GENETIC ANALYSIS OF DYSLEXIA

by

JORDANA TZENOVA

B.Sc. Honours, McGill University 1999

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

Department of Medical Genetics

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 2002

© Jordana Tzenova 2002
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Medical Genetics

The University of British Columbia
Vancouver, Canada

Date August 23, 2002
ABSTRACT

Dyslexia is a complex trait, which manifests as a specific reading and spelling disability that is independent of general intelligence and educational opportunity. A genetic basis for the trait has been clearly established, leading to the identification of eight dyslexia susceptibility loci to date. The aim of this thesis is to gain more insight into the genetic mechanisms underlying dyslexia by looking at variation at a single locus, variation at multiple loci, and by examining a phenotype that is thought to be co-morbid with the trait in a sample of 100 dyslexia families. The results of the first part of the study indicate evidence for linkage of dyslexia to chromosome 1p34-p36. This finding replicates the results from two previous reports, suggesting that a locus in this region may contribute to a significant proportion of dyslexia cases. The results of the second part of the study indicate evidence for the involvement of a multilocus genetic system in dyslexia. Both previously identified dyslexia susceptibility loci that do not have significant single-locus effects on the trait in our sample, and novel candidate gene regions show evidence for gene-interactions and locus heterogeneity. Finally, an examination of handedness genetics in dyslexia families was unable to replicate the findings of a recent genome-wide scan. There was no evidence to suggest that dyslexia susceptibility loci are involved in the genetic component of handedness, however the results implicate a novel region, 6q25.3, in relative hand skill determination.
# TABLE OF CONTENTS

**ABSTRACT** ................................................................................................................................... ii

**LIST OF TABLES** ............................................................................................................................ v

**LIST OF FIGURES** .......................................................................................................................... vii

**ACKNOWLEDGMENTS** .................................................................................................................... viii

**CHAPTER I INTRODUCTION** ......................................................................................................... 1

1.1 DYSLEXIA ................................................................................................................................. 1

1.2 NEUROLOGICAL STUDIES OF DYSLEXIA ................................................................................. 4

1.3 EVIDENCE FOR A GENETIC BASIS IN DYSLEXIA ............................................................... 6

1.4 GENETIC DISSECTION OF COMPLEX TRAITS ..................................................................... 10

1.5 AIM OF THE STUDY .................................................................................................................... 13

**CHAPTER II GENERAL METHODS** ............................................................................................. 15

2.1 SUBJECTS ..................................................................................................................................... 15

2.2 PHENOTYPE DATA .................................................................................................................... 18

2.3 DNA ANALYSIS ........................................................................................................................ 21

2.4 GENERAL STATISTICAL ANALYSES ....................................................................................... 26

**CHAPTER III CONFIRMATION OF A DYSLEXIA SUSCEPTIBILITY GENE**

ON CHROMOSOME 1p ......................................................................................................................... 30

3.1 INTRODUCTION ........................................................................................................................ 30

3.2 METHODS .................................................................................................................................... 32
3.3 RESULTS ........................................................................................................ 35
3.4 DISCUSSION ................................................................................................ 43

CHAPTER IV EVIDENCE FOR LOCUS INTERACTION AND
HETEROGENEITY IN DYSLEXIA ........................................................................ 48

4.1 INTRODUCTION .......................................................................................... 48
4.2 METHODS ..................................................................................................... 49
4.3 RESULTS ....................................................................................................... 53
4.4 DISCUSSION .................................................................................................. 69

CHAPTER V THE SEARCH FOR HANDEDNESS GENES IN FAMILIES WITH
DYSLEXIA .......................................................................................................... 76

5.1 INTRODUCTION .......................................................................................... 76
5.2 METHODS ..................................................................................................... 77
5.3 RESULTS ....................................................................................................... 83
5.4 DISCUSSION .................................................................................................. 86

CHAPTER VI CONCLUSION .............................................................................. 89

REFERENCES .................................................................................................... 91
LIST OF TABLES

Table 1. Description of the different family sets used for analysis by gene-mapping programs. ................................................................. 17

Table 2. Descriptive statistics of the quantitative measures in the sample of (a) all individuals from the 100 dyslexia families and (b) 112 control children. .......... 20

Table 3. A description of the microsatellite markers genotyped on the 100 families. ..... 23

Table 4. Map positions for the 12 markers on chromosome 1p34-p36. ....................... 33

Table 5. Results from single point linkage analysis of dyslexia........................................ 37

Table 6. Results from the family-based association tests using the qualitative phenotype. ........................................................................................................... 40

Table 8. Genetic models used in single locus parametric linkage analysis. .................. 50

Table 9. A list of 33 markers used in GHP analyses, from the 15 regions which indicated suggestive or significant evidence for linkage or association in the set of 100 dyslexia families. ........................................................................................................... 55

Table 10. A list of the 22 chromosomal regions included in the GHP analyses.............. 58

Table 11. Pairs of regions that show significant evidence for epistasis in GHP, by considering the increase in the weighted NPL score, using either the proportional or the discrete weighting schemes. These regions show significant (p<0.01) evidence for epistasis in the increase of the weighted NPL score compared to the single-locus NPL score, for locus 1 (by weighting linkage to locus 2); as well as significant (p<0.05) evidence for epistasis for locus 2 (by weighting linkage to locus 1) ........ 61

Table 12. Pairs of regions that show significant evidence for heterogeneity in GHP, by considering the increase in the weighted NPL score, using either the proportional or
the discrete weighting schemes. These regions show significant (p<0.01) evidence for heterogeneity in the increase of the weighted NPL score compared to the single-locus NPL score, for locus 1 (by weighting linkage to locus 2); as well as significant (p<0.05) evidence for heterogeneity for locus 2 (by weighting linkage to locus 1).

Table 13. Description of handedness status information in the Dyslexia families........... 78

Table 14. Genetic models used in parametric linkage analysis. ........................................... 82

Table 15. A list of all markers with Lod scores greater than 1 for the 47 dyslexia families.
.................................................................................................................................................. 84

Table 16. A list of all markers with Sibpal Linear Regression p-values less than 0.05. .. 85
LIST OF FIGURES

Figure 1. Relative map position of markers used in three studies examining evidence for linkage to chromosome 1p. ................................................................. 31

Figure 2. Multipoint parametric and non-parametric linkage analysis results using the qualitative phenotype. ................................................................................ 38

Figure 3. Multipoint quantitative linkage analysis of dyslexia................................. 42

Figure 4. A schematic of the different clusters formed by GHP epistatic analysis. .... 66

Figure 5. Distribution of handedness codes in the 100 dyslexia families and 47 dyslexia families with at least one left-handed and one right-handed individual. .......... 80
ACKNOWLEDGMENTS

I thank my supervisor, L. Leigh Field, for giving me direction when I needed it and letting me work independently the rest of the time. I acknowledge Bonnie J. Kaplan and Susan Crawford for the phenotypic data and help that they provided. I also thank Wendy Robinson for the lively discussions about science and for keeping the journal club alive. Carolyn Brown provided invaluable advice. I am grateful to Jan Friedman for showing me how to work really hard and giving me an opportunity to present my first talk. I thank Sally Otto for giving me a different perspective on genetics and improving my confidence. I am very grateful to all the members of the Field lab for their help and support. I want to particularly thank Ella and Malgosia for their hard work and having to put up with my nagging for long months. Karla and Zheng Wang were always there to help. Robin pushed me to work hard. Shao, Ravi and Mona helped and cheered me up when I needed it. I would really like to thank Tracey Petryshen as well for all the direction and help that she provided long distance.

On the personal side, I couldn't have done it without Tom. Thanks to Ladan, Emily, Anna and Angus for keeping me sane all this time. The journal club group, the soccer team, the volleyball bunch, Simon, Jen and many others helped me along the way. And finally, I would really like to thank my family.
CHAPTER I INTRODUCTION

1.1 Dyslexia

Developmental dyslexia, first described by Morgan (1896), is a severe difficulty in reading that is independent of general intelligence, educational opportunity, and visual acuity. Dyslexia is defined as a specific reading and spelling disability, but is often associated with problems in writing, composition and other language-related functions. Developmental dyslexia affects approximately 3%-10% of school-age children (Yule and Rutter 1976), although estimates vary depending on the population and the rigor of the diagnosis. Thus, dyslexia is one of the most common childhood learning disorders (Lerner 1989), accounting for 80% of all learning-disability cases (Shaywitz 1998). Dyslexia must be differentiated from reading and spelling disabilities caused by mental retardation, emotional disturbance, lack of opportunity and neurological disorders that include impaired vision, impaired hearing and cerebral palsy (Warnke 1999).

Longitudinal studies indicate that dyslexia is a persistent condition, and not a transient developmental lag (Shaywitz 1998). Most adults compensate for the disorder by visual memory and other means. A higher incidence of dyslexia among males (Critchley 1970) has been observed, however recent reports have found an equal gender ratio (Shaywitz et al. 1990). A number of studies have focused on defining the dyslexia phenotype and examining evidence for a neurological, physiological and genetic basis in the trait.

Dyslexia is associated with several behavioral and physiological conditions. Attention deficit hyperactivity disorder (ADHD) is highly co-morbid with dyslexia (Wilcutt et al. 2000). Anxiety and depression also occur more frequently in dyslexic
individuals than in the general population (Gilger et al. 1992). In addition, parents of children with autism have increased frequency of reading problems (Folstein et al. 1999). A significant proportion of the sons of mothers with Systemic Lupus Erythematosus (SLE) have dyslexia (Lahita 1988). Dyslexic individuals are also more likely to have an increased number of immune deficits (Pennington et al. 1987). Finally, the incidence of left-handed individuals is thought to be higher among dyslexics compared to the general population (Eglinton and Annett 1994), although this finding remains controversial (Crawford et al. 1994).

Dyslexia is defined as a reading disability, however diagnosis is complicated by lack of a generally accepted clinical definition or a standard protocol of measures for diagnosis. Unfortunately, studies use a number of different psychometric tests that measure separate components of reading. To obtain the qualitative dyslexia phenotype individuals are typically scored as dyslexic if their test scores for reading performance are below those expected for their age level. A quantitative dyslexia phenotype may also be derived from the age-standardized test scores for each of the different components of reading.

It is difficult to dissect reading into separate components because of the high correlation between such measures. However, cognitive studies have identified several developmental dyslexia subtypes, which are thought to separate the language processes that contribute to dyslexia (Raskind et al. 2000). Phonological coding dyslexia (PCD), in which some aspect of phonological processing is impaired, is the most common form of dyslexia (Olson et al. 1989). Orthological or orthographic skill problems (Olson et al. 1989) are present in a smaller subset of dyslexic patients. Finally, rapid automatized
naming (RAN; Denckla and Rudel 1976), representing lexical retrieval, is argued to be either part of phonological coding dyslexia or a separate process. In addition, vocabulary and memory skills also contribute to reading and therefore to dyslexia.

There is currently a consensus among investigators in the field that the underlying problem in most dyslexic individuals (80-90%) is a phonological processing deficiency. Phonological coding dyslexia is comprised of several processes, which include phoneme awareness, the ability to decompose words into phonemes, phonological coding, the ability to link phonemes to graphemes, and RAN, the ability to automatize these skills. Most cognitive studies agree with the phonological deficit hypothesis of dyslexia (Stanovich 1988; Frith 1998), which states that a biological abnormality in the brain results in a specific phonological deficiency, which is manifested in problems with phoneme awareness, coding, RAN, and reading comprehension.

In support of the phonological deficit hypothesis, estimates of the incidence of dyslexia vary by language, apparently mirroring the phonemic complexity of language (Miles 2000). The percentage of dyslexic school age children is quite low in Japan and China (1%), higher in Germany (5%), and much higher in English-speaking countries (20%), with an overall median of 7% (Grigorenko 2001). Several hypotheses have been proposed to account for the variation, taking into account differences in culture, educational practices, and phonological depth of language. It has been proposed that the mechanism leading to dyslexia is universal in all languages and is related to a person's phonological skills; however, the manifestation of the phenotype varies, reflecting the phonological demands of the linguistic system (Grigorenko 2001).
1.2 Neurological Studies of Dyslexia

Neurological studies using brain morphology (Rumsey et al. 1986), postmortem brain specimens (Galaburda and Kemper 1979), electrophysiology (Kraus et al. 1996), and functional brain imaging (Rumsey et al. 1992) indicate a neuro-developmental basis for the disorder. Comparisons between the brains of dyslexic individuals and controls using the above techniques indicate a number of functional and morphological differences in several brain regions and pathways. Postmortem brain specimens and brain morphology studies consistently indicate unusual symmetry, or reversed asymmetry in the planum temporale, the area of the brain which is thought to be involved in language function (Cohen et al. 1989; Larsen et al. 1990). Functional brain imaging studies indicate that cortical regions around the left hemispheric language area exhibit altered patterns of activation during reading in dyslexic individuals compared to controls (Paulesu et al. 2001). Thus, abnormal brain lateralization and left hemispheric language area functions have been implicated in the etiology of dyslexia.

A number of studies have shown that dyslexic individuals have a deficit in the magnocellular visual pathway, which processes gross detail stimuli, but not in the parvocellular pathway, which processes colour and fine detail stimuli (Merigan and Maunsell 1993; Witton et al. 1998). In addition, several studies have shown that dyslexics have deficits in auditory processing, which likely arise through a system similar to the magnocellular pathway in visual processing (Tallal et al. 1993; Galaburda et al. 1994). These results are summarized in the magnocellular hypothesis of dyslexia, which contends that reading problems derive from impaired processing, caused by abnormal auditory (Tallal et al. 1993) and visual (Eden et al. 1996) magnocellular pathways.
Dyslexic individuals also appear to have problems with balance (Yap and van der Leij 1994), muscle tone (Fawcett and Nicolson 1999), coordination (Fawcett and Nicolson 1999), and general automatizing skills, not necessarily limited to literacy (Nicolson and Fawcett 1990). The recently proposed cerebellar deficit hypothesis states that a deficit in cerebellar function accounts for these symptoms, and may also account for some of the language related deficiencies (Nicolson et al. 2001). This hypothesis remains contentious, because there is little behavioral overlap between cerebellar syndromes and dyslexia (Zefiro and Eden 2001).

A third hypothesis, the temporal deficit hypothesis, attempts to unite the phonological, auditory, visual and motor skill deficits in dyslexic individuals. This theory states that such deficits arise from the same underlying biological mechanism related to the poor perception of rapidly changing sensory stimuli (Tallal and Piercy 1973; Tallal 1980). According to this theory, dyslexic individuals would not be able to perceive rapid elements of speech (phonemes), which would cause phonological deficits and lead to reading difficulties.

The theories proposed to explain the neurological basis of dyslexia are not mutually exclusive. It is possible that the proposed neurological deficits occur in subgroups of patients and independently cause the dyslexic phenotype, or that mechanisms of different degrees of severity occur together in the same individuals and lead to the dyslexia in combination with each other. For example, it has been suggested that magnocellular neuronal cell lines involved in processing temporal stimuli in all sensory systems are developmentally impaired in dyslexic individuals, and the biological
mechanisms underlying these deficits are due to specific genetic variants (Stein and Walsh 1997; Stein 2001).

### 1.3 Evidence for a Genetic Basis in Dyslexia

Dyslexia has a clear genetic basis, evidence for which comes from population, family, and twin studies. The risk of dyslexia in first-degree relatives of probands exceeds the risk in the general population (Hallgren 1950; Pennington et al. 1991). The risk and the severity of the trait increase with the number of affected parents (Wolf and Melngailis 1994). The monozygotic twin concordance rate has been estimated to be 68%, which is higher than that of dizygotic twins, 38% (DeFries et al. 1987; Stevenson et al. 1987), which suggest that both genetic and environmental factors affect the trait. In addition, monozygotic twins are more similar for continuous measures of dyslexia than dizygotic twins (Olson et al. 1989; Bishop et al. 1996). Heritability estimates for the trait vary from 0.51 for reading comprehension to 0.73 for spelling, and 0.93 for phonological coding (Stevenson et al. 1987), while the orthographic component of dyslexia was found to be non-heritable (Olson et al. 1989).

The mode of inheritance of dyslexia has not been clearly established. Several studies have suggested autosomal dominant, recessive, and polygenic genetic models of inheritance, including genetic heterogeneity (Field and Kaplan 1998). Segregation analysis of dyslexia has been performed on both dyslexic and control family sets by several independent groups. Results from analyses using the qualitative phenotype indicate evidence for a major gene mode of transmission of dyslexia (Pennington et al. 1991), while analyses using quantitative phenotypic measures indicate evidence for one (Lewitter et al. 1980; Gilger et al. 1994) or a few (Wijsman et al. 2000) major genes and
a polygenic component (Lewitter et al. 1980) in the trait. Results indicative of a major
gene mode of transmission could not resolve the mode of inheritance, presenting
evidence for autosomal recessive (Lewitter et al. 1980), dominant (Hallgren 1950;
Pennington et al. 1991), and additive (Lewitter et al. 1980) mechanisms, depending on
the family sample and the mode of ascertainment. In addition, one of the samples studied
(Lewitter et al. 1980) indicates that both genetic and environmental factors are likely to
contribute to the trait. Recent analyses estimating the number of quantitative trait loci
(QTLs) underlying the trait, report 2.4 QTLs involved in the non-word repetition
component and 1.9 QTLs involved in the digit span component of dyslexia (Wijsman et
al. 2000). Dyslexia thus appears to be a complex disorder affected by a number of
susceptibility genes and environmental factors.

Various reports have identified several different chromosomal regions involved in
dyslexia. Most studies use either a qualitative definition of dyslexia or quantitative
reading measures; however inconsistencies in the clinical definition of the phenotype
complicate comparisons across studies (see section 1.4). To date, eight chromosomal
regions have been implicated in dyslexia using parametric and non-parametric linkage
analysis on chromosomes 1p34-p35, 2p15-p16 (DYX3), 3p12-q13, 6p23-p21.3 (DYX2),
6q11.2-q12 (DYX4), 11p15.5, 15q21.1 (DYX1), and 18p11. The first susceptibility locus
for dyslexia was identified on chromosome 15 (Smith et al. 1983) in a sample of nine
families. The results could not be replicated by an independent study of five Danish
families (Bisgaard et al. 1987). Smith et al. (1991) used additional families and found
significant linkage to a region (DYX1), distal to the one previously identified, suggesting
the presence of two dyslexia susceptibility loci on chromosome 15q. Linkage to DYX1
was confirmed by Grigorenko et al. (1997) for the single word reading component of dyslexia in six extended families. Later studies confirmed the results for DYX1 for the spelling component of dyslexia in two independent sets of seven families each (Schulte-Korne et al. 1998; Nothen et al. 1999). Finally, a recent linkage disequilibrium study also confirms the results (Morris et al. 2000).

Subsequently, Rabin et al. (1993) reported suggestive linkage of dyslexia to chromosome 1p34-p36 in nine families. Although these results have been replicated in a sample of eight families (Grigorenko et al. 2001), they have not been replicated in other studies (Smith et al. 1998; Fisher et al. 2002).

Most dyslexia gene-mapping studies have shown significant linkage to one region on chromosome 6p, DYX2, located near the HLA region. Smith et al. (1989) first identified this locus, obtaining weak evidence for linkage in a set of 20 families. Cardon et al. (1994, 1995) conducted a search for QTLs in dyslexia and obtained significant results for DYX2 in 19 families. Grigorenko et al. (1997, 2000) also replicated the chromosome 6p findings in a total set of eight families for several quantitative reading measures. Fisher et al. (1999) replicated the results in a set of 82 nuclear families using quantitative measures of the trait. Finally, Gayan et al. (1999) replicated the linkage to chromosome 6p in a set 79 families, using QTL analyses. The two implications of these findings are first, that chromosome 6p harbors a major susceptibility locus for dyslexia and second, that this locus may be directly related to the higher incidence of immune deficits observed in dyslexic individuals relative to the general population. Turic et al. (2000) examined the second possibility and excluded it, by obtaining a linkage peak distal to the HLA region. However, several reports have not been able to replicate the
chromosome 6p linkage results using both qualitative (Field and Kaplan 1998; Nothen et al. 1999) and quantitative (Petryshen et al. 2000) analyses. This suggests that while DYX2 may be a major gene in the susceptibility to dyslexia, it is not the single predisposing factor, and cannot exclude genetic heterogeneity in the trait.

A locus on chromosome 2p15-p16, DYX3, was also identified as a predisposing locus for dyslexia in a large Norwegian family (Fagerheim et al. 1999) and replicated in a sample of 96 Canadian families (Petryshen et al. 2002). Linkage of dyslexia to a locus on chromosome 6q12 (DYX4) was reported in 96 Canadian families (Petryshen et al. 2001). Another study reported significant linkage to chromosome 3 in one extended family (Nopola-Hemmi et al. 2001). Two recent studies have indicated two new dyslexia susceptibility loci. A large genome-wide scan conducted in three British and American sets of approximately 80 nuclear families each (Fisher et al. 2002) resulted in the identification of a major susceptibility locus on chromosome 18p11. In addition, moderate evidence for linkage of dyslexia to chromosomes 2p15 (DYX3), 3p12-q13, and 11p15 was also reported (Fisher et al. 2002). Finally, a potential locus for dyslexia on chromosome 11p, near the dopamine receptor type 4 gene (DRD4) was recently reported in 100 Canadian families (Hsiung et al. 2002).

Overall, it appears that many genes may contribute to the genetics of dyslexia. The most replicated finding is the chromosome 6p linkage, however not all studies indicate positive results for that region. It is therefore likely that genetic heterogeneity, gene-interactions and environmental factors may be involved in the trait.

There are several chromosomal regions that could be considered as candidate gene regions for dyslexia. The dopamine receptors, transporter and the gamma-
aminobutyric acid (GABA) receptors could be involved in dyslexia, because neurotransmitters and their receptors are important in neural development in the brain. In addition, the serotonergic system may be involved in dyslexia, because serotonin is an important factor in brain development. The serotonin receptors and transporter should therefore also be examined for their potential involvement in dyslexia. Previously identified susceptibility loci in autism, SLE, and ADHD, can also be considered candidate genes in dyslexia because of the potential link between these disorders and dyslexia. Chromosomal region 7q31.3 is another potential dyslexia locus, because it is involved in severe speech disorder (Fisher et al. 1998) and shows evidence for linkage to autism. There are many potential candidate genes for dyslexia in addition to those mentioned above, because of the large number of mechanisms that contribute to reading.

1.4 Genetic Dissection of Complex Traits

Complex traits are phenotypes that arise from multiple genetic and environmental factors, and the interactions among them. Particularly in humans there has been much effort to identify genes that contribute to complex traits.

Gene-mapping studies of genetic traits use linkage analysis and association tests to identify susceptibility loci. Linkage analysis tests for the cosegregation of a disease phenotype and genotypic variation at a chromosomal region. There exist both parametric and non-parametric approaches to linkage analysis. Non-parametric approaches are generally more robust in the study of complex traits (Kruglyak et al. 1996) because the mode of inheritance of the disease does not have to be specified. However, parametric linkage analysis is more powerful, provided the correct genetic model is specified (Abreu et al. 1999). Linkage analysis identifies loci that have a statistically significant effect on
the phenotype, but has low power in identifying genes with only moderate effects. Direct
tests of association on allelic variants of disease loci of modest effect on the trait can be
more powerful (Risch and Merikangas 1996). Association studies test for association
between genotypic variation at a specific locus and the disease phenotype. The tests used
by most studies are either population-based association or family-based association
(FBA). FBA overcomes the problem of population stratification by examining the
transmission of alleles in nuclear families; the frequency of transmitted alleles from
parents to affected children is compared to that of the control, or non-transmitted alleles.
However, the disadvantage of FBA is that a very large sample size is required to detect
significant effects (Risch and Merikangas 1996). Using mainly non-parametric linkage
analysis and association tests, many studies have identified different susceptibility loci
for various complex traits. Most of these loci have a moderate effect on the phenotype
and most often several contributing loci are identified for one disorder.

Many problems hamper the genetic dissection of complex traits. One of the key
issues in complex trait genetics is the accurate definition of the phenotype. It is often
possible to obtain a more homogeneous trait by narrowing down the disease phenotype or
by restricting the patient population (Lander and Shork 1994). Studies that follow such a
procedure are more likely to identify predisposing genes of major effect on complex
traits. For example, major susceptibility genes were identified in colon cancer by
narrowing down the clinical phenotype (Kinzler et al. 1991; Aaltonen et al. 1993), and in
Alzheimer Disease (Goate et al. 1991) and breast cancer (Miki et al. 1994) by specifying
the age of onset. This issue is of great importance to dyslexia genetics for several reasons.
First, different research groups often use different psychometric tests to identify dyslexic
individuals, which complicates comparisons across studies. It is important that independent groups use equivalent phenotypic tests, to assure phenotypic and therefore biological homogeneity of the samples. Second, phenotype measures vary across linkage studies; some studies use a qualitative definition while others use a quantitative definition of dyslexia. Much information is lost by grouping quantitative data into discrete categories, and more power could be gained if quantitative phenotypes could be studied directly. Furthermore, in some studies the qualitative phenotype is based primarily on one quantitative component (usually phonological coding or lifetime diagnosis) of dyslexia, which may not be comparable to the other quantitative measures used, such as vocabulary, spelling and RAN, despite the overlap among such measures. Therefore, evidence for linkage to a locus using the qualitative phenotype may not be equivalent to evidence for linkage to a locus using one separate quantitative component of dyslexia.

Another complication in complex trait genetics is that such traits are likely to involve multiple genetic and environmental factors. It may not be possible to identify a major predisposing factor to the phenotype, without taking into consideration either variation at another locus or a specific environmental component. Thus, it is important to consider gene-mapping methods that take into account gene-gene and gene-environment interactions, as well as genetic heterogeneity.

The hypothesis that multiple genes are involved in susceptibility to complex traits has resulted in several studies that have tried to devise methods to consider the contribution of several genes on the phenotype. Early studies that examined multiple genes in linkage analysis looked at the independent contribution of each locus to the trait (Schork et al. 1993, MacLean et al. 1993). Most proposed models are applicable to locus
heterogeneity or cases in which gene-effects on the trait are additive. Epistatic (gene-interaction) effects in complex traits have not been examined in detail. Numerous reports have found evidence that most complex traits depend upon more than one locus, and the phenotypic variance of the trait cannot be explained by additive gene-effects of the identified loci, or locus heterogeneity between them (Cox et al. 1999; Templeton 2000). Epistatic interactions are apparently much more common than was initially thought (Templeton 2000). Because of the importance of epistasis in complex traits there has been considerable research in recent years that has attempted to produce methods for detecting gene interactions.

1.5 Aim of the Study

In summary, in my opinion there are three principal areas of research on the genetics of dyslexia. First, there is considerable interest in identifying genes of major effect on the dyslexia phenotype. Second, there is increasing interest in identifying how interactions among genes of moderate effects can alter complex phenotypes, such as dyslexia. Third, the link between dyslexia and co-morbid phenotypes, such as ADHD and autism, remains unanswered. The goal of this thesis is to use data on 100 dyslexic families to answer questions in each of these three areas. In particular, Chapter III examines evidence for linkage to one chromosomal region, which has previously been identified as a susceptibility locus in small samples of dyslexia patients. In Chapter IV, I consider genetic variation at multiple loci in conferring susceptibility to the trait by taking into account epistatic and heterogeneity effects at each locus. Finally, in Chapter V, I examine evidence for a genetic component of handedness in families of dyslexic
patients, thus testing whether dyslexia susceptibility loci are involved in lateralization of language functions in the brain.
CHAPTER II GENERAL METHODS

2.1 Subjects

Families with at least two dyslexic siblings were ascertained from schools for learning-disabled children. All subjects were > 8 years of age and gave consent in accordance with the University of Calgary Ethics Review Board (parents gave consent for children where appropriate). One hundred families were included in the study. Three of the ascertained families were later determined to have only one affected sibling, but remained included in the analysis. In addition, the dyslexic siblings in two families were determined to be monozygotic twins and only one twin chosen at random was included in all of the gene-mapping analyses. Finally, one family did not have DNA information available on the second dyslexic siblings, which resulted in the contribution of only one dyslexic child in the analyses.

Of the 100 families (Table 1, set 1), 51 were nuclear families, and 49 were extended pedigrees. Four of the 100 families had a parent of non-European ancestry. There was history of dyslexia on both the paternal and maternal side in 22 bilineal pedigrees (Table 1, set 2). The 100 families consist of 1088 individuals, of which 914 have DNA information available. Of the 1088 individuals 919 participated in the study, the remaining 169 were deceased individuals or declined participation and were only included in order to preserve the links in some of the extended pedigrees. Among the 919 participants, there were 554 (60.3%) adults and 365 (39.7%) children, defined as being greater or less than 18 years of age respectively at the time of assessment. Among the 1088 individuals, there were 587 males, of which 305 had dyslexia, 112 were Normal and
170 were of unknown phenotype; and 501 females, of which 176 had dyslexia, 188 were Normal and 137 were of unknown phenotype.

It was necessary to split the large pedigrees into smaller families in order to carry out some of the statistical analyses. Therefore, a sample of 121 families (Table 1, set 3) was used that satisfied the 16 max bits requirement in Genehunter (GH) and Genehunter Plus (GHP; see section 2.3). The new version of Genehunter (v2.1) could accommodate up to 30 max bits per pedigree, which resulted in an overall sample of 107 families (Table 1, set 4). Finally, a set of 100 nuclear families (Table 1, set 5) was chosen for all of the family association tests. These families represent the 51 nuclear families in the original 100, as well as 49 nuclear families chosen from the 49 extended pedigrees. In addition, the analysis in Genehunter Two-Locus (GHT; see section 2.3) required a 12 max bit pedigree size, and therefore the 100 nuclear families were also used for analyses in GHT. Refer to Table 1 for a description of the different family sets.
Table 1. Description of the different family sets used for analysis by gene-mapping programs.

<table>
<thead>
<tr>
<th>Set</th>
<th>Number of families</th>
<th>Number of subjects</th>
<th>Dyslexic</th>
<th>Normal</th>
<th>Unknown</th>
<th>Analysis used&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>1088</td>
<td>481</td>
<td>300</td>
<td>307</td>
<td>Linkage, SAGE, Solar</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>462</td>
<td>195</td>
<td>121</td>
<td>146</td>
<td>Linkage</td>
</tr>
<tr>
<td>3</td>
<td>121</td>
<td>1034</td>
<td>487</td>
<td>287</td>
<td>260</td>
<td>GH v2, GHP</td>
</tr>
<tr>
<td>4</td>
<td>107</td>
<td>1020</td>
<td>482</td>
<td>284</td>
<td>254</td>
<td>GH v2.1</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>496</td>
<td>310</td>
<td>125</td>
<td>61</td>
<td>AFBAC, ETDT, GHT</td>
</tr>
</tbody>
</table>

<sup>a</sup> For a description of the methods used for analysis refer section 2.3

GH v2 = Genehunter version 2.0

GHP = Genehunter Plus

GH v2.1 = Genehunter version 2.1

AFBAC = Affected Family Based Controls

ETDT = Extended Transmission Disequilibrium Test

GHT = Genehunter Two-Locus
In addition to the individuals in the 100 families, a sample of 112 children with normal reading abilities was ascertained through local public schools. These individuals were used as control subjects for statistical analysis of the quantitative reading measures. The sample consisted of children 8-16 years of age, with a male: female ratio of 2.7:1.

2.2 Phenotype Data

QUANTITATIVE READING MEASURES

All participants in the study underwent a number of psychometric assessment tests assessing the four components of reading: phonological awareness, phonological coding, spelling, and rapid automatized naming (RAN). Of the 1088 individuals included in the study, 850 completed all the phenotypic assessments, while 901 completed at least one of the psychometric tests. Phonological awareness, the ability to recognize and manipulate phonemes, was assessed by the Auditory Analysis Test (Rosner and Simon 1971). Phonological coding, the ability to apply grapheme-phoneme conversion rules to non-words, was tested using the word attack subtests of the Woodcock Reading Mastery Test (Woodcock 1987) and the Woodcock-Johnson Psychoeducational Test-Revised (Woodcock and Johnson 1989). Spelling was assessed with the Wide-Range Achievement Test (Jastak and Wilkinson 1984). RAN speed, the ability to quickly recall and verbalize the names of presented objects, was assessed using the orthographically based Rapid Automatized Naming of Numbers Test (Denckla and Rudel 1976). The sample of 112 control children underwent the same psychometric tests as the dyslexia study subjects.

In addition to the psychometric tests, each subject's age and general cognitive ability was included in the phenotypic assessment. The latter was assessed using a short
The raw scores obtained from the quantitative tests for phonological awareness, phonological coding, spelling and RAN were converted to age-adjusted or standardized scores. Test scores for phonological coding and spelling were standardized according to test norms for all subjects in the dyslexia families. Test scores for phonological awareness and RAN speed were age-adjusted for individuals < 18 years of age using data from all children in the dyslexia families and the 112 controls. Test scores for phonological awareness were converted using the following formula: raw score + [(18 – age) x 0.716], while test scores for RAN speed were converted using the formula: raw score + [(18 – age) x 0.148] (Petryshen 2001). The statistical analyses were performed by Dr. Ming Liu using the SAS package (SAS institute 1990). The test score distributions were approximately normal for most phenotypic measures with the exception of phonological awareness scores, which exhibited a ceiling effect that could not be corrected by data transformations. Descriptive statistics for the adjusted phonological awareness, phonological coding, spelling, and RAN scores from all individuals in the dyslexia families and 112 controls are presented in Table 2. JmpIn v4.1 was used to obtain descriptive statistics on the quantitative measures for the 100 dyslexia families.
Table 2. Descriptive statistics of the quantitative measures in the sample of (a) all individuals from the 100 dyslexia families and (b) 112 control children.

<table>
<thead>
<tr>
<th>Measure</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) 100 Dyslexia families</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonological awareness</td>
<td>895</td>
<td>18.09</td>
<td>5.65</td>
<td>28.81</td>
<td>1.00</td>
<td>-0.90</td>
<td>0.14</td>
</tr>
<tr>
<td>Phonological coding</td>
<td>890</td>
<td>98.77</td>
<td>16.63</td>
<td>149.00</td>
<td>12.00</td>
<td>-0.02</td>
<td>0.89</td>
</tr>
<tr>
<td>Spelling</td>
<td>887</td>
<td>90.98</td>
<td>16.21</td>
<td>124.00</td>
<td>47.00</td>
<td>-0.22</td>
<td>-0.88</td>
</tr>
<tr>
<td>RAN speed</td>
<td>877</td>
<td>3.08</td>
<td>0.62</td>
<td>5.60</td>
<td>1.00</td>
<td>0.17</td>
<td>0.16</td>
</tr>
<tr>
<td>(b) 112 Controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonological awareness</td>
<td>112</td>
<td>23.48</td>
<td>4.00</td>
<td>28.35</td>
<td>9.90</td>
<td>-1.20</td>
<td>1.05</td>
</tr>
<tr>
<td>Phonological coding</td>
<td>112</td>
<td>112.72</td>
<td>13.23</td>
<td>138.00</td>
<td>82.00</td>
<td>-0.25</td>
<td>-0.65</td>
</tr>
<tr>
<td>Spelling</td>
<td>112</td>
<td>105.33</td>
<td>12.00</td>
<td>133.00</td>
<td>74.00</td>
<td>0.08</td>
<td>0.16</td>
</tr>
<tr>
<td>RAN speed</td>
<td>112</td>
<td>3.52</td>
<td>0.54</td>
<td>4.80</td>
<td>2.19</td>
<td>0.08</td>
<td>-0.48</td>
</tr>
</tbody>
</table>

20
DYSLEXIA: A QUALITATIVE TRAIT

Two psychologists reviewed the test results for each subject from the dyslexia families. Phonological coding was the basis for the dyslexia phenotype in this study because it has been reported that majority of children with dyslexia have impaired phonological coding skills (Olson et al. 1989). Therefore, the word attack scores were the primary determinant of phenotype classification, while the remaining test data were used to refine the certainty of the category. Children were diagnosed with dyslexia if the test scores indicated a gap of more than two years between chronological age and performance age. For adults, the full range of test scores on phonological coding, phonological awareness and spelling was used, with particular weight given to the reading history (Field and Kaplan 1998).

2.3 DNA Analysis

GENETIC MARKERS

This study used genotype data from 225 microsatellite markers, which were genotyped on more than 40% of the individuals in the 100 dyslexia families (the majority of markers were genotyped on >90% of subjects). The genetic markers were located throughout the genome (excluding the sex chromosomes), with an average density of 15.08 cM between markers (genetic distance used from Marshfield, research.marshfieldclinic.org). The number of markers genotyped on at more than 40% of the families ranged from 3 to 38 per chromosome. The average density between markers was greater on chromosomes 6q, 6p, 11p, 2p, 1p, which correspond to dyslexia candidate gene regions.
The microsatellite markers used in this study were genotyped by the following individuals: Malgorzata Zapala, Elzbieta Swiergala, Tracey L. Petryshen, Martha L. Hughes, G.-Y. Robin Hsiung, and Norma Schmill de French. Table 3 lists the number of markers per chromosome used in the study. Genetic distances and map order varied depending on the reference map database used. We generally used the Marshfield genetic map, although for certain regions on chromosomes 1p, 2p, 6p, 6q, and 11p the consensus map order was obtained from both physical (www.genome.ucsc.edu) and genetic maps (Marshfield research.marshfieldclinic.org; Genetic Location Database, LDB, cedar.genetics.soton.ac.uk). If the map order differed between reference maps, we either relied on the physical map or used the MultiMap computer software to establish the marker order and genetic distances from our data (see section 3.2).
Table 3. A description of the microsatellite markers genotyped on the 100 families.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Number of chromosome genotyped</th>
<th>Length of chromosome (cM)\textsuperscript{a}</th>
<th>Map span covered by markers (cM)</th>
<th>Average density of markers (cM)</th>
<th>Maximum gap between markers (cM)</th>
<th>Average number of genotyped individuals \textsuperscript{c}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td>289.66</td>
<td>243.00</td>
<td>12.79</td>
<td>51.99</td>
<td>852</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>269.07</td>
<td>243.01</td>
<td>11.05</td>
<td>43.01</td>
<td>847</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>228.14</td>
<td>151.00</td>
<td>16.78</td>
<td>51.14</td>
<td>868</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>211.65</td>
<td>184.98</td>
<td>20.55</td>
<td>65.00</td>
<td>848</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>197.54</td>
<td>185.00</td>
<td>16.82</td>
<td>39.00</td>
<td>819</td>
</tr>
<tr>
<td>6</td>
<td>38</td>
<td>193.14</td>
<td>148.00</td>
<td>3.89</td>
<td>33.14</td>
<td>825</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>181.97</td>
<td>87.00</td>
<td>10.88</td>
<td>47.97</td>
<td>764</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>167.90</td>
<td>75.27</td>
<td>15.05</td>
<td>54.25</td>
<td>786</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>168.98</td>
<td>104.47</td>
<td>34.82</td>
<td>72.48</td>
<td>819</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>173.13</td>
<td>111.00</td>
<td>22.20</td>
<td>65.00</td>
<td>827</td>
</tr>
<tr>
<td>11</td>
<td>24</td>
<td>147.77</td>
<td>147.77</td>
<td>6.17</td>
<td>27.00</td>
<td>819</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>170.66</td>
<td>55.00</td>
<td>18.33</td>
<td>95.00</td>
<td>847</td>
</tr>
<tr>
<td>13</td>
<td>8</td>
<td>114.98</td>
<td>89.00</td>
<td>11.13</td>
<td>26.00</td>
<td>675</td>
</tr>
<tr>
<td>14</td>
<td>16</td>
<td>138.18</td>
<td>123.02</td>
<td>7.69</td>
<td>25.00</td>
<td>572</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Genetic map distances were obtained from Marshfield Genetics (research.marshfieldclinic.org).

\textsuperscript{b} These include gaps from the end of a chromosome to the first genotyped distal marker.

\textsuperscript{c} The value is averaged over all markers genotyped for a given chromosome.
Table 3 (continued)

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Number of genotyped markers</th>
<th>Length of chromosome covered by markers (cM)</th>
<th>Map span average density of markers (cM)</th>
<th>Average gap between markers (cM)</th>
<th>Maximum number of genotyped individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>7</td>
<td>122.14</td>
<td>109.00</td>
<td>15.57</td>
<td>37.06</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>134.12</td>
<td>107.00</td>
<td>26.75</td>
<td>39.79</td>
</tr>
<tr>
<td>17</td>
<td>8</td>
<td>126.46</td>
<td>117.00</td>
<td>14.63</td>
<td>22.99</td>
</tr>
<tr>
<td>18</td>
<td>6</td>
<td>126.00</td>
<td>88.44</td>
<td>14.74</td>
<td>31.44</td>
</tr>
<tr>
<td>19</td>
<td>5</td>
<td>105.02</td>
<td>79.00</td>
<td>15.80</td>
<td>33.99</td>
</tr>
<tr>
<td>20</td>
<td>7</td>
<td>101.22</td>
<td>96.09</td>
<td>13.73</td>
<td>29.99</td>
</tr>
<tr>
<td>21</td>
<td>4</td>
<td>57.77</td>
<td>57.99</td>
<td>14.50</td>
<td>20.00</td>
</tr>
<tr>
<td>22</td>
<td>4</td>
<td>62.31</td>
<td>32.00</td>
<td>8.00</td>
<td>26.31</td>
</tr>
<tr>
<td>Total</td>
<td>225</td>
<td>3425.50</td>
<td>2634.29</td>
<td>15.08</td>
<td>42.62</td>
</tr>
</tbody>
</table>

*a* Genetic map distances were obtained from Marshfield Genetics (research.marshfieldclinic.org).

*b* These include gaps from the end of a chromosome to the first genotyped distal marker.

*c* The value is averaged over all markers genotyped for a given chromosome.
DNA was extracted from each subject from a 15ml blood sample by standard procedure using a salting out method (Miller et al. 1988).

Marker genotyping was performed using both manual and automated methods. For the manual approach, the concentrations of reagents used for the standard PCR reaction were as follows: 25ng dried DNA, 0.6μM forward and reverse primers, 1x buffer N with 2.0mM MgCl₂ (50mM KCl, 10mM TRIS at pH 8.3, 170 μg/ml bovine serum albumin, 0.05% Tween 20, 0.05% Nonidet P-40, 2mM MgCl₂) 200μM of each dATP, dGTP, and dTTP, 2.5μM of dCTP, 0.4MBq α-³²P-dCTP, and 0.5units of Taq DNA polymerase (Gibco) in a total volume of 10μl. PCR amplification was performed using either a Perkin Elmer GeneAmp PCR System 9600 thermocycler or an Ericomp EasyCycler Series thermocycler. The standard PCR reaction used for most of the primers was as follows: denaturation at 94°C for 6 minutes, 28 amplification cycles of 1 minute at 94°C, 2 minutes at 55°C, and 1 minute at 72°C, followed by extension at 72°C for 6 minutes. Following amplification, 5μl loading dye (20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol, 92% formamide) was added to each 10μl sample. 2μl of sample and a 2μl bacteriophage M13mp18 DNA sequencing ladder were loaded onto a 6% acrylamide gel (6% acrylamide, 420mg/ml urea, 0.1% TEMED, 0.25% ammonium persulfate, 1x TBE) and electrophoresed in 1xTBE buffer (10.8g/L TRIS, 5.5 g/L boric acid, 2mM EDTA at pH 8.0) at 1000V. Subsequently, the gels were exposed to Kodak autoradiography films overnight at room temperature and the films were processed with an automated developer. Each gel was read by two lab members, who independently
assigned marker alleles and entered the allele sizes twice into the database. Mismatches were resolved by repeating the PCR reaction and reloading the samples in question. For the automated method, the concentrations of reagents used for the standard PCR conditions were as follows: 25ng dried DNA, 0.04µM M-13-tailed forward primer, 0.04µM reverse primers, 1x buffer N with 2.0mM MgCl₂ (50mM KCl, 10mM TRIS at pH 8.3, 170 µg/ml bovine serum albumin, 0.05% Tween 20, 0.05% Nonidet P-40, 2mM MgCl₂) 200µM of each dATP, dGTP, dTTP, and dCTP, 0.2µM IRDye-700 or -800 (Licor), and 0.25units of Taq DNA polymerase (Gibco). PCR amplification was performed using MJ Research DNA Engine dual-bay DYAD PCR machines PTC-0220. The standard PCR reaction used for most of the primers was as follows: denaturation at 94°C for 5 minutes, 30 amplification cycles of 30 seconds at 94°C, 75 seconds at 55°C, and 15 seconds at 72°C, followed by extension at 72°C for 5 minutes. Following amplification, 2µl of IR² Stop Solution (Licor) were added to each 5µl sample. 1µl per sample and a 1µl IRD700 or IRD800 50-350bp DNA Sizing Standard (Licor) were loaded onto a 5% acrylamide gel (1x FMC Bioproducts Long Ranger Gel Solution, 350mg/ml urea, 0.05% TEMED, 0.05% ammonium persulfate, 1x TBE) and electrophoresed in 1xTBE buffer at 1200V on a LI-COR 4200S-2 Gene ReadIR DNA Analyzer. Marker alleles were assigned by GeneImagR software, confirmed by a lab member, and entered automatically into a database.

2.4 General Statistical Analyses

QUALITATIVE TRAIT LINKAGE ANALYSIS
Single point parametric linkage analysis was conducted using the Fastlink program v4.1P of the Linkage software (Lathrop and Lalouel 1984). The evidence for linkage is assessed using the logarithm of the odds-ratio score (Lod score), which is the log_{10} ratio of the likelihood of the data given the hypothesis of linkage at a certain recombination frequency, divided by the likelihood of no linkage (or a recombination frequency of 50%). A Lod score of 3 or more indicates evidence for linkage between the hypothetical disease locus and the genetic marker examined when only a single comparison is made, which corresponds to an overall p-value of 0.05 (i.e. on average 5% of all linkages will be false positives). The Lod score threshold of 3.3 is indicative of linkage for genome scans, to achieve a genome-wide level of significance of 0.05 (Lander and Kruglyak 1995). Single-point heterogeneity Lod scores (HLOD) were computed using the HOMOG program (Ott 1991) and Genehunter v2.0 (Kruglyak et al. 1996; Pratt et al. 2000) and v2.1 (Markianos et al. 2001). HLOD scores are Lod scores which take into account evidence for interfamilial heterogeneity, and report alpha, the proportion of families linked to the region. Multipoint Lod and HLOD scores were computed using Genehunter v2.0 and 2.1.

Single point non-parametric linkage analysis was first performed on a subset of 351 sibpairs from the dyslexia families using the Sibpal program v3.1 of the SAGE program package (SAGE 1997). The Sibpal program presents both the significance in deviation from 50% allele sharing among affected sibpairs, and the Linear Regression p-value. Sibpal Linear Regression takes into account the allele sharing between all possible pairs of concordant and discordant sibs, and compares it to that expected under no
linkage. A significant difference between the two will result in a significant Linear Regression p-value.

Single-point and multipoint non-parametric linkage analyses were performed using Genehunter v2.0 and v2.1 using the sets of 121 and 107 families corresponding to the different max bits requirements respectively. In Genehunter, evidence for linkage is assessed using the non-parametric Lod score (NPL score). Under the null hypothesis of no linkage, the NPL score asymptotically follows the standardized normal distribution, for a large number of pedigrees of similar size and for complete informativeness of markers. Thus, one can obtain the significance of the observed NPL score by comparing it to the distribution of NPL scores expected under the null hypothesis of no linkage, which is approximately normal.

ASSOCIATION ANALYSIS

Only family-based association tests were used for association analysis on the sample of 100 families. The Affected Family-Based Controls program (AFBAC; Thomson 1995) was used to test for evidence of increased transmission of susceptibility alleles from parents to affected children. The extended transmission disequilibrium test (ETDT; Sham and Curtis 1995) was also used to test the same hypothesis, only using heterozygous parents. In AFBAC, statistical significance is determined using a $\chi^2$ test in a $2 \times n$ contingency table, where $n$ is the number of alleles. One of the assumptions of the $\chi^2$ test is that the calculated $\chi^2$ statistic must be a close approximation to the theoretical $\chi^2$ continuous distribution. In order to obtain an unbiased calculated $\chi^2$, the general rule is that 80% or more of all the expected values must be 5.0 or more. It is therefore sometimes necessary to group rare alleles with expected observations of less than 5.0 into
one category. Analysis was initially conducted without grouping rare alleles into one category. However, for some markers such groupings were required. For ETDT and other programs based on the transmission disequilibrium test, alleles with expected counts of less than 5.0 are automatically excluded from the analysis.
CHAPTER III CONFIRMATION OF A DYSLEXIA SUSCEPTIBILITY GENE ON CHROMOSOME 1p

3.1 Introduction

Rabin et al. (1993) first found suggestive evidence for linkage of dyslexia to chromosome 1p34-p36. Linkage analysis with Rh and two DNA markers (FUCA1 and D1S165) produced a maximum Lod score of 2.33 at D1S165 ($\theta = 0.2$). Concurrently, a German family was identified (Froster et al. 1993), in which the dyslexia phenotype cosegregated with a balanced translocation 46, XY: t(1,2)(1p22;2q31), suggesting presence of a dyslexia susceptibility gene distal to 1p22, possibly in 1p34-p36. More recently, Grigorenko et al. (2001) found linkage to the same region on chromosome 1p, using a quantitative definition of dyslexia. We tested for the presence of a dyslexia gene in this region on chromosome 1p (Figure 1) in a sample of 100 Canadian families using both qualitative and quantitative definitions of the phenotype.
Figure 1. Relative map position of markers used in three studies examining evidence for linkage to chromosome 1p.
3.2 Methods

SUBJECTS AND PHENOTYPE DATA

The study sample and information on both quantitative and qualitative phenotype are described in sections 2.1 and 2.2 respectively.

GENETIC DATA

Genotype data were obtained for 12 closely linked markers in the 1p34-36 chromosomal region (Table 4). The markers were genotyped by Martha Hughes, Elzbieta Swiergala and Malgorzata Zapala using methods described in section 2.3. The number of genotyped individuals ranged from 808 for MYCL1 to 909 for D1S2826, due to the presence of misinheritances and insufficient DNA in some families. Marker allele frequencies were obtained from the parents of one nuclear family per pedigree. Map information was collected from published genetic (Marshfield research.marshfieldclinic.org; LDB cedar.genetics.soton.ac.uk) and physical (http://www.genome.ucsc.edu/ April 2002 draft) maps, to ensure a consensus marker order. Since there were two discrepancies in marker order between the genetic and physical maps (D1S199 and D1S552 order, and D1S1622 and D1S186 order were reversed on the physical map, however there have been several physical map positions for both D1S552 and D1S186 within the last 6 months, two of which correspond to those in the genetic map) we used the 100 families to determine a map order with the Multimap map building software (Matise et al. 1994). The map order from the Multimap results was equivalent to that obtained from the LDB, and was therefore used in this study. The genetic distances used for the multipoint analyses were obtained from LDB sex-averaged map.
Table 4. Map positions for the 12 markers on chromosome 1p34-p36.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Multimap</th>
<th>LDB Position</th>
<th>Marshfield</th>
<th>UCSC</th>
<th>Hetero-zigosity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Position (cM)</td>
<td>(cM)</td>
<td>Position (cM)</td>
<td>Position (bp)</td>
<td></td>
</tr>
<tr>
<td>D1S468</td>
<td>0</td>
<td>2.53</td>
<td>4.22</td>
<td>3538831</td>
<td>0.76</td>
</tr>
<tr>
<td>D1S1612</td>
<td>12.6</td>
<td>4.68</td>
<td>16.22</td>
<td>6183115</td>
<td>0.81</td>
</tr>
<tr>
<td>D1S2667</td>
<td>19.8</td>
<td>6.43</td>
<td>24.68</td>
<td>10772085</td>
<td>0.83</td>
</tr>
<tr>
<td>D1S1597</td>
<td>27.2</td>
<td>9.08</td>
<td>29.93</td>
<td>12450385</td>
<td>0.70</td>
</tr>
<tr>
<td>D1S507</td>
<td>32.4</td>
<td>12.66</td>
<td>33.75</td>
<td>13694930</td>
<td>0.79</td>
</tr>
<tr>
<td>D1S3669</td>
<td>42.0</td>
<td>16.06</td>
<td>37.05</td>
<td>16422026</td>
<td>0.79</td>
</tr>
<tr>
<td>D1S2826</td>
<td>48.1</td>
<td>18.21</td>
<td>41.92</td>
<td>17123220</td>
<td>0.64</td>
</tr>
<tr>
<td>D1S199</td>
<td>53.0</td>
<td>22.53</td>
<td>45.33</td>
<td>18580739</td>
<td>0.84</td>
</tr>
<tr>
<td>D1S552</td>
<td>56.7</td>
<td>22.58</td>
<td>45.33</td>
<td>17890708</td>
<td>0.69</td>
</tr>
<tr>
<td>D1S1622</td>
<td>74.1</td>
<td>32.49</td>
<td>56.74</td>
<td>31187253</td>
<td>0.71</td>
</tr>
<tr>
<td>D1S186</td>
<td>90.6</td>
<td>46.03</td>
<td>67.22</td>
<td>28986324</td>
<td>0.85</td>
</tr>
<tr>
<td>MYCL1</td>
<td>95.7</td>
<td>46.30</td>
<td>NA</td>
<td>34951430</td>
<td>0.87</td>
</tr>
</tbody>
</table>
STATISTICAL ANALYSIS

The analyses of dyslexia as a qualitative trait included both single point and multipoint parametric and non-parametric linkage analysis, and family based association tests as described in section 2.4. Autosomal dominant, recessive and intermediate genetic models were tested in the parametric analysis.

To analyze dyslexia as a quantitative trait I used the variance components (VC) approach implemented in Solar (Almasy and Blangero 1998) and Genehunter (Pratt et al. 2000). The variance components approach subdivides the phenotypic variance into several variance components. The classical technique subdivides the total variance into components due to genotypic and environmental effects (Lange et al. 1976). Hopper and Mathews (1982) extended the method to linkage analysis by modeling an additional variance component for a hypothesized QTL near a marker. Linkage to the QTL is indicated by a significant non-zero value for the QTL variance component. Thus, the model becomes: \( \sigma_p = \sigma_q + \sigma_g + \sigma_e \); where \( \sigma_p \) is the total phenotypic variance, \( \sigma_q \) is the variance due to the QTL, \( \sigma_g \) is the variance due to genetic effects outside of the QTL (i.e. polygenic variance), and \( \sigma_e \) is the variance due to environmental components. Almasy and Blangero (1998) improved the method by an approximation to a multipoint algorithm, thus allowing QTL multipoint linkage analysis to be performed. The VC method is implemented in both Genehunter, for which a family subdivision is necessary for our sample, and in Solar, which accommodates pedigrees of any size. QTL analysis in Genehunter can be performed under four genetic models:

1. The first model assumes only additive variance at both the QTL and polygenic components, and environmental variance:
\[ \sigma_p = \sigma_{qA} + \sigma_{gA} + \sigma_e \]

where \( \sigma_A \) is additive variance.

2. The second model assumes additive and dominant variance at the QTL, additive polygenic variance, and environmental variance:

\[ \sigma_p = \sigma_{qA} + \sigma_{qD} + \sigma_{gA} + \sigma_e; \]

where \( \sigma_D \) is dominant variance.

3. The third model assumes additive QTL variance, additive and dominant polygenic variance, and environmental variance:

\[ \sigma_p = \sigma_{qA} + \sigma_{gA} + \sigma_{gD} + \sigma_e; \]

4. The final model assumes both additive and dominant variance at each the QTL and polygenic components, and environmental variance:

\[ \sigma_p = \sigma_{qA} + \sigma_{qD} + \sigma_{gA} + \sigma_{gD} + \sigma_e; \]

The Lod scores obtained in Genehunter under models 2, 3, and 4 are computed under 2 degrees of freedom (df) to account for both the additive and dominant variance components. Lod scores under 2 df can be converted to Lod scores under 1 df by multiplying Lod (2 df) by 4.6 to obtain the approximate \( \chi^2 \) value under 2df. The corresponding p-value is then used to determine the \( \chi^2 \) value under 1df, which divided by 4.6 results in the Lod score under 1 df (Petryshen 2001).

Multipoint linkage analysis in SOLAR is limited to model 1 (shown above) only. SOLAR is currently unable to accommodate model 2 for multipoint analyses.

### 3.3 Results

*Dyslexia: a qualitative phenotype*
Two-point parametric linkage analysis in the 100 dyslexia families indicated suggestive evidence for linkage of dyslexia to chromosome 1p34-p36. The maximum Lod score was 1.23 ($\theta = 0.3$) at D1S1597 (Table 5), obtained under a recessive model with incomplete penetrance. Using the same genetic model, we performed single point (Table 5) and multipoint parametric linkage analysis using Genehunter. The evidence for linkage to chromosome 1p increased slightly for multipoint analyses, resulting in a maximum HLOD score of 1.53 between markers D1S199 and D1S552 (Figure 2). Sibpair analysis indicated evidence for linkage to markers D1S3669 ($p = 0.008$), D1S1622 ($p = 0.029$), D1S186 ($p = 0.010$) using the linear regression p-values, and markers D1S1597 ($p = 0.006$), D1S507 ($p = 0.005$), and D1S1622 ($p = 0.029$) using only affected sibpairs (Table 5). Two-point NPL scores did not show strong evidence for linkage to the region, however, multipoint non-parametric linkage results indicated linkage in the region with a peak NPL score of 1.62 ($p$-value $= 0.04$) between markers D1S3669 and D1S2826 (Figure 2).
Table 5. Results from single point linkage analysis of dyslexia. P-values and scores that correspond to p-values less than 0.05 are represented in bold.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Position (cM)</th>
<th>Lod score</th>
<th>HLOD score</th>
<th>NPL score&lt;sup&gt;a&lt;/sup&gt;</th>
<th>NPL score&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Sibpal p-value&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Affected sibs p-value&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1S468</td>
<td>2.53</td>
<td>-0.22</td>
<td>0.00</td>
<td>-1.12</td>
<td>-1.30</td>
<td>0.702</td>
<td>0.522</td>
</tr>
<tr>
<td>D1S1612</td>
<td>4.68</td>
<td>-0.26</td>
<td>0.00</td>
<td>0.09</td>
<td>0.44</td>
<td>0.702</td>
<td>0.268</td>
</tr>
<tr>
<td>D1S2667</td>
<td>6.43</td>
<td>0.45</td>
<td>0.28</td>
<td>0.78</td>
<td>0.80</td>
<td>0.169</td>
<td>0.275</td>
</tr>
<tr>
<td>D1S1597</td>
<td>9.08</td>
<td>1.23</td>
<td>0.86</td>
<td>0.69</td>
<td>0.62</td>
<td>0.074</td>
<td><strong>0.006</strong></td>
</tr>
<tr>
<td>D1S507</td>
<td>12.66</td>
<td>0.34</td>
<td>0.13</td>
<td>1.34</td>
<td>0.84</td>
<td>0.169</td>
<td><strong>0.005</strong></td>
</tr>
<tr>
<td>D1S3669</td>
<td>16.06</td>
<td>0.29</td>
<td>0.21</td>
<td>0.32</td>
<td>0.28</td>
<td><strong>0.008</strong></td>
<td>0.126</td>
</tr>
<tr>
<td>D1S2826</td>
<td>18.21</td>
<td>0.22</td>
<td>0.28</td>
<td>1.03</td>
<td>1.23</td>
<td>0.146</td>
<td>0.077</td>
</tr>
<tr>
<td>D1S199</td>
<td>22.53</td>
<td>0.49</td>
<td><strong>1.35</strong></td>
<td><strong>1.74</strong></td>
<td>1.17</td>
<td>0.228</td>
<td>0.071</td>
</tr>
<tr>
<td>D1S552</td>
<td>22.58</td>
<td>0.22</td>
<td>0.14</td>
<td>0.36</td>
<td>0.36</td>
<td>0.077</td>
<td>0.129</td>
</tr>
<tr>
<td>D1S1622</td>
<td>32.00</td>
<td>0.22</td>
<td>0.01</td>
<td>0.67</td>
<td>0.32</td>
<td><strong>0.029</strong></td>
<td><strong>0.029</strong></td>
</tr>
<tr>
<td>D1S186</td>
<td>46.03</td>
<td>-0.33</td>
<td>0.00</td>
<td>-0.27</td>
<td>0.00</td>
<td><strong>0.010</strong></td>
<td>0.131</td>
</tr>
<tr>
<td>MYCL1</td>
<td>46.30</td>
<td>-0.02</td>
<td>0.10</td>
<td>-0.05</td>
<td>0.08</td>
<td>0.162</td>
<td>0.126</td>
</tr>
</tbody>
</table>

<sup>a</sup> NPL score from Genehunter v2.0 (16 max bit), in the set of 121 families.

<sup>b</sup> NPL score from Genehunter v2.1 (30 max bit), in the set of 107 families.

<sup>c</sup> Sibpal Linear Regression p-value, using all pairs of sibs.

<sup>d</sup> Sibpal p-value assessing the deviation from 50% allele sharing between affected sibs.
Figure 2. Multipoint parametric and non-parametric linkage analysis results using the qualitative phenotype.
Linkage analysis was also performed in 22 bilineal pedigrees, in which individuals from both the paternal and maternal side were affected with dyslexia. The obtained Lod scores were not indicative of linkage, possibly resulting from the large reduction of sample size.

Association analysis was performed on a set of 100 nuclear families from each of the 100 dyslexia pedigrees using the qualitative dyslexic phenotype (Table 6). There are two markers that show weak evidence for association, using AFBAC, D1S468 (p = 0.042) and D1S3669 (p = 0.042). There was no evidence for allele-wise association with ETDT. ETDT uses a smaller sample of subjects for the association tests, as it only considers families in which the parents are heterozygotes.
Table 6. Results from the family-based association tests using the qualitative phenotype.

<table>
<thead>
<tr>
<th>Marker</th>
<th>AFBAC p-value rare</th>
<th>AFBAC p-value rare</th>
<th>ETDT p-value(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>alleles ungrouped</td>
<td>alleles combined</td>
<td></td>
</tr>
<tr>
<td>D1S468</td>
<td>0.068</td>
<td>0.042</td>
<td>0.921</td>
</tr>
<tr>
<td>D1S1612</td>
<td>0.252</td>
<td>0.247</td>
<td>0.629</td>
</tr>
<tr>
<td>D1S2667</td>
<td>0.979</td>
<td>0.995</td>
<td>0.607</td>
</tr>
<tr>
<td>D1S1597</td>
<td>0.440</td>
<td>0.440</td>
<td>0.992</td>
</tr>
<tr>
<td>D1S507</td>
<td>0.899</td>
<td>0.833</td>
<td>0.836</td>
</tr>
<tr>
<td>D1S3669</td>
<td>0.043</td>
<td>0.042</td>
<td>0.335</td>
</tr>
<tr>
<td>D1S2826</td>
<td>0.851</td>
<td>0.936</td>
<td>0.856</td>
</tr>
<tr>
<td>D1S199</td>
<td>0.586</td>
<td>0.861</td>
<td>0.295</td>
</tr>
<tr>
<td>D1S552</td>
<td>0.494</td>
<td>0.663</td>
<td>0.351</td>
</tr>
<tr>
<td>D1S1622</td>
<td>0.709</td>
<td>0.788</td>
<td>0.227</td>
</tr>
<tr>
<td>D1S186</td>
<td>0.983</td>
<td>0.966</td>
<td>0.708</td>
</tr>
<tr>
<td>MYCL1</td>
<td>0.188</td>
<td>0.668</td>
<td>0.260</td>
</tr>
</tbody>
</table>

\(^a\) The allele-wise p-value is presented.
Dyslexia: a quantitative phenotype

Using the variance components approach implemented in Genehunter I tested for linkage of the four components of dyslexia to chromosome 1p. The strongest evidence for linkage was obtained under a model with both additive and dominance QTL variance, additive polygenic variance and residual environmental variance. The results (Figure 3) indicate that the maximum LOD score for spelling is 3.97, corresponding to a 1 degree of freedom Lod score of 3.25. The maximum Lod score was obtained between markers D1S199 and D1S1622 and overlaps with the maximum HLOD and NPL scores obtained for the qualitative analyses.
Figure 3. Multipoint quantitative linkage analysis of dyslexia.
I also used the variance components approach implemented in Solar, in which it is not necessary to split up the families. Unfortunately, Solar can only accommodate one model with additive variance at the QTL, additive polygenic variance and environmental variance. The resulting Lod scores were identical to those obtained in Genehunter under the same model (data not shown).

### 3.4 Discussion

Linkage analysis of dyslexia as a qualititative trait indicates suggestive evidence for linkage to chromosome 1p. The evidence for linkage is stronger using the quantitative measures of dyslexia. The Lod scores from the quantitative analyses surpassed the threshold value of $p = 0.01$, recommended for confirmation of linkage (Lander and Kruglyak 1995). Furthermore, the Lod score for spelling approaches the score of 3.3, required for significant genome-wide linkage result.

This is the third study that has found positive evidence for linkage of dyslexia to the region. The conclusions from the current study are more robust because of the considerably larger dataset, 100 families, compared the set of 9 (Rabin et al. 1993) and 8 (Grigorenko et al. 2001) families used by other groups. Perhaps as a result the two previous studies do not detect a strong or well-defined linkage signal, although such a result is not surprising for a complex trait. Probably due to the large sample size, we obtained the strongest overall evidence for linkage to chromosome 1p. This region merits further attention: although Grigorenko et al. (2001) obtain a flat linkage signal, spanning more than 45cM, I was able to localize the linkage to a 10cM region within this map, between markers D1S199 and D1S1622, and future studies may narrow down the region even further.
The use of both qualitative and quantitative phenotype measures also presents an advantage, as these results can easily be compared to other studies. For example, Grigorenko et al. (2001) obtained significant results using quantitative measures, while Rabin et al. (1993) obtained suggestive results using a qualitative phenotype, which is reflected in this study as well. Interestingly, the multipoint peaks obtained for the qualitative and quantitative trait analyses overlap; however the peak for qualitative analysis is much wider. We speculate that this finding may be correlated to the marker heterozygosity as well as the differences in the number of individuals with phenotype data; for example 781 individuals were dyslexic or normal for the qualitative phenotype, while 887 subjects had quantitative data for spelling.

Evidence for linkage using the quantitative trait measures can be compared between our study and that of Grigorenko et al. (2001). They examined phonemic awareness, phonological decoding, rapid naming, single word reading, vocabulary and lifetime diagnosis. Linkage peaks were obtained for different traits at different map locations. In some instances the same quantitative trait (phonological awareness) showed two peaks at the distal and proximal ends of the 45cM region, while in our data there is only one clearly defined peak for two of the traits, and no significant results over the rest of the 45cM region. Phonological coding also maximized over the entire 45cM region in the study of Grigorenko et al (2001), while our results clearly indicate a peak between D1S199 and D1S1622. It should be noted that linkage of the quantitative components of dyslexia to 1p might be due to a locus that has effects on a number of different mechanisms involved in reading. However, conclusions about the particular reading skill
that such a locus may be part of are necessarily tentative, because variance in quantitative measures can affect the QTL linkage analysis (Pennington 1997).

The variance component linkage method assumes multivariate normality of the quantitative phenotypic data. This assumption is probably violated in most complex trait samples, where patients are selected to lie within one tail of the trait distribution. The presence of a major gene and gene-interactions may also lead to non-normality of the data (Allison et al. 1999). As discussed previously (see section 2.2), the quantitative phenotypic measures used in this study appear to be normally distributed.

The results also suggest that there is genetic heterogeneity in the sample. The HLOD scores obtained from Genehunter were greater in some instances than the two-point Lod scores, even though several large families were broken down into smaller pedigrees to perform analyses in Genehunter. The loss in power from the family subdivisions did not correspond to a corresponding decrease in the evidence for linkage, which would be expected if most of the families were linked to chromosome 1p. The HLOD scores also account for interfamilial heterogeneity, further supporting the hypothesis of genetic heterogeneity in this sample. Of all the studies that have obtained positive evidence for linkage to chromosome 1p, not one shows a very strong and well defined linkage signal as in a Mendelian trait. Thus it is likely that there is a gene in this region that contributes to dyslexia but is not the determining factor in the etiology of the trait, as is the case with most complex trait loci.

I also compared the results from the Genehunter analyses between versions 2.0 and 2.1, because version 2.1 can accommodate larger pedigrees, and most of the family structure was preserved. Analyses with Genehunter v2.1 were limited to the qualitative
trait because of computational restrictions. The evidence of linkage to the qualitative trait with both versions is suggestive, but not highly significant, and therefore it is difficult to make a clear comparison. The similarity in the results is not unexpected and suggests lack of intrafamilial heterogeneity in this sample. However, the large proportion of bilineal pedigrees in the dataset suggests that intrafamilial heterogeneity may be present in the data. Accordingly, there is no strong evidence of linkage for analyses on the bilineal pedigrees only.

Family based association tests provide very weak evidence for association. However, it is difficult to draw conclusions from this result. The lack of significance may be due to the fact that the marker alleles examined in this study are not in close linkage disequilibrium with the disease allele at the dyslexia locus in the region. As only markers that are less than 2cM apart are expected to show linkage disequilibrium in humans under certain conditions, linkage disequilibrium is unlikely to yield associations unless several markers are tested very close to the contributing genes.

I searched for candidate genes in the 10cM region between D1S199 and D1S1622, in which our linkage signal maximizes. Two Rhesus (RH) blood group antigen genes localize to the region, RHD and RHCE, as previously discussed (Grigorenko et al. 2001) together with a factor involved in the immune system, the interleukin 22 receptor (IL22R). Other possible candidate genes in the region include genes involved in brain function and neurological development: the opioid receptor delta 1 (OPRD1); the cannabinoid receptor 1 (CNR1), which is primarily expressed in the brain; the nerve injury induced protein 2 (NINJ2), which promotes neurite outgrowth; and the neuronal thread protein AD7C-NTP, overexpression of which may promote apoptosis of neuronal
cells. In addition, an ephrin receptor gene, EPHB2, which belongs to the ephrin family of molecules that mediate developmental processes particularly in the brain, is 3.2 Mb centromeric to D1S199. Finally, two serotonin receptor genes, serotonin receptor 6 (HTR6) and 1D (HTR1D), also localize to the region, centromeric to D1S199 by 35 Kb and 3.5Mb respectively. The HTR6 gene has been previously studied for its involvement in neuro-psychiatric disorders. A polymorphism in the HTR6 gene has been found to be associated with schizophrenia (Tsai et al. 1999) and Parkinson Disease (Messina et al. 2002), but not with bipolar disorder or major depression (Hong et al. 1999). It is interesting to speculate that the HTR6 gene or the HTR1D gene might be involved in dyslexia, since the serotonergic system has been shown to be involved in a number of complex psychiatric and neurological disorders. Furthermore, serotonin is an important regulator of morphogenesis during early central nervous system development, which makes it an excellent candidate dyslexia gene. In addition, studies in rats show that serotonin depletion during synaptogenesis leads to a decreased synapse density and learning deficits (Mazer et al. 1997).

In conclusion, this study confirms and strengthens recent findings of a dyslexia susceptibility gene on chromosome 1p. There are several interesting potential candidate genes in this dyslexia susceptibility region. Future work is needed to identify the exact gene or genes involved in the trait.
CHAPTER IV EVIDENCE FOR LOCUS INTERACTION AND HETEROGENEITY IN DYSLEXIA

4.1 Introduction

Most gene-mapping studies focus on identifying genes conferring a major predisposition to the trait. However, the effects of individual genes may not be independent in predicting complex phenotypes and therefore multiple genetic factors need to be considered simultaneously. Several studies have addressed the importance of considering variation at multiple genes in complex genetic phenotypes and as a result, a number of methods have been proposed to identify gene-interactions or genetic heterogeneity in complex disorders. These include, but are not limited to, the sib-pair method (Dizier and Clerget-Darpoux 1986); the two locus Lod score method (Ott 1991; Schork et al. 1993); homogeneity tests (Ott 1991); the marker association segregation \( \chi^2 \) (MASC) extension for two loci (Dizier et al. 1994); the maximum Lod-score (MLS) method proposed by Risch (1990a, b, c) and extended by Cordell et al. (1995) to account for two or more (Cordell et al. 2000) genes; the family weighting approach in Genehunter-Plus (GHP) developed by Cox et al. (1999); the Genehunter-Two Locus (GHT) extension of Genehunter by Strauch et al. (2000); and the use of Solar (Almasy and Blangero 1998) in developing models of gene-interaction (Mitchell et al. 2000).

While some of these methods have been used by researchers to test for gene-interactions in independent samples and for different traits, to our knowledge, no one has applied more than one method to the same data and compared the results. In this study I test for evidence of gene-interactions and heterogeneity in the susceptibility to dyslexia using
three of the above methods from the literature. I apply the GHP, GHT, and homogeneity tests to test for both gene-interactions and genetic heterogeneity in dyslexia between eight previously identified susceptibility regions and novel candidate gene-regions.

4.2 Methods

SINGLE LOCUS ANALYSIS

Parametric Linkage analysis was performed as described in section 2.4 under eight genetic models, encompassing an autosomal recessive, dominant and intermediate mode of inheritance (Table 8) to increase the likelihood to detect linkage.
Table 8. Genetic models used in single locus parametric linkage analysis.

<table>
<thead>
<tr>
<th>Genetic Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genotype</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Penetrances:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aa</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ab</td>
<td>0</td>
<td>0</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.6</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>bb</td>
<td>0.8</td>
<td>0.8</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.8</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Disease allele b</td>
<td>0.01</td>
<td>0.25</td>
<td>0.001</td>
<td>0.01</td>
<td>0.25</td>
<td>0.01</td>
<td>0.01</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Non-parametric linkage analysis and family-based association were performed as described in section 2.4 for all 225 markers genotyped.

TWO-LOCUS ANALYSES

*Genehunter Plus*

Genehunter-Plus (GHP; Cox et al. 1999) is a method which takes into account individual family evidence of linkage to one chromosomal region when examining evidence for linkage to a second, unlinked chromosomal region. GHP is based on the Genehunter algorithm for assessing evidence for linkage, using the NPL score. The method first conducts single locus linkage analysis to locus 1, then assigns individual family weights to reflect either an epistatic (gene-interaction) or a heterogeneity model, and finally performs linkage analysis at locus 2, resulting in the "weighted" NPL score at locus 2. I first tested for gene-interactions using two different weighting schemes (proportional and discrete) for all pairs of markers. Under the epistatic proportional scheme, each family receives a weight equal to its NPL score for region 1, while assessing the overall NPL score for region 2. For example, if a family receives an NPL score of 3.00 for locus 1, it receives a weight of 3.00 while assessing the overall "weighted" NPL score (for all dyslexia families) to locus 2. Thus, families that show evidence for linkage to the first chromosomal region contribute more towards the final NPL score for the second chromosomal region. For the discrete epistatic scheme all positive familial NPL scores are converted to weights of 1, while all the negative NPL scores are converted to weights of 0. I then tested for locus heterogeneity between all pairs of markers using the two weighting schemes (proportional and discrete). Under the heterogeneity proportional scheme the weights are the negative of the NPL score for
region 1, while assessing evidence for linkage to region 2. For example, if a family obtains a NPL score of -3.00 to locus 1, it receives a weight of 3.00 in the calculation of the overall “weighted” NPL score (for all dyslexia families) at locus 2. In the discrete heterogeneity scheme all positive NPL scores for region 1 are converted to weights of 0 and all the negative NPL scores are converted to weights of 1. I tested for evidence of gene-interaction or heterogeneity among 22 chromosomal regions, using a total of 52 microsatellite markers (average 3.6 markers/region). To assess the evidence for gene-interaction or heterogeneity I calculated |X|, where X is the difference between the single-locus NPL score and the weighted NPL score at locus 2. To assess statistical significance, Cox et al. (1999) show that 4.6*|X| follows a chi-square distribution with 1 degree of freedom.

*Genehunter Two-Locus*

I used Genehunter Two-locus to replicate the heterogeneity results from the GHP analysis. GHT is another extension of Genehunter, which simultaneously considers variation at two loci. A model needs to be specified in the program to account for the penetrances for the nine possible genotype combinations. For each marker I selected the parametric model (Table 1) that resulted in the maximum parametric single locus Lod score. For the non-parametric analyses, the two-locus NPL score is the sum of the corresponding single locus NPL scores. I only selected one pair of markers per region that showed evidence for epistasis or heterogeneity in GHP to replicate these results in GHT. I expected to observe an increase in the NPL score for the pairs of markers that showed evidence for heterogeneity in GHP. The analysis was performed on the set of 100 nuclear families selected from all the dyslexia pedigrees.
Homogeneity tests

I also used a homogeneity test to try and replicate the results from GHP. Using the results from the parametric single-locus linkage analysis, I used the program HOMOG to obtain \( \alpha \), the proportion of families linked to the region. In addition, the HOMOG output includes the familial \( \alpha \), which is the probability of heterogeneity in a given family. I considered all the markers involved in the final results in GHP and chose those, for which the family alphas ranged around 0.5. For each marker I then subdivided the total families into two groups: those with familial alphas > 0.5 and those with alphas < 0.5. I next performed linkage analysis separately in the two family samples for locus 2, which had already showed evidence for gene-interaction or heterogeneity in the GHP results.

4.3 Results

SINGLE LOCUS ANALYSIS

Parametric and non-parametric linkage analysis was performed for 225 microsatellite markers throughout the genome, to assess the evidence for linkage of dyslexia to each chromosomal region. In addition, association tests were also performed on all the markers in the dataset. Suggestive or significant evidence for linkage and association was obtained for 49 markers on 12 chromosomes. Evidence for linkage was considered suggestive if the Lod scores were greater than 1, the NPL score p-values were less than 0.05, the sib-pair Linear Regression p-values were less than 0.05, and the AFBAC or ETDT p-values were < 0.05. Parametric linkage analysis indicated 23 markers on 8 chromosomes in regions 1p, 2p, 4p, 4q, 6q, 11p, 13q, 14q, and 20q, which had Lod scores greater than 3.
scores greater than 1, for genetic models 1, 2 and 5-8 (Table 8). Non-parametric linkage analysis using all affected pedigree members, indicated 25 markers on 9 chromosomes in regions 1p, 2p, 4p, 6q, 7q, 11p, 13q, 14q, and 17p, which had NPL score p-values less than 0.05. Sibpair analysis indicated 23 markers on 7 chromosomes in regions 1p, 2p, 4p, 6q, 11p, 13q, and 14q, which had Linear Regression p-values less than 0.05. Association analyses indicated 13 markers on 8 chromosomes in regions 1p, 2p, 4p, 8p, 11p, 15q, and 16q, which had AFBAC or ETDT p-values less than 0.05. Table 9 represents the 33 markers chosen for GHP analyses from the 15 regions that showed suggestive or significant evidence for linkage or association in the dataset.
Table 9. A list of 33 markers used in GHP analyses, from the 15 regions which indicated suggestive or significant evidence for linkage or association in the set of 100 dyslexia families. Results indicative of suggestive or significant linkage or association are represented in bold.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Region</th>
<th>Position</th>
<th>Lod score</th>
<th>Model</th>
<th>NPL p-value</th>
<th>Sibpal p-value</th>
<th>AFBAC p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1S3669</td>
<td>lp36.13</td>
<td>37.01</td>
<td>0.29</td>
<td>1</td>
<td>0.42</td>
<td>0.002</td>
<td>0.04</td>
</tr>
<tr>
<td>D1S199</td>
<td>lp36.12</td>
<td>45.00</td>
<td>0.49</td>
<td>1</td>
<td>0.04</td>
<td>0.228</td>
<td>0.59</td>
</tr>
<tr>
<td>D1S186</td>
<td>lp34.1</td>
<td>67.00</td>
<td>0.00</td>
<td>2</td>
<td>0.62</td>
<td>0.010</td>
<td>0.19</td>
</tr>
<tr>
<td>D2S2952</td>
<td>2p25.2</td>
<td>18.00</td>
<td>-0.05</td>
<td>5</td>
<td>0.26</td>
<td>0.337</td>
<td>0.05</td>
</tr>
<tr>
<td>D2S1788</td>
<td>2p21</td>
<td>56.00</td>
<td>-0.10</td>
<td>2</td>
<td>0.40</td>
<td>0.018</td>
<td>0.69</td>
</tr>
<tr>
<td>D2S2240</td>
<td>2p16.1</td>
<td>69.77</td>
<td>0.46</td>
<td>5</td>
<td>0.04</td>
<td>0.037</td>
<td>0.85</td>
</tr>
<tr>
<td>D2S2378</td>
<td>2p16.1</td>
<td>70.31</td>
<td>0.25</td>
<td>5</td>
<td>0.22</td>
<td>0.018</td>
<td>0.38</td>
</tr>
<tr>
<td>D2S378</td>
<td>2p15</td>
<td>77.00</td>
<td>1.42</td>
<td>5</td>
<td>0.14</td>
<td>0.009</td>
<td>0.13</td>
</tr>
<tr>
<td>D2S2183</td>
<td>2p15</td>
<td>78.00</td>
<td>0.57</td>
<td>5</td>
<td>0.45</td>
<td>0.050</td>
<td>0.61</td>
</tr>
<tr>
<td>D2S2279</td>
<td>2p15</td>
<td>78.01</td>
<td>-0.02</td>
<td>2</td>
<td>0.75</td>
<td>0.039</td>
<td>0.46</td>
</tr>
<tr>
<td>MSX1</td>
<td>4p16.2</td>
<td>5.02</td>
<td>0.94</td>
<td>2</td>
<td>0.07</td>
<td>0.054</td>
<td>0.01</td>
</tr>
<tr>
<td>DRD5B</td>
<td>4p15.33</td>
<td>24.02</td>
<td>0.98</td>
<td>2</td>
<td>0.03</td>
<td>0.018</td>
<td>0.06</td>
</tr>
<tr>
<td>D4S403</td>
<td>4p15.33</td>
<td>26.00</td>
<td>0.35</td>
<td>2</td>
<td>0.12</td>
<td>0.017</td>
<td>0.66</td>
</tr>
<tr>
<td>D4S2943</td>
<td>4q34.3</td>
<td>190.00</td>
<td>1.02</td>
<td>5</td>
<td>0.14</td>
<td>0.069</td>
<td>0.22</td>
</tr>
<tr>
<td>D6S421</td>
<td>6q12</td>
<td>84.00</td>
<td>1.36</td>
<td>8</td>
<td>0.12</td>
<td>0.021</td>
<td>0.53</td>
</tr>
<tr>
<td>D6S286</td>
<td>6q12</td>
<td>90.01</td>
<td>2.65</td>
<td>1</td>
<td>0.04</td>
<td>0.229</td>
<td>0.56</td>
</tr>
</tbody>
</table>
Table 9 (continued)

<table>
<thead>
<tr>
<th>Marker</th>
<th>Region</th>
<th>Position</th>
<th>Lod score (cM)</th>
<th>Model</th>
<th>NPL p-value</th>
<th>Sibpal p-value</th>
<th>AFBAC p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>D6S251</td>
<td>6q14.3</td>
<td>90.09</td>
<td>2.46</td>
<td>8</td>
<td>0.04</td>
<td>0.039</td>
<td>0.91</td>
</tr>
<tr>
<td>IGF2R</td>
<td>6q25.3</td>
<td>160.00</td>
<td>0.14</td>
<td>8</td>
<td>0.69</td>
<td>0.636</td>
<td>0.03</td>
</tr>
<tr>
<td>D7S2527</td>
<td>7q31.33</td>
<td>129.00</td>
<td>0.19</td>
<td>1</td>
<td>0.05</td>
<td>0.843</td>
<td>0.13</td>
</tr>
<tr>
<td>D8S1771</td>
<td>8p21.2</td>
<td>50.00</td>
<td>0.52</td>
<td>1</td>
<td>0.50</td>
<td>0.443</td>
<td>0.03</td>
</tr>
<tr>
<td>D11S1363</td>
<td>11p15.5</td>
<td>0.01</td>
<td>1.48</td>
<td>5</td>
<td>0.18</td>
<td>0.331</td>
<td>0.79</td>
</tr>
<tr>
<td>DRD4</td>
<td>11p15.5</td>
<td>1.02</td>
<td>2.08</td>
<td>8</td>
<td>0.03</td>
<td>0.005</td>
<td>0.84</td>
</tr>
<tr>
<td>D11S1984</td>
<td>11p15.5</td>
<td>2.10</td>
<td>1.69</td>
<td>8</td>
<td>0.26</td>
<td>0.045</td>
<td>0.04</td>
</tr>
<tr>
<td>D11S2362</td>
<td>11p15.5</td>
<td>8.90</td>
<td>-0.10</td>
<td>5</td>
<td>0.76</td>
<td>0.476</td>
<td>0.009</td>
</tr>
<tr>
<td>D13S800</td>
<td>13q21.32</td>
<td>55.00</td>
<td>1.43</td>
<td>5</td>
<td>0.04</td>
<td>0.670</td>
<td>0.34</td>
</tr>
<tr>
<td>D14S73</td>
<td>14q13.3</td>
<td>96.08</td>
<td>0.06</td>
<td>1</td>
<td>0.64</td>
<td>0.015</td>
<td>0.55</td>
</tr>
<tr>
<td>D14S67</td>
<td>14q31.3</td>
<td>96.10</td>
<td>0.25</td>
<td>2</td>
<td>0.60</td>
<td>0.005</td>
<td>0.69</td>
</tr>
<tr>
<td>PLON</td>
<td>14q31.3</td>
<td>99.00</td>
<td>0.89</td>
<td>2</td>
<td>0.72</td>
<td>0.029</td>
<td>0.27</td>
</tr>
<tr>
<td>D14S81</td>
<td>14q32.11</td>
<td>108.00</td>
<td>1.07</td>
<td>2</td>
<td>0.04</td>
<td>0.032</td>
<td>0.38</td>
</tr>
<tr>
<td>D14S292</td>
<td>14q32.33</td>
<td>134.00</td>
<td>1.12</td>
<td>2</td>
<td>0.18</td>
<td>0.042</td>
<td>0.29</td>
</tr>
<tr>
<td>D15S143</td>
<td>15q21.1</td>
<td>46.06</td>
<td>0.36</td>
<td>2</td>
<td>0.16</td>
<td>0.336</td>
<td>0.04</td>
</tr>
<tr>
<td>D16S404</td>
<td>16p13.3</td>
<td>18.00</td>
<td>0.27</td>
<td>5</td>
<td>0.16</td>
<td>0.165</td>
<td>0.01</td>
</tr>
<tr>
<td>D17S974</td>
<td>17p13.1</td>
<td>22.02</td>
<td>0.69</td>
<td>5</td>
<td>0.03</td>
<td>0.030</td>
<td>0.10</td>
</tr>
<tr>
<td>D20S171</td>
<td>20q13.33</td>
<td>96.00</td>
<td>2.17</td>
<td>4</td>
<td>0.53</td>
<td>0.615</td>
<td>0.75</td>
</tr>
</tbody>
</table>
TWO-LOCUS ANALYSES

*Genehunter-Plus*

Twenty-two chromosomal regions, represented by a total of 52 microsatellite markers were included in the tests for gene interaction (Table 10). These regions were chosen because they were previously identified in the literature as regions containing susceptibility genes, because they showed suggestive evidence for linkage or association in our data, or because they contained candidate genes from a biological perspective.
Table 10. A list of the 22 chromosomal regions included in the GHP analyses.

<table>
<thead>
<tr>
<th>Chromosome Region</th>
<th>Number of markers</th>
<th>Average density (cM)</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>1p34-p36</td>
<td>3</td>
<td>7.00</td>
<td>Dyslexia susceptibility locus. Evidence for linkage and association.</td>
</tr>
<tr>
<td>1p32-p34</td>
<td>2</td>
<td>20.00</td>
<td>Evidence for linkage and association</td>
</tr>
<tr>
<td>3p12-q13</td>
<td>3</td>
<td>3.00</td>
<td>Dyslexia susceptibility locus.</td>
</tr>
<tr>
<td>4p16.3-p15.1</td>
<td>3</td>
<td>6.99</td>
<td>Evidence for linkage and association. DRD5B candidate gene.</td>
</tr>
<tr>
<td>4q35.1</td>
<td>1</td>
<td>NA</td>
<td>Evidence for linkage.</td>
</tr>
<tr>
<td>6p23-p21</td>
<td>4</td>
<td>5.75</td>
<td>Dyslexia susceptibility locus.</td>
</tr>
<tr>
<td>6q13-q15</td>
<td>3</td>
<td>2.03</td>
<td>Dyslexia susceptibility locus. Evidence for linkage.</td>
</tr>
<tr>
<td>6q25.3</td>
<td>1</td>
<td>NA</td>
<td>IGF2R candidate gene Evidence for association</td>
</tr>
<tr>
<td>7q31.3</td>
<td>2</td>
<td>1.00</td>
<td>Candidate locus involved in language disability and autism. Evidence for linkage.</td>
</tr>
<tr>
<td>8p21.2</td>
<td>1</td>
<td>NA</td>
<td>Evidence for association</td>
</tr>
<tr>
<td>Chromosome Region</td>
<td>Number of markers</td>
<td>Average density (cM)</td>
<td>Reason</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------</td>
<td>----------------------</td>
<td>--------</td>
</tr>
<tr>
<td>11p15.5</td>
<td>4</td>
<td>2.23</td>
<td>Dyslexia susceptibility locus. Evidence for linkage.</td>
</tr>
<tr>
<td>11q12-q13</td>
<td>2</td>
<td>6.02</td>
<td>FGF3 and FCER1B candidate genes</td>
</tr>
<tr>
<td>13q21.32</td>
<td>1</td>
<td>NA</td>
<td>Evidence for linkage. Candidate locus involved in autism.</td>
</tr>
<tr>
<td>14q13-q31</td>
<td>3</td>
<td>0.97</td>
<td>Evidence for linkage.</td>
</tr>
<tr>
<td>14q32.1</td>
<td>1</td>
<td>NA</td>
<td>Evidence for linkage.</td>
</tr>
<tr>
<td>14q32.3</td>
<td>2</td>
<td>1.02</td>
<td>Evidence for linkage.</td>
</tr>
<tr>
<td>15q21.1</td>
<td>1</td>
<td>NA</td>
<td>Dyslexia susceptibility locus. Evidence for association.</td>
</tr>
<tr>
<td>16p13.3</td>
<td>1</td>
<td>NA</td>
<td>Evidence for association.</td>
</tr>
<tr>
<td>17p13.1</td>
<td>1</td>
<td>NA</td>
<td>Evidence for linkage.</td>
</tr>
<tr>
<td>18p11</td>
<td>1</td>
<td>NA</td>
<td>Dyslexia susceptibility locus.</td>
</tr>
<tr>
<td>20q13.33</td>
<td>1</td>
<td>NA</td>
<td>Evidence for linkage.</td>
</tr>
</tbody>
</table>
Once the 22 chromosomal regions were identified, 52 markers were chosen to span them at an average density of 3.6 cM. Next, GHP was performed under the four different weight schemes to test for gene-interactions and heterogeneity among the 52 markers. This resulted in 2652 pairwise tests.

I identified 40 pairs of regions for which there was evidence for gene-interactions, and 28 pairs of regions for which there was evidence for heterogeneity at the significance level of 0.01, using either the discrete or proportional weighting scheme. Next, I only considered the regions for which there was significant evidence for gene-interaction or heterogeneity in both directions (i.e. weighting evidence for linkage at locus 1, while assessing linkage at locus 2 and the reverse). As a result, there were 16 pairs of regions that showed significant evidence for epistasis (Table 11), and 22 pairs of regions with significant evidence for heterogeneity (Table 12).
Table 11. Pairs of regions that show significant evidence for **epistasis** in GHP, by considering the increase in the weighted NPL score, using either the proportional or the discrete weighting schemes. These regions show significant (p<0.01) evidence for epistasis in the increase of the weighted NPL score compared to the single-locus NPL score, for locus 1 (by weighting linkage to locus 2); as well as significant (p<0.05) evidence for epistasis for locus 2 (by weighting linkage to locus 1).

<table>
<thead>
<tr>
<th>Set</th>
<th>Locus 1</th>
<th>Locus 2</th>
<th>Marker pairs</th>
<th>Increase in NPL with evidence for epistasis (weighted NPL p)</th>
<th>Increase in NPL score for markers at locus 1 (weighted NPL p)</th>
<th>Increase in NPL score for markers at locus 2 (weighted NPL p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1p34-p36</td>
<td>2p25-p16</td>
<td>1</td>
<td>0.00819</td>
<td>0.00118</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1p32-p34</td>
<td>2p25-16</td>
<td>6</td>
<td>0.00092</td>
<td>0.00199</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1p32-p34</td>
<td>7q31.3</td>
<td>2</td>
<td>0.00948</td>
<td>0.01901</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2p25-p16</td>
<td>6p23-21</td>
<td>2</td>
<td>0.00209</td>
<td>0.00525</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2p25-p16</td>
<td>6q25.3</td>
<td>1</td>
<td>0.00766</td>
<td>0.03160</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2p25-p16</td>
<td>17p13.1</td>
<td>1</td>
<td>0.00160</td>
<td>0.04288</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>6p23-p21</td>
<td>1p34-p36</td>
<td>1</td>
<td>0.00709</td>
<td>0.03995</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>6p23-21</td>
<td>6q12-q15</td>
<td>2</td>
<td>0.00047</td>
<td>0.03811</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>6p23-p21</td>
<td>7q31.3</td>
<td>6</td>
<td>0.00953</td>
<td>0.01042</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>6q25.3</td>
<td>20q13.33</td>
<td>1</td>
<td>0.00709</td>
<td>0.02060</td>
<td></td>
</tr>
</tbody>
</table>

* This represents the weighted NPL score p-value, which showed the most significant increase in the weighted compared to the single locus NPL score for all pairs of markers tested for that region pair.
Table 11 (continued)

<table>
<thead>
<tr>
<th>Set</th>
<th>Locus 1</th>
<th>Locus 2</th>
<th>Marker pairs</th>
<th>Increase in NPL with evidence for epistasis</th>
<th>Increase in NPL score for markers at locus 1&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Increase in NPL score for markers at locus 2&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>7q31.3</td>
<td>8p.21.2</td>
<td>1</td>
<td>0.00593</td>
<td>0.00932</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>11p15.5</td>
<td>1p34-p36</td>
<td>1</td>
<td>0.00119</td>
<td>0.01540</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>11p15.5</td>
<td>2p25-p16</td>
<td>1</td>
<td>0.00190</td>
<td>0.01321</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>14q13-q31</td>
<td>1p32-p34</td>
<td>1</td>
<td>0.00043</td>
<td>0.02351</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>14q13-q31</td>
<td>7q31.3</td>
<td>1</td>
<td>0.00701</td>
<td>0.02998</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>14q32.1</td>
<td>15q21.1</td>
<td>1</td>
<td>0.00927</td>
<td>0.03469</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> This represents the weighted NPL score p-value, which showed the most significant increase in the weighted compared to the single locus NPL score for all pairs of markers tested for that region pair.
Table 12. Pairs of regions that show significant evidence for **heterogeneity** in GHP, by considering the increase in the weighted NPL score, using either the proportional or the discrete weighting schemes. These regions show significant (p<0.01) evidence for heterogeneity in the increase of the weighted NPL score compared to the single-locus NPL score, for locus 1 (by weighting linkage to locus 2); as well as significant (p<0.05) evidence for heterogeneity for locus 2 (by weighting linkage to locus 1).

<table>
<thead>
<tr>
<th>Set</th>
<th>Locus 1</th>
<th>Locus 2</th>
<th>Marker pairs</th>
<th>Increase in NPL with evidence for heterogeneity (weighted NPL p)</th>
<th>Increase in NPL score for markers at locus 1a (weighted NPL p)</th>
<th>Increase in NPL score for markers at locus 2a (weighted NPL p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ip32-p34</td>
<td>l1q12-q13</td>
<td>1</td>
<td>0.00489</td>
<td>0.03178</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Ip32-p34</td>
<td>14q32.1</td>
<td>1</td>
<td>0.00389</td>
<td>0.03948</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2p25-p16</td>
<td>Ip34-p36</td>
<td>1</td>
<td>0.00745</td>
<td>0.03510</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2p25-p16</td>
<td>7q31.3</td>
<td>5</td>
<td>0.00311</td>
<td>0.00343</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2p25-p16</td>
<td>8p21.2</td>
<td>1</td>
<td>0.00494</td>
<td>0.02813</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2p25-p16</td>
<td>11p15.5</td>
<td>1</td>
<td>0.00070</td>
<td>0.01241</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2p25-p16</td>
<td>14q14-q31</td>
<td>3</td>
<td>0.00093</td>
<td>0.00351</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>6p23-p21</td>
<td>Ip32-p34</td>
<td>2</td>
<td>0.00819</td>
<td>0.01648</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>6p23-p21</td>
<td>2p25-p16</td>
<td>2</td>
<td>0.00927</td>
<td>0.00086</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>6p23-p21</td>
<td>14q13-31</td>
<td>2</td>
<td>0.00528</td>
<td>0.00278</td>
<td></td>
</tr>
</tbody>
</table>

*a This represents the weighted NPL score p-value, which showed most significant increase in the weighted compared to the single locus NPL score for all pairs of markers tested for that region pair.*
<table>
<thead>
<tr>
<th>Set</th>
<th>Locus 1</th>
<th>Locus 2</th>
<th>Marker pairs with evidence for heterogeneity</th>
<th>Increase in NPL score for markers at locus 1&lt;sup&gt;a&lt;/sup&gt; (weighted NPL p)</th>
<th>Increase in NPL score for markers at locus 2&lt;sup&gt;a&lt;/sup&gt; (weighted NPL p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>6p23-p21</td>
<td>17p13.1</td>
<td>1</td>
<td>0.00398</td>
<td>0.01292</td>
</tr>
<tr>
<td>12</td>
<td>6p23-p21</td>
<td>6q25.3</td>
<td>3</td>
<td>0.00766</td>
<td>0.00561</td>
</tr>
<tr>
<td>13</td>
<td>7q31.3</td>
<td>6q25.3</td>
<td>2</td>
<td>0.00169</td>
<td>0.01048</td>
</tr>
<tr>
<td>14</td>
<td>7q31.3</td>
<td>20q13.33</td>
<td>1</td>
<td>0.00257</td>
<td>0.00213</td>
</tr>
<tr>
<td>15</td>
<td>11p15.5</td>
<td>1p32-p34</td>
<td>2</td>
<td>0.00064</td>
<td>0.01284</td>
</tr>
<tr>
<td>16</td>
<td>11p15.5</td>
<td>6p23-p21</td>
<td>1</td>
<td>0.00716</td>
<td>0.03948</td>
</tr>
<tr>
<td>17</td>
<td>11p15.5</td>
<td>14q13-p31</td>
<td>1</td>
<td>0.00762</td>
<td>0.01336</td>
</tr>
<tr>
<td>18</td>
<td>14q13-q31</td>
<td>1p34-p36</td>
<td>1</td>
<td>0.00540</td>
<td>0.02520</td>
</tr>
<tr>
<td>19</td>
<td>14q13-q31</td>
<td>2p25-p16</td>
<td>1</td>
<td>0.00470</td>
<td>0.02716</td>
</tr>
<tr>
<td>20</td>
<td>14q13-q31</td>
<td>4q35.1</td>
<td>1</td>
<td>0.00398</td>
<td>0.01141</td>
</tr>
<tr>
<td>21</td>
<td>14q13-q31</td>
<td>6q25.3</td>
<td>2</td>
<td>0.00389</td>
<td>0.02037</td>
</tr>
<tr>
<td>22</td>
<td>20q13.33</td>
<td>1p34-p36</td>
<td>1</td>
<td>0.00192</td>
<td>0.00716</td>
</tr>
</tbody>
</table>

<sup>a</sup> This represents the weighted NPL score p-value, which showed most significant increase in the weighted compared to the single locus NPL score for all pairs of markers tested for that region pair.
The GHP results provide evidence that gene-interactions and heterogeneity are quite common in dyslexia. The most reliable findings are likely those in which more than one marker pair provides evidence for gene-interactions or heterogeneity, such as chromosomal pairs 1p32-p34 – 2p25-p16, 1p32-p34 – 7q31.3, 2p25-p16 – 6p23-p21, 6p23-p21 – 6q12-q15, and 6p23-p21 – 7q31.3 for epistasis. For heterogeneity, the corresponding regions are 2p25-p16 – 7q31.3, 2p25-p16 – 14q13-q31, 6p23-p21 – 1p32-p34, 6p23-p21 – 2p25-p16, 6p23-p21 – 14q13-q31, 6p23-p21 – 6q25.3, 7q31.3 – 6q25.3, 11p15.5 – 1p32-p34, 14q13-q31 – 6q25.3. Naturally, this comparison cannot be used for regions which were only represented by a single marker, such as regions 4q35.1, 8p21.2, 13q21.32, 15q21.2, 16p13.3, 17p13.1, 18p11, and 20q13.32. It is interesting that some region pairs appear in both interaction and heterogeneity results: 1p34-p36 – 2p25-p16, 2p25-p16 – 6p23-p21, and 2p25-p16 – 11p15.5. In these cases, it is always different marker pairs from the regions that show evidence for either interaction or heterogeneity. In analyzing the GHP epistasis results a framework of interactions seems to be formed (Figure 4). In the heterogeneity results, there are many complex pairings between regions and no clear clusters or separate networks can be observed.
Figure 4. A schematic of the different clusters formed by GHP epistatic analysis.
Correction for multiple testing

I used several methods to carry out corrections for multiple testing from the GHP results in the current study. First, I only considered regions that showed evidence for epistasis and heterogeneity at the p-value threshold of 0.01, and reduced those to pairs of regions that only showed evidence for interaction or heterogeneity at \( p < 0.05 \) in the reciprocal.

Second, I used two Bonferroni corrections for all the tests that I had performed in GHP given a particular genetic model. I first corrected for all the tests conducted given a particular weighting scheme within a model. This resulted in an overall number 2652 tests performed per model. The resulting p-value threshold, corresponding to \( p = 0.05 \), was corrected \( p = 0.000019 \), which gave no significant results.

The second Bonferroni correction I used was to take into account all the tests involving different chromosomal regions (as opposed to single markers). The reasoning for this method was that many of the markers within the same chromosomal region are in close linkage, and therefore not all the tests involving single markers are independent. There were 462 tests involving different chromosomal regions. Once that was performed, there were two regions that were near the level of significance corresponding to 0.05 (corrected \( p = 0.00014 \)) for epistasis only, 1p32-p34 – 14q13-q31 and 6p23-p21 – 6q12-q15.

Finally, I conducted a test to assess the overall significance of the results, i.e. whether there is overall significant evidence for locus interaction or heterogeneity in dyslexia, following the procedure proposed by Field et al. (1984). To assess the overall significance of the results, a Bonferroni correction is appropriate in cases when one test...
is significant, out of N tests performed. However, it is not appropriate to use a
Bonferroni correction in cases when multiple tests are significant (Field et al. 1984).
Field et al. (1984) propose an alternate method, which obtains the overall corrected p-
value by using the binomial probability. If X number of tests are significant at the Type I
error rate of $\alpha$ (0.01 in this case) out of N tests performed, the corrected p value $P_a [X of
N]$ is the probability that X of N tests are significant at $\alpha$ level by chance. This can then
be expressed as the binomial probability:

$$P(X \text{ of } N) = 1 - \sum_{i=0}^{X-1} \binom{N}{i} \alpha^i (1 - \alpha)^{N-i}$$

Using this method, the overall p-value for interactions was calculated to be
5.43$x10^{-14}$, using 40 significant pairs of regions (at p = 0.01) out of 924 tested (462 pairs
of regions and two weighting schemes). The corresponding p-value for the heterogeneity
results is 4.21$x10^{-7}$, using 28 significant pairs of regions out of 924. Thus, there appears
to be highly significant evidence for epistasis and heterogeneity in the sample of 100
dyslexia families.

Genehunter Two-Locus

Genehunter-Two Locus (GHT), a method that identifies primarily genetic
heterogeneity developed by Strauch et al. (2000), was used to replicate the heterogeneity
results from GHP. I performed the analysis using only one pair of markers per region
identified as giving significant evidence for heterogeneity in both directions (Table 9) in
GHP. There was only one pair of markers that resulted in a significant heterogeneity NPL
score in GHT, D2S2183 and D7S489, representing 2p25-p16 and 7q31.3. The NPL score
obtained for the pair was 1.89, corresponding to a p-value of 0.024. The corresponding
regions in the GHP results (Table 12) show evidence for genetic heterogeneity with five
pairs of markers, where each pair is composed of one marker at 2p25-p16 and one marker at 7q31.3.

Homogeneity Tests

Using this method I tried to replicate the 16 and 22 pairs of regions that showed significant evidence for gene-interaction and heterogeneity, respectively, in both directions in the GHP results. However, not many of the 52 markers chosen in the GHP analysis had family alpha values around 0.5. Thus only 15 tests were conducted. There was only one result which produced a maximum Lod score in the family subsets of greater than 1. Evidence for linkage at PLON, at 14q13-q31, increased from 0.89 for the 100 dyslexia families, to 1.48 for 73 families, which had alphas less than 0.5 for D1S199, at 1p34-p36. This confirms the heterogeneity result observed in GHP involving the same regions (Table 12).

4.4 Discussion

Several studies have shown that methods that consider variation at multiple loci are more powerful in disease-gene mapping than single-locus analysis both for parametric (Schork et al. 1993) and non-parametric (Knapp et al. 1994; Cordell et al. 1995) linkage analysis and association methods (Dizier et al. 1994). This is the first study of dyslexia that has examined multilocus variation in predisposition to the trait by using several different methods proposed in the literature. The results provide support for the involvement of previously reported dyslexia loci of major effect and identify several regions that might contain genes of moderate effect in the trait. Among the regions that show consistently significant evidence for epistasis and heterogeneity are loci 1p34-p36, 2p25-p16, and 6p23-p21, which have been identified as susceptibility loci in dyslexia in
independent family samples. These results indicate that locus heterogeneity is present in our sample, as has been suggested by previous studies of dyslexia from our group (Field and Kaplan 1998; Petryshen et al. 2000, 2001, 2002).

One study has previously briefly examined evidence for gene-interactions in a set of 9 dyslexic families (Grigorenko et al. 2001) using only GHP. The authors only considered two dyslexia susceptibility regions, 1p34-p36 and 6p23-p21, for which there was evidence for single-locus linkage in their data. We are able to replicate their findings of interaction between the two regions, suggesting that such an interaction may be important, at least in a subset of dyslexic families.

There was also significant evidence for gene-interaction and heterogeneity at chromosomal region 7q31.3. This region is an important candidate gene region in dyslexia because of the presence of a number of genes in the region involved in related disabilities. These include the SPCH1 gene, which is involved in language impairment (Fisher et al. 1998), and an autism susceptibility locus (IMGSAC 1998). In addition, a recent study of patients with autism that incorporated language phenotypes into the trait, found significant evidence for linkage to chromosomes 7q31.3 and 13q21.32 (Bradford et al. 2001). Both regions have previously been implicated in autism, however the significance of the results was much increased when language phenotypes were taken into account. It is interesting that in our sample of dyslexic patients, both regions 7q31.3 and 13q21.32 are involved in either suggestive single-locus evidence for linkage (13q21.32, Table 9), or gene-interactions and heterogeneity (7q31.3). The results of the current study therefore complement recent studies of autism, and imply that the link between autism and dyslexia suggested by the high frequency of language-related
problems in parents of autistic children may be explained by genetic variation at the 7q31.3 and 13q21.32 loci. Furthermore, the results demonstrate the importance of taking into account interactions between genes because genome-wide scans have not shown evidence that a locus in this region of chromosome 7 is involved in dyslexia (Cardon et al. 1994, 1995; Noppola-Hemmi et al. 2001; Fisher et al. 2002). It is possible that a locus at chromosome 7q31.3 is involved in both autism and dyslexia, but that the phenotypic effects will depend on the variation at other loci in the genome.

Markers on chromosome 14q also show significant evidence for gene-interactions and heterogeneity within our sample. Interestingly, this region has been implicated in type 1 diabetes (Field et al. 1996) and might therefore be linked to the higher proportion of immune deficits observed in dyslexic patients. However, this result should be treated with caution, because only approximately 60-70% of all individuals were genotyped for markers in this area, compared to 90-100% for most of the other markers included in the analyses in this chapter.

Chromosome 6p21-p23, a major predisposition locus for dyslexia, is also often significant in the interaction and heterogeneity results. However, previous work on the same families indicated that there is no evidence for single locus linkage or association to the region (Field and Kaplan 1998; Petryshen et al. 2000). Chromosome 6p might therefore be involved in dyslexia predisposition in our sample, but its effect on the phenotype is not independent of genetic variation at other loci. The results provide support for the rarely-tested hypothesis that many genes of modest single-locus effects may be of major phenotypic significance, but will not be detected unless taken into
consideration with genotypic variance at other loci across the genome (Frankel and Schork 1996; Phillips 1999; Templeton 2000).

Although most agree that epistatic interactions are important in determining the phenotype of many common diseases, the best way to detect epistasis remains unclear. A number of methods have been suggested, including, but not limited to, the candidate gene approach and oligogenic methods. The candidate gene approach (Templeton 2000), which is undertaken in the current study, selects candidate interaction or heterogeneity regions based on single locus results prior to conducting tests for epistasis and heterogeneity. The GHP method and the homogeneity tests both use the candidate gene approach, and are therefore based on prior examination of single-locus effects. Critics have suggested that this approach will lead to a bias towards the inclusion of loci with main or moderate effects on the trait. There is the possibility that many loci will go undetected, because their effects on the phenotype can only be observed within “an epistatic framework” (Frankel and Schork 1996).

Another set of studies proposes oligogenic methods to take into account variation across the genome. These methods conduct tests of gene-interaction without prior knowledge of single-locus effects. Programs that use this method include the GHT used in this study, the MLS method (Risch et al. 1990a, b, c) extended to account for gene-interactions (Cordell et al. 1995, 2000), the MLS method implemented in TLINK (Ott 1991), and the variance components method for quantitative traits used by SOLAR (Almasy and Blangero 1998; Mitchell et al. 2000). These approaches can be used on both candidate gene regions and genome-wide data. A number of methods are available to conduct tests of epistasis across the genome. Although this might be the most desirable
approach because it is unnecessary to first select candidate regions, there are a number of problems with current methods. Most importantly, it remains unclear how to determine the significance of the results statistically.

A central problem to the treatment of epistasis is obtaining statistical significance of the results. Significance tests are made difficult by the large number of tests conducted to account for the complexity of all possible combinations of interacting genes and the necessity to control for the number of tests. There are generally two methods to show statistical significance of a result in gene-mapping studies. Most studies use analytical approaches that calculate a formal statistic, which is then approximated to a normal distribution. Investigators choose threshold values for the statistic (e.g. Lod of 3.3), which correspond to genome-wide p-values of 0.05. Unfortunately, this approach is dependent on several key assumptions, such as normality of the statistic and the pooled independent meioses, which are often violated, especially when sample sizes are not large (Lander and Shork 1994). A better approach is to use simulation to estimate statistical significance of the results. In one of the gene-interaction methods applied in this study, GHP, Cox et al. (1999) propose both a formal statistic and a simulation method to evaluate the statistical significance of the results. It would be of great interest to apply both methods to the data presented in this report, and compare the number of significant epistatic and heterogeneity pairs as a result. Unfortunately, because of time constraints, the simulation approach was not used in the current study, but should be applied in future studies addressing the same questions.

Although most statistical research on gene-interactions has concentrated on how to detect significant effects, it is also of particular importance to be able to estimate the
power of the analyses (i.e. the sample size required to detect a significant effect given the
data). Very large samples are necessary to detect epistatic effects, and it is possible that a
sample size of 100 families, such as the one used in this study only detects the
interactions of highly significant effects. Statistical significance will also depend strongly
on specifying the correct epistatic model. Many different models of interactions can be
specified in different programs. Assumptions about the interaction model may be
problematic analytically, because of the large number of parameters that need to be
estimated to accommodate such a model (Frankel and Schork 1996).

The results of this study not only demonstrate that some regions appear to be
consistently important in gene-interactions, but also that different programs sometimes
give different results even when they are confronted with the same data (e.g. compare the
results of GHP with GHT and HOMOG). To date there has been insufficient research to
identify why the different methods give different results, or on the circumstances under
which it is more appropriate to use a particular program. Until such work has been
conducted, the results of the current study suggest that it is important to take advantage of
a number of the methods that are available to ascertain whether the results are consistent.
Much work is needed to synthesize the different methods.

Finally, this study only examines evidence for gene-interactions between pairs
of loci, while in reality gene-interactions may often involve greater than two loci.
Unfortunately, it would be necessary to examine hundreds to millions of possible gene-
combinations when testing for epistasis in a genome-wide scan. However, the goal of
epistasis research should be toward a method that is able to detect interaction among any
number of loci.
Epistasis appears to be a nearly universal component of the genetic architecture of most common traits and has profound evolutionary and clinical implications (Phillips 1999). In the genetic dissection of complex traits, most studies to date have not dealt with epistatic interactions, but rather with major susceptibility genes. Our ability to detect epistasis in most complex traits is hindered by the difficulties in treating epistasis statistically. There are only a few studies so far that have examined gene-interaction in complex traits, and clearly more work needs to be done in this area. Further research should focus on proposing models of epistatic effects and determining criteria for modifying statistically significant threshold values when using different approaches to detect epistasis in humans. Alternative methods in detecting epistatic effects may also be developed by extending current single-locus linkage analysis methods and developing methods that are not based on the linkage algorithm (see Nelson et al. 2001; Ritchie et al. 2001). Problems with the current methodology notwithstanding, the present study provides support for previous single-locus studies, suggests a number of novel candidate regions involved in the trait, and indicates evidence for a multilocus genetic system in the predisposition to dyslexia. The results provide direction for future research on dyslexia to narrow the number of regions that merit further study.
CHAPTER V THE SEARCH FOR HANDEDNESS GENES IN FAMILIES WITH DYSLEXIA

5.1 Introduction

Functional asymmetry for language in the brain was first inferred by Broca (1861) and Wernicke (1874). The planum temporale (PT) is an area in the temporal lobe of the brain, which is thought to be involved in language function (Geschwind and Levitsky 1968). The PT is larger on the left side of the brain in most control subjects, which is termed cerebral dominance (Geschwind and Miller 2001). The leftward asymmetry is reduced in individuals with dyslexia and schizophrenia, and in left-handed individuals. Several studies have proposed that abnormal cerebral laterality may be related to dyslexia, which has given rise to a number of theories about aberrant hemispheric communication and abnormal cortical development in certain language areas in the brain (see Habib 2000). Handedness is proposed to be strongly correlated to language lateralization (Geshwind and Miller 2001) and has been used in previous studies as a measure of such cerebral lateralization (Laval et al. 1998). The large majority (99%) of right-handed individuals show left hemisphere language localization (Geschwind 1970; Benson 1986). About 60% of left-handed individuals show left-hemisphere language dominance, 30% show bilateral lateralization and 10% show right hemispheric language localization (Geschwind and Miller 2001). Thus, a number of reports have examined the biology and genetics of handedness and have extended the findings to cerebral lateralization, which is thought to underlie relative hand status determination.
Family, twin and adoption studies have provided significant evidence for a genetic contribution to handedness (Annett 1974). The prevalence of left-handed individuals in the general population is approximately 11%, with more left-handed individuals among males than females (Laval et al. 1998). Although the exact genetic mechanisms contributing to handedness remain unknown, several major gene models have been proposed (Annett and Kilshaw 1984; McManus 1992). Two previous reports have directly examined evidence for a genetic basis in handedness. The first searched for a link between handedness and schizophrenia (Laval et al. 1998), while the second analyzed handedness in a set of families with dyslexia (Francks et al. 2002). Three regions have been implicated in handedness to date, on chromosomes Xq21, 2p11, and 17p11 (Laval et al. 1998; Francks et al. 2002). The recent genome-wide scan for handedness genes in a large sample of dyslexic sib-pairs reported the two loci on chromosomes 2p and 17p (Francks et al. 2002). This study will attempt to replicate the results in an independent sample of dyslexic families.

5.2 Methods

SUBJECTS AND GENOTYPE DATA

The 100 families ascertained for the Dyslexia project as described in section 2.1 also had information collected regarding handedness status. Of the 1088 members in the 100 families, 831 individuals had information on handedness status (Table 13). In total, 47 families had at least one left-handed and one right-handed individual per pedigree (Table 13). Linkage analysis was performed in the set of 47 families, comprising of 707 subjects. Genotypes were obtained as described in section 2.3.
Table 13. Description of handedness status information in (a) 100 Dyslexia families and (b) 47 Dyslexia families, each with at least one left-handed and one right-handed individual.

<table>
<thead>
<tr>
<th>Set</th>
<th>Number of subjects</th>
<th>Subjects with handedness data</th>
<th>All Subjects</th>
<th>Dyslexic Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>LH(^a)</td>
<td>RH(^b)</td>
<td>LH(^a)</td>
</tr>
<tr>
<td>(a) 100 Dyslexia families</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>1088</td>
<td>831</td>
<td>85 (10%)</td>
<td>746 (90%)</td>
</tr>
<tr>
<td>Males</td>
<td>587</td>
<td>449</td>
<td>55 (12%)</td>
<td>394 (88%)</td>
</tr>
<tr>
<td>Females</td>
<td>501</td>
<td>382</td>
<td>30 (8%)</td>
<td>354 (92%)</td>
</tr>
<tr>
<td>(b) 47 Families</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>707</td>
<td>515</td>
<td>85 (17%)</td>
<td>430 (83%)</td>
</tr>
<tr>
<td>Males</td>
<td>367</td>
<td>263</td>
<td>55 (21%)</td>
<td>208 (79%)</td>
</tr>
<tr>
<td>Females</td>
<td>340</td>
<td>252</td>
<td>30 (12%)</td>
<td>222 (88%)</td>
</tr>
</tbody>
</table>

\(^a\) LH = Left-handed

\(^b\) RH = Right-handed
PHENOTYPE DATA

Phenotype data on handedness was collected according to the methods proposed in Crovitz and Zener (1962). Subjects were asked to fill out a questionnaire assessing performance in four tasks: to print, to throw a ball, to draw a picture, and to hold scissors. Each item was scored on a five-point scale: always left, usually left, either hand, usually right, and always right. Always left was scored as 4, and always right was scored as 0. Altogether, each subject received a pooled score out of 16, where 0 indicates strongly right-handed status and 16 indicates strongly left-handed. Individuals who scored between 0 and 7 were assigned as right-handed, while those who score between 8 and 16 were assigned as left-handed. Figure 5 represents the distribution of the handedness scores in the sets of 100 and 47 dyslexia families.
Figure 5. Distribution of handedness codes in the 100 dyslexia families and 47 dyslexia families with at least one left-handed and one right-handed individual each.
STATISTICAL ANALYSES

Parametric Linkage analysis was performed as described in section 2.4 under eight genetic models (Table 14) including autosomal dominant, recessive, intermediate and genetic models proposed in the literature to increase the likelihood to detect linkage. Left-handed status was coded as affected. Non-parametric sib-pair linkage analysis was performed using the Sibpal program of the SAGE package (SAGE 1997).
Table 14. Genetic models used in parametric linkage analysis.

<table>
<thead>
<tr>
<th>Genetic Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penetrances:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aa</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.75</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.75</td>
</tr>
<tr>
<td>ab</td>
<td>0.9</td>
<td>0.1</td>
<td>0.5</td>
<td>0.75</td>
<td>0.9</td>
<td>0.1</td>
<td>0.5</td>
<td>0.75</td>
</tr>
<tr>
<td>bb</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9</td>
<td>0.98</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9</td>
<td>0.98</td>
</tr>
<tr>
<td>Disease allele b</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Disease allele b frequency
5.3 Results

Results from markers that indicated suggestive evidence for parametric and non-parametric linkage analysis are shown in Table 15 and 16. IGF2R marker on chromosome 6q25.3 showed suggestive evidence for both parametric linkage analysis and sibpair analysis.
<table>
<thead>
<tr>
<th>Chromosome region</th>
<th>Position (cM)</th>
<th>Marker</th>
<th>Maximum Lod Score (Theta)</th>
<th>Model</th>
<th>Number of genotyped individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>2p15</td>
<td>80.00</td>
<td>D2S1337</td>
<td>1.19</td>
<td>2,4</td>
<td>580</td>
</tr>
<tr>
<td>4q35.1</td>
<td>190.00</td>
<td>D4S2943</td>
<td>1.21</td>
<td>4</td>
<td>435</td>
</tr>
<tr>
<td>6q16.3-q27</td>
<td>143.01</td>
<td>D6S310</td>
<td>1.14</td>
<td>8</td>
<td>506</td>
</tr>
<tr>
<td>6q25.3</td>
<td>160.00</td>
<td>IGF2R</td>
<td>1.62</td>
<td>6,2</td>
<td>377</td>
</tr>
<tr>
<td>10q11.2</td>
<td>69.00</td>
<td>RBP3</td>
<td>1.14</td>
<td>6</td>
<td>572</td>
</tr>
<tr>
<td>14q11.2</td>
<td>12.00</td>
<td>D14S742</td>
<td>1.02</td>
<td>4</td>
<td>500</td>
</tr>
<tr>
<td>14q11.2</td>
<td>28.02</td>
<td>TCRD</td>
<td>1.17</td>
<td>5</td>
<td>295</td>
</tr>
</tbody>
</table>

Table 15. A list of all markers with Lod scores greater than 1 for the 47 dyslexia families.
Table 16. A list of all markers with Sibpal Linear Regression p-values less than 0.05.

<table>
<thead>
<tr>
<th>Chromosome region</th>
<th>Position (cM)</th>
<th>Marker</th>
<th>Linear Regression p (left-handed Affected)</th>
<th>Number of genotyped sibpairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>4p13-p12</td>
<td>61.00</td>
<td>GABAR1</td>
<td>0.05</td>
<td>597</td>
</tr>
<tr>
<td>6q25.3</td>
<td>160.00</td>
<td>IGF2R</td>
<td>0.0026</td>
<td>350</td>
</tr>
<tr>
<td>11q25</td>
<td>131.00</td>
<td>D11S912</td>
<td>0.017</td>
<td>557</td>
</tr>
<tr>
<td>12q21.33</td>
<td>95.00</td>
<td>D12S1064</td>
<td>0.024</td>
<td>539</td>
</tr>
<tr>
<td>15q26</td>
<td>116.00</td>
<td>D15S87</td>
<td>0.0066</td>
<td>128</td>
</tr>
</tbody>
</table>
**5.4 Discussion**

A genome-wide screen of handedness status was conducted in a sample of 47 dyslexic families with at least one right-handed and one left-handed individual each. The strongest evidence for linkage using parametric and non-parametric methods was obtained for the IGF2R gene, on chromosome 6q25.3. The IGF2R gene has been previously implicated in cognitive ability (Chorney et al. 1998), and for that reason was selected in the dyslexia study. Also, as shown in Chapter IV, IGF2R shows evidence for association with dyslexia and for both gene-interaction and heterogeneity. Therefore, IGF2R is an interesting region in the genetic basis of handedness and deserves further study in independent samples.

Two markers on chromosome 14q11 also show suggestive evidence for parametric linkage to handedness status. This region on chromosome 14 was also part of the gene-interaction and heterogeneity results in Chapter IV. It may be related to the higher proportion of immune deficits in dyslexics, as this region is also implicated in SLE (Moser et al. 1998) which is a disorder of the immune system. It should be noted that one of the models that have been proposed to explain handedness status, the GBG (Geschwind-Behan-Galaburda) model states that an underlying biological factor affects hand status determination (Geschwind and Behan 1982; Galaburda 1989). The model also speculates that left-handed status, dyslexia, and immune deficits are linked by a common mechanism (Kaplan and Crawford 1994). Therefore the findings in this study may support such a hypothesis, and merit further research in future studies.

It is possible to compare the results to those from the previous handedness genome-scan (Francks et al. 2002) performed in an independent set of dyslexic families.
Francks et al. (2002) reported significant evidence for linkage of handedness to chromosomes 2p and 17. I was unable to replicate these findings. However, the methods used in our study were slightly different. Francks et al. (2002) used quantitative data for handedness and had therefore greater power to detect loci of moderate effects on the trait. In contrast, I used a qualitative phenotype measure, which makes it more likely that the results point only toward genes of major effect. It is very interesting that Francks et al. (2002) did not detect a linkage signal near the IGF2R gene, which presents most evidence for linkage in our data. Handedness is a complex trait, and is likely to be affected by a large number of genetic and environmental factors. Therefore, it is not surprising that such findings cannot be replicated across independent samples.

According to some hypotheses, the same genetic mechanisms and therefore the same genetic loci may be involved in both dyslexia and handedness. Surprisingly, none of the previously reported dyslexia susceptibility loci show evidence for parametric or non-parametric linkage to handedness. The only suggestive result was the Lod score for marker D2S1337, which is at the proximal end of the DYX3 susceptibility region for dyslexia. Interestingly, Francks et al (2002) also obtained weak evidence for linkage to chromosome 2p.

It is possible to speculate that a link exists between dyslexia, schizophrenia, and left-handed status, involving the morphology and function of the PT even if there is no evidence in the literature for high co-morbidity of these disorders. Interestingly, one of the dyslexia susceptibility loci recently identified is located on chromosome 18p11. There is evidence that the same region is involved in predisposition to schizophrenia in several family sets (Fang et al. 1995; Hampson et al. 1999). The genetics of handedness has
previously been examined in relation to schizophrenia genetics, implicating a locus on the X chromosome. Unfortunately, our sample does not have genotype data on markers spanning the sex chromosomes.

In conclusion, a genome-side scan of handedness has been completed in a sample of dyslexic patients. The most significant finding of the study is the suggestive linkage to the IGF2R gene in handedness status determination. Handedness is a complex trait, in which non-genetic factors are likely to play a significant role. Therefore, future studies are needed to replicate this finding.
CHAPTER VI CONCLUSION

The aim of this study was to provide an insight into the complex genetic mechanisms that underlie dyslexia. I have explored three aspects of dyslexia – the effect of one particular locus on the trait, the possibility of a multilocus mechanism in dyslexia, and the genetic relationship between handedness and dyslexia.

The results reported in Chapter III indicate that a locus on chromosome 1p34-p36 significantly contributes to dyslexia. This finding is very interesting, especially when considered with previous reports of weak and significant linkage to the same region in two independent dyslexia family sets. The implications of the three studies are that there is a gene in this area that has a moderate, but significant effect on dyslexia. It is likely that either variation at that gene accounts for a proportion of the dyslexic cases, or that the gene indirectly contributes to the trait by an additive or interactive mechanism.

The results presented in Chapter IV show evidence for both locus heterogeneity and interaction among four dyslexia susceptibility loci identified to date and novel candidate gene loci. Although the novelty and complexity of this area of study precludes any definitive answers, this work presents exciting new results on the genetics of dyslexia. Epistatic interactions are thought to be very common in most traits, but are especially difficult to identify in humans. The methods applied in this work are currently rudimentary, but provide very interesting results which should be studied in more detail. This area of genetics deserves much more work and I believe will become a more common theme in the future.

Chapter V examines the genetic relationship between handedness and dyslexia. While many studies discuss the genetics of handedness, there are very few studies that
examine the genetic basis of handedness. Because of the purported higher incidence of left-handed individuals among dyslexic compared to the general population, and the involvement of similar regions in the brain for both handedness determination and dyslexia, previous research has suggested a genetic link between handedness and dyslexia. However, the results do not support such a conclusion for previously identified dyslexia loci, but indicate the involvement of IGF2R which may be a significant factor in both traits.

Future work in the genetics of dyslexia should focus on several areas. First, it is necessary to narrow down the previously identified susceptibility regions for dyslexia and identifying candidate genes within these regions. Second, it is important to study gene-interactions and heterogeneity in independent dyslexia samples. Interactions among more than two loci should also be examined. Future studies are needed to develop more sophisticated methods that take into account variation at more than one or two loci. Finally, examining subsets of dyslexic patients, which exhibit behavioral overlaps with other co-morbid phenotypes, such as autism, ADHD, depression and potentially schizophrenia may also lead to the identification of novel susceptibility genes in dyslexia.
REFERENCES


Interaction of HLA and immunoglobulin antigens in type 1 (insulin-dependent) diabetes. Diabetologia 27:504-8


Localisation of a gene implicated in a severe speech and language disorder. Nat Genet 18:168-70


Larsen JP, Hoien T, Lundberg I, Odegaard H (1990) MRI evaluation of the size and
symmetry of the planum temporale in adolescents with developmental dyslexia. Brain
Lang 39:289-301

Lathrop GM, Lalouel JM (1984) Easy calculations of lod scores and genetic risks on

Laval SH, Dann JC, Butler RJ, Loftus J, Rue J, Leask SJ, Bass N, Comazzi M, Vita A,
(1998) Evidence for linkage to psychosis and cerebral asymmetry (relative hand skill)

Adolesc Psychiatry 28:326-331

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

MacLean CJ, Sham PC, Kendler KS (1993) Joint linkage of multiple loci for a complex

Markianos K, Daly MJ, Kruglyak L (2001) Efficient Multipoint Linkage Analysis

maps using an expert system (MultiMap): a human genome linkage map. Nat Genet
6:384-90

Serotonin depletion during synaptogenesis leads to decreased synaptic density and


Risch N (1990b) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am J Hum Genet 46:229-41


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


Wernicke C (1874) Der ophasische symptomenkomplex. Eine Psychologische Studie auf anatomische basis. Max Cohn und Weigert, Bresiam


