another, thus the statistical tests are not independent. Thus, Bonferoni correction for multiple testing would probably be overly conservative, and therefore uncorrected significance levels are presented for all analyses (with the exception of empirical *P* values from TRIMHAP LD analyses).

Chapter Three: Absence of Significant Linkage Between Phonological Coding Dyslexia and Chromosome 6p23-21.3 (*DYX2*) Using Quantitative-Trait Methods

3.1. Introduction

While Dr. Field's and Dr. Kaplan's dyslexia linkage study was in its early stages, Cardon et al. (1994) reported linkage between a composite measure of reading disability and a 2cM region on chromosome 6p21.3 (named *DYX2*). This finding was later supported by Grigorenko et al. (1997), who reported evidence for linkage between phonological awareness and microsatellite markers slightly telomeric to the region suggested by Cardon and colleagues. Confirmation of these reports was provided by Fisher et al. (1999) and Gayán et al. (1999), who reported linkage of phonological and orthographic components of reading to chromosome 6p regions consistent with that suggested by the previous studies. Recently, linkage disequilibrium was reported between reading disability and a 1.5Mb region that overlaps the above 6p regions except that of Cardon et al. (1994) (Turic et al. 2000). In contrast, in a sample of 79 families containing at least two affected siblings, Dr. Field's laboratory was unable to find evidence for linkage between phonological coding dyslexia (PCD) analyzed as a qualitative (affected, unaffected, uncertain) phenotype and markers spanning *DYX2*, using either parametric or sibpair linkage methods. In addition, linkage disequilibrium could not be detected between PCD and the markers in this region (Field and Kaplan 1998). Since three of the positive linkage studies employed *quantitative* measures of reading disability, in contrast to the *qualitative* PCD phenotype, while the fourth analyzed *single* qualitative measures of component reading skills, in contrast to the *composite* nature of the PCD phenotype, it is possible that use of the PCD phenotype was the cause of the null findings. To address this issue, I present here a reanalysis of linkage to the 6p25-p21.3 region in our sample of 79 families using quantitative-trait locus (QTL) sibpair linkage analyses and variance-component analyses of four measures of reading disability. The results confirm absence of significant linkage to *DYX2*, indicating that the DYX2 locus does not contribute to dyslexia in these families and suggests that our sample may contain different genetic subtypes of dyslexia compared to other samples.

3.2. Methods

3.2.1. Subjects

As described in Section 2.2, the sample consisted of 79 families (45 nuclear families, 34 extended pedigrees) each containing at least two dyslexic siblings. 76 of the families were of European ancestry, while 3 were non-European. There were a total of 615 individuals who had DNA sampled and underwent a battery of psychometric testing to assess four components of reading: phonological awareness, phonological coding, spelling, and RAN speed (see Section 2.4). Each individual was also assessed for estimated IQ, and adults were assessed for reading history. The results of the reading tests (except RAN speed) and reading history in adults were used to assign an affected, unaffected, or uncertain diagnosis of PCD (see Section 2.5), resulting in 53% of the subjects diagnosed as affected, 33% as unaffected, and 14% as uncertain.

3.2.2. Descriptive statistics of reading measures

Descriptive statistics of the phonological awareness, phonological coding, spelling, and RAN speed measures, and estimated IQ, were determined as described in Section 2.6. Statistics were calculated in three samples of subjects used in this study: all members of the 79 families that were used for variance-component analyses, and two sibpair samples that were used for sibpair linkage analyses (one sample containing sibpairs of all ages and a subsample containing sibpairs <18 years of age).

3.2.3. Pearson correlation analysis

Pearson correlation analyses of the PCD phenotype and the phonological awareness, phonological coding, spelling, and RAN speed measures were performed as described in Section 2.6. Three data sets were investigated: all members of the 79 families, a sample of sibpairs of all ages, and a sample of sibpairs <18 years of age.

3.2.4. Markers, genotyping and marker map

DNA from each individual was genotyped with technical assistance using manual methods (see Section 2.8) for the following microsatellite markers: F13A1,

D6S89, D6S299, D6S105, TNFB, D6S291, and GLP1R. These markers span a 43cM region on chromosome 6p25-p21.3 spanning *DYX2*. In particular, D6S105 was reported to be significantly linked to reading disability in the studies by Cardon et al. (1994), Fisher et al. (1999), and Gayán et al. (1999), whereas D6S299 demonstrated the most significant linkage to phonological awareness in the study by Grigorenko et al. (1997). D6S105 and D6S299 are very polymorphic (heterozygosity = 0.77 and 0.79, respectively), providing high power to detect linkage. The marker map employed for linkage analyses (Figure 3.1) was derived from published genetic marker maps, and is the same as that used by Field and Kaplan (1998).

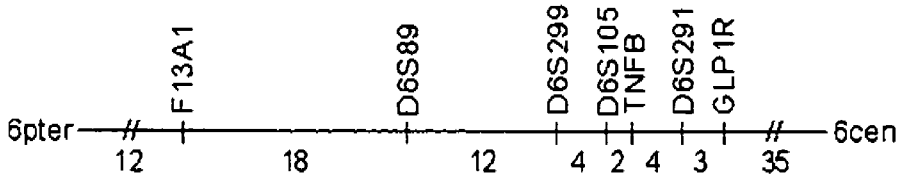


Figure 3.1 Genetic marker map of chromosome 6q25-p21.3 markers, with intermarker distances (cM) indicated.

3.2.5. Quantitative-trait locus linkage analysis

Two-point QTL sibpair linkage analysis of the reading measures was performed using the SIBPAL program, a genetic model-free simple linear regression method, as described in Section 2.15.1.1. Multipoint QTL sibpair linkage analysis was performed using the nonparametric rank-sum method in the MAPMAKER/SIBS program, as described in Section 2.15.1.2. Two samples were investigated in the two-point and multipoint analyses: a broad sample of 79 nuclear families containing 241 sibpairs of all ages (144 independent pairs), and a subsample of 68 nuclear families containing 165 sibpairs <18 years of age (112 independent pairs) (see Section 2.14.2). Table 3.1 indicates the number of nuclear families of various sibship sizes, in both the all-ages sibpair sample and the <18 years of age sibpair sample. The large majority of nuclear families consisted of

two or three siblings, with only a few larger five- and seven-sibling nuclear families that were not likely to distort the linkage results. Separate analysis of the <18 years of age sibpair sample was performed because ceiling effects occurred in adults in the phonological coding and spelling psychometric tests. Thus, these measures may not be as reliable in adults as in children.

Table 3.1

Distributions of nuclear families of various sibship sizes in the all-ages sibpair sample and the <18 years of age sibpair sample used for *DYX2* QTL sibpair linkage analyses

	Number of nuclear families	
Sibship size:	all-ages sample	<18 sample[a]
Two sibs	37	32
Three sibs	25	29
Four sibs	13	6
Five sibs	3	1
Seven sibs	1	0
Total:	79	68

[a] Restriction to siblings <18 years of age resulted in the exclusion of 11 nuclear families, thus the total number of nuclear families in this sample was 68, and changed the distribution of sibship sizes in the remaining nuclear families.

Maximum-likelihood variance-component linkage analysis of the complete sample of 79 families (i.e., including extended pedigrees) was performed using the GENEHUNTER program, as described in Section 2.15.2. Note that five pedigrees were too large to be handled by the program and were divided into subpedigrees, probably reducing the power to detect linkage. Three models were tested: Model 1 included QTL additive variance, polygenic additive variance, and environmental variance; the others were variations of Model 1, with either QTL dominance variance added (Model 2) or polygenic dominance variance added (Model 3).

3.3. Results

3.3.1. Descriptive statistics of reading measures

Table 3.2 shows the descriptive statistics for the phonological awareness, phonological coding, spelling, and RAN speed measures and estimated IQ. The variability in each reading measure was quite large, yielding adequate power to detect linkage. Additional evidence that variability in these quantitative measures was sufficient to detect linkage is that strong evidence for linkage was detected to chromosome 6q11.2-q12 by use of these same measures, as discussed in Chapter Four. With the exception of phonological awareness, the distributions of the reading measures were nearly normal, and since standard data transformations were not able to generate more normal distributions for any of the measures, untransformed data were utilized in QTL linkage analyses.

3.3.2. Pearson correlation analysis

To better understand the relationship between the PCD phenotype and the quantitative reading measures, Pearson correlation analysis was performed using all members of the 79 families. As shown in Table 3.3, the correlation coefficients (r) between PCD, the phonological coding measure, and the spelling measure were substantial at 0.73 to 0.77. The correlations between PCD and the phonological awareness and RAN speed measures were moderate ($r = 0.46$ and $r = 0.50$, respectively). Phonological awareness and RAN speed were also not highly correlated with phonological coding or spelling (range, $r = 0.34$ to 0.52). Pearson correlation analyses utilizing the all-ages sibpair sample and the <18 years of age sibpair sample yielded similar results to those obtained with all members of the 79 families (Table 3.3).

Table 3.2

Descriptive statistics of the reading measures and estimated IQ in all members of the 79 families (*N*=615), the all-ages sibpair sample (*N*=241), and the <18 years of age sibpair sample (*N*=165)

Trait	Mean	SD	Min	Max	Skewness	Kurtosis
Phonological awareness[a]						
All members	17.05	5.67	1	28.44	-0.79	-0.07
All ages sibpairs	17.02	5.56	2.65	28.44	-0.48	-0.58
<18 years sibpairs	16.96	5.75	2.65	28.44	-0.41	-0.69
Phonological coding[b]						
All members	98.03	17.28	12	149	-0.03	1.01
All ages sibpairs	89.69	15.11	46	149	0.53	1.14
<18 years sibpairs	88.52	14.65	46	129	0.33	0.63
Spelling[b]						
All members	90.85	16.33	47	124	-0.20	-0.90
All ages sibpairs	81.36	14.14	47	115	0.51	-0.29
<18 years sibpairs	80.51	13.86	47	115	0.66	-0.02
RAN speed[a]						
All members	3.08	0.62	1.09	5.00	0.16	-0.13
All ages sibpairs	2.96	0.56	1.45	4.64	0.23	0.04
<18 years sibpairs	2.99	0.56	1.59	4.64	0.33	0.12
Estimated IQ[b]						
All members	103.61	12.30	48	141	-0.007	0.72
All ages sibpairs	102.49	13.99	59	141	0.25	0.16
<18 years sibpairs	102.78	14.71	59	141	0.25	-0.03

[a] Data are raw scores for adults and age-adjusted scores for subjects <18 years of age.

[b] Data are standard scores according to test norms (normal population mean = 100, standard deviation =15).

Table 3.3

Pearson correlation coefficient matrix of PCD, phonological awareness, phonological coding, spelling, and RAN speed using all members of the 79 families, the all-ages sibpair sample, and the <18 years of age sibpair sample

	PCD	Phonological awareness	Phonological coding	Spelling
Phonological awareness				
All members	0.50			
All ages sibpairs	0.39			
<18 years sibpairs	0.40			
Phonological coding				
All members	0.73	0.52		
All ages sibpairs	0.65	0.54		
<18 years sibpairs	0.62	0.56		
Spelling				
All members	0.77	0.49	0.75	
All ages sibpairs	0.67	0.51	0.71	
<18 years sibpairs	0.69	0.51	0.71	
RAN speed				
All members	0.45	0.34	0.43	0.46
All ages sibpairs	0.44	0.39	0.47	0.52
<18 years sibpairs	0.51	0.42	0.54	0.55

NOTE — all correlations are significant at $P < 0.0001$.

3.3.3. Quantitative-trait locus sibpair linkage analysis

QTL sibpair linkage analyses did not detect significant evidence for a locus influencing reading disability in the DYX2 region, supporting the previous linkage results of our laboratory using a qualitative PCD phenotype. Table 3.4 shows *P* values for two-point simple linear regression of the squared sibpair trait difference on the

estimated proportion of alleles IBD at each marker. In both the all-ages sibpair sample and the <18 years of age sibpair sample, none of the regressions were significant at P <0.05 for any of the quantitative traits with any of the markers tested. Although P values were nearly significant for linkage between the spelling trait and TNFB (P = 0.10 in the all-ages sibpairs and P = 0.07 in the <18 years of age sibpairs), the results with D6S105 (located 2cM telomeric to TNFB; Figure 3.1) provided no supportive evidence for linkage to spelling ability.

Table 3.4

P Values for *DYX2* SIBPAL simple linear regression analysis of the all-ages sibpair sample ("all") and the <18 years of age sibpair sample ("<18")

		Phonological awareness		Phonological coding		Spelling		RAN speed	
Marker	cM[a]	all	<18	all	<18	all	<18	all	<18
F13A1	18	0.27	0.20	0.86	0.40	0.65	0.25	0.54	0.16
D6S89	12	0.55	0.72	0.13	0.37	0.66	0.68	0.99	0.96
D6S299	4	0.62	0.79	0.26	0.29	0.50	0.24	0.82	0.93
D6S105	2	0.70	0.75	0.93	0.56	0.63	0.18	1.00	0.97
TNFB	4	0.78	0.87	0.44	0.58	0.10	0.07	0.65	0.78
D6S291	3	0.59	0.63	0.33	0.81	0.33	0.50	0.45	0.77
GLP1R	-	0.65	0.79	0.35	0.65	0.24	0.49	0.50	0.70

[a] Genetic distance between marker and the marker below.

Multipoint nonparametric QTL sibpair linkage analyses also did not find significant evidence for linkage for any of the quantitative traits across the region. As shown in Figure 3.2a, analyses of the all-ages sibpair sample for the phonological awareness measure found maximum Z scores of 0.30 (with weighting of sibpairs; P = 0.38) and 0.42 (with no weighting of sibpairs; P = 0.34) at D6S105. However, analyses of phonological coding, spelling, and RAN speed resulted in maximum Z scores that

were mostly negative across the region (Figure 3.2b,c,d). As shown in Figure 3.3a and c, analyses of the <18 years of age sibpair sample were slightly more significant, with maximum Z scores for phonological awareness of 0.53 (unweighted; $P = 0.30$) and 0.70 (weighted; $P = 0.24$) at D6S105, and for spelling of 1.24 (unweighted; $P = 0.11$) and 1.02 (weighted; $P = 0.15$) at D6S299. Analyses of phonological coding and RAN speed measure resulted in negative Z scores across the region (Figure 3.3b,d).

3.3.4. Variance-component linkage analysis

As shown in Figure 3.4, variance-component analyses of the four reading measures also failed to find significant evidence for the DYX2 locus. Under Model 1 (QTL and polygenic additive variance, environmental variance) and Model 3 (QTL additive variance, polygenic additive and dominance variance, environmental variance), where there is no dominance variance for the QTL, the lod scores across the region for each of the four quantitative measures were essentially zero (Figure 3.4a,c). However, under Model 2 (QTL additive and dominance variance, polygenic additive variance, environmental variance), where there is dominance variance at the QTL, weak evidence was found for a locus affecting spelling, phonological coding, and RAN speed (Figure 3.4b). A peak maximum lod score of 0.82 was found in the region of TNFB and D6S291 for the spelling measure, and a lesser maximum lod score of 0.60 occurred between D6S89 and D6S299. Analysis of the phonological coding measure identified a peak maximum lod score of 0.42 between D6S89 and D6S299, whereas analysis of RAN speed found a peak maximum lod score of 0.40 at D6S299. The phonological awareness measure had a maximum lod <0.1 across the 6p region. However, because Model 2 included QTL additive variance and dominance variance, the analysis is under two degrees of freedom (df). The results under one df (to allow comparison to the conventional critical lod score of 3.3) were estimated at 0.45 for spelling, 0.17 for phonological coding, and 0.16 for RAN speed. Nonetheless, while the results of analyses using the phonological coding and spelling measures were weak, they were consistent with the results of two-point sibpair linkage analyses, where *P* values were lower for these same markers with the respective quantitative measure (Table 3.4).

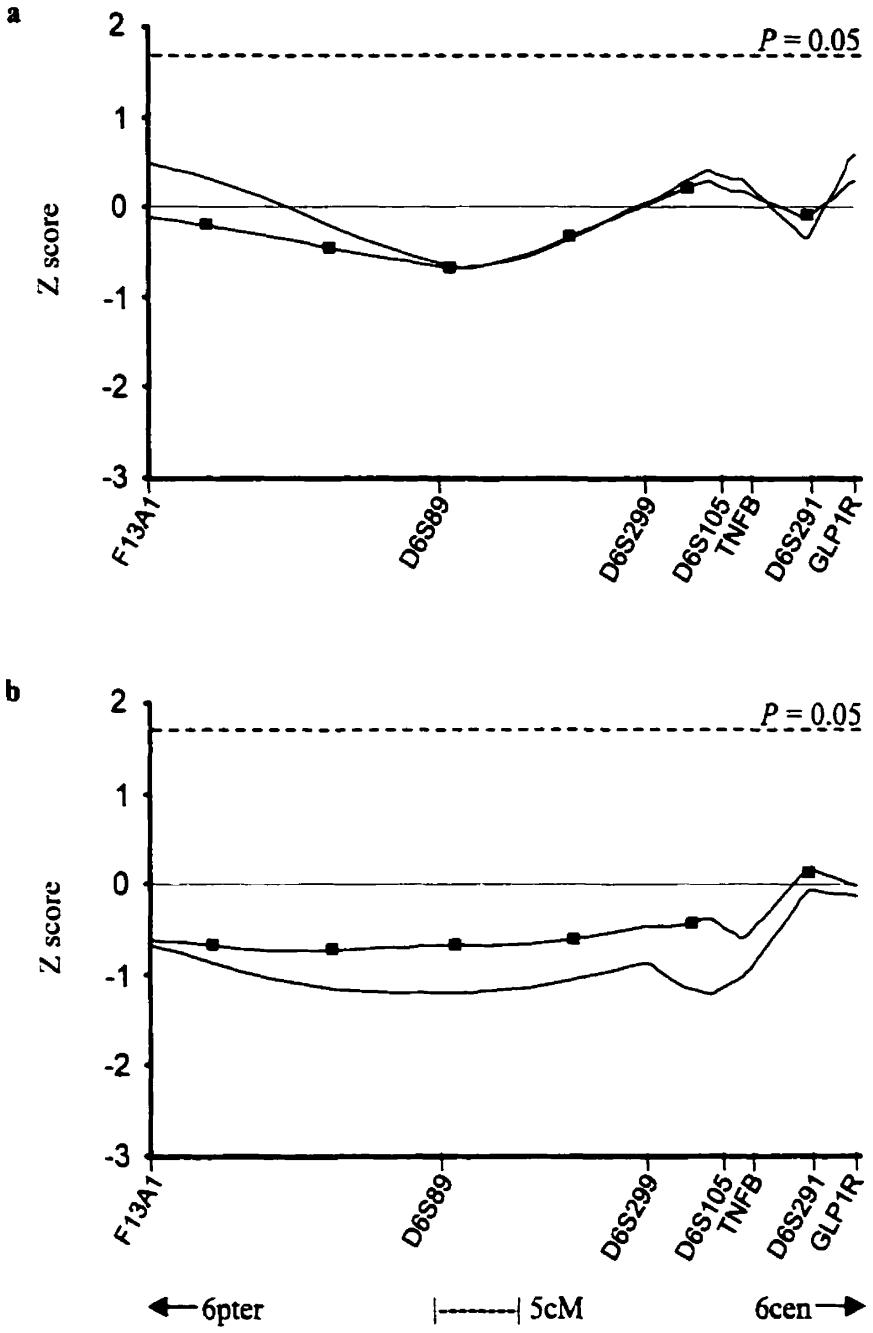


Figure 3.2 Z-score curves from *DYX2* MAPMAKER/SIBS nonparametric sibpair linkage analysis of **a)** phonological awareness, **b)** phonological coding, **c)** spelling, and **d)** RAN speed, using a sample containing sibpairs of all ages. Analyses were performed without weighting (——) and with weighting (—■—) of multiple sibships.

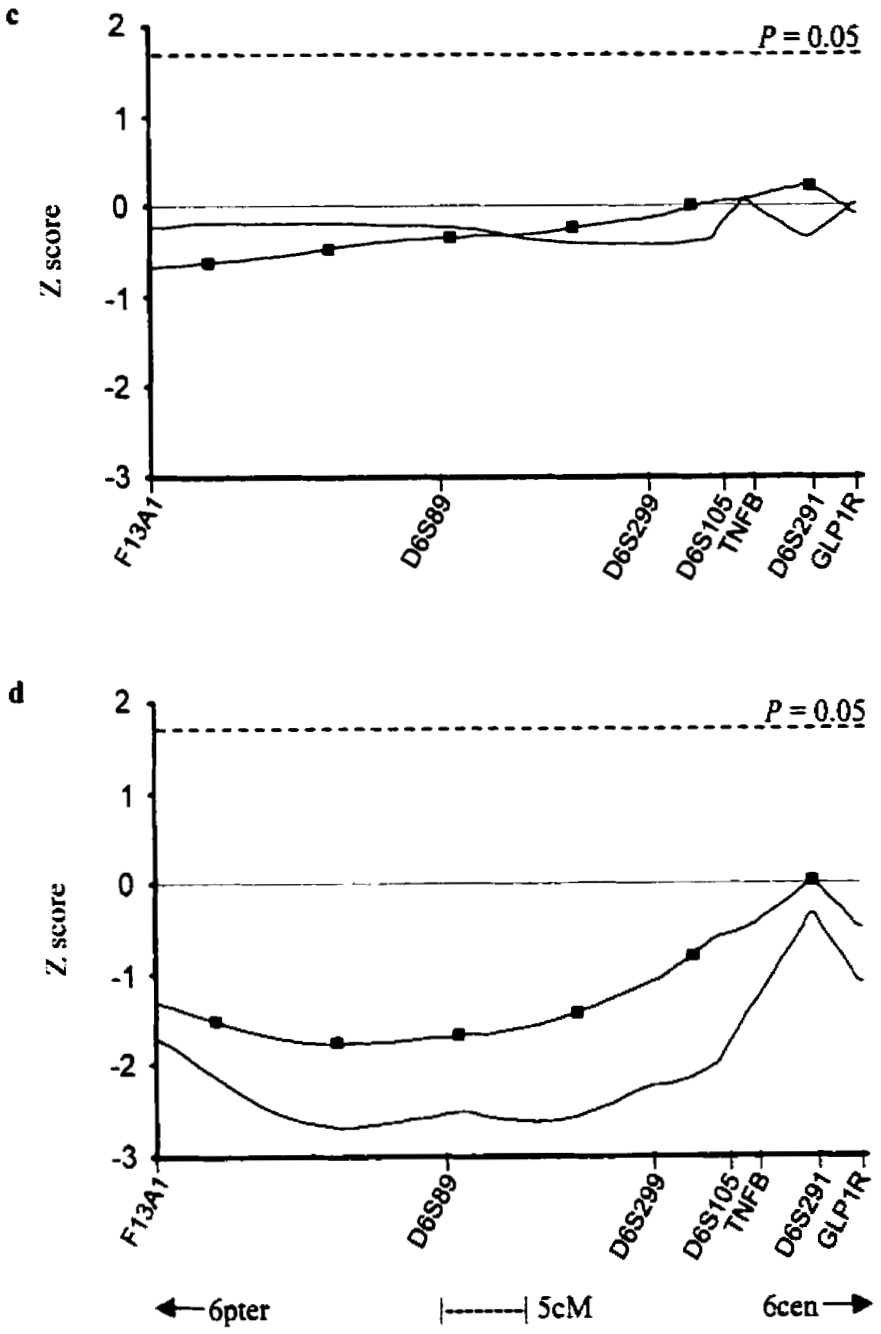


Figure 3.2 (cont'd)

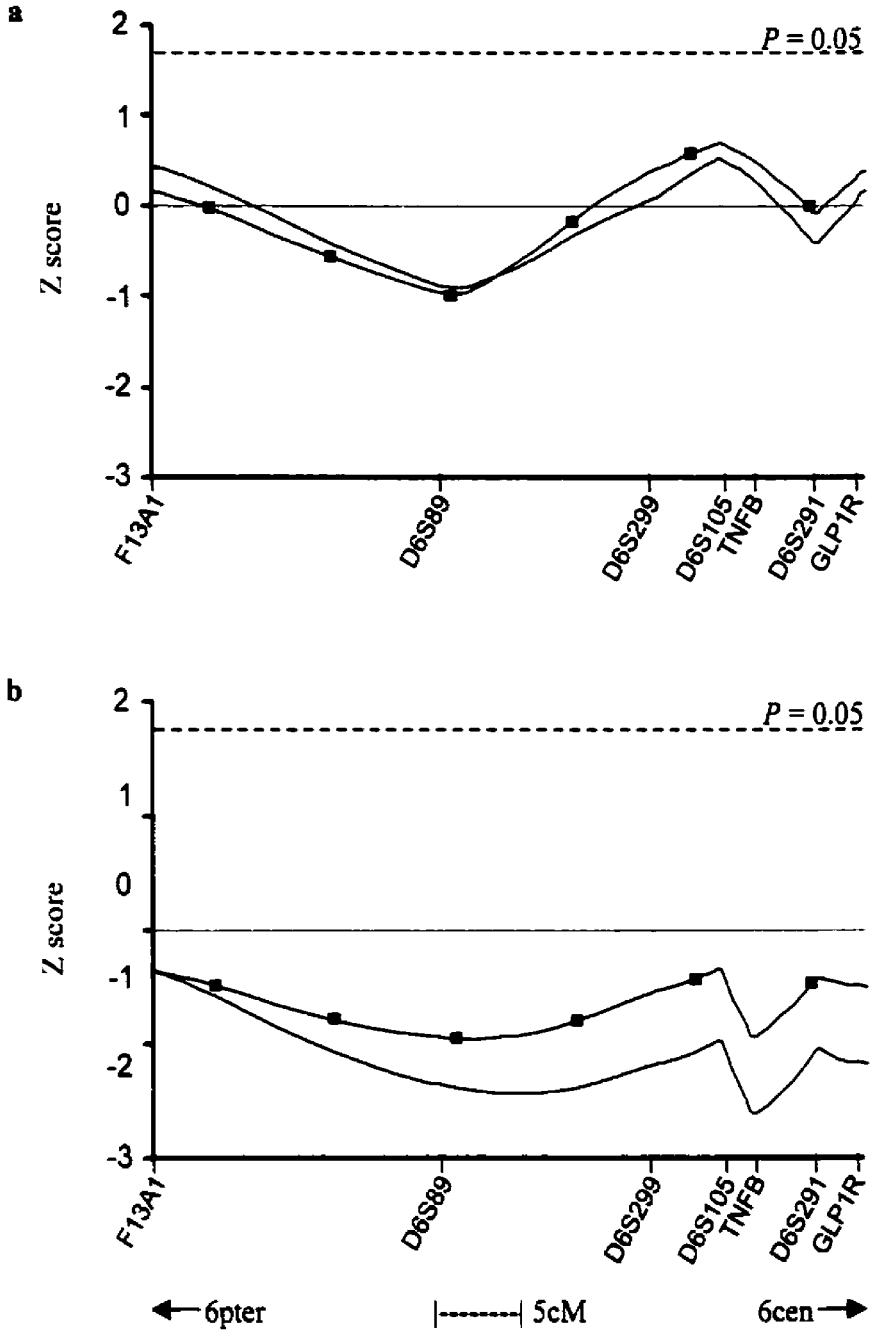


Figure 3.3 Z-score curves from *DYX2* MAPMAKER/SIBS nonparametric sibpair linkage analysis of **a)** phonological awareness, **b)** phonological coding, **c)** spelling, and **d)** RAN speed, using a sample containing sibpairs <18 years of age. Analyses were performed without weighting (——) and with weighting (—■—) of multiple sibships.

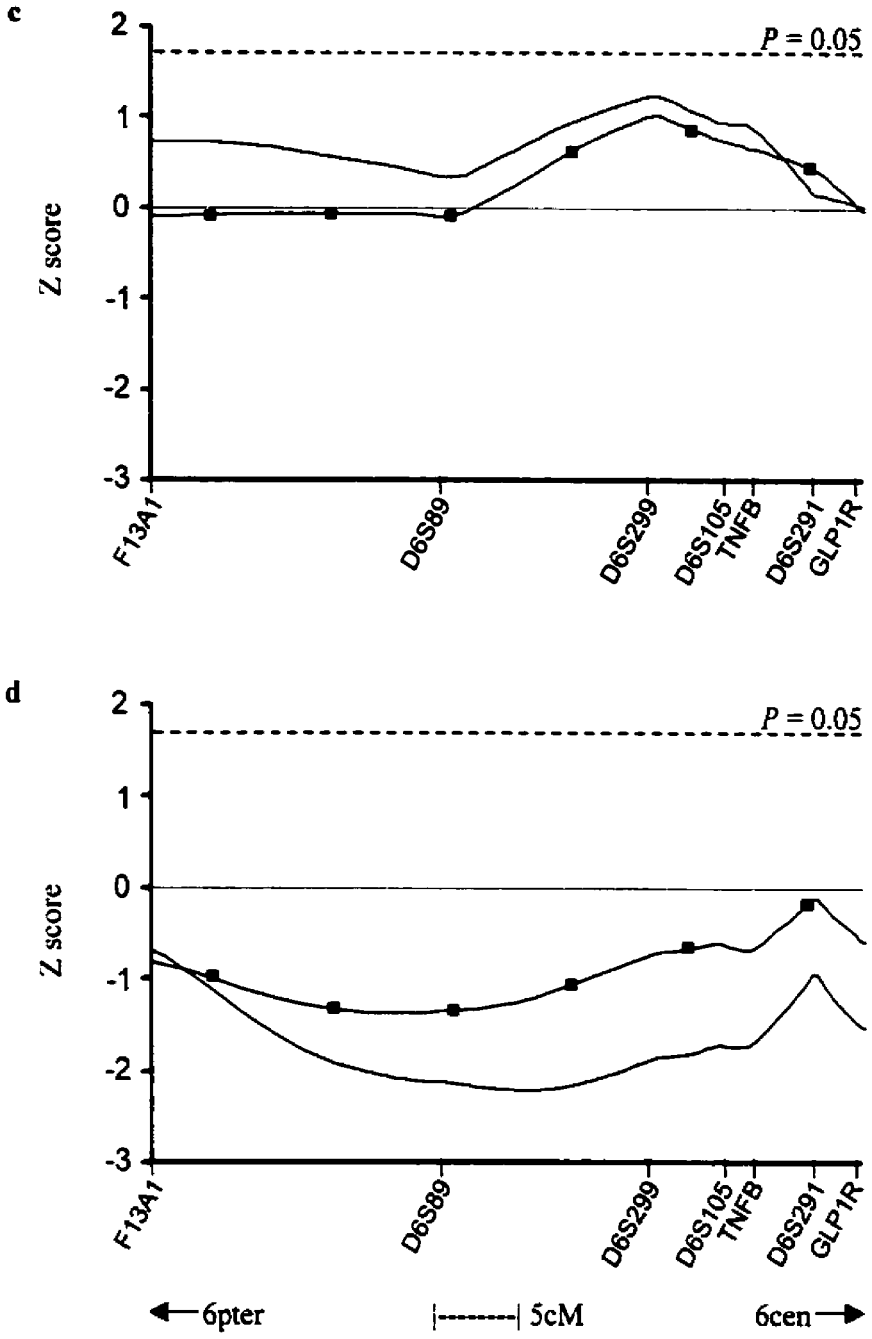


Figure 3.3 (cont'd)

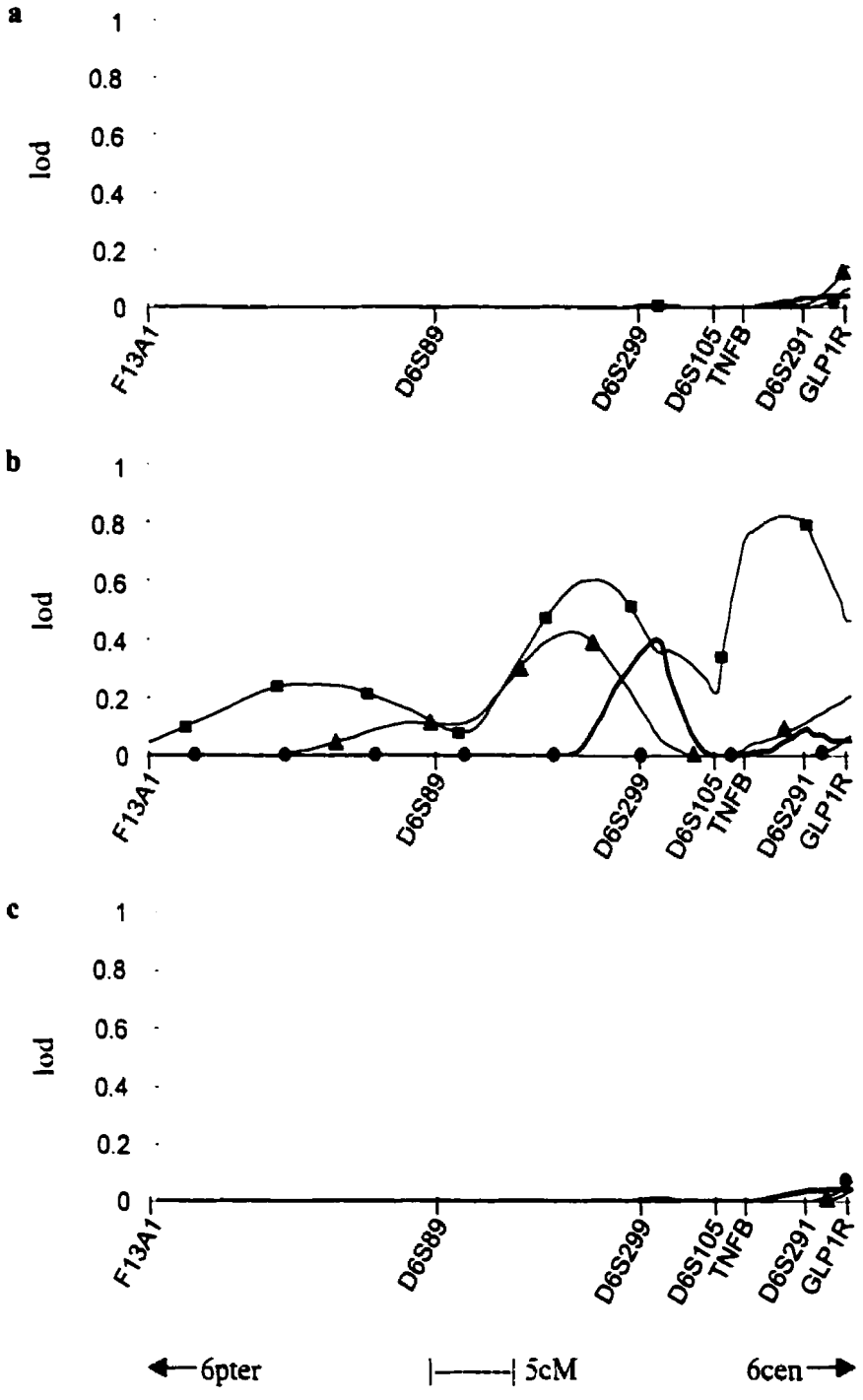


Figure 3.4 Lod-score curves from *DYX2* GENEHUNTER variance-component linkage analysis of phonological awareness (●), phonological coding (▲), spelling (■), and RAN speed (—). Analysis was performed under **a)** Model 1, **b)** Model 2, **c)** Model 3.

3.4. Discussion

The previous study by Field and Kaplan (1998), in which evidence for linkage to the chromosome 6p21.3 region (*DYX2)* was not found, utilized a qualitative PCD phenotype based on quantitative reading data. Since it has been demonstrated that partitioning a quantitative trait into a qualitative phenotype may be less powerful for detecting linkage than using the quantitative data directly (Wijsman and Amos 1997), it was possible that the PCD phenotype lacked power to detect linkage to *DYX2*. Furthermore, the PCD phenotype was based on a composite of dyslexia-related components (phonological awareness, phonological coding, and spelling) that other groups have analyzed separately and have found various strengths of linkage to the DYX2 region. Thus, the possibility existed that use of a composite measure of dyslexia may be less powerful than analyzing single measures of component skills involved in reading ability. Other dyslexia researchers (Fisher et al. 1999) therefore questioned whether the absence of linkage between PCD and *DYX2* in our families was a true nonreplication, or whether the inability to detect linkage was due to the qualitative and composite nature of the PCD phenotype that was used.

To better understand the nature of the PCD phenotype, correlation analysis was performed with four quantitative reading measures. Correlations between PCD, phonological coding, and spelling in all members of the families were quite substantial (r = 0.73-0.77), indicating that the qualitative PCD phenotype was an accurate indicator of reading disability, thus lending credibility to the previous findings of no linkage between dyslexia and *DYX2*. Although the high correlation between PCD and phonological coding was anticipated, since the PCD diagnosis was based primarily on the results of the phonological coding tests, the high correlation to spelling was somewhat unexpected, given that spelling is thought to be comprised of orthographic as well as phonological components, and given that it was only used to assist in diagnosis of PCD. However, the strong correlation between the phonological coding and spelling traits (r = 0.75) indicated that spelling ability significantly involves phonological skills. The low RAN speed correlations that were observed were consistent with the RAN test being a purely orthographic task, whereas the other reading measures and PCD were

comprised primarily of a phonological component. The reason for the modest correlations between phonological awareness and PCD, phonological coding, and spelling may be that the higher-level phonological skills were only partially dependent on phonological awareness skill. Thus, the results of these correlation analyses indicate that the PCD phenotype was an accurate indicator of reading disability, and were not likely to have been the cause of the previous null linkage findings.

The results of these sibpair and variance-component linkage analyses using quantitative measures of reading disability support the previous findings of no evidence for linkage to chromosome 6p23-p21.3 using a qualitative PCD phenotype (Field and Kaplan 1998). Although these quantitative results provide *weak* evidence for a locus affecting reading in this region, they are far from statistically significant. The reason for the lack of significant linkage to *DYX2* in this sample of families, when other studies have found significant linkage and linkage disequilibrium, remains unclear. Analyses were performed using the same markers that showed significant linkage in the other studies and using accredited psychometric tests that, in some cases, were identical to those used in the other studies (e.g. Grigorenko et al. 1997 used the same phonological awareness test). Also, the phenotypic measures had sufficient variability to allow for detection of linkage, and the large number of families and sibpairs in the present study should have conferred high power to detect linkage. In addition, the same QTL sibpair linkage method and variance-component method were used as in one of the previous studies (Fisher et al. 1999). It is therefore proposed that the most likely explanation for the inability to detect linkage to *DYX2* in this sample is that the studies with positive linkage results were enriched for subtypes of dyslexia that were not well-represented in this sample, either due to chance or varying ascertainment criteria. The ascertainment scheme used in the present study specified that at least two siblings met the criteria of having PCD, thus a larger proportion of highly-familial major gene forms of dyslexia may have been selected than ascertainment schemes based on a single dyslexic proband with no specific requirement for a dyslexic sibling (e.g. Cardon et al. 1994; Gayán et al. 1999). In other words, the presence of different genetic forms of dyslexia in our sample compared to other samples may have prevented detection of the DYX2 locus.

Chapter Four: Evidence for a Dyslexia Susceptibility Locus (DYX4) on Chromosome 6q11.2-q12

4.1. Introduction

As discussed in Chapter Three, while the Field and Kaplan linkage study was in its early stages, linkage between dyslexia and chromosome 6p21.3 was reported (Cardon et al. 1994). Thus, markers spanning and flanking the 6p region were investigated to try to replicate this finding. While linkage to 6p21.3 was not detected (Field and Kaplan 1998; Chapter Three), one marker which is centromeric to this region exhibited weak evidence for linkage, thus markers located centromeric to 6p21.3 and on 6q were investigated for evidence for a dyslexia susceptibility locus in this region. I present the results of parametric and nonparametric linkage analyses and linkage disequilibrium analyses of a set of core markers followed by a more dense set of markers on chromosome 6q. The results provide evidence for a dyslexia susceptibility locus on chromosome 6q11.2-q12, which has been named DYX4 by the Human Gene Nomenclature Committee.

4.2. Methods

4.2.1. Subjects

As described in Section 2.2, the study sample consisted of 96 families (46 nuclear families, 50 extended pedigrees) of European descent, with each family containing at least two dyslexic siblings (with the exception of three families that had only one affected member, and three families that had two or more affected members that were not siblings). There were a total of 902 individuals who had DNA sampled and 884 underwent psychometric testing to assess four components of reading: phonological awareness, phonological coding, spelling, and RAN speed (see Section 2.4). Each individual was also assessed for estimated IQ, and adults (>18 years of age) were assessed for reading history (the 18 individuals who did not undergo psychometric testing were diagnosed based on a clear reading history). The results of the reading tests (except RAN speed) and reading history in adults were used to assign an affected, unaffected, or uncertain diagnosis of PCD (see Section 2.5), with the exception of 18

subjects who were diagnosed based on a clear reading history, resulting in 52% of individuals diagnosed as affected with PCD, 33% as unaffected, and 15% as uncertain.

4.2.2. *Descriptive statistics of reading measures*

Descriptive statistics of the phonological awareness, phonological coding, spelling, and RAN speed measures, and estimated IQ, were determined as described in Section 2.6. Statistics were calculated in three samples of subjects used in this study: all 902 members of the 96 families, a sample of 336 sibpairs of all ages, and a sample of 227 sibpairs <18 years of age.

4.2.3. Pearson correlation analysis

Pearson correlation analyses of the phonological awareness, phonological coding, spelling, and RAN speed measures in all members of the 96 families were performed as described in Section 2.6.

4.2.4. Markers and genotyping

Microsatellite markers on chromosome 6q were selected from the following published maps: the Genetic Location Database (LDB) composite map, in which marker locations were determined using available genetic, radiation hybrid, and physical mapping data, (http://cedar.genetics.soton.ac.uk/public_html/ldb.html; November 1999 update; Collins et al. 1996), the Cooperative Human Linkage Center (CHLC) sex-averaged genetic map (http://lpg.nci.nih.gov/CHLC; Version 4.0 map; Murray et al. 1994), the Genethon sex-averaged genetic map (http://www.genethon.fr; Dib et al. 1996), and the Marshfield sex-averaged genetic map (http://research.marshfieldclinic.org/genetics; 1998 update; Broman et al. 1998). Primer sequences were obtained from the Genome Database (GDB) (http://www.gdb.org). Marker genotyping was performed with technical assistance by manual and automated methods, as described in Section 2.8. Genotyping accuracy was checked by analyzing marker haplotypes for excessive numbers of recombination events between marker loci in a small interval. Marker allele frequencies for linkage analyses were calculated from

the parents of one nuclear family selected from each of the 96 pedigrees, as described in Section 2.9. The published heterozygosity (a measure of the degree of polymorphism) and the published map locations for each of the chromosome 6q markers are shown in Table 4.1.

Table 4.1

Published heterozygosities and map locations for the chromosome 6q markers

Marker	heterozygosity	LDB[a]	CHLC[b]	Genethon[b]	Marshfield[b]
D6S1960	0.65	58.61	97.2	-	76.62
D6S294	0.81	64.71	99.4	78.8	78.85
D6S257	0.85	64.76	-	80.0	79.92
D6S402	0.85	65.27	100.8	81.7	80.99
D6S430	0.88	65.91	101.4	82.4	81.52
D6S965	1.00	69.14	-	-	82.59
D6S254	0.66	68.38	-	-	82.59
D6S455	0.75	68.76	-	83.6	82.59
D6S421	0.64	69.32	-	85.0	84.15
D6S280	0.70	70.31	106.2	87.7	87.29
D6S286	0.78	71.64	-	90.0	89.83
D6S460	0.82	71.88	-	90.0	89.83
D6S251	0.78	86.24	108.3	-	90.43
D6S445	0.71	76.39	-	91.8	91.34
D6S1270	0.67	77.72	-	-	92.85
D6S1570	0.79	97.15	-	99.0	99.01
D6S252	0.69	108.03	-	-	102.18

[a] Composite location (calculated using available genetic, physical, RH mapping information).

[b] Genetic location (cM).

4.2.5. Marker mapping

Since the marker order and intermarker genetic distances from the published maps are determined by genotyping a limited number of families, the order of markers and intermarker distances are not well determined, as indicated by different genetic map orders in Figure 4.1. Thus, to generate an accurate marker map for the markers in Table 4.1 (referred to as the "core" markers), marker orders from the published genetic and composite maps were confirmed and discrepancies were resolved by sequence-tagged site (STS) content mapping. As described in Section 2.10, an overlapping set of YAC clones were selected from the Whitehead Institute/MIT Center for Genome Research (WICGR) STS map (Hudson et al. 1995) and were obtained from the MRC Genome Resource Facility in Toronto, Ontario. YAC clones were located on adjacent contigs WC6.8, WC6.9, WC6.10, WC6.11, and WC6.12. The locations of markers on the WICGR STS map were confirmed by PCR amplification of each marker on YAC clone DNA. Markers D6S965, D6S286 and D6S251 were also mapped on the WICGR radiation hybrid map, as described in Section 2.11. Genetic distances were determined from analysis of the 96 families by the MultiMap genetic map building program, as described in Section 2.12.

An additional 52 markers spanning and flanking the core markers were selected from published maps for use in linkage disequilibrium (LD) analyses (hence, are referred to as the "LD" markers). The locations of many of these markers were also determined by STS content mapping, as described above. With the recent availability of the human genome sequence (International Human Genome Sequencing Consortium 2001), however, the marker order from the genome sequence was utilized for linkage disequilibrium analyses, although some markers were not mapped on the genome sequence, thus their STS content mapping locations assisted in placing them relative to the other markers.

4.2.6. Qualitative linkage analysis of PCD

Initial linkage analyses of the core markers were performed using the qualitative PCD phenotype (affected, uncertain, or unaffected diagnosis) by parametric and

nonparametric methods. Two-point parametric linkage analysis was performed using the FASTLINK program, as described in Section 2.14.1.1. To increase the likelihood of detecting linkage, analyses were performed using eight genetic models (Table 4.2), shown by many studies to be a mathematically valid method to detect linkage (Elston 1989; Clerget-Darpoux and Bonaïti-Pellié 1992; Hodge and Elston 1994, Greenberg et al. 1998). The models ranged across recessive, intermediate, and dominant modes of inheritance, all with reduced penetrance (allowing subjects with a PCD-susceptible genotype to have an unaffected phenotype due to reduced disease penetrance or misdiagnosis).

Table 4.2

Inheritance models used in *DYX4* two-point parametric linkage analyses of PCD

Model	1	2	3	4	5	6	7	8
Penetrances: aa	0	0	0	0	0	0	0	0
ab	0	0	0.4	0.4	0.4	0.6	0.8	0.8
bb	0.8	0.8	0.6	0.6	0.6	0.8	1.0	1.0
Disease allele b frequency	0.01	0.25	0.001	0.01	0.25	0.01	0.01	0.05

Multipoint parametric linkage analysis was performed using the GENEHUNTER program, as described in Section 2.14.1.2. Analyses were performed under model 8 only (Table 4.2), since this model produced the most significant two-point linkage results. Lod scores under linkage heterogeneity (hlod scores) are reported since they provided more significant evidence for linkage than lod scores without heterogeneity, suggesting genetic heterogeneity in the sample. The analyses were performed using the sex-averaged map derived from MultiMap analysis of the families, and confirmational analyses were performed using the Marshfield sex-averaged genetic map.

Multipoint nonparametric linkage (NPL) analysis was performed using the GENEHUNTER program, as described in Section 2.14.3, using the sex-averaged map

derived from MultiMap analysis of the families, with confirmatory analyses using the Marshfield sex-averaged genetic map.

Genetic heterogeneity testing was performed by the HOMOG program, as described in Section 2.14.4, using the core markers that had a two-point lod score >2.

4.2.7. Quantitative-trait locus linkage analysis of reading measures

After the initial qualitative linkage analyses were performed, a study was published demonstrating that quantitative analysis may be more powerful than analysis of a qualitative trait (Wijsman and Amos 1997). It was therefore decided to perform multipoint quantitative-trait locus (QTL) linkage analyses of the core markers using each of the reading measures that had been assessed in each subject (phonological awareness, phonological coding, spelling, and RAN speed). The analyses were performed using the sex-averaged genetic map derived from MultiMap analyses of the families, and confirmational analyses were performed using the Marshfield genetic map. Maximum-likelihood variance-component linkage analyses were performed on the dataset of 96 families (i.e., including extended pedigrees) using the GENEHUNTER program, as described in Section 2.15.2. Four models were tested: Model 1 included QTL additive variance, polygenic additive variance, and environmental variance; the others were variations of Model 1, with either QTL dominance variance added (Model 2), polygenic dominance variance added (Model 3), or both QTL and polygenic dominance variance added (Model 4). Multipoint QTL sibpair linkage analyses were performed using the nonparametric (rank-sum test) option in the MAPMAKER/SIBS program, as described in Section 2.15.1.2. As outlined in Section 2.14.2, sibpair analyses were performed using a sample of sibpairs of all ages from 96 nuclear families (336 sibpairs, 188 independent sibpairs) and using a subsample of 81 nuclear families with sibpairs <18 years of age (227 sibpairs, 145 independent sibpairs), in which the psychometric test ceiling effects were reduced. Table 4.3 indicates the numbers of nuclear families of various sibship sizes, in both the all-ages sibpair sample and the <18 years of age sibpair sample. The large majority of nuclear families consisted of two or three siblings, with only a few larger nuclear families that were not likely to distort the linkage results.

Table 4.3

Distributions of sibship sizes in the all-ages sibpair sample and the <18 years of age sibpair sample used for *DYX4* QTL sibpair linkage analyses

	Number of nuclear families	
Sibship size:	all-ages sample	<18 years sample[a]
Two sibs	39	32
Three sibs	34	37
Four sibs	16	9
Five sibs	5	3
Seven sibs	1	-
Eight sibs	1	-
Total:	96	81

[a] Restriction to siblings <18 years of age resulted in the exclusion of 15 nuclear families, thus the total number of nuclear families in this sample was 81, and changed the distribution of sibship sizes in the remaining nuclear families.

4.2.8. Haplotype analysis

Chromosome 6 core marker haplotypes for each individual were determined using the SimWalk2 program, as described in Section 2.16, using the marker map derived from MultiMap analysis of the families. Each family was investigated for sharing of a common haplotype (or haplotypes) anywhere across 6q11.2-q12 markers (from D6S965 to D6S251) amongst all PCD affected individuals within the family, indicating that the family had a higher likelihood of carrying the dyslexia susceptibility locus located in this region. Unaffected or uncertain family members were also permitted to share the affected haplotype to allow for incomplete disease penetrance. These families were designated as "linked", and their affected haplotypes were investigated for recombination breakpoints to attempt to narrow the candidate susceptibility region.

4.2.9. Linkage disequilibrium analysis

Single-marker LD analysis of PCD was performed using the family-based AFBAC program, as described in Section 2.1.1. The dataset of 96 nuclear families was analyzed for associations between PCD and the core chromosome 6q markers. In an attempt to reduce genetic heterogeneity in the sample, separate analyses were also performed on the subset of nuclear families from pedigrees with a high probability of linkage to this region. These 32 linked nuclear families were investigated for linkage disequilibrium to the core markers, as well as the additional 52 LD markers spanning and flanking the core markers.

Multiple-marker haplotype LD analysis was performed using the TRIMHAP program, as described in Section 2.17.2. This program analyzes extended pedigrees for significant association of haplotypes with the disease by determining empirical significance levels for the disease locus being located between each marker in the haplotype. Note that this method does not confound linkage with association. Haplotypes with $P < 0.05$ were also investigated using the HAPMAX program, which analyzes parent-affected child trios for significant association of haplotypes with the disease by chi-squared testing, as described in Section 2.17.2. The dataset of 96 families was analyzed for associations between PCD and two-marker and three-marker haplotypes of the core chromosome 6q markers (larger haplotypes could not be analyzed due to program constraints). The linked nuclear families were investigated for LD to two-marker haplotypes of the core markers and the 52 LD markers. Haplotypes larger than two markers were not analyzed due to the need for correct marker order, which was not determined for the LD markers. Note that although the TRIMHAP program allows analysis of extended pedigrees, only nuclear families were genotyped for the LD markers, thus only nuclear families were used in these analyses.

LD analysis of restriction length fragment polymorphisms (RFLPs) of two candidate genes located in the 6q region, the serotonin receptor genes *HTR1β* and *HTR1E*, was performed on the linked nuclear families. As described in Section 2.13, the following RFLPs were selected from the literature, genotyped in the linked nuclear families, and analyzed using AFBAC: *HTR1β* T-261G and *HTR1β* G861C (silent

variant). In addition, *HTR1β* G-511T, *HTR1β* del-179/-178, *HTR1β* T371G and *HTR1E* C531T were genotyped in the linked nuclear families; however, these variants were not polymorphic in this sample and were not analyzed for LD with PCD.

4.3. Results

4.3.1. Descriptive statistics of reading measures

Table 4.4 shows the descriptive statistics of the phonological awareness, phonological coding, spelling, and RAN speed measures (after age-adjustment or conversion to standard scores) and estimated IQ in three samples: all members of the 96 families, and the all-ages sibpair sample and <18 years of age sibpair sample used for QTL sibpair linkage analyses. All reading measures exhibited adequate variance and were therefore expected to provide sufficient power to detect linkage. With the exception of phonological awareness, the distributions of the reading measures were nearly normal, and since standard data transformations were not able to generate more normal distributions for any of the measures, untransformed data were utilized in QTL linkage analyses. In comparison to the descriptive statistics of these measures in the control subjects (see Table 2.2), the means of each measure are lower in the family data set. Note that estimated IQ was not utilized for any linkage analyses.

4.3.2. Pearson correlation analysis

Pearson correlation analyses were performed to clarify the relationships between the quantitative reading measures (phonological awareness, phonological coding, spelling, and RAN speed). As shown in Table 4.5, the Pearson correlation coefficient (r) between phonological coding and spelling was substantial at $r = 0.74$, suggesting that spelling ability is strongly influenced by phonological skills. Phonological awareness was moderately correlated to phonological coding and spelling ($r = 0.54$ and 0.51, respectively), suggesting some overlap between these skills. RAN speed was not highly correlated with any of the other reading measures (range, $r = 0.34$ to 0.42).

Table 4.4

Descriptive statistics of the reading measures and estimated IQ in all members of the 96 families (*N*=902), the all-ages sibpair sample (*N*=336), and the <18 years of age sibpair sample (*N*=227)

Trait	Mean	SD	Min	Max	Skewness	Kurtosis
Phonological awareness[a]						
All members	18.09	5.67	1.00	28.81	-0.92	0.19
All-ages sibpairs	17.49	5.83	2.65	28.79	-0.51	-0.59
<18 years sibpairs	17.53	5.95	2.65	28.79	-0.42	-0.71
Phonological coding[b]						
All members	98.84	16.60	12	149	-0.03	0.93
All-ages sibpairs	91.02	15.25	46	149	0.50	1.05
<18 years sibpairs	90.02	14.92	46	139	0.42	0.84
Spelling[b]						
All members	91.08	16.16	47	124	-0.21	-0.90
All-ages sibpairs	82.16	14.54	47	119	0.46	-0.43
<18 years sibpairs	81.53	14.33	47	119	0.63	-0.16
RAN speed[a]						
All members	3.09	0.62	1.00	5.60	0.16	0.18
All-ages sibpairs	2.98	0.59	1.00	4.71	0.12	0.31
<18 years sibpairs	3.02	0.56	1.59	4.71	0.29	0.16
Estimated IQ[b]						
All members	104	12.5	48	146	0.01	0.69
All-ages sibpairs	104	13.7	59	141	0.22	0.43
<18 years sibpairs	104	14.2	59	141	0.17	0.32

[a] Data are raw scores for adults and age-adjusted scores for subjects <18 years of age.

[b] Data are standard scores according to test norms (normal population mean = 100, standard deviation =15).

Table 4.5

Pearson correlation coefficient matrix of phonological awareness, phonological coding, spelling, and RAN speed using all members of the 96 families

	Phonological awareness	Phonological coding	Spelling
Phonological coding	0.54	-	-
Spelling	0.51	0.74	-
RAN speed	0.34	0.40	0.42

NOTE — all correlations are significant at $P < 0.0001$.

4.3.3. Marker mapping

The results of STS content mapping of the core markers and the 52 LD markers are shown in Figure 4.1. Note that core markers D6S294 and D6S965 were not mapped due to failure to amplify on any of the YAC clones tested, and that D6S1960, D6S1570, and D6S252 could not be mapped since they are located outside of the set of YAC clones (the overlapping set of clones was selected to cover most of the 6q region of interest). The STS content mapping results corresponded well to published genetic maps, with the exception of the position of D6S251. This marker is placed telomeric to D6S1270 on the LDB composite map but is centromeric to D6S445 on the Marshfield genetic map. However, STS content mapping placed both D6S251 and D6S445 on YAC clones 844-H-5, 914-A-12, and 956-F-11, thus positioning D6S251 very close to D6S445. Note that some of the LD markers were also not mapped due to PCR failure or being located outside the set of YAC clones.

The results of radiation hybrid mapping placed D6S286 on the WICGR RH map in a position that corresponded well to its location on published genetic maps, 16.6cR from GATA11F10 with lod 1.92 (odds 83:1) relative to the next most likely placement. D6S965 was placed 12.2cR from GATA11F10 with lod 2.98 (odds 955:1), thus placing it close to D6S286, which is discrepant from the LDB and Marshfield genetic maps. The second most likely location of D6S965, however, was 65.6cR from WI-5488, which

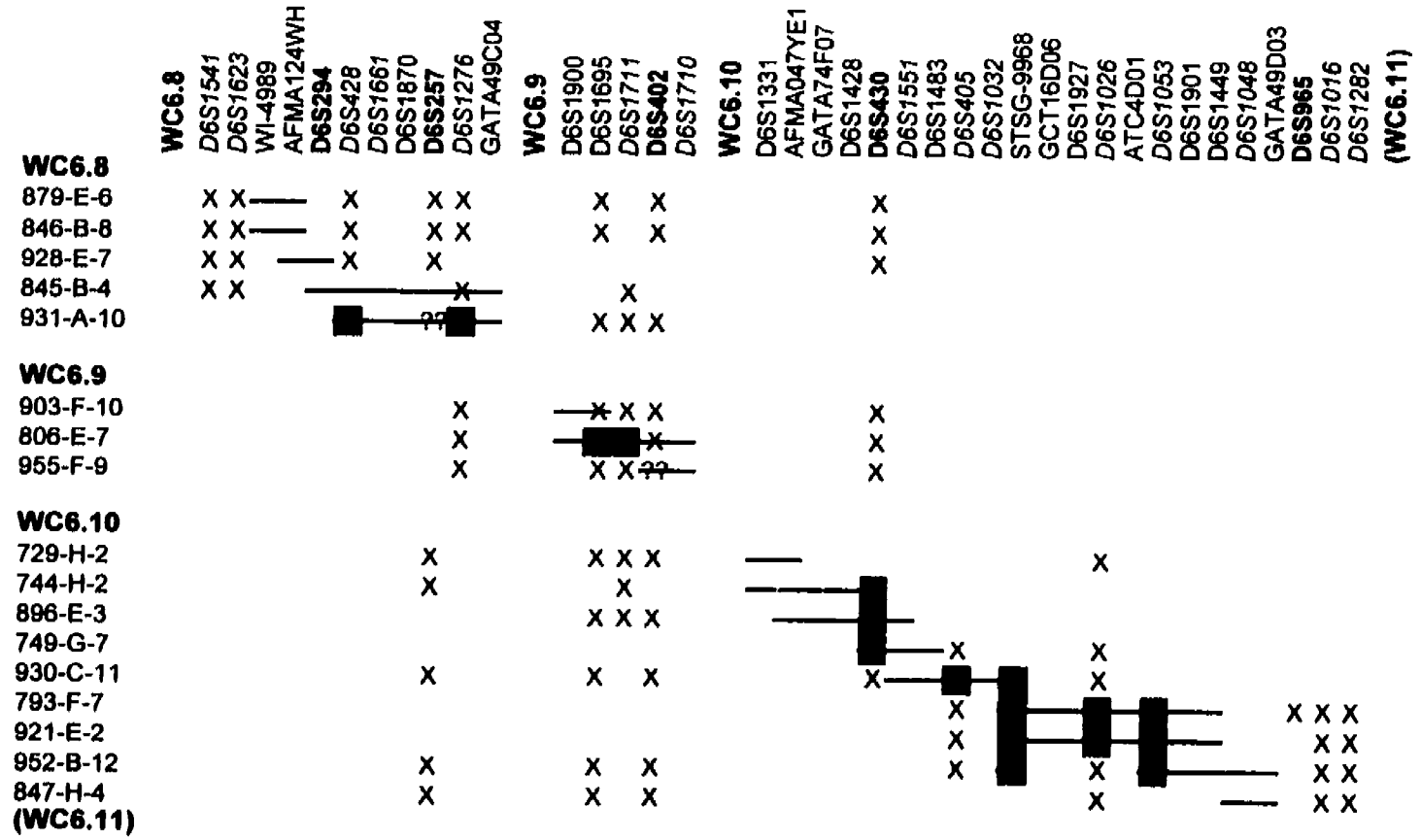


Figure 4.1 Results of STS content mapping of chromosome 6q core markers (**bold**) and LD analysis markers (*italics*). Markers in regular font were not used in any analyses. YAC clones in each contig (WC6.8, WC6.9, etc.) are shown as horizontal lines. Filled box = positive amplification of marker on YAC clone, "X" = negative amplification, "??" = inconclusive amplification results.

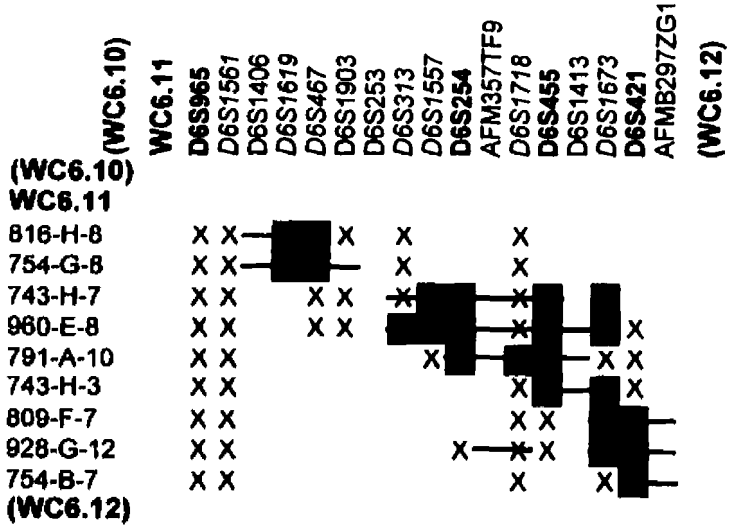


Figure 4.1 (cont'd)

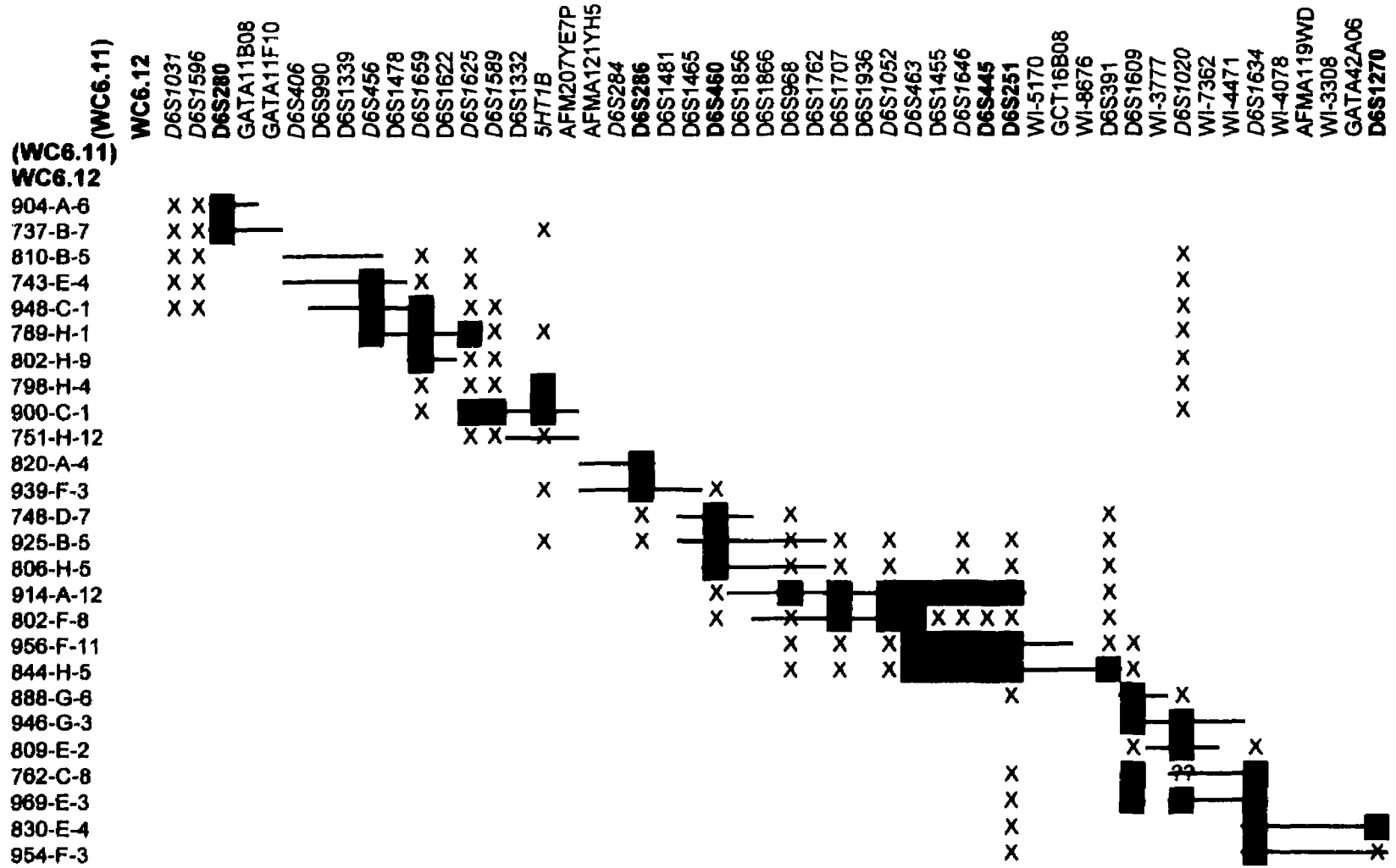


Figure 4.1 (cont'd)

places it near D6S430 in agreement with the genetic maps. Furthermore, the location of D6S965 on the human genome sequence corresponds to its location on the Marshfield map, thus this location was taken as correct. As mentioned above, the location of D6S251 is discrepant between the LDB and Marshfield genetic maps. Radiation hybrid mapping placed D6S251 3.9cR from WI-3966 with lod 1.94 (odds 87:1), thus positioning D6S251 between D6S460 and D6S445. This is in agreement with STS content mapping, where D6S251 and D6S445 map to the same YAC clones and thus are very close, and with the human genome sequence location of D6S251.

The sex-averaged genetic map of the core markers, with intermarker distances determined by MultiMap genetic mapping of the families, is shown in Figure 4.2.

Figure 4.3 shows the map of LD markers relative to the core markers, with the LD marker order determined from the human genome sequence (for most markers) or from the LDB composite map.

4.3.4. Qualitative linkage analysis of PCD

The results of two-point parametric linkage analyses suggested linkage between PCD and markers on chromosome 6q11.2-q12. As shown in Table 4.6, the maximum lod score was found at marker D6S251 ($Z_{max} = 2.82$, $\theta = 0.25$) under a dominant model with a common disease allele (b) frequency of 5% (corresponding to a population prevalence of 10% if completely penetrant) and reduced penetrances of 0, 0.8, and 1.0 for the aa, ab, and bb genotypes, respectively (model 8). Three other markers had maximum lod scores >2 under the same model (D6S254 $Z_{max} = 2.49$, D6S965 $Z_{max} = 2.39$, D6S286 $Z_{max} = 2.47$). Z_{max} scores >2 were also obtained under two other similar models (models 6 and 7). These linkage results surpass the recommended threshold for suggestive linkage (lod = 1.9), but fall short of the threshold to establish significant evidence for linkage (lod = 3.3) (Lander and Kruglyak, 1995). The markers span an 11cM region (Figure 4.2), or ~13Mb according to the human genome sequence (International Human Genome Sequencing Consortium 2001), on chromosome 6q11.2-q12 according to the LDB chromosome 6 "gmap", in which marker cytogenetic locations were determined using cytogenetic band fractional length data (Francke 1994).

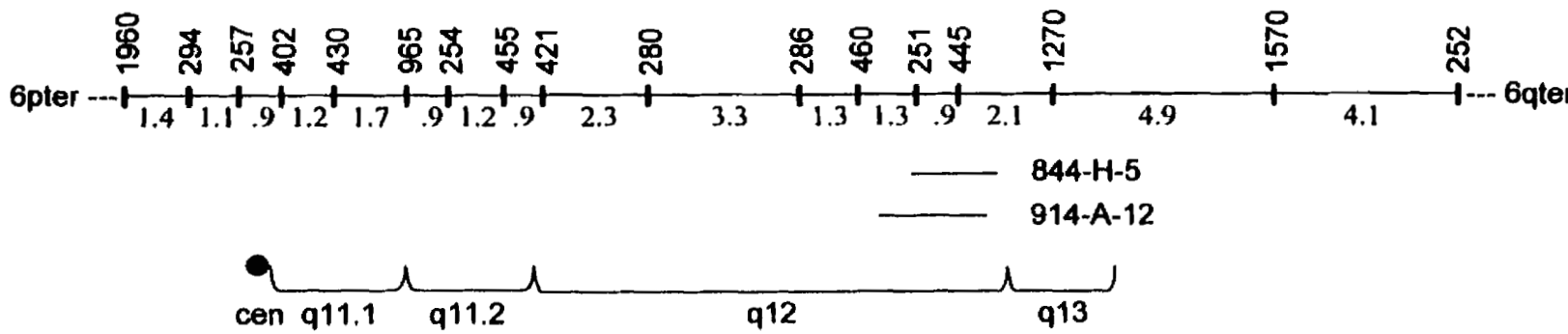


Figure 4.2 Sex-averaged genetic marker map of chromosome 6q core markers. Intermarker distances (cM) were derived from analysis of the pedigrees. YAC clones 844-H-5 and 941-A-12, used in STS content mapping of D6S251, are indicated below the map. Cytogenetic band information is from the Genetic Location Database "gmap".

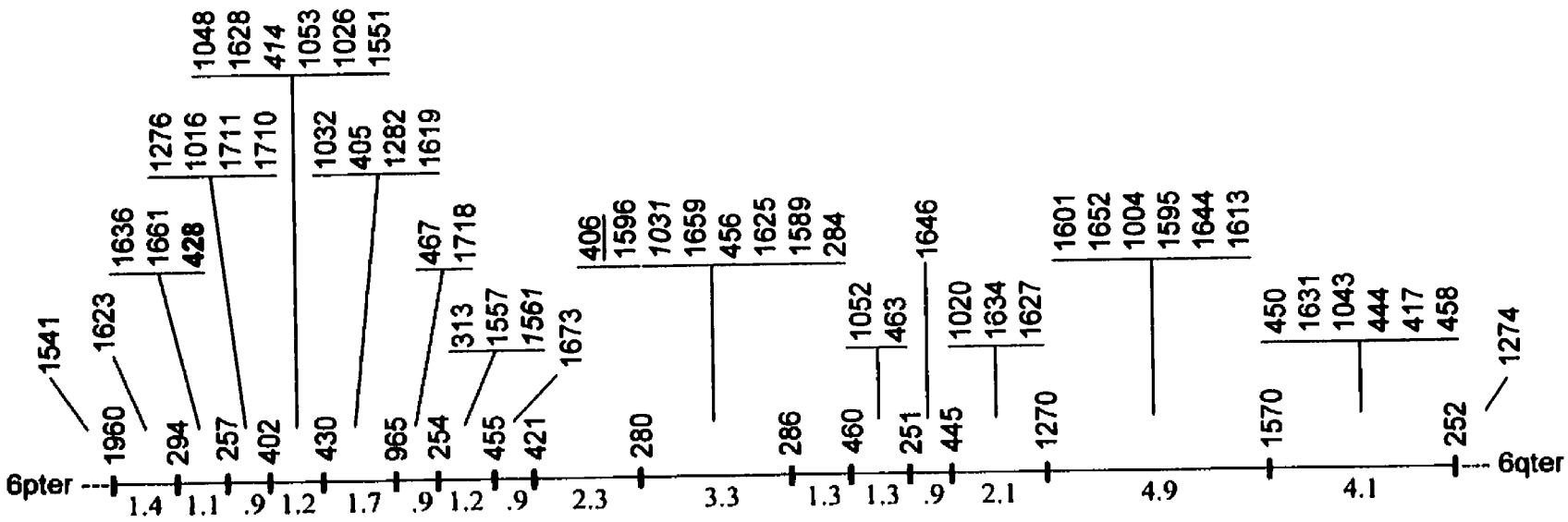


Figure 4.3 Chromosome 6q markers used for linkage disequilibrium analysis (indicated above the "core" markers), with locations according to the human genome sequence or, if necessary, the Genetic Location Database composite map STS location (underlined, or bold if confirmed by STS content mapping) or genetic location (italics).

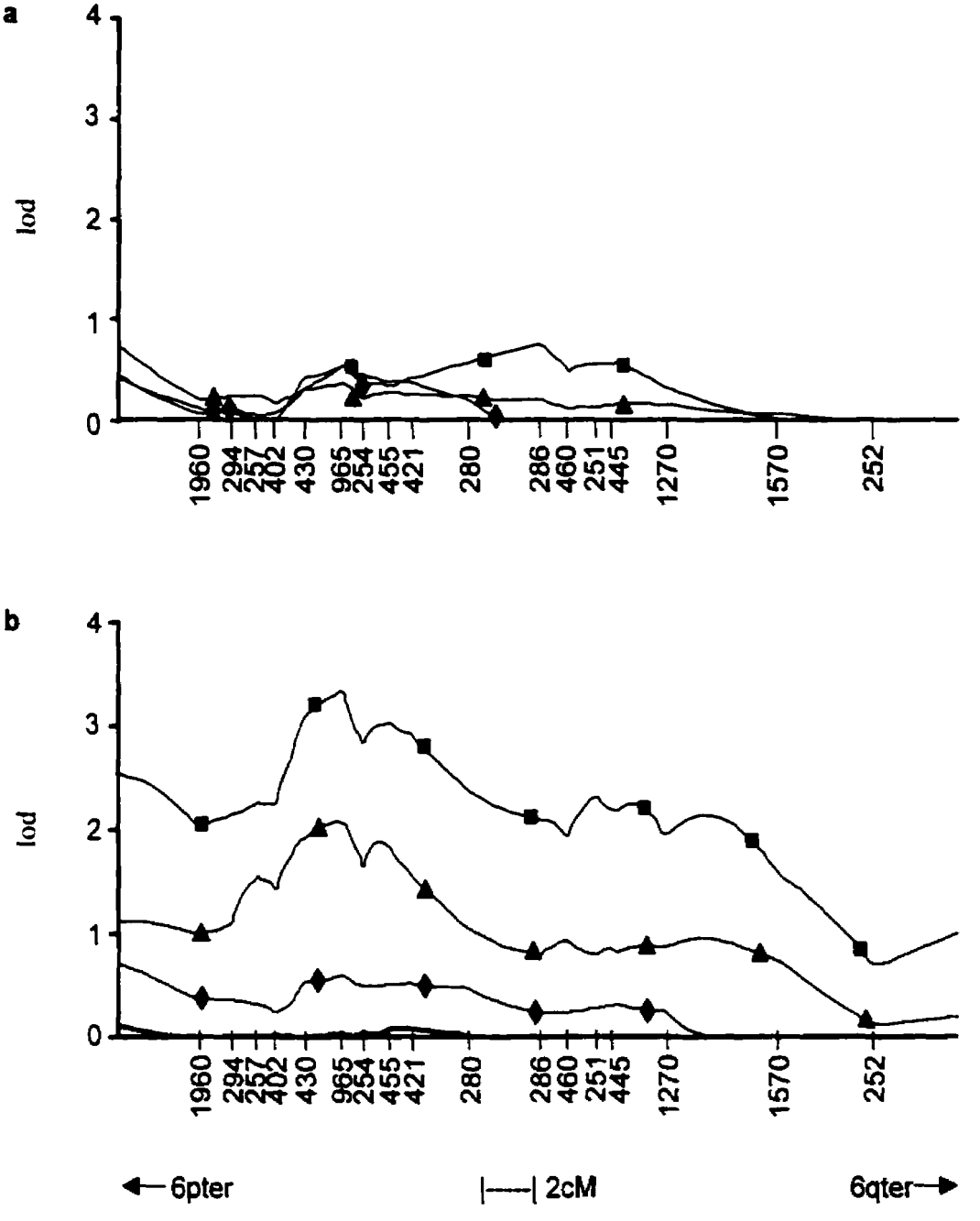


Figure 4.6 Lod-score curves from *DYX4* GENEHUNTER variance-component linkage analysis of phonological awareness (—), phonological coding (▲), spelling (■), and RAN speed (♦). Analysis was performed under **a)** Model 1, **b)** Model 2, **c)** Model 3, **d)** Model 4.

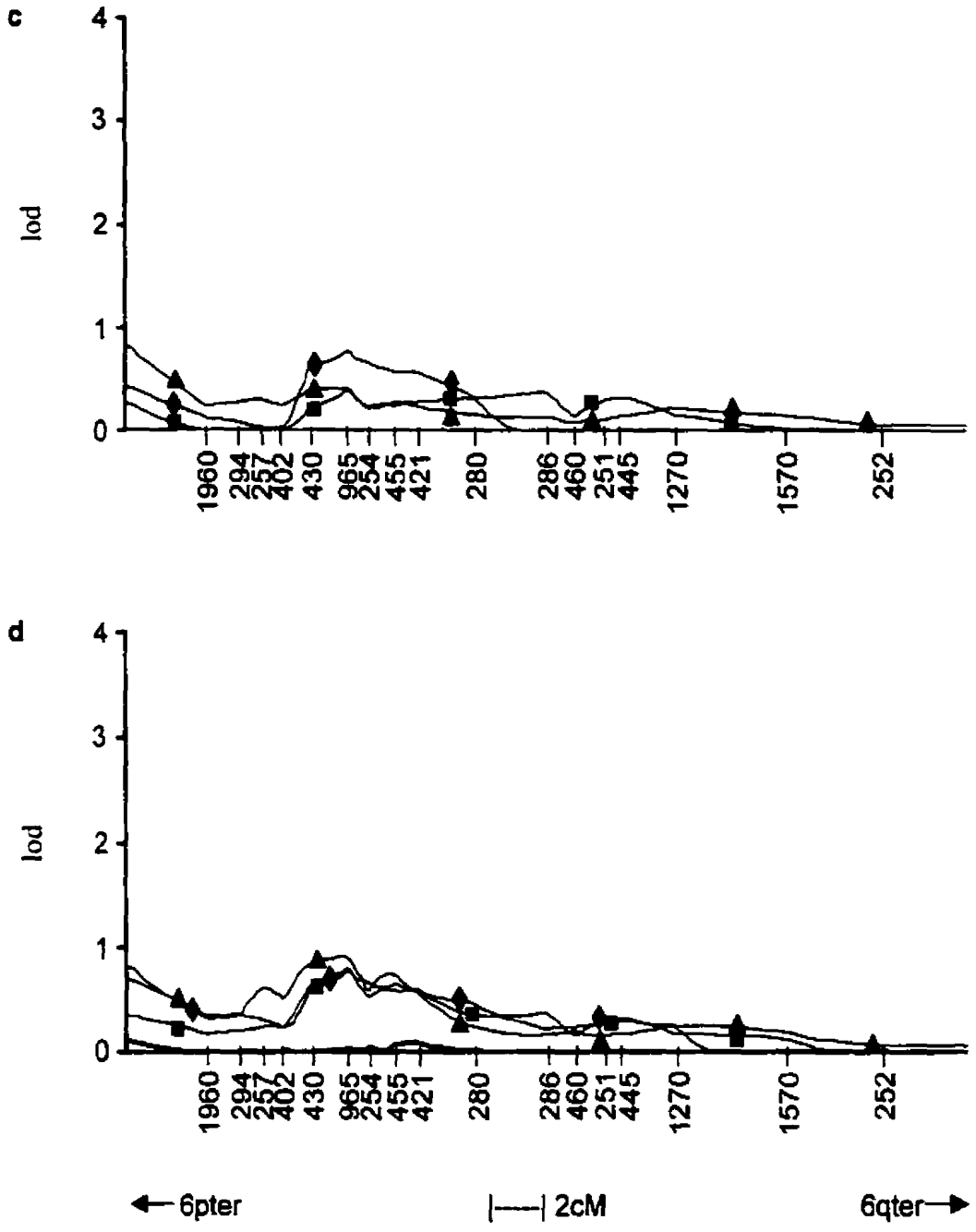


Figure 4.6 (cont'd)

Multipoint linkage analyses also suggested linkage to the chromosome 6q region. As shown in Figure 4.4, multipoint parametric analysis using model 8 identified a peak maximum hlod of 1.58 between D6S280 and D6S286. As shown in Figure 4.5, multipoint NPL analysis identified a peak NPL Z_{all} score of 2.21, corresponding to P = 0.012, at D6S460, with the interval from D6S286 to D6S445 significant at P <0.05.

4.3.5. Genetic heterogeneity testing

Genetic heterogeneity testing using the HOMOG program was performed under model 8 (Table 4.2) using the four markers with two-point Z_{max} scores >2 (D6S254, D6S965, D6S286, and D6S251). As shown in Table 4.7, the likelihood ratios (L^*) of linkage heterogeneity versus no linkage ranged from 50 to 699 for these markers, which corresponds to 0.05> P >0.001 (with 2 degrees of freedom for the independent parameters α and θ). A likelihood ratio >2000, corresponding to P <0.0001, is the recommended criterion to establish heterogeneity when there is no prior significant evidence for linkage (e.g. Z_{max} <3). Thus, significant evidence for linkage with heterogeneity could not be established. However, investigation of lod scores of these four markers in each pedigree found that some families had positive lod scores in this region, while other families clearly demonstrated lack of linkage by very negative lod scores, suggesting that heterogeneity between families (interfamilial heterogeneity) exists, but cannot be detected with the current methods.

Note that under the hypothesis of linkage and heterogeneity (H_2), the proportion of families linked to this region (α) ranges from 0.6 to 1.0. However, because the evidence for linkage heterogeneity is not significant, and because genetic heterogeneity testing requires assumptions that cannot be verified (e.g. all mutations of all genes are equally penetrant), the estimated α's are probably inaccurate (Whittemore and Halpern 2001).

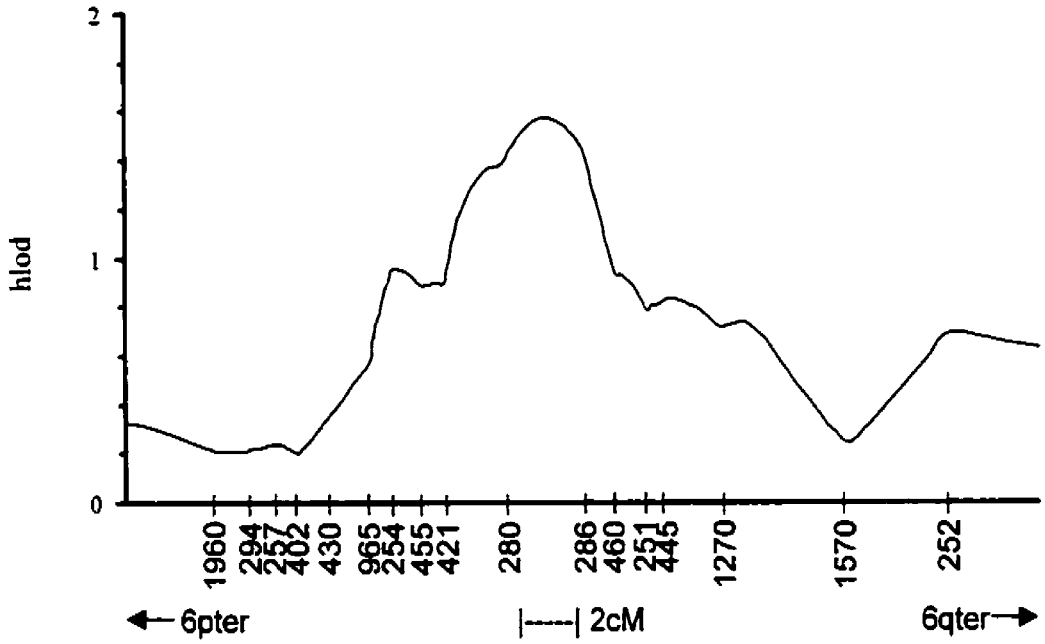


Figure 4.4 Hlod-score curve from *DYX4* GENEHUNTER multipoint parametric linkage analysis of PCD.

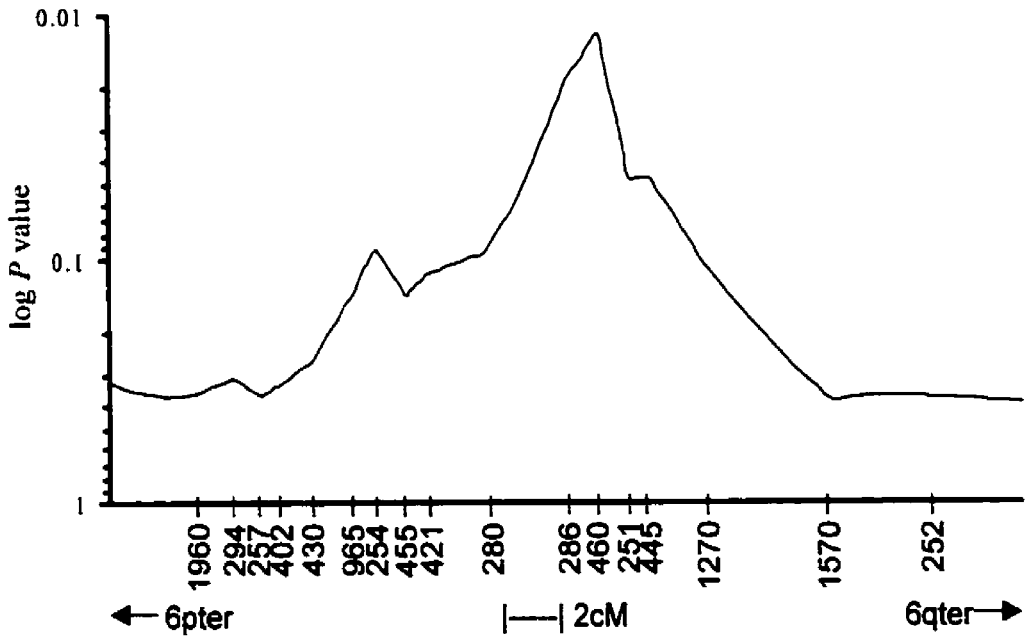


Figure 4.5 Log *P*-value curve from *DYX4* GENEHUNTER multipoint nonparametric linkage analysis of PCD.

Table 4.7

Results of genetic heterogeneity testing of D6S254, D6S965, D6S280, and D6S251

	Maximum $\log_e L(\alpha,\theta\hat{})$[a]		L^*		
Marker	H_1: linkage, homogeneity	H_2: linkage, heterogeneity	H_1 vs. H_0[b]: linkage	H_2 vs. H_1: heterogeneity	H_2 vs. H_0[b]: linkage heterogeneity
D6S254	5.82 (1, .22)	5.82 (1, .22)	337	1.0	337
D6S965	3.90 (1, .28)	3.90 (1, .28)	50	1.0	50
D6S286	5.73 (1, .26)	6.31 (.6, .18)	308	1.79	550
D6S251	6.52 (1, .26)	6.55 (.9, .24)	683	1.02	699

[a] Natural log of the maximum likelihood [$L(\alpha,\theta\hat{})$] of the hypothesis.

[b] L of $H_0 = 1$.

4.3.6. Quantitative-trait locus linkage analysis of reading measures

Multipoint QTL linkage analyses of the phonological awareness, phonological coding, spelling, and RAN speed measures were performed using variance-component and sibpair approaches. As shown in Figure 4.6b, variance-component analyses under Model 2. which included QTL dominance variance, found evidence for linkage to spelling (peak lod = 3.34 at D6S965; 1 lod confidence interval from D6S402 to D6S280) and phonological coding (peak lod = 2.08 at D6S965). This analysis was performed under 2df, thus the peak lod scores under 1df are approximately equivalent to 2.6 (spelling) and 1.5 (phonological coding), for comparison to traditional lod score analysis. RAN speed and phonological awareness had lod scores <0.6 and near zero across the region, respectively. Analyses without QTL or polygenic dominance variance (Model 1), with polygenic dominance variance (Model 3), or with both QTL and polygenic dominance variance (Model 4), provided little evidence for linkage (Figure 4.6a.c.d. respectively).

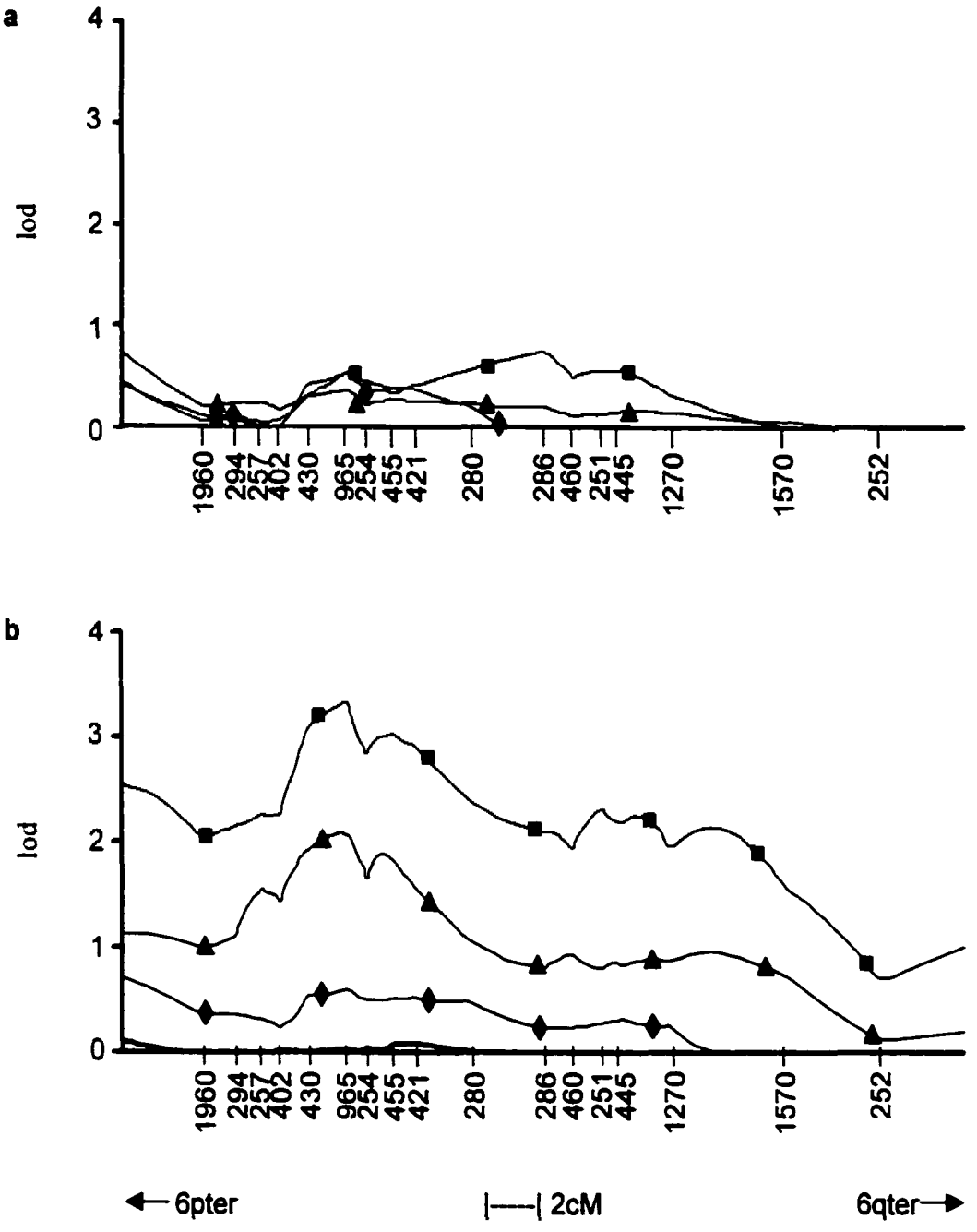


Figure 4.6 Lod-score curves from *DYX4* GENEHUNTER variance-component linkage analysis of phonological awareness (—), phonological coding (▲), spelling (■), and RAN speed (♦). Analysis was performed under **a)** Model 1, **b)** Model 2, **c)** Model 3, **d)** Model 4.

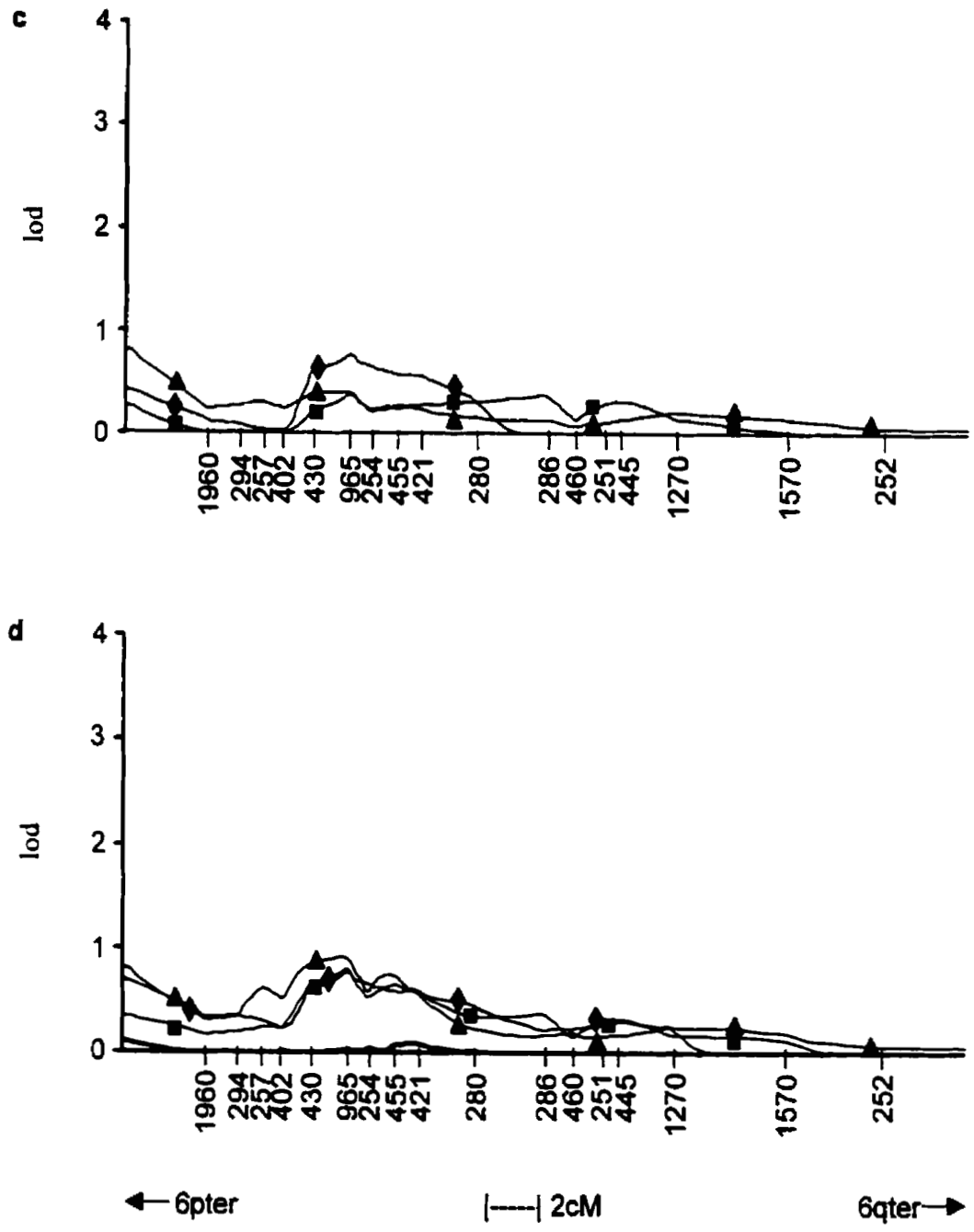


Figure 4.6 (cont'd)

Multipoint nonparametric QTL sibpair analyses of sibpairs of all ages found moderate evidence for linkage to phonological awareness (unweighted Z = 1.95, P = 0.026 near D6S455; weighted Z = 0.94, P = 0.17 at D6S965; Figure 4.7a) and spelling (unweighted Z = 2.66, P = 0.0039; weighted Z = 1.49, P = 0.068 at D6S286; Figure 4.7c). Phonological coding and RAN speed were not significant (P <0.05, unweighted and weighted; Figure 4.7b,d, respectively). Analyses of sibpairs <18 years of age, in which psychometric test ceiling effects were reduced, provided stronger evidence for linkage to the 6q region. The most significant results were found in analyses of spelling (unweighted Z = 3.27, P = 0.00053; weighted Z = 1.80, P = 0.036 at D6S286; Figure 4.8c), and moderate results were found with phonological awareness (unweighted Z = 2.10, P = 0.018 at D6S455; weighted Z = 0.92, P = 0.18 near D6S254; Figure 4.8a) and phonological coding (unweighted Z = 2.13, P = 0.017 at D6S257; weighted Z = 1.13, P = 0.13 near D6S402; Figure 4.8b). RAN speed results were not significant (Figure 4.8d).

4.3.7. Haplotype analysis

Haplotype analysis identified 32 linked pedigrees in which all affected family members share a common haplotype(s) anywhere across D6S965 to D6S251 (Figure 4.9). Recombination breakpoints between D6S280 and D6S286 in pedigrees 1948, 2005, and 3919, and between D6S286 and D6S460 in pedigree 1932, identified a consensus region between D6S280 and D6S460 that potentially harbours the *DYX4* gene (delimited by vertical lines, Figure 4.9). This region spans ~ 4.6cM, or 6.9Mb according to the human genome sequence (International Human Genome Sequencing Consortium 2001). A breakpoint between D6S280 and D2S251 in pedigree 1954 could not be localized because the recombinant individual was homozygous for D2S286 and D2S460, thus localization of this breakpoint could further refine the consensus region. While the 6.9Mb consensus region is more likely to contain the *DYX4* gene than elsewhere on 6q11.2-q12, the critical breakpoint was only found in pedigree 1932, in which two affected siblings share a common haplotype, which is expected to occur by chance at ~50% probability. Thus, these siblings may not actually carry the *DYX4* gene, thus this breakpoint and the resulting consensus region should be regarded with caution.

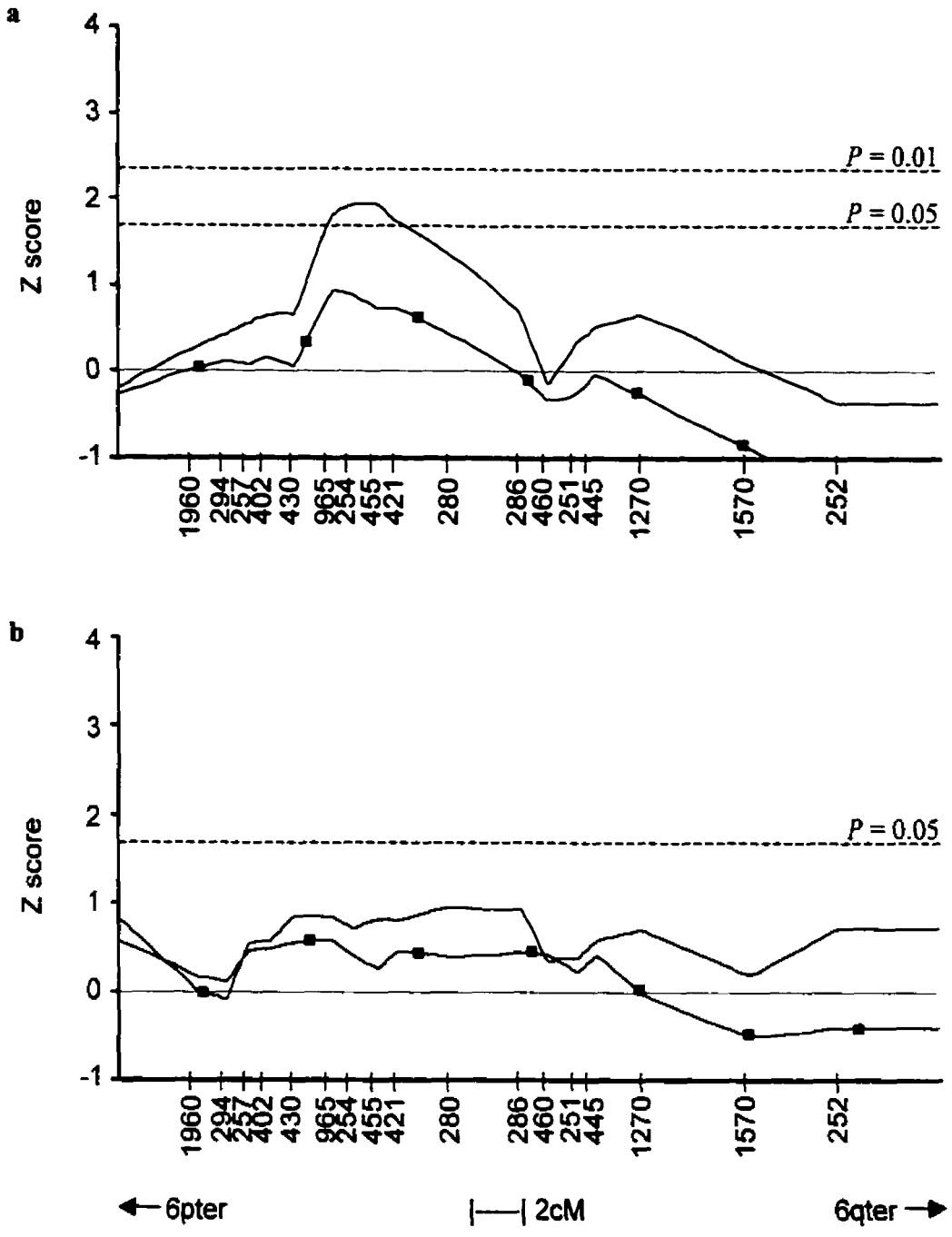


Figure 4.7 *Z*-score curves from *DYX4* MAPMAKER/SIBS nonparametric sibpair linkage analysis of **a)** phonological awareness, **b)** phonological coding, **c)** spelling, and **d)** RAN speed, using a sample containing sibpairs of all ages. Analyses were performed without weighting (——) and with weighting (—■—) of multiple sibships.

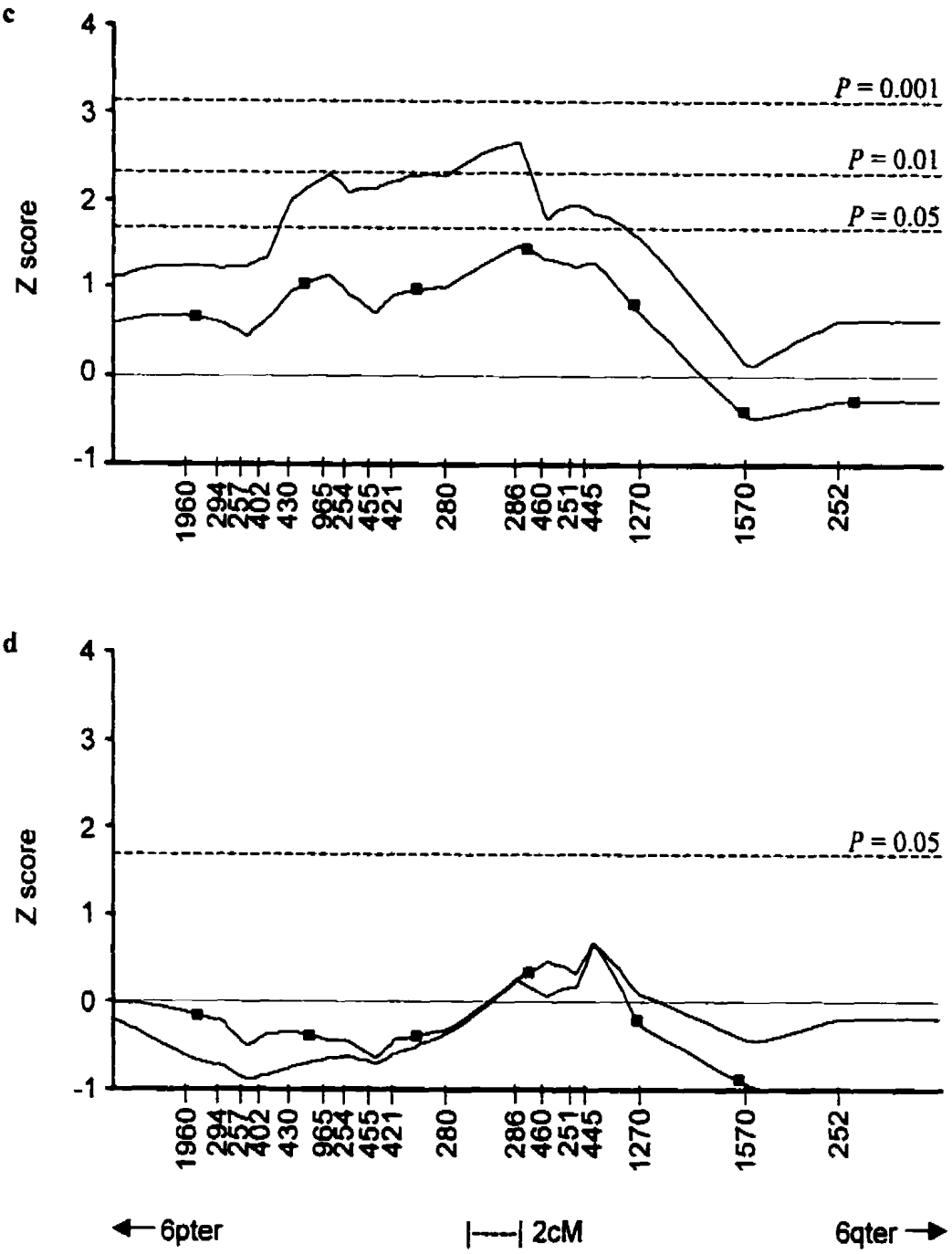


Figure 4.7 (cont'd)

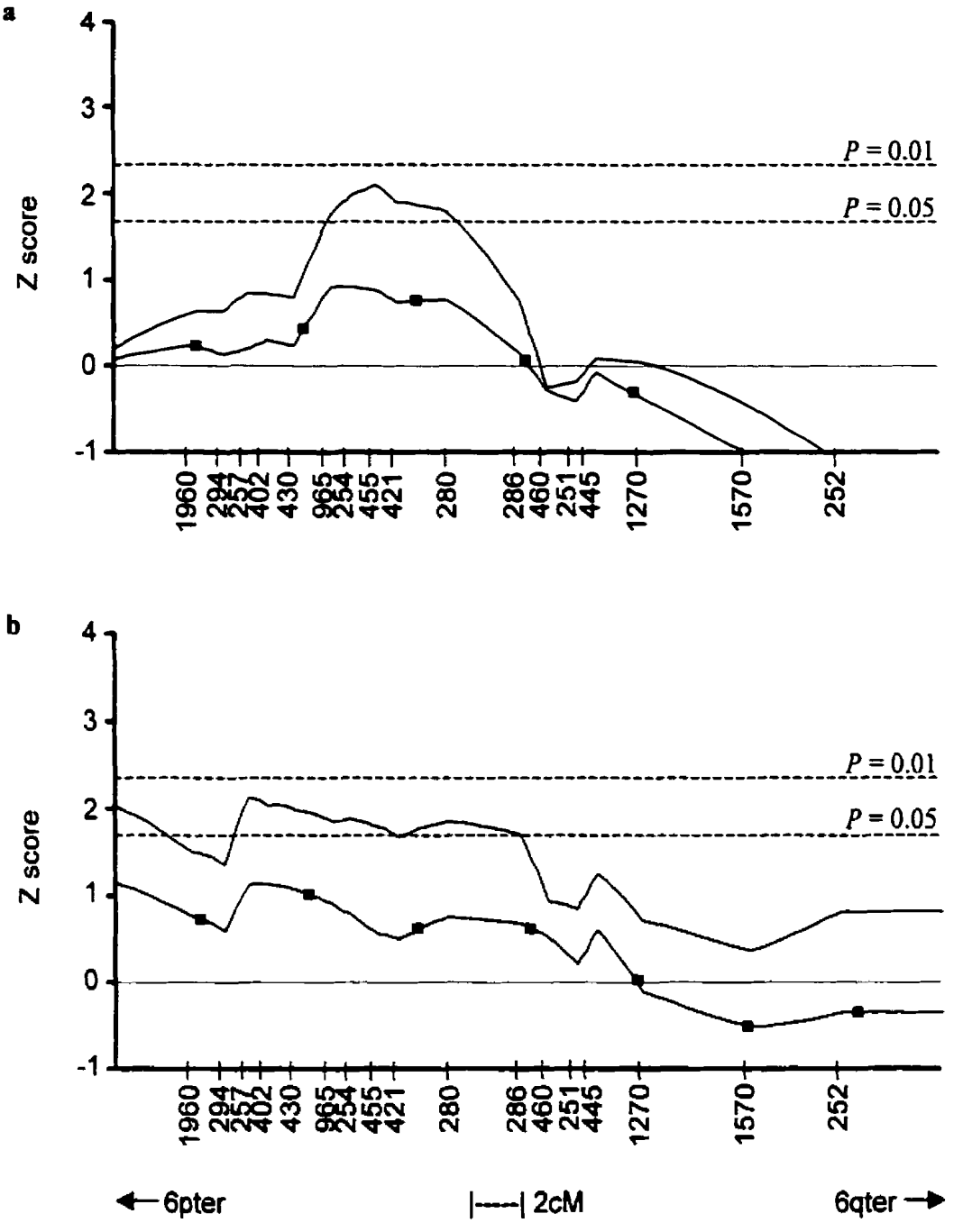


Figure 4.8 Z-score curves from *DYX1* MAPMAKER/SIBS nonparametric sibpair linkage analysis of **a)** phonological awareness, **b)** phonological coding, **c)** spelling, and **d)** RAN speed, using a sample containing sibpairs <18 years of age. Analyses were performed without weighting (——) and with weighting (—■—) of multiple sibships.

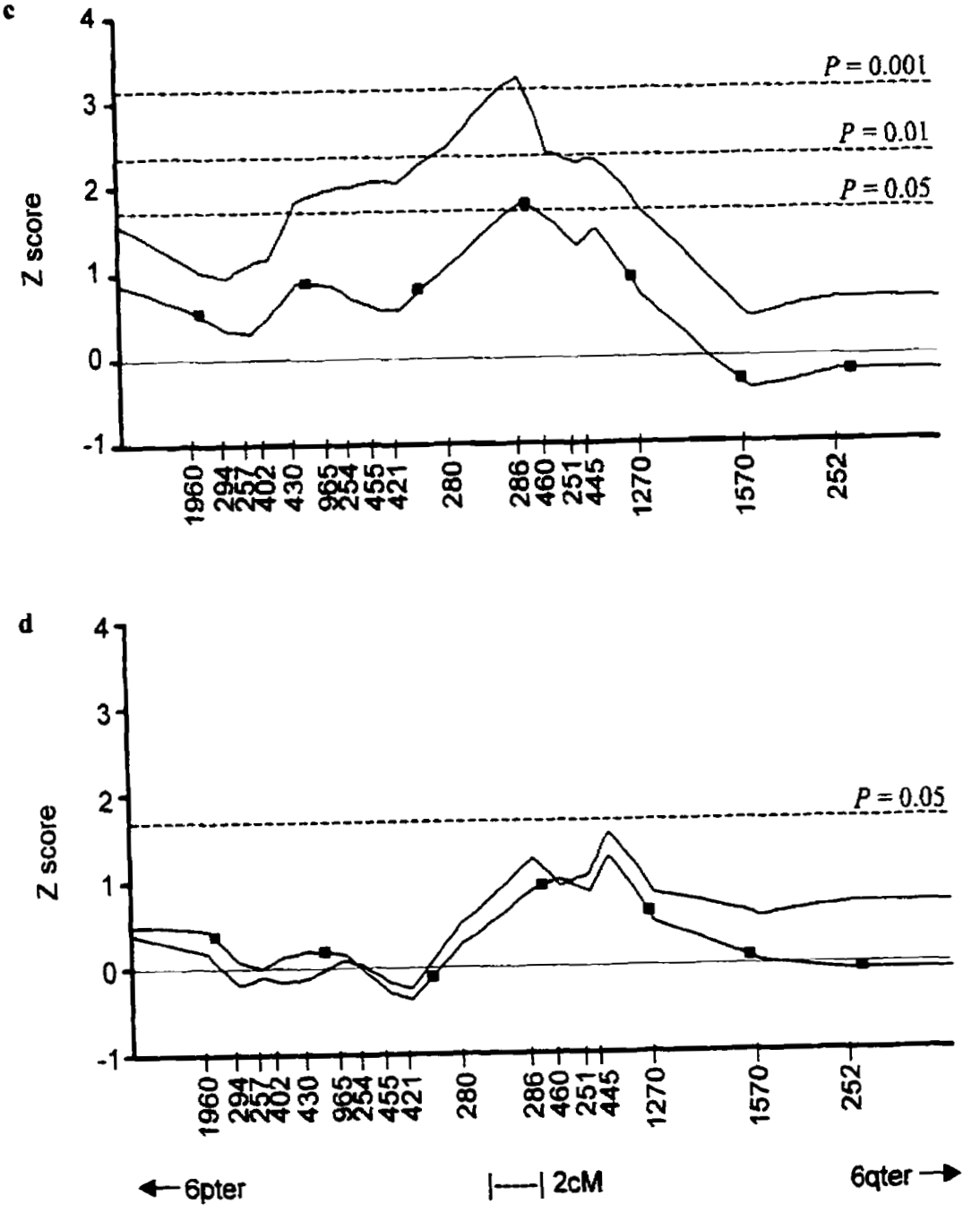


Figure 4.8 (cont'd)

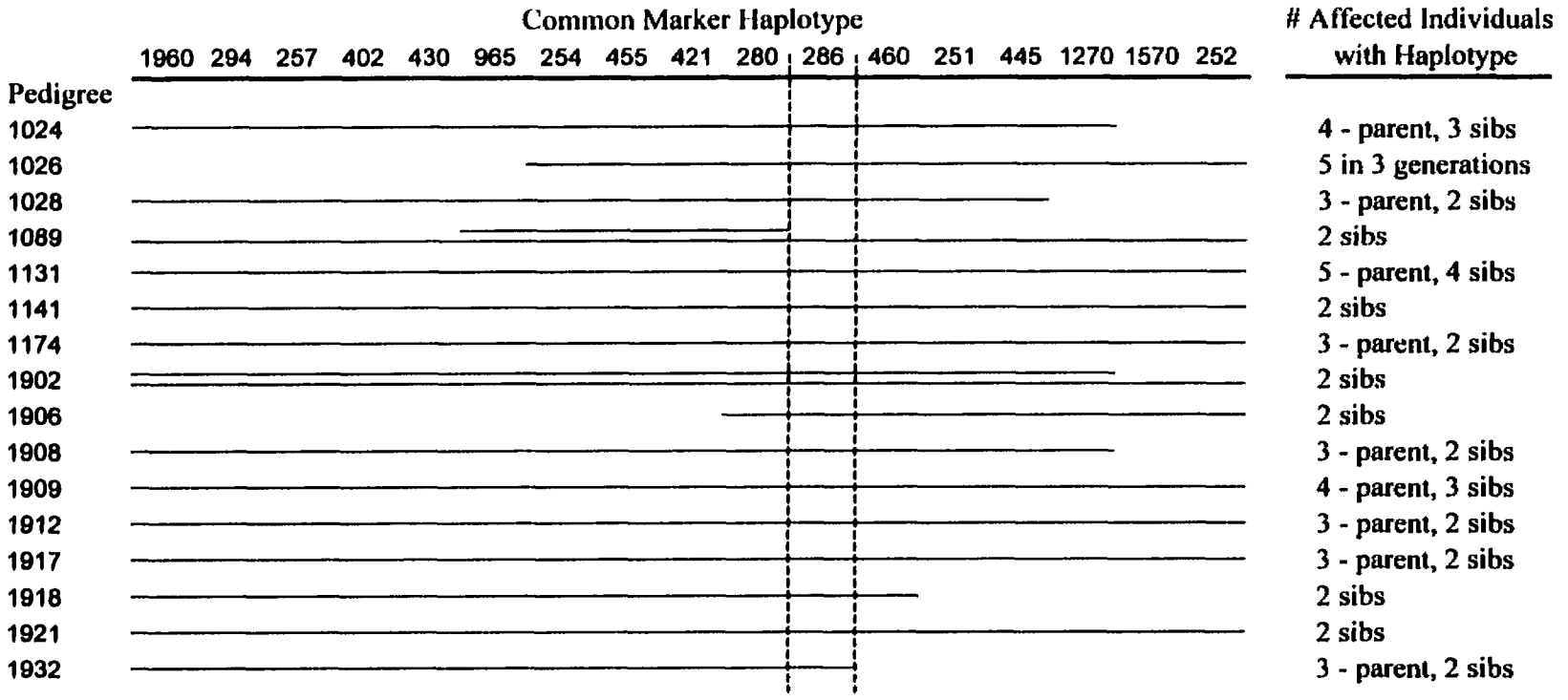


Figure 4.9 Chromosome 6q haplotype(s) shared by affected individuals within each "linked" pedigree. Vertical lines delimit the consensus region. Note that affected individuals in some pedigrees shared more than one founder haplotype, indicated by more than one haplotype line for that pedigree.

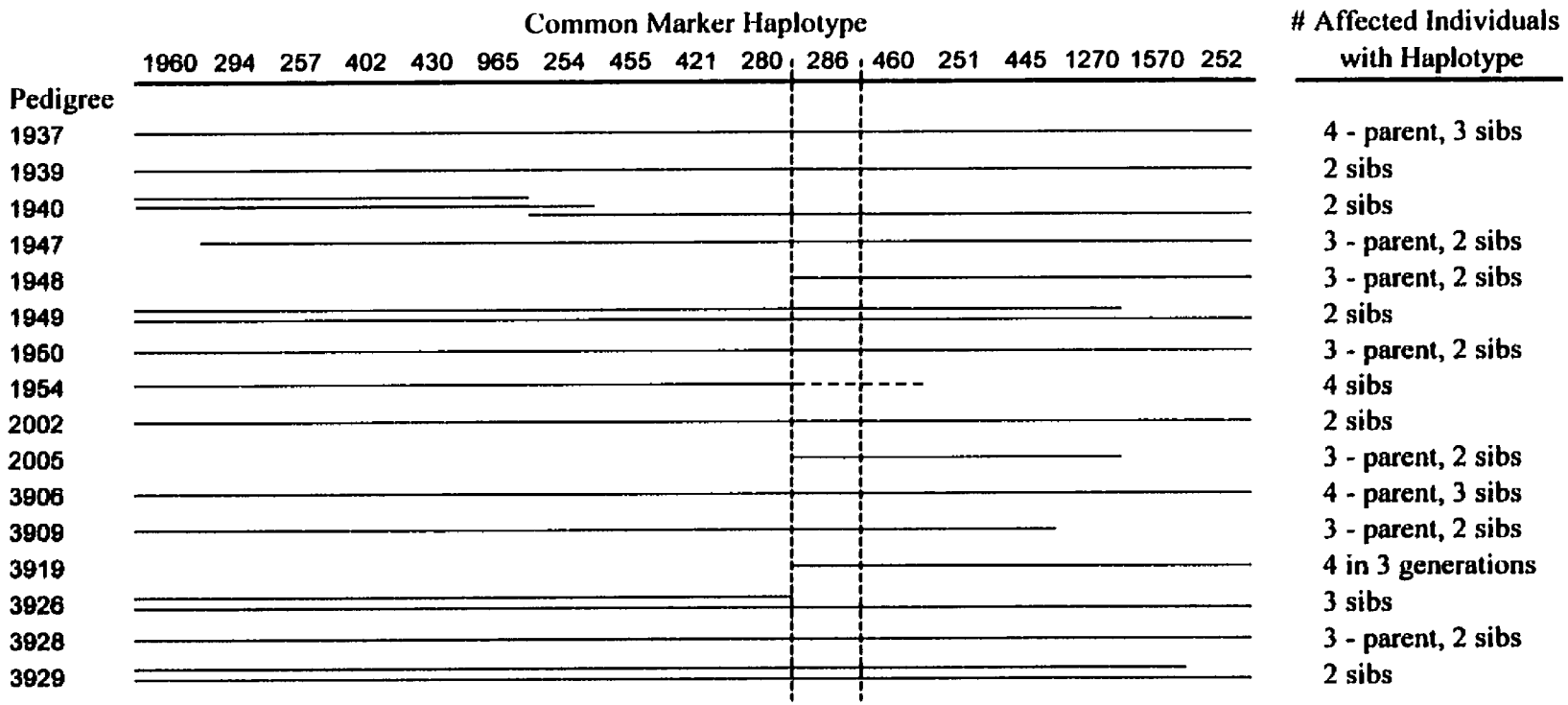


Figure 4.9 (cont'd)

4.3.8. Linkage disequilibrium analysis

As shown in Table 4.8, single-marker AFBAC LD analysis of the 96 nuclear families did not detect associations between PCD and any of the core chromosome 6q markers (all $P > 0.05$ under simplex and multiplex analysis options). TRIMHAP LD analysis using two-marker and three-marker haplotypes of core markers detected significant associations with D6S257/D6S402 haplotypes ($P = 0.012$) and D6S257/D6S402/D6S430 haplotypes ($P = 0.027$ and $P = 0.034$ for disease gene locations in the intervals between the markers). Note that TRIMHAP P values are empirical, thus correction for multiple testing is not required. HAPMAX LD analysis supported the significant D6S257/D6S402 association, with chi-squared $P = 0.0096$, but analysis of D6S257/D6S402/D6S430 haplotypes could not be performed due to program constraints. The physical distance between these PCD associated markers is: D6S257 – 7.4Mb – D6S402 – 4.2Mb – D6S430 (International Human Genome Sequencing Consortium 2001), well outside the ~100kb average range of LD for European-descended populations reported in recent studies (Collins et al. 1999; Abecasis et al. 2001; Reich et al. 2001).

AFBAC single-marker LD analyses of the 32 linked nuclear families using the core markers and the 52 LD markers found much stronger linkage disequilibrium with PCD (Table 4.9). Significant associations were found with adjacent markers D6S1551 (simplex $P = 0.050$) and D6S430 (simplex $P = 0.047$), and with D6S1020 (multiplex $P = 0.026$), located 14cM away (near D6S445). Given the large number of markers investigated, however, type I error is expected to occur with two or three markers. Nonetheless, the proximity of D6S430 and D6S1551 (<1cM or ~400kb) suggests that these associations may indeed be real. TRIMHAP analyses of two-marker haplotypes in the linked families detected a significant association between haplotypes of markers D6S1711 and D6S1710 ($P = 0.049$), which are located ~250kb apart and ~4Mb from D6S1551 and D6S430. However, HAPMAX analysis did not support this association (chi-squared $P = 0.42$), although this P value is probably less accurate under high degrees of freedom than the TRIMHAP empirical value.

Table 4.8

Linkage disequilibrium *P* values from analyses of chromosome 6q core markers in the 96 families

	AFBAC *P* value[a]		TRIMHAP *P* value[b]	
Marker	simplex	multiplex	two-marker haplotype	three-marker haplotype
D6S1960	0.32	0.96		
			0.54	0.51
D6S294	0.72	0.72		
			0.55	0.53
D6S257	0.90	0.42		
			0.012 (0.0096[c])	**0.027**
D6S402	0.11	0.47		
			0.16	**0.034**
D6S430	0.56	0.16		
			0.25	0.15
D6S965	0.34	0.91		
			0.30	0.81
D6S254	0.88	0.44		
			0.24	0.47
D6S455	0.87	0.073		
			0.10	0.14
D6S421	0.63	0.87		
			0.23	0.14
D6S280	0.36	0.95		
			0.41	0.51
D6S286	0.46	0.22		
			0.75	0.83
D6S460	0.63	0.47		
			0.82	0.85
D6S251	0.93	0.33		
			0.53	0.76
D6S445	0.65	0.68		
			0.49	0.59
D6S1270	0.49	0.053		
			0.23	0.43
D6S1570	0.41	0.24		
			0.29	0.37
D6S252	0.74	0.75		

[a] chi-squared *P* value

[b] empirical *P* value

[c] HAPMAX chi-squared *P* value

Table 4.9

Linkage disequilibrium *P* values from analyses of all chromosome 6q markers (core markers and 52 additional markers) in the linked families

	AFBAC *P* value[a]		TRIMHAP *P* value[b]
Marker	simplex	multiplex	two-marker haplotype
1541	0.20	0.79	
			0.67
1960	0.35	0.43	
			0.94
1623	0.66	0.48	
			0.97
294	0.94	0.16	
			0.97
1636	0.41	0.081	
			0.73
1661	0.38	0.94	
			0.86
428	0.25	0.47	
			0.91
257	0.34	0.49	
			0.70
1276	0.73	0.59	
			0.76
1016	0.69	0.63	
			0.22
1711	0.53	0.43	
			0.049 (0.42[c])
1710	0.40	0.31	
			0.069
402	0.20	0.48	
			0.83
1048	0.89	0.77	
			0.35
1628	0.83	0.92	
			0.24
414	0.16	0.73	
			0.31
1053	0.22	0.71	

[a] chi-squared *P* value; [b] empirical *P* value; [c] HAPMAX chi-squared *P* value

Table 4.9 (cont'd)

	AFBAC *P* value[a]		TRIMHAP *P* value[b]
Marker	simplex	multiplex	two-marker haplotype
1053	0.22	0.71	
			0.68
1026	0.69	0.25	
			0.93
1551	**0.050**	0.59	
			0.44
430	**0.047**	0.12	
			0.60
1032	0.35	0.71	
			0.94
405	0.77	0.36	
			0.70
1282	0.86	0.64	
			0.57
1619	0.079	0.10	
			0.55
965	0.53	0.35	
			0.90
467	0.75	0.59	
			1.00
1718	0.65	0.94	
			0.88
254	0.43	0.41	
			0.90
313	0.18	0.41	
			0.84
1557	0.31	0.48	
			0.78
1561	0.52	0.42	
			0.66
455	0.16	0.064	
			0.52
1673	0.063	0.056	
			0.29
421	0.81	0.91	
			0.98
280	0.48	0.70	

[a] chi-squared *P* value; [b] empirical *P* value; [c] HAPMAX chi-squared *P* value

Table 4.9 (cont'd)

	AFBAC *P* value[a]		TRIMHAP *P* value[b]
Marker	simplex	multiplex	two-marker haplotype
280	0.48	0.70	
			0.80
406	0.25	0.40	
			0.34
1596	0.96	0.50	
			0.078
1031	0.30	0.73	
			0.40
1659	0.67	0.85	
			0.076
456	0.44	0.10	
			0.78
1625	0.70	0.77	
			0.73
1589	0.99	0.70	
			0.62
284	0.33	0.58	
			0.69
286	0.48	0.60	
			0.94
460	0.90	0.75	
			0.67
1052	0.58	0.79	
			0.75
463	0.90	0.47	
			0.86
251	0.28	0.39	
			0.58
1646	0.49	0.83	
			0.45
445	0.36	0.55	
			0.57
1020	0.50	**0.026**	
			0.31
1634	0.66	0.68	
			0.61
1627	0.84	0.54	

[a] chi-squared *P* value; [b] empirical *P* value; [c] HAPMAX chi-squared *P* value

Table 4.9 (cont'd)

	AFBAC *P* value[a]		TRIMHAP *P* value[b]
Marker	simplex	multiplex	two-marker haplotype
1627	0.84	0.54	
			0.21
1270	0.88	0.29	
			0.12
1601	0.51	0.53	
			0.48
1652	0.38	0.80	
			0.85
1004	0.35	0.71	
			0.48
1595	0.30	0.18	
			0.15
1644	0.21	0.43	
			0.37
1613	0.47	0.88	
			0.80
1570	0.50	0.51	
			0.34
450	0.34	0.58	
			0.20
1631	0.084	0.31	
			0.14
1043	0.058	0.12	
			0.48
444	0.73	0.88	
			0.99
417	0.32	0.92	
			0.55
458	0.95	0.95	
			0.83
252	0.89	0.50	
			0.68
1274	0.81	0.48	

[a] chi-squared *P* value; [b] empirical *P* value; [c] HAPMAX chi-squared *P* value

AFBAC linkage disequilibrium analysis of *HTR1β* G861C and T-261G polymorphisms in the linked nuclear families did not detect significant associations with PCD under either simplex or multiplex analysis options ($P = 0.63$-1.0; Table 4.10).

Table 4.10

AFBAC linkage disequilibrium results from analyses of *HTR1β* G861C and *HTR1β* T-261G polymorphisms in the linked nuclear families

		transmitted			non-transmitted				
RFLP	allele	obs	exp	freq	obs	exp	freq	Chi-square	*P*
simplex:									
G861C	G	11	11	0.306	11	11	0.306	0	1.00
	C	25	25	0.694	25	25	0.694	0	
T-261G	T	30	29	0.536	28	29	0.500	0.0690	0.71
	G	26	27	0.464	28	27	0.500	0.0741	
multiplex:									
G861C	G	10	10.5	0.312	7	6.5	0.350	0.0529	0.78
	C	22	21.5	0.688	13	14.5	0.650	0.0257	
T-261G	T	27.5	26.4	0.539	15	16.1	0.484	0.1139	0.63
	G	23.5	24.57	0.461	16	14.9	0.516	0.1226	

4.4. Discussion

Linkage analysis of microsatellite markers on chromosome 6 has found suggestive evidence for a dyslexia susceptibility locus on 6q11.2-q12, which has been assigned the name DYX4 by the Human Gene Nomenclature Committee. Evidence for linkage was derived from analyses of a qualitative PCD phenotype by two-point and multipoint parametric linkage analysis methods (under a dominant model with reduced disease penetrance and a common disease allele frequency) and multipoint NPL analysis. QTL linkage analyses of separate reading measures using variance-component

and nonparametric sibpair approaches also provided evidence for a reading disability locus in this region.

While the linkage results surpassed the recommended threshold for suggestive linkage in a genome screen of lod = 1.9, they fell short of the conventional significance threshold of lod = 3.3 (Lander and Kruglyak 1995). However, obtaining a lod score >3.3 in linkage studies of complex traits is often very difficult owing to such factors as incomplete penetrance, phenocopies, genetic heterogeneity, polygenic inheritance, and diagnostic uncertainty (Lander and Schork 1994). Furthermore, there is disagreement as to the appropriateness of using genome screen significance thresholds for studies that are not genome screens, such as the present study (Witte et al. 1996; Elston 1997,1998). Regardless of whether these linkage results are statistically "significant" or merely "suggestive", these findings clearly warrant further attention, and other investigators should attempt replication of these findings in their dyslexia family samples which would confirm and potentially refine this locus.

Linkage disequilibrium analysis also found evidence for the DYX4 dyslexia susceptibility locus. Analysis of all of the PCD families found significant associations with marker haplotypes spanning an 11Mb region at 6q11.1, slightly centromeric to the region of linkage. Furthermore, significant associations were detected with single markers and marker haplotypes spanning 4Mb within this region in analyses of linked families that were more genetically homogeneous than the entire sample of families. Although the single-marker AFBAC results were not corrected for multiple testing, hence some false associations were expected, the haplotype TRIMHAP results were empirical and thus are considered accurate. In spite of this, the fact that LD was detected over such a large region is cause for concern, given that LD is thought to only extend ~100kb in populations such as the one studied (Collins et al. 1999; Abecasis et al. 2001; Reich et al. 2001). However, the associated markers are located near the centromere where recombination is reduced, thus the extent of LD would be expected to be greater than at other chromosomal regions. Additional investigation of this region, possibly with single nucleotide polymorphisms (SNPs), which occur approximately every 1,000bp in the genome, is therefore required to substantiate these LD findings.

Analysis of marker haplotypes identified linked families in which all affected members within the family share a common DYX4 haplotype. Recombination breakpoints in the haplotypes of these families defined a 6.9Mb consensus region between D6S280 and D6S460 that has a high probability of harbouring the *DYX4* gene, although a critical breakpoint was found in only one pedigree that may not carry the *DYX4* gene (the family was classified as linked based on only two affected siblings), thus the consensus region should be regarded with caution. A conservative boundary based on a breakpoint in another family is between D6S280 and D6S251, a region of 8.8Mb.

In this report, the advice of one authority in the field (Elston 1997,1998) was followed and precise linkage results that were not adjusted for multiple testing were presented. The rationale is that traditional Bonferoni correction for multiple testing assumes that the statistical tests are independent, which is not the case here since the reading disability phenotypes are correlated with each other, the genetic markers are located near each other, and several of the methods of linkage analysis are similar to one another. Thus, Bonferoni correction for multiple testing would probably be overly conservative in this situation. Thus, while it is acknowledged that correction for multiple testing is necessary in this case and would decrease the linkage results that have been presented, there is currently no appropriate method for correcting for partially dependent tests such as those utilized in this study. In the case of LD analyses, AFBAC results were also not corrected for multiple testing; however, TRIMHAP results were empirical and thus multiple testing correction was not required.

Most of the linkage analyses employed in this study utilized subdivided pedigrees or a smaller data set selected from the complete sample of families, thus the results were expected to be lower than the "true" results that could have been obtained had it been possible to analyze the complete sample. In particular, the multipoint parametric and NPL analyses of the PCD phenotype, and the variance-component analyses of the quantitative reading measures, were performed using the GENEHUNTER program, which has a number of advantages over other multipoint methods (most notably, a relatively fast linkage algorithm) but is restricted to moderately

sized pedigrees. This restriction necessitated the splitting of 10 large pedigrees in the sample into subpedigrees, which probably reduced the power to detect linkage. Also, the multipoint nonparametric sibpair analysis method used in this study only utilized nuclear families and thus has reduced power to detect linkage compared to analysis of extended pedigrees (Wijsman and Amos 1997). Therefore, the linkage results from multipoint parametric, NPL, variance-component, and sibpair analyses were likely lower than the values that might have been obtained using complete pedigrees.

With regard to the QTL linkage analyses of the reading measures, a few points are worthy of discussion. First, the maximum-likelihood variance-component linkage approach employed in this study has an underlying assumption of phenotypic multivariate normality, which is probably violated in genetic studies where the sample is selected to contain a large proportion of affected individuals. Other factors may also lead to non-normal phenotypic data, such as the presence of a major gene or certain types of gene-environment interaction (Allison et al. 1999). Simulation studies have shown that some types of phenotypic nonnormality, particularly skewness and leptokurtosis (symmetric extreme-tailed distribution), produce type I error rates in excess of the nominal levels (Allison et al. 1999; Pratt et al. 2000). While the distributions of the phonological coding and spelling measures used in this study did not greatly deviate from normality, caution is still warranted when interpreting their variance-component results. However, the fact that sibpair linkage analyses of these reading measures also supported the presence of a locus in the 6q11.2-q12 region, as did parametric and NPL analyses of the PCD phenotype, suggests that the variance-component results for phonological coding and spelling are in fact reliable. It is intriguing, though, that the variance-component approach did *not* detect any evidence for linkage to phonological awareness, when the moderate correlations between this measure and phonological coding and spelling suggest that a common genetic basis may underlie these reading skills. However, the fact that sibpair analyses, which make no assumption of phenotypic normality, *did* find moderate evidence for linkage to phonological awareness suggests that the poor variance-component results may be due to the marked non-normality observed in this measure (due to severe psychometric test ceiling effects). Thus, rather

than producing potentially false-positive linkage results, as simulation studies predict for nonnormal phenotypes, the phonological awareness measure may have generated false-negative linkage results in variance-component analyses. A second point to be addressed is that in sibpair analyses of the reading measures, analysis of a restricted sample of sibpairs less than 18 years of age found stronger evidence for linkage to phonological awareness, phonological coding, and spelling than analyses of sibpairs of all ages. This observation indicates that adult ceiling effects in the psychometric reading tests may have reduced the variance in the measures, thereby reducing the power to detect linkage in sibpair samples including adults. One final comment regarding the quantitative analyses is that since variance in the phenotypic measures and other factors can affect QTL linkage analyses (Pennington 1997), it would be premature to conclude that the DYX4 locus is involved in particular reading skills (i.e., phonological coding and spelling, in which there was evidence for linkage), but is not involved in other skills (i.e., phonological awareness or RAN speed, in which there was little evidence for linkage). Rather, it can only be concluded that this gene appears to be involved in some aspect of the reading process.

Heterogeneity testing detected only moderate (but non-significant) evidence for genetic heterogeneity in the sample, however this may have simply been due to sample size or other factors and not because heterogeneity does not exist. The fact that some families had positive lod scores at the 6q region, while other families clearly demonstrated lack of linkage by very negative lod scores, indicates that heterogeneity between families (interfamilial heterogeneity) is highly probable. Furthermore, heterogeneity testing does not detect heterogeneity *within* families (intrafamilial heterogeneity), which is presumably a major factor in the sample given the large proportion of bilineal pedigrees. Thus, although heterogeneity testing was not able to detect *significant* evidence for genetic heterogeneity in the sample, both interfamilial and intrafamilial heterogeneity are likely present.

Genetic heterogeneity would have had a severe impact on both multipoint linkage analyses and LD analyses. With regard to multipoint linkage analyses, heterogeneity reduces the power to detect linkage because the flanking markers inhibit

the recombination fraction estimate from increasing, thereby reducing or excluding linkage across the whole interval (Xu et al. 1998). This is in contrast to two-point linkage analyses, where heterogeneity can be partially ameliorated by an increased recombination fraction estimate (note in Table 4.6 that two-point lod scores >2 had theta values ~0.25). It should be noted that although GENEHUNTER accommodates interfamilial heterogeneity by the calculation of hlod scores, thereby moderating the effect of heterogeneity, this program does not accommodate intrafamilial heterogeneity. With regard to LD analyses, genetic heterogeneity may have decreased the ability to detect especially strong associations between PCD and markers spanning chromosome 6q11.2-q12. In analyses of nuclear families selected from the 96 pedigrees, families that do not carry the *DYX4* gene could have masked LD present in the families that do carry this gene. For this reason, separate association analyses were performed on a subset of linked families defined by sharing of a single DYX4 haplotype among affected family members. However, even though selection of these linked families probably reduced the genetic heterogeneity in the sample, many of the pedigrees were classified as linked on the basis of only two or three affected siblings (who are expected to share on average 50% of their genes by chance) and may not actually carry the *DYX4* gene. Thus, interfamilial heterogeneity may still have existed in the linked subset of families, and would have obscured strong marker associations with PCD. Note that strong LD may also not have been detected due to small sample size (particularly for the linked family sample) or due to the occurrence of multiple identical or heterogeneous founder mutations in the *DYX4* gene, which is highly possible for a common disorder in an ethnically-mixed sample such as the one used in this study. In addition, sparse marker coverage around the locus may have prevented the detection of strong LD, which is quite likely given that recent studies have found that the extent of detectable LD in north-Europe descended populations is ~100kb (Collins et al. 1999; Abecasis et al. 2001; Reich et al. 2001), whereas the markers used in the present study were on average 600kb apart.

It should be stressed that it is not likely that this positive finding on 6q is merely detecting the known 6p dyslexia locus. Although factors such as genetic heterogeneity

and phenotypic variability between studies may "shift" a susceptibility gene localization, the 6p21.3 dyslexia region has shifted only 15cM among reports, while the 6q11.2-q12 region is at least 40cM away from the 6p21.3 region. Furthermore, it is reasonable to assume that a number of genes are involved in a complex skill such as reading, and the likelihood of two genes being located on the same relatively large chromosome would be nearly as great as being located on different chromosomes.

It is possible that a common predisposing gene may underlie various behavioral or psychological disorders, given that these types of disorders may share a similar etiology of disruption in brain development and/or function. Since parents of autistic children appear to have an increased frequency of reading problems (Folstein et al. 1999), it is intriguing that a study (Phillippe et al. 1999) found evidence for an autism susceptibility locus on chromosome 6q near D6S283, which is located approximately 17cM telomeric to D6S251, the marker that produced the highest two-point lod score with PCD. The marker density used in the autism report was relatively sparse in this region, with the nearest markers to D6S283 being 20cM centromeric and 10cM telomeric (markers D6S286 and D6S261, respectively). Thus, analysis using additional markers could shift the autism gene localization into the 6q11.2-q12 region where suggestive evidence for a dyslexia locus was found, opening up the possibility of a common susceptibility gene for both disorders.

There are a number of candidates for the *DYX4* dyslexia susceptibility gene. Located in the vicinity of D6S257 to D6S430, where haplotype LD analysis found several significant associations, is a protein tyrosine phosphatase gene (*PTP4A1*) and a novel gene (*KIAA0244*) with sequence similarity to human transcription factor *TFIIS*. Since phosphatases and transcription factors are involved in a wide range of cellular processes, it is entirely possible that they are involved in neural development and/or function. Within the 6.9Mb consensus region between D6S280 and D6S460 identified from recombination breakpoints in linked families, candidate genes include $p43^{gndd}$, a novel gene being studied by Dr. Kye-Young Lee at the University of Calgary, who has found evidence for a role for this gene in neural differentiation (personal communication), and the serotonin receptor gene *HTR1β* (Jin et al. 1992), which is a

good candidate based on the involvement of neurotransmitters in brain development (Mattson 1988; Meier et al. 1991; Levitt et al. 1997). The serotonin receptor gene *HTR1E* is also located in the 6q11.2-q12 region, although outside of the consensus region (Levy et al. 1994). Even though linkage disequilibrium between PCD and *HTR1β* and *HTR1E* polymorphisms was not detected in this sample, this does not rule out a role for either of these genes in dyslexia. Other 6q11.2-q12 candidates include the gamma-aminobutyric acid (GABA) receptor rho-subunit genes, *GABRR1* and *GABRR2* (Cutting et al. 1992), and the cannabinoid receptor gene (*CNR1*) (Hoehe et al. 1991), which is a strong candidate based on evidence that the endogenous cannabinoid system plays a role in neural development (see review by Fernández-Ruiz et al. 2000).

To conclude, this study has found evidence for a dyslexia susceptibility locus on chromosome 6q11.2-q12, named DYX4. Replication in other dyslexia family samples is obviously required to confirm the presence of a dyslexia susceptibility gene in this region, and to better localize the gene to facilitate fine mapping. Ultimately, identification of genes involved in dyslexia will increase our understanding of the biological basis of this disorder, and will lead to earlier diagnosis and improved treatment of children at high risk for dyslexia.

Chapter Five: Confirmation of the *DYX3* Dyslexia Susceptibility Gene on Chromosome 2p15-p16

5.1. Introduction

A linkage study by Fagerheim et al. (1999) identified a dyslexia locus (*DYX3*) on chromosome 2p15-p16 in a large Norwegian family with autosomal dominant transmission of dyslexia. Parametric linkage analyses using three diagnostic schemes found significant evidence for linkage in this family (maximum lod = 4.3 at D2S378), which was supported by nonparametric linkage analysis ($P = 0.0009$ between D2S2352 and D2S1337). Furthermore, identification of a three-marker haplotype cosegregating with dyslexia in the family defined a 2cM region between D2S2352 and D2S1337 that likely harbours the *DYX3* gene. Replication of this linkage in other families would confirm the existence of the locus and potentially assist in its localization; however, there have been no reports of replication to date. I therefore investigated our sample of dyslexia families for linkage and/or linkage disequilibrium between PCD and chromosome 2p15-p16 markers. The results provide independent evidence for the DYX3 locus, and support the localization of *DYX3* within the 2cM interval reported in the original Norwegian linkage study.

5.2. Methods

5.2.1. Subjects

As described in Section 2.2, the study sample consisted of 96 families (46 nuclear families, 50 extended pedigrees) of European descent, with each family containing at least two dyslexic siblings. There were a total of 902 individuals who had DNA sampled and 884 underwent psychometric testing to assess four components of reading: phonological awareness, phonological coding, spelling, and RAN) speed. Each individual was also assessed for estimated IQ, and adults were assessed for reading history (see Section 2.4). The results of the reading tests (except RAN speed) and reading history in adults were used to assign an affected, unaffected, or uncertain

diagnosis of phonological coding dyslexia (PCD) (see Section 2.5), resulting in 52% of individuals diagnosed as affected with PCD, 33% as unaffected, and 15% as uncertain.

5.2.2. Markers, genotyping and marker map

The following seven microsatellite markers spanning the DYX3 region were selected from the Fagerheim et al. (1999) report: D2S1352, D2S2352, D2S378, D2S2279, D2S2183, D2S1337, and D2S393. Automated genotyping was performed with technical assistance using a LI-COR 4200S-2 Gene ReadIR DNA Analyzer, as described in Section 2.8. Marker allele frequencies were calculated from the parents of one nuclear family per pedigree, as described in Section 2.9. The Genethon genetic map (Dib et al. 1996) was utilized for intermarker order and distances (Figure 5.1). Note that the Genethon marker order corresponded to the human genome sequence (International Human Genome Sequencing Consortium 2001).

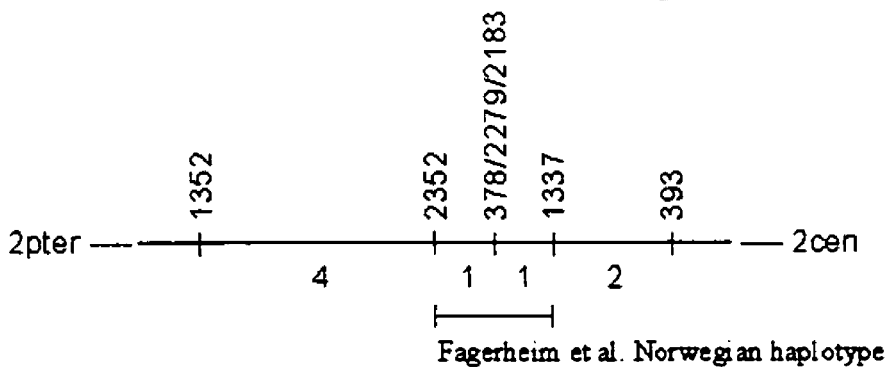


Figure 5.1 Genetic marker map of chromosome 2p15-p16 markers, with intermarker distances (cM) indicated.

5.2.3. Linkage analysis

Two-point parametric linkage analysis of PCD was performed using FASTLINK, as described in Section 2.14.1.1, and multipoint parametric linkage analysis under genetic heterogeneity was performed using the GENEHUNTER program, as described in Section 2.14.12. Two-point and multipoint analyses were performed under a model

with 1% disease allele frequency and phenocopy and penetrance rates of 0.04 and 0.99 for males and 0.01 and 0.85 for females (Table 5.1). These values were selected to match those in the models used in the Fagerheim et al. (1999) report. Multipoint nonparametric linkage (NPL) analysis of PCD was performed using GENEHUNTER by analyzing all affected family members simultaneously, as described in Section 2.14.3. Note that it was necessary to subdivide 10 large pedigrees for all GENEHUNTER analyses, likely reducing the power to detect linkage.

Table 5.1

Inheritance model used for *DYX3* parametric linkage analysis of PCD

	male	female
Penetrance (aa, ab, bb)	0.04, 0.99, 0.99	0.01, 0.85, 0.85
Disease allele b frequency	0.01	0.01

5.2.4. Haplotype analysis

Marker haplotypes were determined using the GENEHUNTER program, as described in Section 2.16. Each family was investigated for sharing of a common haplotype (or haplotypes) containing D2S378, D2S2279, or D2S2183 (the markers comprising the 2cM cosegregating haplotype in the Norwegian family) amongst all PCD affected individuals within the family, indicating that these families had a higher likelihood of carrying the *DYX3* gene. Unaffected or uncertain family members were also permitted to share the affected haplotype to allow for incomplete disease penetrance. These families were designated as "linked", and their affected haplotypes were investigated for recombination breakpoints to attempt to narrow the candidate susceptibility region. In addition, the affected haplotypes of the linked families were investigated for the specific haplotype found segregating in the Norwegian family. Since marker allele size is dependent on PCR marker primers and the genotyping system, three DNA samples from the Norwegian family (obtained from Dr. Toril Fagerheim) were

genotyped and the specific alleles in the Norwegian affected haplotype were determined to allow direct comparison to the linked families' haplotypes.

5.2.5. Linkage disequilibrium analysis

Family-based linkage disequilibrium (LD) analysis of PCD using two-marker and three-marker haplotypes, which yield more power to detect LD than single markers (Martin et al. 2000a; Akey et al. 2001), was performed using the TRIMHAP program, as described in Section 2.17.2. This program analyzes extended pedigrees for significant association of haplotypes with the disease by determining empirical significance levels for the disease locus being located between each marker in the haplotype, and the method does not confound linkage with association. Haplotypes with $P < 0.05$ were also investigated using the HAPMAX program, which analyzes parent-affected child trios for significant association of haplotypes with the disease by chi-squared testing, as described in Section 2.17.2. LD analyses were performed on the sample of 96 families, as well as a subsample of linked families that were more likely to carry *DYX3* mutations, and thus were more genetically homogeneous, than the sample of 96 families.

5.3. Results

5.3.1. Linkage analysis

Results of two-point and multipoint parametric linkage analyses provided weak evidence for linkage between PCD and the DYX3 region. As shown in Table 5.2, the maximum two-point lod score was found at marker D2S1352 ($Z_{max} = 0.77$, $\theta = 0.3$). Multipoint analysis detected a peak hlod score of 0.07 at D2S1352, as shown in Figure 5.2. Multipoint nonparametric linkage analysis, however, provided much stronger evidence for linkage to the DYX3 region. The peak NPL Z_{all} score was 2.33 at D2S1352, corresponding to $P = 0.0087$ (shown in Figure 5.3), thus surpassing the recommended value of $P = 0.01$ to claim significant linkage in a replication study (Lander and Kruglyak 1995).

Table 5.2

Results of *DYX3* two-point parametric linkage analysis of PCD

Marker	Z_{max} (θ)
D2S1352	0.77 (0.3)
D2S2352	0.03 (0.4)
D2S378	0.14 (0.4)
D2S2279	0 (0.5)
D2S2183	0.32 (0.4)
D2S1337	0.29 (0.4)
D2S393	0.44 (0.4)

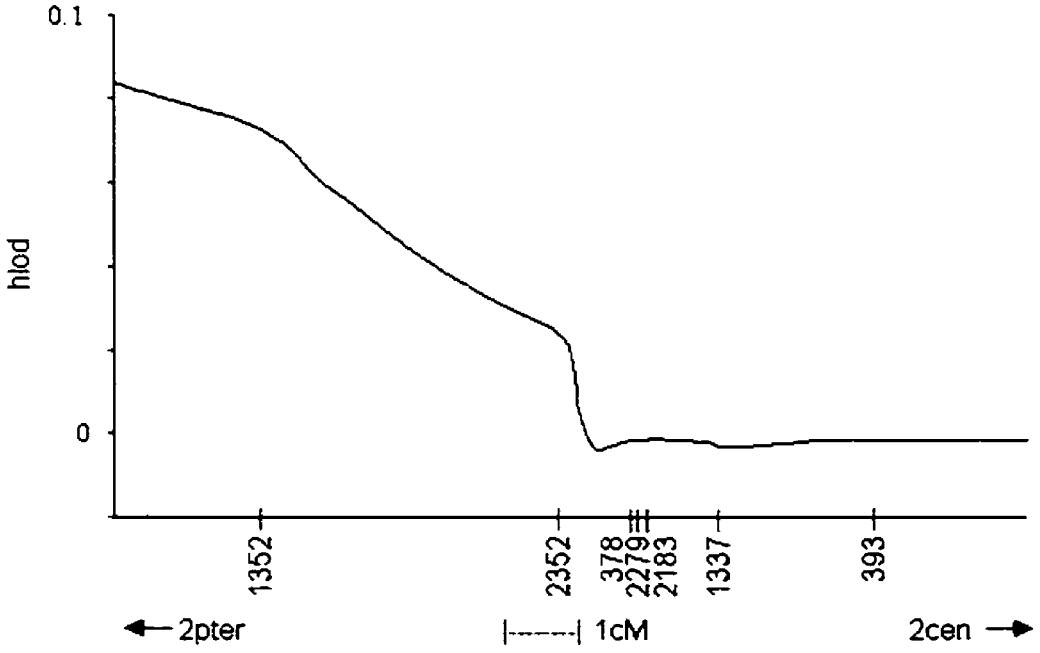


Figure 5.2 Hlod-score curve from *DYX3* GENEHUNTER multipoint parametric linkage analysis of PCD.

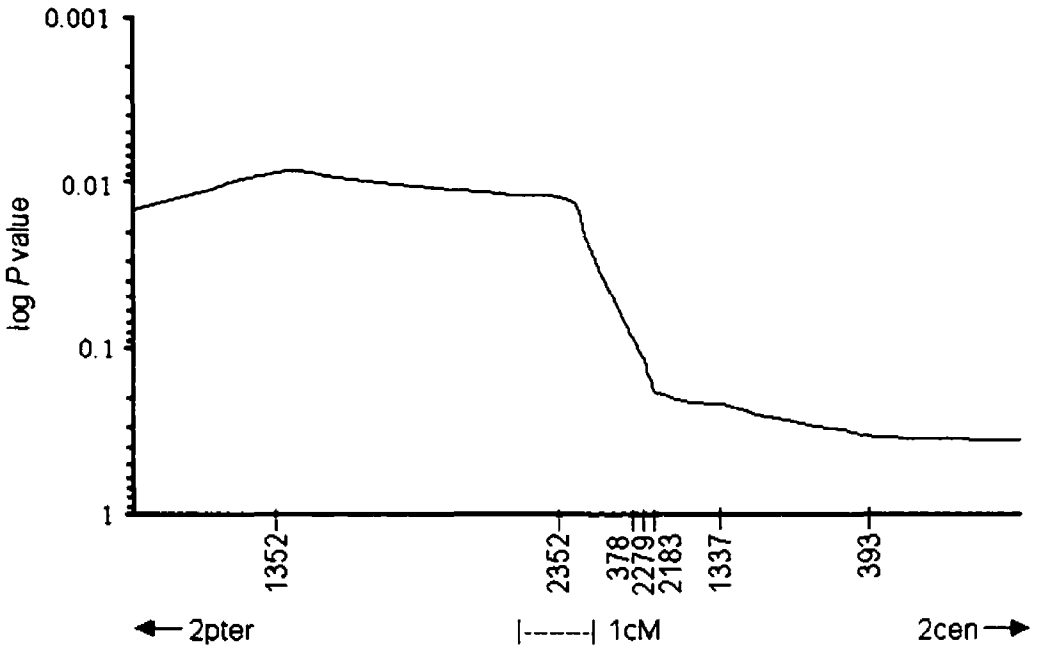


Figure 5.3 Log *P*-value curve from *DYX3* GENEHUNTER multipoint nonparametric linkage analysis of PCD.

5.3.2. Haplotype analysis

Haplotype analysis identified 35 linked pedigrees in which affected family members share a common haplotype (or haplotypes) containing at least one of D2S378, D2S2279, and D2S2183. The Norwegian affected haplotype was not found in any of the families; however, the D2S378 allele in the Norwegian haplotype is rare in our sample (observed in only two of 877 individuals). As shown in Figure 5.4, analysis of recombinations in haplotypes of the linked families identified a recombination breakpoint between D2S2279 and D2S2183 in pedigree 1921. In addition, a breakpoint somewhere between D2S378 and D2S2183 was identified in pedigree 1020. It was not possible to exactly localize this breakpoint because the recombinant individual was homozygous for D2S2279. Nevertheless, these two breakpoints identified a consensus region between D2S378 and D2S2183. According to the Genethon genetic map (Figure

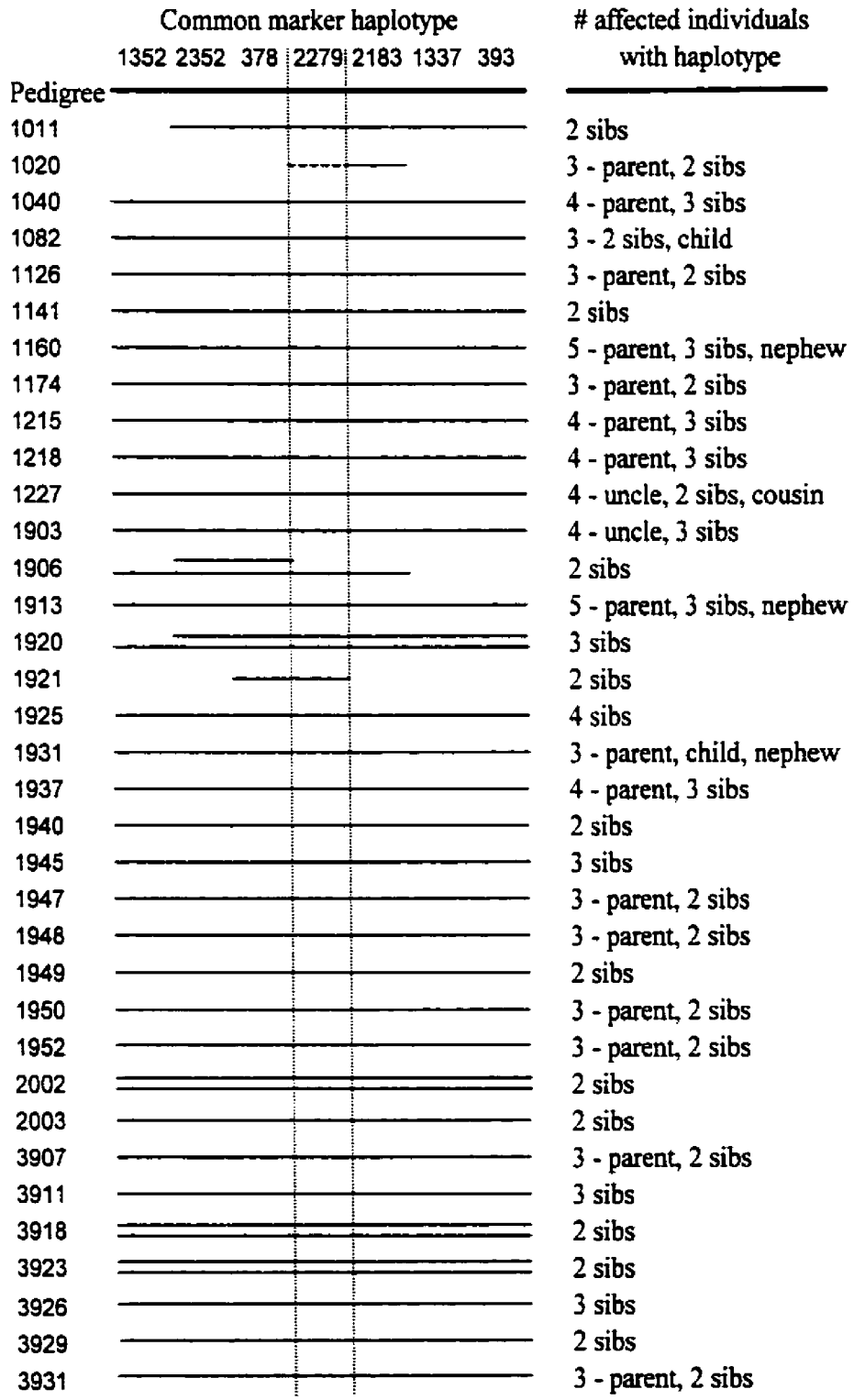


Figure 5.4 Chromosome 2p15-p16 haplotype(s) shared by affected individuals within each linked pedigree. Vertical lines delimit the consensus region.

5.1), these markers are located at the same genetic position (i.e., they cannot be resolved using genetic mapping techniques). However, according to the human genome sequence (International Human Genome Sequencing Consortium 2001), these markers span a 1.9Mb region. Depending on the exact location of the breakpoint in pedigree 1020, the consensus region may be even smaller. It should be pointed out that while this consensus region is more likely to contain the *DYX3* gene than elsewhere on 2p15-p16, the breakpoints were found in only two pedigrees that each contain two affected siblings sharing a common haplotype, which is expected to occur by chance at ~50% probability. Thus, these siblings may not actually carry the *DYX3* gene, thus the breakpoints in these families, and the resulting consensus region, should be regarded with caution.. Breakpoints in other families (1011 and 1906) defined a much larger consensus region from D6S1352 to D6S1337, a region spanning 9.6Mb.

5.3.3. Linkage disequilibrium analysis

As shown in Table 5.3, TRIMHAP LD analysis of the 96 pedigrees did not find any significant associations between PCD and chromosome 2p15-p16 markers. However, analysis of the subsample of 35 linked families found significant associations to haplotypes consisting of D2S2352/D2S378 ($P = 0.021$), D2S378/D2S2279 ($P = 0.017$), D2S1337/D2S393 ($P = 0.049$), and D2S2352/D2S378/D2S2279 ($P = 0.029$). Investigation of these associations using HAPMAX found significant results with D2S2352/D2S378 haplotypes ($P = 0.029$) and D2S2352/D2S378/D2S2279 haplotypes ($P = 0.00094$), as shown in Table 5.3. However, due to the large number of haplotypes analyzed (hence the high degrees of freedom), these HAPMAX chi-squared P values are considered estimates only. Note that no *specific* haplotypes were found associated with PCD.

Table 5.3

TRIMHAP linkage disequilibrium empirical *P* values from analysis of *DYX3* two-marker and three-marker haplotypes in the 96 families and in the linked families

	96 families		linked families	
Marker	two-marker haplotype	three-marker haplotype	two-marker haplotype	three-marker haplotype
D2S1352	0.24	0.46	0.85	0.66
D2S2352	0.41	0.47	**0.021 (0.029[a])**	**0.029 (0.00094[a])**
D2S378	0.15	0.12	**0.017** (0.15[a])	**0.020**
D2S2279	0.084	0.091	0.33	0.26
D2S2183	0.61	0.41	0.18	0.20
D2S1337	0.67	0.39	**0.049** (0.17[a])	0.12
D2S393				

[a] HAPMAX *P* value

5.4. Discussion

Investigation of chromosome 2p15-p16 markers in a sample of Canadian families with PCD has found independent evidence for the DYX3 dyslexia susceptibility locus originally identified in a Norwegian family by Fagerheim et al. (1999). Evidence for linkage was derived from multipoint NPL analysis, where the results surpassed the recommended $P = 0.01$ to claim significant linkage in a replication study (Lander and Kruglyak 1995). Parametric linkage analyses using the genetic model developed by Fagerheim et al. (1999), however, found only very weak evidence for linkage, suggesting that the model (based on Norwegian prevalence rates) may not be the most appropriate model for our Canadian sample.

Significant linkage disequilibrium was detected between PCD and several DYX3 marker haplotypes in analyses of a group of linked families, in which affected

individuals within each family shared a common DYX3 haplotype, thus the sample was more genetically homogeneous than the complete sample of families. Most of the significant associations were to haplotypes containing markers D2S2352, D2S378 and D2S2279. These latter two markers comprise part of the three-marker haplotype that cosegregates with dyslexia in the Norwegian family, thus these results provide support for *DYX3* being located near these markers.

It was interesting that the Norwegian affected haplotype was not found in any of our families. However, the D2S378 allele in the Norwegian haplotype was rare in this sample, occurring in only two of 877 individuals, suggesting that the Norwegian haplotype may be population- or family specific. Similarly, the fact that no *specific* haplotypes were found associated with PCD might be due to the presence of multiple independent mutations associated with different haplotypes in this ethnically mixed sample.

Analysis of recombination breakpoints in affected haplotypes of linked families identified a consensus region between D2S378 and D2S2183 where *DYX3* is most likely located. These markers span a 1.9Mb interval according to the human genome sequence, although depending on the exact location of one critical breakpoint, the consensus region may be even smaller. This consensus region is within the 2cM interval defined by the Norwegian haplotype, which is approximately 5.4Mb in size (International Human Genome Sequencing Consortium 2001). Thus, the present findings considerably refine the location of *DYX3* within this region. One gene mapped within the consensus region called vaccinia virus B1R kinase related kinase 2 (*VRK2*), a putative serine/threonine protein kinase (Nezu et al. 1997), is a potential candidate gene since kinases are involved in a variety of cellular functions.

In conclusion, linkage and linkage disequilibrium analyses of a large Canadian family sample have provided independent evidence for the DYX3 dyslexia locus on chromosome 2p15-p16. As there have not yet been any other reports replicating the *DYX3* linkage, these findings contribute significantly towards substantiating the existence of the *DYX3* gene. In addition, associations with haplotypes consisting of markers D2S378 and D2S2279, and the identification of a 1.9Mb consensus region

between D2S378 and D2S2183, verify the location of *DYX3* near these markers, as reported in the original Norwegian study, and thereby warrant the continued investigation of this region in order to identify the *DYX3* gene.

Chapter Six: Involvement of Neurotransmitter Receptor Genes in Susceptibility to Dyslexia

6.1. Introduction

As discussed in Chapter One, dyslexia co-occurs with attention-deficit/hyperactivity disorder (ADHD), anxiety, and depression more often than expected by chance (Gilger et al. 1992; Shaywitz et al. 1995; Willcutt and Pennington 2000). In addition, parents of autistic children have an increased frequency of reading problems (Folstein et al. 1999), suggesting a link between dyslexia and autism. These findings suggest that genes with pleiotropic effects may be involved in these conditions, therefore genes implicated in ADHD, anxiety, depression, autism, or other behavioural conditions are candidate dyslexia susceptibility genes.

Numerous studies have been carried out to investigate a role for dopamine receptor and dopamine transporter genes in ADHD, since the most common treatment for ADHD is methylphenidate (Ritalin), which primarily acts on the dopaminergic system (see review by Challman and Lipsky 2000). Significant associations have been reported between ADHD and the dopamine transporter gene, *DAT1* (Cook et al. 1995; Gill et al. 1997; Waldman et al. 1998; Barr et al. 2001), and between ADHD and the dopamine D4 receptor gene, *DRD4* (LaHoste et al. 1996; Rowe et al. 1998, Smalley et al. 1998, Swanson et al. 1998; Barr et al. 2000; McCracken et al. 2000; Sunohara et al. 2000; Curran et al. 2001). Gamma-aminobutyric acid (GABA) receptor genes may also be involved in behavioural disorders, with evidence for linkage disequilibrium between autistic disorder and the GABA(A) receptor beta3 subunit gene, *GABRB3* (Cook et al. 1998; Martin et al. 2000b), and the GABA(A) receptor alpha5 subunit gene, *GABRA5*, implicated in unipolar disorder and bipolar disorder (Oruc et al. 1997; Papadimitriou et al. 1998). The above findings are not surprising given that neurotransmitter receptors are important for neural development (Mattson 1988; Meier et al. 1991; Levitt et al. 1997), a process that appears to be abnormal in the above disorders (Zametkin and Liotta 1998; Courchesne 1997; Nemeroff 1998) and in dyslexia, as discussed in Chapter One.

I therefore investigated microsatellite markers near or within several candidate GABA receptor genes, dopamine receptor genes, and the dopamine transporter gene for linkage (using parametric and genetic model-free sibpair methods) and linkage disequilibrium (using a family-based method) with phonological coding dyslexia (PCD). The results provide evidence for the involvement of neurotransmitter receptor genes in susceptibility to PCD, thereby suggesting a relationship between these genes and the atypical brain development and function associated with dyslexia.

6.2. Methods

6.2.1. Subjects

As described in Section 2.2, the study sample consisted of 83 families (43 nuclear families, 40 extended pedigrees) of European descent, with each family containing at least two dyslexic siblings (with the exception of five families that had only one affected sibling). There were a total of 805 individuals who had DNA sampled and underwent psychometric testing to assess four components of reading: phonological awareness, phonological coding, spelling, and RAN speed (see Section 2.4). Each individual was also assessed for estimated IQ, and adults were assessed for reading history. The results of the reading tests (except RAN speed) and reading history in adults were used to assign an affected, unaffected, or uncertain diagnosis of PCD (see Section 2.5), resulting in 50% of individuals diagnosed as affected with PCD, 34% as unaffected, and 16% as uncertain.

6.2.2. Markers, genotyping and marker map

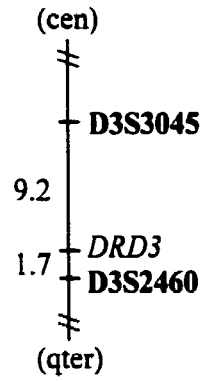
Microsatellite markers within or near candidate GABA receptor genes, dopamine receptor genes, and the dopamine transporter gene (*DAT1*) were selected and primer sequences obtained from the Genome Database (GDB) (http://www.gdb.org). The markers for the GABA receptor candidate genes were: GABRA1 (intragenic marker), GABRA5 (intragenic marker), GABRB1 (intragenic marker) (note that *GABRA2* is located within 0.6cM of *GABRB1*), GABRB3 (intragenic marker) (note that *GABRG3* is located ~10cM from *GABRB3*), and D5S529 and D5S621 (flanking *GABRB2*, *GABRG2*, and *GABRA6*).

The markers for the dopamine receptor candidate genes were: D5S211 (near *DRD1*), DRD2 (intragenic marker), D3S2460 and D3S3045 (flanking *DRD3*), D11S1363 (near *DRD4*), DRD5 (intragenic marker), and the marker for the dopamine transporter candidate gene (*DAT1*) was D5S807. The cytogenetic location, heterozygosity (a measure of the degree of polymorphism), and the candidate gene being tested for each marker are shown in Table 6.1. Automated genotyping was performed with technical assistance using a LI-COR 4200S-2 Gene ReadIR DNA Analyzer, as described in Section 2.8. Published marker allele frequencies were obtained from the GDB and were used for linkage analyses. The Genetic Location Database (LDB) genetic marker maps showing the selected markers are shown in Figure 6.1.

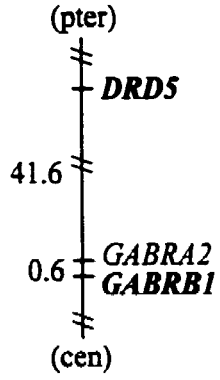
Table 6.1

Cytogenetic location, heterozygosity, and candidate gene tested for each of the GABA receptor gene, dopamine receptor gene, and dopamine transporter gene markers

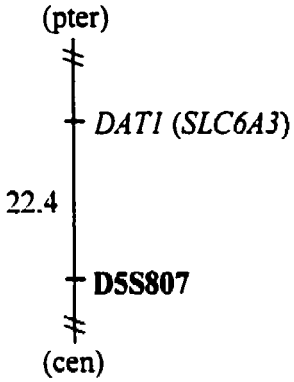
Marker	cytogenetic location	heterozygosity	candidate gene tested
D3S3045	3q13.3	0.82	*DRD3*
D3S2460	3q13.3	0.76	*DRD3*
DRD5	4p15.3-p15.1	0.78	*DRD5*
GABRB1	4p13-p12	0.69	*GABRB1*, *GABRA2*
D5S807	5p15.3	0.76	*DAT1*
D5S529	5q34-q35	0.74	*GABRG2*, *GABRB2*, *GABRA6*
D5S621	5q34-q35	0.64	*GABRG2*, *GABRB2*, *GABRA6*
GABRA1	5q34-q35	0.76	*GABRA1*
D5S211	5q34-q35	0.73	*DRD1*
D11S1363	11p15.5	0.60	*DRD4*
DRD2	11q22.2-q22.3	0.68	*DRD2*
GABRA5	15q11-q13	0.78	*GABRA5*
GABRB3	15q11-q13	0.82	*GABRB3*, *GABRG3*



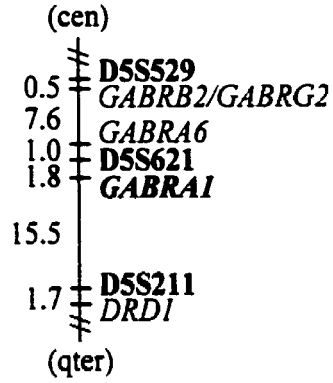
Chromosome 3q13.3



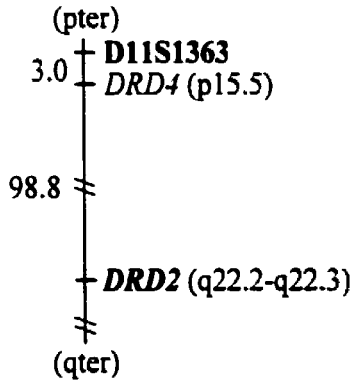
Chromosome 4p15.3-p12



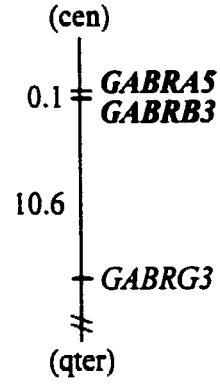
Chromosome 5p15.5



Chromosome 5q34-q35



Chromosome 11



Chromosome 15q11-q13

Figure 6.1 Genetic Location Database genetic marker maps of chromosomes 3q13.3, 4p15.3-p12, 5p15.3, 5q34-q35, 11, and 15q11-q13, with intermarker distances (cM) indicated (maps not to scale). Genotyped markers are shown in bold.

6.2.3. Linkage analysis

Two-point parametric linkage analysis of PCD was performed using FASTLINK from the LINKAGE programs, as described in Section 2.14.1.1. The complete sample of 83 families was analyzed (i.e., including extended pedigrees). To increase the likelihood of detecting linkage, analyses were performed using eight genetic models (Table 6.2). The models ranged across recessive, intermediate, and dominant modes of inheritance, all with reduced penetrance (allowing subjects with a PCD susceptible genotype to have an unaffected phenotype due to reduced disease penetrance or misdiagnosis). Note that these models are the same as those employed for two-point parametric linkage analyses discussed in Chapter 4.

Two-point sibpair linkage analysis of the PCD phenotype was performed using the SIBPAL program, as described in Section 2.14.2.1. This program tests for genetic linkage by performing traditional Haseman-Elston linear regression of the squared sibpair trait difference on the estimated proportion of alleles shared IBD by the sibpair for each marker locus. Since sibpair linkage methods analyze nuclear families only, one nuclear family containing two PCD siblings was selected (described in Section 2.14.2) from each of the 78 pedigrees that contained an affected sibpair (as mentioned in Section 6.2.1, five of the 83 pedigrees contained only one PCD sibling). The distribution of sibship sizes in the nuclear families is shown in Table 6.3. The sample consisted of 305 sibpairs of all ages, of which 163 sibpairs were independent. Analyses were performed without weighting of multiple sibships, since most families had only two or three siblings and thus the results would probably not be overly biased by linkage in multiple sibships. Note that a subsample of nuclear families containing sibpairs <18 years of age was not analyzed (as was done for some of the investigations in other chapters), since the analysis was performed to confirm parametric linkage results in which individuals of all ages were analyzed.

Table 6.2

Results of two-point parametric linkage analysis of candidate gene markers

Model	1	2	3	4	5	6	7	8
Penetrances (aa, ab, bb)	0, 0, 0.8	0, 0, 0.8	0, 0.4, 0.6	0, 0.4, 0.6	0, 0.4, 0.6	0, 0.6, 0.8	0, 0.8, 1.0	0, 0.8, 1.0
Disease allele b freq	0.01	0.25	0.001	0.01	0.25	0.01	0.01	0.05
Marker	Z_{max} (θ)[a]							
D3S3045	0 (0.5)	0 (0.5)	0.09 (0.4)	0.09 (0.4)	0.18 (0.3)	0.21 (0.4)	0.35 (0.4)	0.40 (0.4)
D3S2460	0 (0.5)	0.00 (0.4)	0.06 (0.4)	0.04 (0.4)	0.15 (0.3)	0.24 (0.4)	0.54 (0.4)	0.41 (0.4)
DRD5	**1.22 (0.3)**	**1.25 (0.3)**	0 (0.5)	0 (0.5)	0.82 (0.2)	0 (0.5)	0 (0.5)	0.27 (0.3)
GABRB1	0 (0.5)	0 (0.5)	0 (0.5)	0 (0.5)	0.07 (0.3)	0 (0.5)	0.09 (0.3)	0.32 (0.3)
D5S807	0 (0.5)	0 (0.5)	0 (0.5)	0 (0.5)	0 (0.5)	0 (0.5)	0 (0.5)	0 (0.5)
D5S529	0 (0.5)	0.02 (0.4)	0 (0.5)	0 (0.5)	0 (0.5)	0 (0.5)	0 (0.5)	0 (0.5)
D5S621	0.16 (0.4)	0.20 (0.4)	0.49 (0.3)	0.57 (0.3)	0.27 (0.2)	0.28 (0.3)	0.09 (0.4)	0.02 (0.4)
GABRA1	0 (0.5)	0 (0.5)	0 (0.5)	0 (0.5)	0 (0.5)	0 (0.5)	0 (0.5)	0 (0.5)
D5S211	0 (0.5)	0 (0.5)	0.02 (0.4)	0 (0.5)	0 (0.5)	0 (0.5)	0 (0.5)	0 (0.5)
D11S1363	0 (0.5)	0 (0.5)	0.72 (0.3)	0.83 (0.3)	**1.44 (0.05)**	**1.15 (0.3)**	**1.76 (0.3)**	**2.17 (0.2)**
DRD2	0.14 (0.4)	0.21 (0.4)	0 (0.5)	0 (0.5)	0.09 (0.3)	0 (0.5)	0 (0.5)	0 (0.5)
GABRA5	0 (0.5)	0.10 (0.4)	0 (0.5)	0.00 (0.4)	0.08 (0.3)	0.05 (0.4)	0.02 (0.4)	0 (0.5)
GABRB3	0.23 (0.4)	0.33 (0.3)	0.21 (0.4)	0.27 (0.4)	0.27 (0.3)	0.35 (0.4)	0.38 (0.4)	0.46 (0.4)

[a] for all models, Z scores were calculated using MLINK at θ = 0, 0.05, 0.1, 0.2, 0.3, 0.4, and 0.5.

Table 6.3

Distribution of nuclear families of various sibship sizes used for sibpair linkage analysis of candidate gene markers

Sibship size:	Number of nuclear families
Two sibs	29
Three sibs	25
Four sibs	17
Five sibs	5
Seven sibs	1
Eight sibs	1
Total:	78

6.2.4. Linkage disequilibrium analysis

Family-based linkage disequilibrium (LD) analysis of each of the markers was performed using the AFBAC program, as described in Section 2.17.1. Analysis was performed on the 78 nuclear families that contained two or more affected siblings, using either the first affected sibling (simplex analysis option) or the first two affected siblings with weighted transmissions (multiplex analysis option). Analyses were performed with grouping of rare alleles (<5% observed frequency).

6.3. Results

6.3.1. Linkage analysis

Table 6.2 shows the results of two-point parametric linkage analyses of the markers under the eight inheritance models tested. Evidence for linkage to PCD was found with marker D11S1363 near *DRD4* ($Z_{max} = 2.2$, $\theta = 0.2$) under a dominant model with reduced penetrance (model 8). Z_{max} scores > 1.2 were also found with this marker under other dominant and intermediate models (models 5, 6, and 7; Table 6.2). A Z_{max} score of 2.2 is

considered suggestive evidence for linkage, since it surpasses the Lander and Krugylak (1995) recommended "suggestive" critical lod of 1.9 but does not reach the "significant" threshold of lod = 3.3, and it also passes the critical lod >2 to establish suggestive linkage to a mendelian trait (Morton 1955). Moderate evidence for linkage was also found to the *DRD5* gene intragenic marker, with Z_{max} = 1.2 and 1.3 under two similar recessive models (models 1 and 2; Table 6.3). None of the other markers showed any evidence for linkage to PCD, with Z_{max} <1 under all inheritance models.

Genetic model-free two-point sibpair linkage analyses confirmed the parametric linkage results. As shown in Table 6.4, simple linear regression analysis detected strong evidence for linkage between PCD and D11S1363, with $P = 0.0041$. Furthermore, concordant unaffected sibpairs had significantly increased IBD allele sharing at D11S1363 (58% sharing, $P = 0.0012$), and discordant sibpairs had significantly decreased IBD allele sharing (45% sharing, $P = 0.017$). Sibpair linkage analysis of the DRD5 intragenic marker found moderate evidence for linkage, with a simple linear regression $P = 0.071$, and no significant skewing in sibpair IBD allele sharing. Interestingly, there was significantly skewed IBD allele sharing for D5S807 and DRD2 in some sibpair types; however, there was no evidence for linkage between PCD and these markers by simple linear regression analysis.

6.3.2. Linkage disequilibrium analysis

AFBAC LD analysis detected significant associations between PCD and the DRD5 intragenic marker (simplex $P = 0.019$) and D5S529, near the *GABRB2/GABRG2* cluster (simplex $P = 0.027$), as shown in Table 6.5. However, association was not detected with D11S1363 (simplex $P = 0.98$, multiplex $P = 0.770$), or with any of the other candidate gene markers.

Table 6.4

Results of two-point sibpair linkage analysis of candidate gene markers

		Proportion of shared IBD alleles (*P* value)		
Marker	Simple linear regression *P* value	concordant unaffected [a]	discordant [b]	concordant affected [c]
D3S3045	0.76	0.46 (0.75)	0.51 (0.57)	0.48 (0.79)
D3S2460	0.18	0.46 (0.75)	0.45 (0.059)	0.49 (0.71)
DRD5	0.071	0.52 (0.38)	0.46 (0.10)	0.52 (0.22)
GABRB1	0.19	0.49 (0.56)	0.48 (0.29)	0.52 (0.19)
D5S807	0.93	**0.63 (0.0097)**	0.53 (0.87)	**0.46 (0.028)**
D5S529	0.90	0.50 (0.50)	0.55 (0.93)	0.50 (0.53)
D5S621	0.48	0.44 (0.82)	0.52 (0.73)	0.53 (0.096)
GABRA1	0.89	0.53 (0.29)	0.55 (0.93)	0.49 (0.63)
D5S211	0.40	0.49 (0.56)	0.50 (0.51)	0.51 (0.27)
D11S1363	**0.0041**	**0.58 (0.0012)**	**0.45 (0.017)**	0.52 (0.17)
DRD2	0.92	0.48 (0.66)	**0.56 (0.026)**	0.51 (0.23)
GABRA5	0.17	0.54 (0.27)	0.48 (0.25)	0.51 (0.31)
GABRB3	0.37	0.49 (0.57)	0.48 (0.22)	0.49 (0.67)

[a] Concordant unaffected sibpairs, $N = 24$.

[b] Discordant sibpairs, $N = 95$.

[c] Concordant affected sibpairs, $N = 171$.

Table 6.5

AFBAC linkage disequilibrium *P* values from analysis of candidate gene markers

Marker	simplex	multiplex
D3S3045	0.62	0.75
D3S2460	0.81	0.73
DRD5	**0.019**	0.16
GABRB1	0.16	0.96
D5S807	0.27	0.51
D5S529	**0.027**	0.62
D5S621	0.69	0.83
GABRA1	0.12	0.16
D5S211	0.66	0.84
D11S1363	0.98	0.70
DRD2	0.36	0.45
GABRA5	0.38	0.31
GABRB3	0.88	0.97

6.4. Discussion

Linkage and LD analysis of candidate GABA receptor, dopamine receptor, and dopamine transporter gene markers in our sample of Canadian families has found evidence that neurotransmission genes play a role in susceptibility to PCD. The findings are not entirely surprising given the evidence for the involvement of *DAT1* and *DRD4* in attention deficit/hyperactivity disorder, and *GABRB3* in autism, two disorders that are interrelated with dyslexia. Furthermore, neurotransmission is known to be important for proper neural development, a process that is abnormal in dyslexia, thus genes involved in this process might be expected to be involved in the manifestation of dyslexia.

Highly suggestive evidence for linkage was detected between PCD and marker D11S1363, located near the *DRD4* gene on chromosome 11p15.5. The strongest parametric linkage results were found under a dominant model with reduced penetrance, and sibpair

linkage analysis results supported this linkage. Investigation of IBD allele sharing for this marker found significantly increased allele sharing in unaffected sibpairs and decreased sharing in discordant (unaffected-affected) sibpairs. These results suggest that a disease gene near D11S1363, which could potentially be the *DRD4* gene, has an allele with a protective effect such that individuals who carry this allele are protected from susceptibility to dyslexia. However, if unaffected sibpairs received a "protective" allele from one of their parents, affected sibpairs would be expected to have increased sharing of the other "nonprotective" allele transmitted from that parent, and although there was slightly >50% allele sharing of D11S1363 in affected sibpairs, it was not statistically significant. Interestingly, LD was not detected between PCD and D11S1363. However, this marker is located approximately 3cM telomeric to *DRD4*, which is outside the ~100kb range of LD that is observed in populations of north-Europe descent, such as this Canadian sample (Collins et al. 1999; Abecasis et al. 2001; Reich et al. 2001). Thus, tests for LD with a marker very close to or, ultimately, within *DRD4* must be performed to either endorse or rule out a role for this gene in PCD, and investigation of other genes near D11S1363 may also be in order.

Moderate evidence for both linkage and LD was found with a marker within the DRD5 locus on chromosome 4p15.3-p15.1. Parametric linkage analysis under a recessive model with reduced penetrance provided the strongest evidence for linkage, although the lod score did not meet the recommended threshold of lod = 1.9 to claim suggestive evidence for linkage (Kruglyak and Lander 1995). The results of sibpair linkage analysis also fell short of statistical significance. However, the fact that significant association was detected with the DRD5 marker supports the involvement of *DRD5* in PCD. Since LD testing can be much more sensitive than tests of linkage for loci having small effects in a population (Risch and Merikangas 1996), *DRD5* may be a minor dyslexia susceptibility gene. Evidence was also found for LD (but not linkage) between PCD and marker D5S529, located 0.5cM telomeric to the GABRB2 and GABRG2 loci on chromosome 5q34-q35, suggesting that one (or possibly both) of the *GABRB2* or *GABRG2* genes may also be a minor dyslexia susceptibility gene.

Further LD analysis of the above candidate genes, ultimately using haplotypes of markers within or spanning the genes, which is more powerful than analysis of single variants (Martin et al. 2000b; Akey et al. 2001), is necessary to substantiate the involvement of *DRD4*, *DRD5*, *GABRB2*, and *GABRG2* in PCD susceptibility. While this study did not find evidence for involvement of the other candidate GABA receptor, dopamine receptor, or dopamine transporter genes in PCD, factors such as genetic heterogeneity and small gene effects may have hampered detection of linkage. Furthermore, while the markers were generally close enough to the candidate genes to be within a detectable range of linkage (with the possible exception of D5S807, which is located 22cM from *DAT1*), only the intragenic candidate gene markers might potentially be in LD with dyslexia mutations, since the average range of LD in this sample is expected to be ~100kb. Thus, it would be inappropriate to conclude that the candidate genes whose markers did not show linkage or LD to PCD in this study are not involved in susceptibility to dyslexia, only that the current study was unable to provide support for this hypothesis.

To conclude, this study has found evidence implicating the dopamine receptor genes *DRD4* and *DRD5*, and possibly the GABA receptor *GABRB2* or *GABRG2* gene, in susceptibility to dyslexia. As there are no reports in the literature regarding the involvement of specific dopaminergic or gabaminergic genes in dyslexia, these findings contribute significantly towards the identification of dyslexia genes, which ultimately will increase our understanding of the biological basis of this disorder and lead to earlier diagnosis and treatment of children at risk for dyslexia.

Chapter Seven: Discussion and Future Perspectives

7.1. Genetic analysis of complex traits: challenges and issues

The identification of genes involved in complex traits is hampered by a number of factors such as phenotypic uncertainty, genetic heterogeneity, polygenic inheritance, and phenocopies (Lander and Schork 1994). A number of steps were therefore taken to minimize these problems and to increase the probability of detecting dyslexia susceptibility loci. Several important issues regarding the analysis of complex traits also manifested themselves during the course of this project, namely appropriate statistical significance levels, replication of linkage findings, and optimal study designs.

7.1.1. The dyslexia phenotype

One factor that is problematic in genetic studies of dyslexia is phenotypic (or diagnostic) uncertainty. Because many dyslexic individuals have deficits in a number of reading components in addition to their key problem with phonological skills, it is difficult to precisely characterize the dyslexia phenotype. Some investigators define dyslexia by a discrepancy between prose reading and general intelligence (e.g. Smith et al. 1983; Morris et al. 2000a,b). However, standard IQ tests include reading components, and since dyslexic individuals would be expected to have depressed reading scores, IQ tests that include reading are probably not an accurate indicator of intelligence in reading disabled individuals. Also, as reading requires many skills, the use of a discrepancy between reading performance and intelligence to identify dyslexic individuals would result in a heterogeneous mixture of disorders in the sample (e.g. people with phonological deficits, orthographic deficits, etc.). Each of these reading disorders may have a different genetic causation, further confounding the identification of susceptibility loci. Hence, most genetic studies of dyslexia instead utilize specific reading components to assess reading disability. Since there is general agreement that the majority of reading disabled individuals have deficits in phonological skills, this component is usually assessed in genetic studies (e.g. Grigorenko et al. 1997, 2000; Fisher et al. 1999, Gayán et al. 1999). However, the level of phonological skill that has been used to define

reading disability often differed between studies, with some studies measuring phonological awareness (at the lowest level), and others assessing higher-level phonological coding or single word reading (which draws upon phonological skills and orthographic skills). Many groups also evaluated other reading components in the study of dyslexia, such as orthographic coding, spelling, and RAN (e.g. Cardon et al. 1994; Grigorenko et al. 1997, 2000; Fisher et al. 1999, Gayán et al. 1999). In contrast, one research group utilized only spelling disability in the study of dyslexia, claiming that spelling deficits are more heritable than reading deficits (Schulte-Körne et al. 1998b; Nöthen et al. 1999). Because the variance and reliability of the reading phenotypes can affect the ability to detect linkage and/or linkage disequilibrium to a susceptibility locus (Pennington 1997), the use of different phenotypes by different research groups is probably the main reason that identified loci have not been confirmed in all studies.

The phonological coding dyslexia (PCD) phenotype utilized in the studies in this thesis was developed based on the most recent reading disability research when the Field and Kaplan study was initiated. At that time, there was a consensus in the literature (as there is today) that the central cause of the reading problems in nearly all dyslexic individuals is difficulty understanding the phonemic nature of language, and is thought to be most pronounced at the level of phonological coding (the use of grapheme-phoneme rules to sound out unfamiliar words). This skill was therefore assessed in all subjects in the Field and Kaplan study, and was used as the primary determinant of an affected/unaffected/uncertain PCD diagnosis. Since cognitive studies demonstrated a significant overlap between phonological coding and phonological awareness (the perception and oral manipulation of phonemes), phonological awareness was also measured for each subject (in fact, there is debate as to which of the two phonological components is the core deficit in dyslexia [Torgesen et al. 1994; Blachman 1994]). Research also revealed that most dyslexic individuals have a persistent problem with spelling that is resistant to remediation, probably due to the involvement of phonological skills in spelling, thus each subject was also assessed for spelling ability. The phonological awareness and spelling test scores were then used to assist in PCD diagnosis. Because the reading tests were designed for children, and because dyslexic

adults may use compensatory strategies to read and thus may not be detected by reading tests, adults were also assessed for reading history, which was critical in determining PCD diagnosis for adults. Therefore, PCD was a specific phenotype defined by the central deficit observed in dyslexia, and resulted in a sample of families with a relatively homogeneous reading disorder (as far as our current knowledge and methods are able to determine). The fact that linkage and LD were detected to the DYX3 and DYX4 regions and to candidate dopamine receptor and GABA receptor genes using the PCD phenotype suggests that it is an accurate indicator of dyslexia.

While carrying out this thesis project, a paper was published that demonstrated that analysis of a quantitative (continuous) trait is often more powerful than analysis of a qualitative (discrete) phenotype (Wijsman and Amos 1997). This brought into question whether the PCD phenotype had less power to detect linkage than the quantitative measures on which it was based. In addition, the heightened focus on the use of quantitative-trait locus (QTL) linkage analysis methods for human complex trait studies prompted the use of QTL methods to separately analyze each of the reading measures that were used for PCD diagnosis (phonological awareness, phonological coding, and spelling) and a measure of RAN speed. It should be pointed out that reading experts disagree with partitioning reading into separate components, arguing that many correlated reading skills are required for skilled reading and they should all be considered together when assessing reading disability. One option that would satisfy this argument would be to determine a composite quantitative measure of reading ability based on several reading components, as was done for the Cardon et al. (1994) linkage study. Discriminant analysis could be used to select a subset of useful variables from the set of reading components and determine a composite measure. However, this analysis is fairly complicated and, if done incorrectly, may result in a composite measure that does not accurately reflect reading ability. Thus, a simpler option was to separately analyze each reading measure, keeping in mind that positive linkage findings for some traits but not for other correlated traits does not mean that different genes are involved in different reading components. Rather, some reading measures may have higher variance than others, possibly due to reading test designs, or due to differences in variability in the population for those reading skills. Higher

variance would result in higher power to detect linkage by QTL methods, potentially leading to stronger linkage results compared to a reading measure with lower variance. This was in fact observed in the *DYX4* linkage study, in which the spelling measure demonstrated stronger linkage than the correlated phonological awareness and phonological coding traits. From these results one can conclude that the *DYX4* gene is involved in some aspect of reading, but it would be erroneous to conclude that *DYX4* is involved in spelling ability but not the other skills, since the stronger linkage to spelling is most probably due to higher variance in the measure.

In contrast to the assumption that analysis of the quantitative reading measures would be more powerful than analysis of the PCD phenotype, only very weak linkage was detected to the chromosome 6p21.3 DYX2 region by QTL methods (whereas no linkage was detected with PCD; Field and Kaplan 1998). This result led to the conclusion that linkage to *DYX2* does not exist in this sample, probably because the sample contains a larger proportion of highly familial major gene forms of dyslexia (as a result of our strict ascertainment scheme) compared to other dyslexia researchers' samples. Stronger linkage to the DYX4 region was also not detected in analyses of the quantitative reading measures compared to analyses of PCD. While the results of QTL variance-component and sibpair analyses both supported linkage between dyslexia and the DYX4 region, the results were not statistically significant, just as for linkage analyses of the PCD phenotype. It was possible, however, that QTL analysis did not detect stronger linkage because of the need to subdivide large pedigrees (for variance-component analysis) or to use smaller nuclear families (for sibpair analysis). This probably reduced the power to detect linkage compared to the power that was available in the complete sample used for two-point parametric analysis of PCD, which yielded the strongest PCD results of all qualitative linkage analyses employed in the *DYX4* study.

As mentioned above, one of the reading components assessed in the subjects in the Field and Kaplan study was RAN speed. At the time the study was initiated, cognitive studies had demonstrated that many dyslexic individuals have deficits in RAN speed, but the relationship between these deficits and the phonological coding deficits

was unclear. Thus, the results of RAN speed tests were not used in determining PCD diagnosis. Research since that time has provided evidence for the independence of RAN speed and phonological deficits (Wolf 1999), and the results presented in this thesis support this independence. Firstly, the phenotypic correlations between RAN speed and the phonologically-based measures (phonological awareness, phonological coding, and spelling) were low. Furthermore, *DYX4* QTL linkage analyses found positive linkage findings for the phonologically-based measures (although the strength of linkage depended on the particular QTL linkage analysis method), but no linkage was found between the DYX4 region and RAN speed, regardless of the method. While it is possible that low informativeness of the RAN speed measure may have prevented the detection of linkage, the variance of this measure appeared to be adequate (at least to the same degree as the other measures), thus linkage between the DYX4 region and RAN speed should have been detected if RAN speed indeed shares a genetic basis with the phonological reading components.

7.1.2. Genetic heterogeneity

Dyslexia, like many other complex traits, is characterized by genetic heterogeneity. Not only is there locus heterogeneity, where different genes are involved in conferring disease risk in different families and populations, as evidenced by significant linkage to different chromosomal regions in different samples, but also there is undoubtedly allelic heterogeneity, where different mutations in a single gene confer disease susceptibility. Locus heterogeneity affects the detection of linkage, since linkage in some families might be negated by other unlinked families, and may also affect the detection of LD, since LD with a locus will only occur in the (potentially small) proportion of families linked to that locus. Allelic heterogeneity, however, does not affect linkage detection, but does hinder the detection of LD, since each mutation in the gene will be associated with a different haplotype descended from the founder in which the mutation arose. To overcome these analytical obstacles, it has been proposed that extremely large sample sizes will be needed to identify complex trait genes (Altshuler et al. 2000, Rao 2001). Alternatively, focusing on isolated populations, which are generally more

genetically homogeneous than North American and European populations, may be successful in identifying complex trait genes specific to these populations (Chapman and Thompson 2001; Peltonen et al. 2000). Another approach that is gaining interest for the identification of complex trait loci is meta-analysis, a variety of statistical procedures to synthesize the results of independent linkage studies (Gu et al. 2001).

Genetic heterogeneity was apparent in the Field and Kaplan dyslexia sample, even though the results of formal genetic heterogeneity testing were not statistically significant. The fact that linkage could not be detected to the chromosome 6p21.3 DYX2 locus by analysis of the PCD phenotype (Field and Kaplan 1998) or quantitative reading measures (Chapter 3), when this locus has been detected in five independent samples (Cardon et al. 1994; Grigorenko et al. 1997,2000; Fisher et al. 1999; Gayán et al. 1999; Turic et al. 2000), strongly suggests that our sample contains different genetic forms of dyslexia compared to other samples. In contrast, strong linkage and LD to the chromosome 6q11.2-q12 DYX4 region was found in this sample (Chapter Four), demonstrating that the sample had sufficient power for the detection of dyslexia loci. While LD was detected to several markers and haplotypes in a defined region, allelic heterogeneity at the DYX4 locus was indicated by the fact that none of the families shared a common haplotype with any other family. It will be interesting to see whether other groups are able to detect linkage to the DYX4 locus in their samples, or whether locus heterogeneity between samples will prevent replication of *DYX4*. The Field and Kaplan sample, however, did appear to possess the same dyslexia predisposing gene as found in another sample, namely the DYX3 locus identified in a large Norwegian family (Fagerheim et al. 1999). Significant linkage was found to *DYX3* by NPL analysis, and significant LD was also detected with haplotypes in this region (Chapter Five). However, allelic heterogeneity was indicated by the fact that the Norwegian affected haplotype was not found in the sample and, furthermore, the families linked to this region did not share a common DYX3 haplotype.

7.1.3. Significance levels in studies of complex traits

One of the difficulties in interpreting the results of a complex trait study is sorting out the true positive signals from the false positive signals. This difficulty arises because of

the multiple tests that are often employed in complex trait studies. Linkage and LD tests are often carried out numerous times using several methods and models, and for multiple phenotypes. Use of the conventional 5% significance level when interpreting the results of each of these tests will lead to an increased probability of false positives. Instead, the significance level must be adjusted to reflect the fact that multiple tests were performed. Traditional Bonferoni correction for multiple testing assumes that the tests are independent, which is generally not the case in genetic studies. In the studies presented in this thesis, the microsatellite markers that were investigated were located near each other, the multiple linkage tests and LD tests that were performed were similar to one another, and the dyslexia phenotypes (PCD and the quantitative reading measures) were, for the most part, correlated with each other. Thus, Bonferoni correction would be overly conservative, and thus is not appropriate for these studies or for genetic studies in general. An alternative way to control the false positive rate when multiple tests are performed is to adopt a stringent significance level, as proposed by Lander and Kruglyak (1995). For a genome scan of dense markers analyzed using parametric linkage methods, the recommended lod score to claim suggestive linkage is 1.9, and to claim significant linkage, the lod is 3.3. While these stringent critical lod scores certainly reduce the chance of a false positive linkage result, they also compromise the detection of true linkage, since a lod $\geq$3.3 is extremely difficult to achieve for a complex trait due to incomplete penetrance, phenocopies, genetic heterogeneity, and polygenic inheritance (Lander and Schork 1994). In other words, the application of stringent significance levels increases the probability of false *negative* results, or type II error (i.e., no evidence for linkage when linkage exists). Lander and Kruglyak (1996) also argue that results should be corrected for the large number of tests performed in a genome scan, even if such a large number of tests was not performed, since one *would* undertake such a dense scan to detect linkage, if necessary. Others (Witte et al. 1996; Elston 1997,1998) strongly argue against this reasoning, and instead recommend that precise linkage and LD results be reported (not adjusting for multiple tests) to allow interpretation in light of the particular study. In addition, because one expects more than one gene to be involved in the etiology of a complex trait, focusing on one or a few extreme results might not provide an accurate picture of the multifactorial situation (Witte et al.

1996). An alternative method to control the rate of false positive results in studies of complex traits is to perform computer simulations to determine empirical significance levels for the particular study being performed, rather than using pre-determined significance levels that may not be appropriate (Ott 1999, p79). However, few linkage or LD programs include this option as of yet, leading investigators to perform in-house computer simulations, which may cause difficulties when comparing results between different studies. It is expected that demand by the scientific community will result in the implementation of computer simulations in many linkage and LD programs, hopefully resolving the issue of how to properly control the rate of false positive results in complex trait studies. And finally, it is possible that the width of a lod score peak may be useful for determining the significance of a linkage result, since true positive peaks are generally wider than false positive peaks (Terwilliger et al. 1997).

Regarding the studies presented in this thesis, multiple testing correction of lod scores and *P* values was not performed, as suggested by Witte et al. (1996) and Elston (1997,1998), yet the recommended critical lod scores of Lander and Kruglyak (1995) were considered when interpreting the linkage results. This was done more to abide by the current practice in the literature, where use of these stringent significance levels is common, than because of agreement with these thresholds. With regard to the LD analyses, the number of tests performed by AFBAC analysis was considered when interpreting the LD results, particularly for *DYX4* analysis where almost 70 markers were investigated. Fortunately, haplotype LD analysis using TRIMHAP determined empirical significance levels, thus these results could be viewed as accurate.

7.1.4. Replication of linkage findings

Replication is commonly required for accepting a positive finding, even if the original finding surpassed stringent significance thresholds. However, failure to replicate a positive finding in a different study is not necessarily proof that the original finding was false. Suarez et al. (1994) demonstrated by computer simulation that if several loci each with a modest effect are implicated in a disease, then linkage will be difficult to detect and replicate, and that failure to replicate can be due to heterogeneity or to the

statistical consequences of attempting to map several genes involved in a complex trait. For example, if there are six unlinked disease loci for a disorder, and the power to detect any one of them is 20%, then the power to detect at least one of them is $1 - (0.8)^6 =$ 74%. Thus, there may be a high probability of detecting one of the six susceptibility loci, but in a replication study looking for a specific previously-detected locus, the power is only 20%. Suarez et al. (1994) concluded that unless the replication sample is much larger than the original sample, replication will often fail. Other strategies to improve the chances of replication are to carry out the replication study on the same underlying population from which the original sample was drawn, or to match the samples for other characteristics such as age, gender, ethnicity, etc. (Rao and Gu 2001).

Two replication studies were presented in this thesis: investigation of the chromosome 6p21.3 DYX2 region (Chapter Three) and the chromosome 2p15-p16 DYX3 region (Chapter Five). As already mentioned during the discussion of genetic heterogeneity, linkage to the DYX2 locus could not be replicated, even though this locus has been replicated by four independent groups (Grigorenko et al. 1997,2000; Fisher et al. 1999; Gayán et al. 1999; Turic et al. 2000). It was concluded that genetic heterogeneity between samples is probably the reason for this failed replication. Low power to detect this locus (i.e., small genetic effect) is not likely given that numerous studies have been able to replicate the finding. The DYX3 locus was replicated in our sample, however, and this is the first reported replication of *DYX3*, thus substantiating the existence of this locus.

7.1.5. Study design

Optimal study designs play a critical role for successful mapping of genes involved in complex traits. While some factors that influence the success of a study are out of the investigator's control, other factors can be manipulated to improve the chance of identifying disease loci. One aspect of the study design, selection of a specific phenotype, has already been discussed in detail above. Other factors, such as the ascertainment scheme, family structure, and analytical procedures, for example, are equally critical for successfully mapping genes for complex disorders.

One strategy to identify loci involved in susceptibility to a complex trait is to focus on high-risk families with multiple affected members. The original Field and Kaplan study design was to ascertain nuclear families consisting of two parents and two or more affected offspring. While this scheme was selected to allow for linkage analysis by affected sibpair methods, it probably also had the effect of minimizing the frequency of phenocopies in the sample compared to single proband ascertainment. Many of the ascertained families had affected parents and siblings in addition to the two affected siblings, and many families had affected extended relatives who were invited to participate in the study. As a result, a large proportion of the families appeared to exhibit dominant transmission of dyslexia, with many affected members in each generation. Thus, the "two affected siblings" ascertainment scheme inadvertently selected families with multiple affected members, and therefore probably selected highly familial, major gene forms of dyslexia that may have increased our chances of detecting major loci. However, enrichment of the sample for major gene forms of dyslexia probably also prevented the detection of linkage to specific loci, such as DYX2 on chromosome 6p21.3, which has been detected by five other groups. Thus, the ascertainment scheme can have serious consequences on the outcome of a linkage study, whether increasing or decreasing the chance of detecting particular loci, and should be considered when interpreting linkage results between studies that utilized different sampling schemes.

As discussed above, the original Field and Kaplan study design was to ascertain nuclear families with two or more affected siblings, however other family structures were included in the study, such that there was an approximately 1:1 ratio of nuclear families and extended pedigrees in the sample. This raises the issue of which family structure is optimal for linkage analysis of complex traits. Extended pedigrees are more informative than nuclear pedigrees in terms of power to detect linkage and accuracy of gene localization, even when the samples contain identical numbers of individuals (Wijsman and Amos 1997). The increase in power probably occurs because of the additional meioses available in the large pedigrees and the increased ability to determine phase (which alleles were received as a haplotype from one parent), whereas this information must be inferred from sibships in nuclear families. Large pedigrees also allow one to look for segregation of a

candidate gene mutation with the disorder, a helpful piece of evidence when claiming that a candidate gene is the true disease susceptibility gene. However, large pedigrees have a higher chance of different mutations being introduced through married-in individuals. In the Field and Kaplan sample, approximately one-third of the pedigrees had married-in members who were either affected or had a family history of dyslexia. While these bilineal pedigrees were still informative for linkage analyses, they probably hampered the detection of loci relative to linear pedigrees. The size of a pedigree is also directly proportional to the computational complexity of current multipoint linkage analysis algorithms. As a result, extremely large pedigrees may need to be divided into smaller subpedigrees (as had to be done for all GENEHUNTER analyses reported in this thesis), which reduces the power for detecting linkage, and raises the question of whether the effort required to collect these large pedigrees was well spent. Ascertainment of smaller families with affected sibpairs, on the other hand, is easier and less expensive, which may allow the collection of a larger sample, and offset the decreased linkage power of nuclear families. Methods have specifically been developed for linkage analysis of this type of family (e.g. affected sibpair [ASP] linkage analysis). Ultimately, the key factor when determining the optimal family structure to employ in a genetic study is the true mode of disease inheritance, which is generally not known. Thus, there may not be one optimal family structure for analysis of a particular complex trait. Rather, the best strategy may be to collect both nuclear and extended families, as was done in the Field and Kaplan study, and to recognize the particular strengths and limitations inherent to each family type.

While the ascertainment scheme and family structure are important aspects of an optimal study design, the success of a complex trait study is also dependent on the analysis methods that are used. Every method of linkage analysis is optimal for different modes of inheritance, different family structures, and different disease prevalences. Parametric linkage analyses carry the most power to detect linkage, however they require an assumed inheritance model that is not easy to specify for a complex trait of unknown inheritance. Nonparametric (genetic model-free) methods based on IBD allele sharing in affected family members have therefore been developed, but they have lower power to detect linkage and require a larger sample size compared to parametric analysis under the correct genetic

model. Similarly, sibpair linkage analysis (or affected sibpair [ASP] linkage analysis) is a genetic model-free method specifically designed for nuclear families. This method is more likely to detect recessive loci than dominant loci because, for dominant traits, only one parent is informative for linkage (i.e., transmitting the disease allele, which is detected by increased sibpair IBD allele sharing), whereas for recessive traits, both parents are informative (Ott 1999, p274). One aspect of genetic model-free methods that is often overlooked, however, is that they are less affected by heterogeneity than parametric analysis. This is because for parametric analysis, linkage in some families can be negated by other unlinked families, whereas for IBD sharing methods, linkage will not be excluded by heterogeneity, although larger sample sizes will be required to detect genes. As mentioned above in discussion of the dyslexia phenotype, QTL linkage methods should prove more successful in identifying complex trait loci than qualitative methods, due to the increased informativeness of quantitative measures. Blangero et al. (2001) demonstrated that the success of a QTL linkage study is dependent on the disease prevalence, since for common diseases with 15% or higher prevalence (thus higher than the dyslexia prevalence of 3-10%; Lerner 1989), extended pedigrees have markedly greater power for mapping loci by QTL methods than affected sibpair samples, requiring approximately half the number of individuals to achieve 80% power to detect a QTL. Conversely, affected sibpairs were found to be more powerful than extended pedigrees for detecting QTLs for rare diseases. While most of this discussion has focused on linkage analysis methods, LD analysis methods are also optimized for different family structures and different genotype information. Most LD methods in current use analyze nuclear families for single-marker associations with a qualitative phenotype (e.g. AFBAC; Thomson 1995). Methods have recently become available that investigate parent-affected child trios (e.g. HAPMAX; http://www.uwcm.ac.uk/uwcm/mg/download) or extended pedigrees (e.g. TRIMHAP; MacLean et al. 2000) for associations between a qualitative phenotype and multiple-marker haplotypes, which have more power to detect LD than analysis of single markers (Martin et al. 2000a; Akey et al. 2001). To conclude, the particular linkage and LD methods utilized in a study are critical for successfully detecting complex trait loci. However, each method is optimal for different family structures, different modes of inheritance, different disease

prevalences, and different phenotypes. Thus, these factors must all be given careful consideration when designing a complex trait study to maximize the chance of detecting susceptibility loci.

A variety of linkage and LD methods were utilized in the studies presented in this thesis. Since families containing two or more affected siblings were ascertained, sibpair methods could be employed. However, these methods were not able to utilize the entire sample, since half of the families were extended pedigrees but only one nuclear family from each of these pedigrees could be used in sibpair analyses. Other more powerful methods that analyze larger pedigrees were therefore employed, for example, parametric and nonparametric linkage analysis and TRIMHAP LD analysis. While use of a variety of methods increased the chances of detecting evidence for linkage and/or LD, it also may have increased the chance of false positive linkages, thus correction for multiple testing may have been necessary. However, Bonferoni correction is not appropriate since it assumes that the tests are completely independent, and at this time there is no acceptable method for correcting for partially dependent tests such as those used in these studies. This highlights the need for new methods and discussion in the literature to resolve the issue of how to appropriately correct for testing by multiple analytical methods.

Parametric linkage analysis requires an inheritance model that is unknown for dyslexia, thus eight models representing recessive, intermediate, and dominant inheritance with a variety of disease allele frequencies were used for the investigations of *DYX4* and the candidate GABA receptor, dopamine receptor, and dopamine transporter genes. While this strategy probably increased the chance of detecting linkage, it also raised the thorny issue of how to appropriately correct for testing multiple models. The two-point lod scores obtained under models with the same penetrances but different disease allele frequencies were generally about the same, indicating that disease allele frequency has very little impact on linkage results, as found in simulation studies (Xu et al. 1998). Thus, analysis using fewer models (for example, one recessive, one intermediate, and one dominant) might still provide a high chance of detecting linkage, and only necessitate minor corrections for multiple testing, if any. Because of the

multiple testing issue, only one model was employed in parametric linkage analyses of the DYX3 locus. This model was similar to that used in the original Norwegian linkage study and was based on Norwegian prevalence rates (Fagerheim et al. 1999). However, very weak evidence for linkage was found in our sample using this model, whereas nonparametric linkage analysis detected significant linkage to *DYX3*. Thus, this model was apparently not the best model to use for our Canadian sample, and recent work by Jordana Tzenova in Dr. Field's laboratory has found much stronger linkage to the DYX3 region under one of the dominant models in the set of eight models, indicating that use of a standard set of models might be more successful in detecting loci in our sample.

7.2. Future Perspectives

7.2.1. The DYX4 locus

Strong evidence for a dyslexia susceptibility locus on chromosome 6q11.2-q12 (*DYX4*) was found by linkage and LD analysis, as discussed in Chapter Four. However, the linkage results did not pass conventional significance thresholds, thus replication in other dyslexia family samples is required to confirm the presence of this locus. Even if significant linkage had been obtained, replication would probably still be required before the scientific community would regard the locus as a true positive.

The DYX4 region must be refined to a single, small candidate interval (~1Mb) before fine mapping of the gene is feasible. The results presented in Chapter Four identified two candidate DYX4 intervals. The first interval spans ~4Mb region on 6q11.1, in the vicinity of D6S1711 to D6S430, and was identified based on significant associations with markers in this region in linked families (markers covering a larger region were also associated in all of the families). The second candidate region spans ~6.9Mb on 6q12, between D6S280 and D6S460, and was identified based on recombination breakpoints in haplotypes of linked families. Therefore, additional investigation of both of these regions must be performed to determine which, if either, is more likely to harbour the *DYX4* gene. One method that could be used in this determination would be to genotype additional microsatellite markers and single nucleotide polymorphisms (SNPs), which are found every 1kb on average, in the two

candidate regions. Haplotype LD analysis, which is more powerful than single marker LD analysis, (Martin et al. 2000a; Akey et al. 2001), could then be performed with the hope that haplotypes in one of the candidate regions will demonstrate strong associations with PCD, and the associated region will be small enough to allow fine mapping of *DYX4*. If the 6.9Mb consensus region between D6S280 and D6S460 is identified as the more likely location of *DYX4*, this region can also be further refined using information already at hand. Genotype information for the linked families is available for eight markers in this region that were used for LD analysis, thus after establishing the correct order of these markers (e.g. by STS content mapping and/or the human genome sequence), haplotypes can be determined for the five families known to have breakpoints in this region, and localization of the breakpoints in these more informative haplotypes may further refine the consensus region. Replication in independent samples may also refine the location of the DYX4 locus, thereby facilitating fine mapping of the gene.

To identify the *DYX4* gene, cosegregation of a candidate gene variant with dyslexia in a large family or several families and/or LD between dyslexia and the mutation must first be demonstrated to implicate the gene in dyslexia. Since it is possible that several sequence variants will occur in a candidate gene, to identify the particular disease-predisposing variant (or variants, since allelic heterogeneity is possible), functional variants should be given priority when carrying out the segregation or LD studies. When identifying functional variants in a candidate gene, it should be kept in mind that sequence variants located in regulatory regions and splice sites may affect gene expression and thus are potential disease-predisposing mutations, in addition to coding sequence variants that alter amino-acids. Regarding cosegregation studies in families, unaffected individuals will potentially possess the variant as a result of incomplete disease penetrance, and affected individuals will potentially not possess the variant due to phenocopy or intrafamilial genetic heterogeneity (particularly for dyslexia, where there is a high proportion of bilineal families). Regarding LD studies of the sequence variant, to increase the chances of detecting LD, strategies may be used such as subdividing the sample based on evidence for linkage (e.g. families with lod scores over a certain value, or families in which affected members share a common haplotype), and

employing LD methods that can analyze the largest family size in the sample (e.g. using TRIMHAP to analyze extended pedigrees, rather than using AFBAC to analyze only one or two affected children per pedigree).

Because cosegregation of, or LD between, dyslexia and a candidate gene mutation may occur when the variant is in tight linkage disequilibrium with the true disease-causing mutation (which may be in a different nearby gene), functional and mutational studies must be performed to prove that the candidate gene is involved in dyslexia. For example, significant alteration of the normal expression of the gene in dyslexic individuals can be investigated by measuring mRNA expression levels by Northern blotting or reverse transcriptase PCR, or by measuring protein expression levels by Western blotting, fluorescence-activated cell sorting (FACS), or immunocytochemistry. Demonstrating a functional change in the protein product might also prove that the candidate gene variant is involved in dyslexia, for example by assays for ligand specificity, ion channel permeability, or enzyme kinetics, depending on the protein's function, using cells from dyslexic individuals. A candidate gene may also be proven to have a role in dyslexia by generation of a similar phenotype in transgenic animals. The "transgenic" animal model may not possess the same neurobiological phenotype as humans due to different biology or divergent contribution of the orthologous gene (and obviously a mouse will not have the reading disabled phenotype), however, alteration in some critical aspect of normal brain development and/or function in the animal model may be evidence that directs additional studies to support a role for the candidate gene in dyslexia.

7.2.2. The DYX3 locus

Confirmatory evidence for the DYX3 locus on chromosome 2p15-p16 was found by linkage and LD studies, as discussed in Chapter Five. Furthermore, a 1.9Mb consensus region was identified between D2S378 and D2S2183, which supports the location of *DYX3* reported in the original Norwegian study. While this consensus region is small enough to make fine mapping of *DYX3* feasible, the consensus region might be further refined using two methods. Firstly, additional markers saturating the consensus

region may be genotyped in the two families known to have recombination breakpoints in this region and, after determining the correct marker order (e.g. by STS content mapping and/or the human genome sequence) and determining haplotypes, the breakpoints may be localized in these more informative haplotypes, potentially reducing the size of the candidate interval. This information would be particularly useful given the current uncertainty in the breakpoint location in one of these families. Secondly, the DYX3 candidate interval may be further refined by detecting strong LD between PCD and SNP haplotypes spanning a small region, since LD is expected to extend only ~100kb in this sample (Collins et al. 1999; Abecasis et al. 2001; Reich et al. 2001). Regardless of whether the DYX3 candidate region is narrowed any further, fine mapping of DYX3 is feasible. A putative serine-threonine kinase, *VRK2*, is already mapped to this region, and with the availability of sequence of this region (International Human Genome Sequencing Consortium 2001), gene prediction tools and other technologies can be used to identify additional candidate genes. Sequence variants in candidate genes may then be investigated for cosegregation in families or for LD with PCD, followed by functional and mutational studies (as discussed above for *DYX4*), with the ultimate goal of identifying the *DYX3* gene.

7.2.3. Candidate GABA receptor, dopamine receptor, and dopamine transporter genes

Linkage and LD studies found evidence implicating the dopamine receptor genes *DRD4* and *DRD5*, and possibly the GABA receptor *GABRB2* or *GABRG2* gene, in susceptibility to dyslexia, as discussed in Chapter Six. While strong linkage was detected with a marker located near *DRD4*, LD with this marker was not detected, however this could be because the marker is located ~3cM from *DRD4* and is not in LD. Therefore, tests for LD must be performed with sequence variants within *DRD4* such as SNPs or the 48bp repeat in exon 3 (which is currently being investigated by Dr. Robin Hsiung in Dr. Field's laboratory), and ultimately with functional variants in *DRD4*. LD was detected with a marker in the *DRD5* gene, and with a marker very close to *GABRB2* and *GABRG2*. Functional variants in each of these genes should also be investigated for LD with PCD. Sequence variants in any of the above genes that are found to be associated with PCD could also be investigated for cosegregation in families to provide further evidence for the gene in

dyslexia susceptibility. And, as discussed in detail above with regard to *DYX4*, functional and mutational studies of the gene would finally be required to prove a role in dyslexia.

Linkage and LD were not detected between PCD and markers for several other candidate GABA receptor and dopamine receptor genes and the dopamine transporter gene. However, this may have been due to factors such as genetic heterogeneity and small gene effects, and not because these genes are not involved in dyslexia. Several of the markers were outside of the ~100kb detectable range of LD expected in this sample (Collins et al. 1999; Abecasis et al. 2001; Reich et al. 2001), thus markers within these genes (preferably SNP haplotypes) should be investigated for LD before dismissing these genes as having no involvement in dyslexia

7.2.4. The future of complex trait studies

It is clear that a multistrategy approach is required for studies of complex traits, since no single method is sufficient or optimal. While linkage analyses have had great success for mapping single gene traits, it is clear that they are limited for mapping complex traits, although strategies such as the use of a specific disease phenotype and studying more homogeneous isolated populations may improve the chance of detecting linkage to complex trait loci. The development of more powerful analytical techniques, such as multipoint linkage methods and meta-analysis methods, will hopefully improve the success rate of linkage studies.

Linkage disequilibrium studies for the discovery of complex trait genes might, under certain circumstances, be more powerful than linkage analyses and may be capable of detecting loci with small effect (Risch and Merikangas 1996). However, recent studies indicate that LD extends over a relatively small region of approximately 100kb in north-Europe descended populations (Collins et al. 1999; Abecasis et al. 2001; Reich et al. 2001), and an even smaller region of approximately 5kb in older, African populations (Reich et al. 2001). Thus, LD studies will be most successful if a dense map of SNP markers (located ~1kb apart in the genome) are utilized and analyzed as haplotypes, which display stronger LD than single SNPs and can better localize complex trait genes (Martin et al. 2000a; Akey et al. 2001). Genome scans at a density of 100kb

will only be feasible for large centres and consortiums, thus it may be practical for most researchers to focus LD studies on candidate genes, more of which will become available as the Human Genome Project nears completion.

In addition to linkage and LD studies, other approaches, such as investigation of animal models of complex traits, DNA microarray screens of large numbers of genes for differential gene expression, and population-level surveys of human variation, will all contribute to uncovering the genetics of complex disorders such as dyslexia. Identification of the genes involved in dyslexia will ultimately lead to a greater understanding of the biological basis of this disorder, and will result in better diagnostic methods for children at high risk of developing this disorder and better treatments for people affected with dyslexia.

References

Abecasis GR, Noguchi E, Heinzmann A, Traherne JA, Bhattacharyya S, Leaves NI, Anderson GG, Zhang Y, Lench NJ, Carey A, Cardon LR, Moffatt MF, Cookson WO (2001) Extent and distribution of linkage disequilibrium in three genomic regions. *Am J Hum Genet* 68:191-7

Akey J, Jin L, Xiong M (2001) Haplotypes vs. single marker linkage disequilibrium tests: what do we gain? *Eur J Hum Genet* 9:291-300

Allison DB, Neale MC, Zannolli R, Schork NJ, Amos CI, Blangero J (1999) Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure. *Am J Hum Genet* 65:531-44

Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl M-C, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, Tuomi T, Gaudet D, Hudson TJ, Daly M, Groop L, Lander ES (2000) The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nature Genet* 26:76-80

Barr CL, Wigg K, Bloom S, Schachar R, Tannock R, Roberts W, Malone M, Kennedy JL (2000) Further evidence from haplotype analysis for linkage of the dopamine D4 receptor gene and attention-deficit hyperactivity disorder. *Am J Med Genet* 96:262-7

Barr CL, Xu C, Kroft J, Feng Y, Wigg K, Zai G, Tannock R, Schachar R, Malone M, Roberts W, Nothen MM, Grunhage F, Vandenbergh DJ, Uhl G, Sunohara G, King N, Kennedy JL (2001) Haplotype study of three polymorphisms at the dopamine transporter locus confirm linkage to attention-deficit/hyperactivity disorder. *Biol Psychiatry* 49:333-9

Behan PO, Geschwind N (1985) Dyslexia, congenital anomalies, and immune disorders: the role of the fetal environment. *Ann N Y Acad Sci* 457:13-8

Bisgard ML, Eiberg H, Møller N, Niebuhr E, Mohr J (1987) Dyslexia and chromosome 15 heteromorphisms: negative lod score in a Danish material. *Clin Genet* 32:118-9

Blachman B (1994) What have we learned from longitudinal studies of phonological processing and reading and some unanswered questions: a response to Torgesen, Wagnere, and Rashotte. *J Learning Disabil* 2:287-91

Blangero J, Williams JT, Almasy L (2001) Variance component methods for detecting complex trait loci. In Rao CD, Province MA (eds.) Genetic dissection of complex traits. Academic Press, London, UK: pp.151-181

Bloedel JR (1993) 'Involvement in' versus 'storage of'. *Trends Neurosci* 16:451-2

Bowers PG, Swanson LB (1991) Naming speed deficits in reading disability: multiple measures of a singular process. *J Exp Child Psychol* 51:195-219

Brietmeyer BG (1980) Unmasking visual masking: a look at the "why" behind the veil of the "how". *Psychol Rev* 87:52-69

Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic maps: Individual and sex-specific variation in recombination. *Am J Hum Genet* 63:861-9

Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC (1994) Quantitative trait locus for reading disability on chromosome 6. *Science* 266:276-9; erratum 268:1553

Challmann TD, Lipsky JJ (2000) Methylphenidate: its pharmacology and uses. *Mayo Clin Proc* 75:711-21

Chapman NH, Thompson EA (2001) Linkage disequilibrium mapping: the role of population history, size, and structure. In Rao CD, Province MA (eds.) Genetic dissection of complex traits. Academic Press, London, UK: pp.413-37

Clerget-Darpoux F, Bonaïti-Pellié C (1992) Strategies based on marker information for the study of human diseases. *Ann Hum Genet* 56:145-53

Collins A, Frezal J, Teague J, Morton NE (1996) A metric map of humans: 23,500 loci in 850 bands. *Proc Natl Acad Sci USA* 93:14771-5

Collins A, Lonjou C, Morton NE (1999) Genetic epidemiology of single-nucleotide polymorphisms. *Proc Natl Acad Sci USA* 96:15173-7

Cook EH Jr, Courchesne RY, Cox NJ, Lord C, Gonen D, Guter SJ, Lincoln A, Nix K, Haas R, Leventhal BL, Courchesne E (1998) Linkage-disequilibrium mapping of autistic disorder, with 15q11-13 markers. *Am J Hum Genet* 62:1077-83

Cook EH Jr, Stein MA, Krasowski MD, Cox NJ, Olkon DM, Kieffer JE, Leventhal BL (1995) Association of attention-deficit disorder and the dopamine transporter gene. *Am J Hum Genet* 56:993-8

Cornelissen P, Richardson A, Mason A, Fowler S, Stein J (1995) Contrast sensitivity and coherent motion detection measured at photopic luminance levels in dyslexics and controls. *Vision Res* 35:1483-94

Cottingham Jr RW, Idury RM, Schaffer AA (1993) Faster sequential genetic linkage computations. *Am J Hum Genet* 53:252-63

Courchesne E (1997) Brainstem, cerebellar and limbic Neuroanatomical abnormalities in autism. *Curr Opin Neurobiol* :269-78

Crawford SG, Kaplan BJ, Kinsbourne M (1994) Are families of children with reading difficulties at risk for immune disorders and nonrighthandedness? *Cortex* 30:281-92

Critchley M (1970) The dyslexic child. Charles C Thomas, Springfield, IL

Curran S, Mill J, Sham P, Rijsdijk F, Marusic K, Taylor E, Asherson P (2001) QTL association analysis of the *DRD4* exon 3 VNTR polymorphism in a population sample of children screened with a parent rating scale for ADHD symptoms. Am J Med Genet 105:387-93

Cutting GR, Curristin S, Zoghbi H, O'Hara BF, Seldin MF, Uhl GR (1992) Identification of a putative γ-aminobutyric acid (GABA) receptor subunit rho_2 cDNA and colocalization of the genes encoding rho_2 (GABRR2) and rho_1 (GABRR1) to human chromosome 6q14-q21 and mouse chromosome 4. *Genomics* 12:801-6

Dausset J, Ougen P, Abderrahim H, Billault A, Sambucy JL, Cohen D, Le Paslier D (1992) The CEPH YAC library. *Behring Inst Mitt* 91:13-20

DeFries JC and Alarcón M (1996) Genetics of specific reading disability. *Ment Retard Dev Disabil Res Rev* 2:39-47

DeFries JC, Fulker DW, LaBuda MC (1987) Evidence for a genetic aetiology in reading disability of twins. *Nature* 329:537-9

Demb JB, Boynton GM, Best M, Heeger DJ (1998) Psychophysical evidence for a magnocellular pathway deficit in dyslexia. *Vision Res* 38:1555-9

Denckla MB, Rudel RG (1974) Rapid "automatized" naming of pictured objects, colors, letters and numbers by normal children. *Cortex* 10:186-202

——— (1976) Rapid automatized naming (R.A.N.): dyslexia differentiated from other learning disabilities. *Neuropsychologia* 14:471-9

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380:152-4

Eden GF and Zeffiro TA (1998) Neural systems affected in developmental dyslexia revealed by functional neuroimaging. *Neuron* 21:279-82

Ehri L (1995) Phases of development in learning to read words by sight. *J Res Read* 18:116-25

Elston RC (1989) Man bites dog? The validity of maximizing lod scores to determine mode of inheritance. *Am J Med Genet* 34:487-8

——— (1997) Algorithms and inferences: the challenge of multifactorial diseases. *Am J Hum Genet* 60:255-62

——— (1998) Methods of linkage analysis – and the assumptions underlying them. *Am J Hum Genet* 63:931-4

Fagerheim T, Raeymaekers P, Tonnessen FE, Pedersen M, Tranebjaerg L, Lubs HA (1999) A new gene (*DYX3*) for dyslexia is located on chromosome 2. *J Med Genet* 36:664-9

Faraone SV, Biederman J, Lehman BK, Keenan K, Norman D, Seidman LJ, Kolodny R, Kraus I, Perrin J, Chen WJ (1993) Evidence for the independent familial transmission of attention deficit hyperactivity disorder and learning disabilities: results from a family genetic study. *Am J Psychiatry* 150:891-5

Fawcett AJ, Nicolson RI (1999) Performance of dyslexic children on cerebellar and cognitive tests. *J Motor Behav* 31:68-78

Fawcett AJ, Nicolson RI, Dean P (1996) Impaired performance of children with dyslexia on a range of cerebellar tasks. *Ann Dyslexia* 46:259-83

Felton RH, Naylor CE, Wood FB (1990) Neuropsychological profile of adult dyslexics. *Brain Lang* 39:485-97

Fernández-Ruiz J, Berrendero F, Hernández ML, Ramos JA (2000) The endogenous cannabinoid system and brain development. *Trends Neurosci* 23:14-20

Field LL, Kaplan BJ (1998) Absence of linkage of phonological coding dyslexia to chromosome 6p23-p21.3 in a large family data set. *Am J Hum Genet* 63:1448-56; erratum 64:334

Finucci JM, Childs B (1981) Are there really more dyslexic boys than girls? In Ansara A, Geshwind N, Galaburda A, Albert M, Gartrell N (eds.) Gender Differences in Dyslexia. Orton Dyslexia Society, Towson, MD: pp.1-9

Finucci JM, Guthrie JT, Childs AL, Abbey H, Childs B (1976) The genetics of specific reading disability. *Ann Hum Genet* 40:1-23

Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans Royal Soc Edinburgh* 52:399-433

Fisher SE, Marlow AJ, Lamb J, Maestrini E, Williams DF, Richardson AJ, Weeks DE, Stein JF, Monaco AP (1999) A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. *Am J Hum Genet* 64:146-56

Folstein SE, Santangelo SL, Gilman SE, Piven J, Landa R, Lainhart J, Hein J, Wzorek M (1999) Predictors of cognitive test patterns in autism families. *J Child Psychol Psychiatry* 40:1117-28

Francke U (1994) Digitized and differentially shaded human chromosome ideograms for genomic applications. *Cytogenet Cell Genet* 65:206-19

Frith U (1998) Cognitive deficits in developmental disorders. *Scand J Psychol* 39:191-5

Frost R (1998) Toward a strong phonological theory of visual word recognition: true issues and false trials. *Psychol Bull* 123:71-99

Galaburda AM, Menard MT, Rosen GD (1994) Evidence for aberrant auditory anatomy in developmental dyslexia. *Proc Natl Acad Sci USA* 91:8010-3

Gayán J, Smith SD, Cherny SS, Cardon LR, Fulker DW, Brower AM, Olson RK, Pennington BF, DeFries JC (1999) Quantitative-trait locus for specific language and reading deficits on chromosome 6p. *Am J Hum Genet* 64:157-64

Georgiewa P, Rzanny R, Hopf JM, Knab R, Glauche V, Kaiser WA, Blanz B (1999) fMRI during word processing in dyslexic and normal reading children. *Neuroreport* 10:3459-65

Gilger JW, Pennington BF, DeFries JC (1991) Risk for reading disability as a function of parental history in three family studies. *Reading Writing* 3:205-17

Gilger JW, Pennington BF, DeFries JC (1992) A twin study of the etiology of comorbidity: attention-deficit hyperactivity disorder and dyslexia. *J Am Acad Child Adolesc Psychiatry* 31:343-8

Gill M, Daly G, Heron S, Hawi Z, Fitzgerald M (1997) Confirmation of association between attention deficit hyperactivity disorder and a dopamine transporter polymorphism. *Mol Psychiatry* 2:311-313

Glickstein M (1993) Motor skills but not cognitive tasks. *Trends Neurosci* 16:450-1

Gough PB, Walsh MA (1991) Chinese, Phoenicians, and the orthographic cipher of English. In Brady SA, Shankweiler DP (eds.) Phonological processes in literacy. Erlbaum, Hillsdale, NJ: pp.199-209

Greenberg DA, Abreu P, Hodge SE (1998) The power to detect linkage in complex disease by means of simple LOD-score analyses. *Am J Hum Genet* 63:870-9

Grigorenko EL, Wood FB, Meyer MS, Hart LA, Speed WC, Shuster A, Pauls DL (1997) Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. *Am J Hum Genet* 60:27-39

Grigorenko EL, Wood FB, Meyer MS, Pauls DL (2000) Chromosome 6p influences on different dyslexia-related cognitive processes: further confirmation. *Am J Hum Genet* 66:715-23

Grigorenko EL, Wood FB, Meyer MS, Pauls JED, Hart LA, Pauls DL (1998) Linkage studies suggest a possible locus for dyslexia near the Rh region on chromosome 1. *Behav Genet* 28:470

Gu C, Province MA, Rao DC (2001) Meta-analysis for model-free methods. In Rao CD, Province MA (eds.) Genetic dissection of complex traits. Academic Press, London, UK: pp.255-72

Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillett D, Muselet D, Prud'Homme JF, Dib C, Auffray C, Morissette J, Weissenbach J, Goodfellow PN (1996) A radiation hybrid map of the human genome. *Hum Mol Genet* 5:339-46

Hallgren B (1950) Specific dyslexia ("congenital word blindness"): a clinical and genetic study. *Acta Psychiatr Scand* 65:1-287

Halpern J, Whittemore AS (1999) Multipoint linkage analysis: a cautionary note. *Hum Hered* 49:194-6

Hari R, Kiesila P (1996) Deficit of temporal auditory processing in dyslexic adults. *Neurosci Lett* 205:138-40

Hodge SE, Elston RC (1994) Lods, wrods, and mods: the interpretation of lod scores calculated under different models. *Genet Epidemiol* 11:329-42

Hoehe MR, Caenazzo L, Martinez MM, Hsieh WT, Modi WS, Gershon ES, Bonner TI (1991) Genetic and physical mapping of the human cannabinoid receptor gene to chromosome 6q14-q15. *New Biol* 3:880-5

Holligan C, Johnston RS (1988) The use of phonological information by good and poor readers in memory and reading tasks. *Mem Cognit* 16:522-32

Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, del Bosque-Plata L, Horikawa Y, Oda Y, Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hanis CL, Bell GI (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 26:163-75; erratum 26:502

Hudson T, Stein L, Gerety S, Ma J, Castle A, Silva J, Slonim D, Baptista R, Kruglyak L, Xu S, Hu X, Colbert A, Rosenberg C, Reeve-Daly MP, Rozen S, Hui L, Wu X, Vestergaard C, Wilson K, Bae J, Maitra S, Ganiatsas S, Evans C, DeAngelis M, Ingalls K, Nahf R, Horton L, Oskin M, Collymore A, Ye W, Kouyoumjian V, Zemsteva I, Tam J, Devine R, Courtney D, Renaud M, Nguyen H, O'Connor T, Fizames C, Faure S, Gyapay G, Dib C, Morissette J, Orlin J, Birren B, Goodman N, Weissenbach J, Hawkins T, Foote S, Page D, and Lander E. (1995) An STS-Based Map of the HumanbGenome. *Science* 270:1945-54, with supplementary data from the

Whitehead Institute/MIT Center for Genome Research, Human Genetic Mapping Project, Data Release 12 (July 1997)

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860-921

Ito M (1993) Movement and thought: identical control mechanisms by the cerebellum. *Trends Neurosci* 16:448-50

Jastak S, Wilkinson GS (1984) The wide range achievement test - revised. Jastak Associates, Wilmington, DE

Jin H, Oksenberg D, Ashkenazi A, Peroutka SJ, Duncan AMV, Rozmahel R, Yang Y, Mengod G, Palacios JM, O'Dowd BF (1992) Characterization of the human 5-hydroxytryptamine$_{1B}$ receptor. *J Biol Chem* 267:5735-8

Johnston RS, Rugg MD, Scott T (1987) Phonological similarity effects, memory span and developmental reading disorders: the nature of the relationship. *Br J Psychol* 78:205-11

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58:1347-63

Kruglyak L, Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 57:439-54

Kubova Z, Kuba M, Peregrin J, Novakova V (1996) Visual evoked potential evidence for magnocellular system deficit in dyslexia. *Physiol Res* 45:87-9

LaHoste GJ, Swanson JM, Wigal SB, Glabe C, Wigal T, King N, Kennedy JL (1996) Dopamine D4 receptor gene polymorphism is associated with attention deficit hyperactivity disorder. *Mol Psychiatry* 1:121-4

Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci USA* 84:2363-7

Lander ES, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 11:241-7

Lander ES, Kruglyak L (1996) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 12:357-358

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. *Science* 265:2037-48

Lappalainen J, Dean M, Charbonneau L, Virkkunen M, Linnoila M, Goldman D (1995) Mapping of the serotonin 5-$HT_{1D\beta}$ autoreceptor gene on chromosome 6 and direct analysis for sequence variants. *Am J Med Genet* 60:157-61

Lathrop GM, Lalouel JM (1984) Easy calculations of LOD scores and genetic risks on small computers. *Am J Hum Genet* 36:460-5

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus analysis in humans. *Proc Natl Acad Sci USA* 81:3443-6

Lathrop GM, Lalouel JM, White RL (1986) Construction of human genetic linkage maps: likelihood calculations for multilocus analysis. *Genet Epidemiol* 3:39-52

Lehmkuhle S. Garzia RP, Turner L, Hash T, Baro JA (1993) A defective visual pathway in children with reading disability. *N Engl J Med* 328:989-96

Leiner HC, Leiner AL, Dow RS (1993) Cognitive and language functions of the human cerebellum. *Trends Neurosci* 16:444-7; comment 16:453-4

Lerner JW (1989) Educational interventions in learning disabilities. *J Am Acad Child Adolesc Psychiatry* 28:326-31

Levitt P, Harvey JA, Friedman E, Simansky K, Murphy EH (1997) New evidence for neurotransmitter influences on brain development. *Trends Neurosci* 20:269-74

Levy FO, Holtgreve-Grez H, Taskén K, Solberg R, Ried T, Gudermann T (1994) Assignment of the gene encoding the 5-HT_{1E} serotonin receptor (S31) (locus HTR1E) to human chromosome 6q14-q15. *Genomics* 22:637-40

Lewitter FI, DeFries JC, Elston RC (1980) Genetic models of reading disability. *Behav Genet* 10:9-30

Livingstone MS, Rosen GD, Drislane FW, Galaburda AM (1991) Physiological and anatomical evidence for a magnocellular defect in developmental dyslexia. *Proc Natl Acad Sci USA* 88:7943-7

Lovegrove WJ, Bwoling A, Badcock B, Blackwood M (1980) Specific reading disability: differences in contrast sensitivity as a function of spatial frequency. *Science* 210:439-40

MacLean CJ, Martin RB, Sham PC, Wang H, Straub RE, Kendler KS (2000) The trimmed-haplotype test for linkage disequilibrium. *Am J Hum Genet* 66:1062-75

Martin ER, Lai EH, Gilbert JR, Rogala AR, Afshari AJ, Riley J, Finch KL, Stevens JF, Livak KJ, Slotterbeck BD, Slifer SH, Warren LL, Conneally PM, Schmechel DE, Purvis I, Pericak-Vance MA, Roses AD, Vance JM (2000a) SNPing away at complex diseases: analysis of single-nucleotide polymorphisms around APOE in Alzheimer disease. *Am J Hum Genet* 67:383-94

Martin ER, Menold MM, Wolpert CM, Bass MP, Donnelly SL, Ravan SA, Zimmerman A, Gilbert JR, Vance JM, Maddox LO, Wright HH, Abramson RK, DeLong GR, Cuccaro ML, Pericak-Vance MA (2000) Analysis of linkage disequilibrium in gamma-aminobutyric acid receptor subunit genes in autistic disorder. *Am J Med Genet* 96:43-8

Matise TC, Perlin M, Chakravarti A (1994) Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. *Nat Genet* 6:384-90

Mattson MP (1988) Neurotransmitters in the regulation of neuronal cytoarchitecture. *Brain Res Rev* 13:179-212

McAnally KI, Stein JF (1996) Auditory temporal coding in dyslexia. *Proc R Soc Lond B* 263:961-5

McCracken JT, Smalley SL, McGough JJ, Crawford L, Del'Homme J, Cantor RM, Liu A, Nelson SF (2000) Evidence for linkage of a tandem duplication polymorphism upstream of the dopamine D4 receptor gene (*DRD4*) with attention deficit hyperactivity disorder (ADHD). *Mol Psychiatry* 5:531-6

Meier E, Hertz L, Schousboe A (1991) Neurotransmitters as developmental signals. *Neurochem Int* 19:1-15

Merigan WH and Maunsell JHR (1993) How parallel are the primate visual pathways? *Ann Rev Neurosci* 16:369-402

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215

Moats LC (1995) Spelling: developmental disability and instruction. York Press, Baltimore, MD

Morris DW, Robinson L, Turic D, Duke M, Webb V, Milham C, Hopkin E, Pound K, Fernando S, Easton M, Hamshere M, Williams N, McGuffin P, Stevenson J, Krawczak M, Owen MJ, O'Donovan MC, Williams J (2000a) Family-based association mapping provides evidence for a gene for reading disability on chromosome 15q. *Hum Mol Genet* 9:843-8

Morris DW, Robinson L, Turic D, Duke M, Webb V, Southgate L, Hamshere M, Williams N, McGuffin P, Rowen L, Stevenson J, Krawczak M, Owen MJ, Williams J, O'Donovan MC (2000b) Linkage disequilibrium and positional candidate gene analyses of a susceptibility locus for reading disability on chromosome 15q. *Am J Med Genet* 96:471

Morton NE (1955) Sequential tests for the detection of linkage. *Am J Hum Genet* 7:277-318

Murray JC, Buetow KH, Weber JL, Ludwigsen S, Scherpbier-Heddema T, Manion F, Quillen J, Sheffield VC, Sunden S, Duyk GM, Weissenbach J, Gyapay G, Dib C, Morrissette J, Lathrop GM, Vignal A, White F, Matsunami N, Gerken S, Melis R, Albertsen H, Plaetke R, Odelberg S, Ward D, Dausset J, Cohen D, Cann H (1994) A comprehensive human linkage map with centimorgan density. *Science* 265:2049-54

Nemeroff CB (1998) The neurobiology of depression. *Sci Am* 28:42-9

Nezu J, Oku A, Jones MH, Shimane M (1997) Identification of two novel human putative serine/threonine kinases, VRK1 and VRK2, with structural similarity to vaccinia virus B1R kinase. *Genomics* 45:327-31

Nicolson RI, Fawcett AJ (1990) Automaticity: a new framework for dyslexia research? *Cognition* 35:159-82

Nicolson RI, Fawcett AJ, Berry EL, Jenkins IH, Dean P, Brooks DJ (1999) Association of abnormal cerebellar activation with motor learning difficulties in dyslexic adults. *Lancet* 353:1662-7

Nöthen MM, Erdmann J, Shimron-Abarbanell D, Propping P (1994) Identification of genetic variation in the human serotonin 1Dß receptor gene. *Biochem Biophys Res Commun* 205:1194-200

Nöthen MM, Schulte-Korne G, Grimm T, Cichon S, Vogt IR, Muller-Myhsok B, Propping P, Remschmidt H (1999) Genetic linkage analysis with dyslexia: evidence for linkage of spelling disability to chromosome 15. *Eur Child Adolesc Psychiatry* 8 Suppl 3:56-9

Olson R, Wise V, Conners F, Rack J, Fulker D (1989) Specific deficits in component reading and language skills: genetic and environmental influences. *J Learning Disabil* 22:339-48

Oruc L, Verheyen GR, Furac I, Ivezic S, Jakovljevic M, Raeymaekers P, Van Broeckhoven C (1997) Positive association between the GABRA5 gene and unipolar recurrent major depression. *Neuropsychobiology* 36:62-4

Ott J (1991) Analysis of human genetic linkage, 2nd edition. Johns Hopkins University Press, Baltimore, MD

——— (1999) Analysis of human genetic linkage, 3rd edition. Johns Hopkins University Press, Baltimore, MD

Papadimitriou GN, Dikeos DG, Karadima G, Avramopoulos D, Daskalopoulou EG, Vassilopoulos D, Stefanis CN (1998) Association between the GABA(A) receptor alpha5 subunit gene locus (GABRA5) and bipolar affective disorder. *Am J Med Genet* 81:73-80

Peltonen L, Palotie A, Lange K (2000) Use of population isolates for mapping complex traits. *Nat Rev Genet* 1:182-90

Pennington BF (1997) Using genetics to dissect cognition. *Am J Hum Genet* 60:13-6

Pennington BF, Gilger JW, Pauls D, Smith SA, Smith SD, DeFries JC (1991) Evidence for major gene transmission of developmental dyslexia. *JAMA* 266:1527-34

Pennington BF, Smith SD, Kimberling WJ, Green PA, Haith MM (1987) Left-handedness and immune disorders in familial dyslexics. *Arch Neurol* 44:634-9

Philippe A, Martinez M, Guilloud-Bataille M, Gillberg C, Rastam M, Sponheim E, Coleman M, Zappella M, Aschauer H, Van Maldergem L, Penet C, Feingold J, Brice A, Leboyer M (1999) Genome-wide scan for autism susceptibility genes. *Hum Mol Genet* 8:805-12

Pratt SC, Daly MJ, Kruglyak L (2000) Exact multipoint quantitative-trait linkage analysis in pedigrees by variance components. *Am J Hum Genet* 66:1153-7

Rabin M, Wen XL, Hepburn M, Lubs HA, Feldman E, Duara R (1993) Suggestive linkage of developmental dyslexia to chromosome 1p34-p36. *Lancet* 342:178

Rao DC (2001) Genetic dissection of complex traits: an overview. In Rao CD, Province MA (eds.) Genetic dissection of complex traits. Academic Press, London, UK: pp.13-34

Rao DC, Gu C (2001) False positives and false negatives in genome scans. In Rao CD, Province MA (eds.) Genetic dissection of complex traits. Academic Press, London, UK: pp.487-98

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. *Nature* 411:199-204

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. *Science* 273:1516-7

Rosner J, Simon DP (1971) The auditory analysis test: an initial report. *J Learn Dis* 4:382-92

Rowe DC, Stever C, Giedinghagen LN, Gard JM, Cleveland HH, Terris ST, Mohr JH, Sherman S, Abramowitz A, Waldman ID (1998) Dopamine DRD4 receptor polymorphism and attention deficit hyperactivity disorder. *Mol Psychiatry* 3:419-26

Rumsey JM, Nace K, Donohue B, Wise D, Maisog JM, Andreason P (1997) A positron emission tomographic study of impaired word recognition and phonological processing in dyslexic men. *Arch Neurol* 54:562-73

S.A.G.E. (1997) Statistical Analysis for Genetic Epidemiology, Release 3.1. Computer program package available from the Department of Epidemiology and Biostatistics, Rammelkamp Center for Education and Research, MetroHealth Campus, Case Western Reserve University, Cleveland

SAS Institute (1990) SAS release 6.4. Cary, NC.

Sawyer DL, Krishnamani RS, Hannig VL, Farcia M, Kim JK, Haines JL, Phillips III JA (1998) Genetic analysis of phonologic core deficit dyslexia (PCDD). *Am J Hum Genet* 63:A307

Schaffer AA, Gupta SK, Shriram K, Cottingham Jr. RW (1994) Avoiding recomputation in genetic linkage analysis. *Hum Hered* 44:225-37

Schulte-Korne G, Deimel W, Bartling J, Remschmidt H (1998a) Auditory processing and dyslexia: evidence for a specific speech processing deficit. *NeuroReport* 9:337-40

Schulte-Körne G, Grimm T, Nöthen MM, Müller-Myhsok B, Cichon S, Vogt IR, Propping P, Remschmidt H (1998b) Evidence for linkage of spelling disability to chromosome 15. *Am J Hum Genet* 63:279-82

Seidenberg MS, McClelland JL (1989) A distributed developmental model of word recognition and naming. *Psychol Rev* 96:447-52

Shapleske J, Rosselll SL, Woodruff PWR, David AS (1999) The planum temporale: a systematic, quantitative review of its structure, functional and clinical significance. *Brain Res Rev* 29:26-49

Shaywitz BA, Fletcher JM, Shaywitz SE (1995) Defining and classifying learning disabilities and attention-deficit/hyperactivity disorder. *J Child Neurol* 10:S50-7

Shaywitz SE, Shaywitz BA, Fletcher JM, Escobar MD (1990) Prevalence of reading disability in boys and girls. *JAMA* 264:998-1002

Shaywitz SE, Shaywitz BA, Pugh KR, Fulbright RK, Constable RT, Mencl WE, Shankweiler DP, Liberman AM, Skudlarski P, Fletcher JM, Katz L, Marchione KE, Lacadie C, Gatenby C, Gore JC (1998) Functional disruption in the organization of the brain for reading in dyslexia. *Proc Natl Acad Sci USA* 95:2636-41

Shimron-Abarbanell D, Nöthen MM, Erdmann J, Propping P (1995) Lack of genetically determined structural variants of the human serotonin-1E (5-HT_{1E}) receptor protein points to its evolutionary conservation. *Brain Res Mol Brain Res* 29:387-90

Simos PG, Breier JI, Fletcher JM, Foorman BR, Bergman E, Fishbeck K, Papanicolaou AC (2000) Brain activation profiles in dyslexic children during non-word reading: a magnetic source imaging study. *Neurosci Lett* 290:61-5

Smalley SL, Bailey JN, Palmer CG, Cantwell DP, McGough JJ, Del'Homme MA, Asarnow JR, Woodward JA, Ramsey C, Nelson SF (1998) Evidence that the dopamine D4 receptor is a susceptibility gene in attention deficit hyperactivity disorder. *Mol Psychiatry* 3:427-30; erratum 4:100

Smith SD, Kelley PM, Brower AM (1998) Molecular approaches to the genetic analysis of specific reading disability. *Hum Biol* 70:239-56

Smith SD, Kimberling WJ, Pennington BF (1991) Screening for multiple genes influencing dyslexia. *Read Writ Interdisc* 3:285-98

Smith SD, Kimberling WJ, Pennington BF, Lubs HA (1983) Specific reading disability: identification of an inherited form through linkage analysis. *Science* 219:1345-7

Smith SD, Kimberling WJ, Shugart YY, Ing PS, Pennington BF (1989) Genetic linkage analysis of 20 families with specific reading disability. *Am J Hum Genet* 45:A65

Sobel E, Lange K (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am J Hum Genet* 58:1323-37

Spreen O (1988) Learning disabled children growing up: a follow-up into adulthood. Oxford University Press, Oxford, UK

Stanovich KE (1988) Explaining the differences between the dyslexic and the garden-variety poor reader: the phonological-core variable-difference model. *J Learn Dis* 21:590-604

Stein J and Walsh V (1997) To see but not to read; the magnocellular theory of dyslexia. *Trends Neurosci* 20:147-52

Stevenson J, Graham P, Fredman G, McLoughlin V (1987) A twin study of genetic influences on reading and spelling ability and disability. *J Child Psychol Psychiatry* 28:229-47

Sunohara GA, Roberts W, Malone M, Schachar RJ, Tannock R, Basile VS, Wigal T, Wigal SB, Schuck S, Moriarty J, Swanson JM, Kennedy JL, Barr CL (2000) Linkage of the dopamine D4 receptor gene and attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 39:1537-42

Swanson JM, Sunohara GA, Kennedy JL, Regino R, Fineberg E, Wigal T, Lerner M, Williams L, LaHoste GJ, Wigal S (1998) Association of the dopamine receptor D4 (DRD4) gene with a refined phenotype of attention deficit hyperactivity disorder (ADHD): a family-based approach. *Mol Psychiatry* 3:38-41

Tallal P (1980) Auditory temporal perception, phonics, and reading disabilities in children. *Brain Lang* 9:182-98

Tallal P, Miller S, Fitch RH (1993) Neurobiological basis of speech: a case for the preeminance of temporal processing. *Ann N Y Acad Sci* 682:27-47

Terwilliger JD, Shannon WD, Lathrop GM, Nolan JP, Goldin LR, Chase GA, Weeks DE (1997) True and false positive peaks in genomewide scans: applications of length-biased sampling to linkage mapping. *Am J Hum Genet* 61:430-8

Thomson G (1995) Mapping disease genes: family-based association studies. *Am J Hum Genet* 57:487-98

Torgesen JK, Wagner RK, Rashotte CA (1994) Longitudinal study of phonological processing in reading. *J Learning Disabil* 27:276-86

Turic D, Morris DW, Robinson L, Duke M, Grierson A, Raha-Chowdhury R, Southgate L, Webb V, Hamshere M, Williams B, McGuffin P, Stevenson J, Krawczak M, Owen MJ, O'Donovan MC, Williams J (2000) Linkage disequilibrium mapping provides evidence for a susceptibility gene for reading disability on chromosome 6p. *Am J Med Genet* 96:556

Van Orden GC, Pennington BF, Stone GO (1990) Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychol Rev* 97:488-522

Waldman ID, Rowe DC, Abramowitz A, Kozel ST, Mohr JH, Sherman SL, Cleveland HH, Sanders ML, Gard JMC, Stever C (1998) Association and linkage of the dopamine transporter gene and attention-deficit hyperactivity disorder in children: heterogeneity owing to diagnostic subtype and severity. *Am J Hum Genet* 63:1767-1776

Wechsler D (1974) Manual for the Wechsler intelligence scale for children - revised. Psychological Corp., New York, NY

——— 1981. Wechsler adult intelligence scale - revised. Psychological Corp., New York, NY

Whittemore AS, Halpern J (2001) Problems in the definition, interpretation, and evaluation of genetic heterogeneity. *Am J Hum Genet* 68:457-465

Wijsman EM, Amos CI (1997) Genetic analysis of simulated oligogenic traits in nuclear and extended pedigrees: summary of GAW10 contributions. *Genet Epidemiol* 14:719-35

Willcutt EG, Pennington BF (2000) Psychiatric comorbidity in children and adolescents with reading disability. *J Child Psychol Psychiat* 41:1039-48

Willcutt EG, Pennington BF, DeFries JC (2000) Twin study of the etiology of comorbidity between reading disability and attention-deficit/hyperactivity disorder. *Am J Med Genet* 96:293-301

Witte JS, Elston RC, Schork NJ (1996) Genetic dissection of complex traits. *Nat Genet* 12:355-6

Witton C, Talcott JB, Hansen PC, Richardson AJ, Griffiths TD, Rees A, Stein JF, Green GGR (1998) Sensitivity to dynamic auditory and visual stimuli predicts nonword reading ability in both dyslexic and normal readers. *Curr Biol* 8:791-7

Wolf M (1999) What time may tell: towards a new conceptualization of developmental dyslexia. *Ann Dyslexia* 49:3-28

Wolf M, Bally H, Morris R (1986) Automaticity, retrieval processes, and reading: a longitudinal study in average and impaired readers. *Child Dev* 57:988-1000

Wolff PH (1993) Impaired temporal resolution in developmental dyslexia. *Ann N Y Acad Sci* 682:87-103

Woodcock RW (1987) Woodcock reading mastery tests. American Guidance Service, Circle Pines, MI

Woodcock RW, Johnson MB (1989) Woodcock-Johnson psychoeducational battery - revised. DLM Teaching Resources, Allen, TX

Xu J, Meyers DA, Pericak-Vance MA (1998) Lod score analysis. In Haines JL, Pericak-Vance MA (eds.) Approaches to gene mapping in complex human diseases. Wiley-Liss, New York, NY: pp.253-272

Yap RL, van der Leij A (1994) Testing the automatization deficit hypothesis of dyslexia via a dual-task paradigm. *J Learning Disabil* 27:660-5

Zahalkova M, Vrzal V, Kloboukova E (1972) Genetical investigations in dyslexia. *J Med Genet* 9:48-52

Zametkin AJ, Liotta W (1998) The neurobiology of attention-deficit/hyperactivity disorder. *J Clin Psychiatry* 59 Suppl 7:17-23