DETECTING GENES FOR DEVELOPMENTAL DYSLEXIA ON CHROMOSOME 6p

NATALIE ALEXANDRA COPE BSc

DEPARTMENT OF PSYCHOLOGICAL MEDICINE

WALES COLLEGE OF MEDICINE

CARDIFF UNIVERSITY

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DETECTING GENES FOR DEVELOPMENTAL DYSLEXIA ON CHROMOSOME 6p

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ....

Date 14/07/06

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed ....

Date 14/07/06

STATEMENT 2

I hereby give consent for my thesis, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ....

Date 14/07/06

CONTENTS

Chapter Two: The Genetic Basis of Developmental Dyslexia

Chapter Three: Materials and Methods

Chapter Four: Positional Candidate Gene Association Studies Between Developmental Dyslexia and Chromosome 6p: The Identification of *KIAA0319* as a Susceptibility Gene for Developmental Dyslexia

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Prof. Julie Williams and Prof. Mick O'Donovan for their supervision and help during my PhD and for giving me the opportunity to do a PhD in the department. I would also like to thank all the members of Psychological Medicine for their help and advice, with particular thanks to Denise, Elaine and Darko, for answering my constant flow of questions, and to the members of 'our office' for providing a nice environment in which to work.

Many thanks have to go to all the children and families that took part in this research – without their involvement this research would not be possible.

Finally I would like to thank my mum, dad and family for all their love, help and encouragement throughout school and university and for their support throughout my PhD.

PUBLICATIONS

1) N Cope, D Harold, G Hill, V Moskvina, J Stevenson, P Holmans, M J Owen, M C O'Donovan, J Williams (2005) Strong Evidence That KIAA0319 on Chromosome 6p Is a Susceptibility Gene for Developmental Dyslexia. American Journal of Human Genetics 76(4): 581-591

2) N Cope, G Hill, M van den Bree, V Moskvina, M Owen, J Williams, M O'Donovan (2005) No Support for Association Between Dyslexia Susceptibility 1 Candidate 1 and Developmental Dyslexia *Molecular Psychiatry* 10(3): 237-238

ABBREVIATIONS

A	Adenine
ADHD	Attention Deficit Hyperactivity Disorder
ALDH5A1	Succinate-Semialdehyde Dehydrogenase (gene - human)
ANOVA	Analysis of Variance
APC	Antigen Presenting Cell
APD	Auditory Processing Disorder
APE1	Apurinic/Apyrimidinic Endonuclease I
APM	Affected Pedigree Member
BAS	British Ability Scales
BASR	British Ability Scale Single Word Reading Score
BASS	British Ability Scale Spelling Score
BF	Properdin Factor
bp	Base Pairs
BSA	Bovine Serum Albumin
C.	Cysteine
C	Cytosine
°C	Degrees Celsius
C6orf62	Chromosome 6 Open Reading Frame 62 (gene - human)
CD	Central Domain
cDNA	Complementary Deoxyribonucleic Acid
CEPH	Centre d'Etude du Polymorphisme Humain
CFRS	Colorado Family Reading Study
CI	Confidence Interval
CLS	Connecticut Longitudinal study
CM	Centimorgan
CNS	Central Nervous System
CRP	Colorado Reading Project
CTLs	Cytotoxic T Lymphocytes
Da	Daltons
DCAMKL1	Doublecortin and CaM Kinase-Like 1 (gene – human)
DCDC2	Doublecortin Domain-Containing 2 (gene - human)
dCTP	Deoxycytidine Triphosphate
DCX	Doublecortin (gene – human)
DCX	Doublecortin Domain Containing Proteins
DD	Developmental Dyslexia
ddNTPs	Dideoxynucleotide Triphosphate
df	Degrees of Freedom
dHPLC	Denaturing High Performance Liquid Chromatography
DMSO	Mimethyl Sulphoxide
DNA	Deoxyribonucleic Acid
dNTPs	Deoxynucleotide Triphosphate
dNTPs	Deoxynucleotide Triphosphate
dsDNA	Double-Stranded Deoxyribonucleic Acid
DSM	Diagnostic and Statistical Manuel
DTI	Diffusion Tensor Imaging
DYX1	Region on chromosome 15q implicated in DD susceptibility by linkage and association studies
DYX1C1	Dyslexia Susceptibility 1 Candidate 1 (gene - human)

DYX2	Region on chromosome 6p implicated in DD susceptibility by linkage and association studies
DYX3	Region on chromosome 2p implicated in DD susceptibility by linkage and association studies
DYX4	Region on chromosome 6q implicated in DD susceptibility by linkage and association studies
DYX5	Region on chromosome 3 implicated in DD susceptibility by linkage and association studies
DYX6	Region on chromosome 18p implicated in DD susceptibility by linkage and association studies
DYX7	Region on chromosome 11p implicated in DD susceptibility by linkage and association studies
DYX8	Region on chromosome 1p implicated in DD susceptibility by linkage and association studies
DYX9	Region on chromosome Xq implicated in DD susceptibility by linkage and association studies
DZ	Dizygotic
E	Glutamic Acid
EDTA	Ethylenediaminetetraacetic Acid
EGFP	Enhanced Green Fluorescent Protein
EKN1	Dyslexia Susecptibility 1 Candidate 1 (protein - human)
EKN1	Dyslexia Susceptibility 1 Candidate 1 (gene - human)
Elk-1	p62 Ternary Complex Factor
ExoI	Exonuclease I
F	Frequency
FAM	Fluorescein
FBAT	Family Based Association Test
[F]ddNTP	Fluorescently Labeled Dideoxynucleotide Triphosphate
FISH	Fluorescence *In Situ* Hybridisation
fMRI	Functional Magnetic Resonance Imaging
FTD	Familial Frontotemporal Dementia
G.	Glycine
G	Guanine
gDNA	Genomic Deoxyribonucleic Acid
GH	Growth Hormone
GLO / GLO1	Glyoxylase 1
GLUE	Graphical Interface
GPLD1	Glycosylphosphatidylinositol-Specific Phospholipase D1 (gene - human)
GRR	Glycine Rich Region
h^2_g	Heritability
HCL	Hydrochloride
HLA	Human Leukocyte Antigen
HLH	Helix-Loop-Helix
HLOD	Logarithm of the Odds Assuming Heterogeneity
HMG	High Mobility Group
HSTF	Heat Shock Gene Transcription Factor
HW	Hardy Weinberg
I	Isoleucine
IBD	Identical by Descent

ICD	International Classification of Diseases
ID4	DNA-Binding Inhibitor Protein 4 (gene – human)
ID4	DNA-Binding Inhibitor Protein 4 Protein
Ig	Immunoglobulin
IQ	Intelligence Quotient
IS	Inattention symptoms
ITIM	Immunoreceptor Tyrosine-Based Inhibitor Motif
k	Correction Factor
KAAG1	Kidney-Associated Antigen 1 (gene - human)
kb	Kilobase
kDa	Kilo Daltons
KIAA0319	*KIAA0319* (gene – human)
KIAA0319	KIAA0319 (protein – human)
KIAA0396	FEM-1-Like death receptor binding (gene – human)
KIAA0396	FEM-1-Like death receptor binding protein
KIAA1299	SH2-B β-signalling (gene – human)
KIAA1299	SH2-B β-signalling protein
µl	Microlitre
LD	Linkage Disequilibrium
LEA	Local Education Authorities
LOD	Logarithm of the Odds
Mb	Megabases
MDP	Mg^{2+}/Mn^{2+} -Dependant Phosphodiesterases
MEG	Magnetoencephalopathy
MFLOD	Model-Free Logarithm of the Odds
mg	Milligrams
Mg^{2+}	Magnesium Ion
$MgCl_2$	Magnesium Chloride
ml	Millilitres
Mn^{2+}	Manganese Ion
MRC	Medical Research Council
MRI	Magnetic Resonance Imaging
mRNA	Messenger Ribonucleic Acid
MRS	Magnetic Resonance Spectroscopy
MRS2L	Magnesium Homeostasis Factor (gene - human)
ms	Milliseconds
MZ	Monozygotic
NARA	Neale analysis of reading
NARA	Neale Analysis of Reading Ability
NCBI	National Center for Biotechnology Information
NF- κB	Nuclear Factor – Kappa B (protein – human)
ng	Nanograms
nm	Nanometers
NPL	Non Parametric Linkage
NWR	Non-word reading score
NWRT	Non-word repetition task
OC	Orthographic Coding
OR	Odds Ratio
P	Proline
PCR	Polymerase Chain Reaction

PD	Phoneme deletion
PDP	Parallel Distributed Processing
PET	Positron Emission Tomography
PHJT	Pseudohomophone judgement task
PIQ	Performance Intelligence Quotient
PKD	Polycystic Kidney Disease
PKD1	Polycystic Kidney Disease 1 (protein – human)
PL	Placental Lactogen
PLA2G4B	Phospholipase A_2, Group IVB (gene – human)
PLCB2	Phospholipase C β 2 (gene – human)
pmol	Picomolar
pPRL	Pituitary Prolactin
PRL	Prolactin (gene – human)
PRL	Prolactin (protein – human)
PRLR	Prolactin Receptor
QTL	Quantitative Trait Locus
RAN	Rapid Automated Naming
RD	Specific Reading Disability
RDN	Rapid Digit Naming
READ	Single Word Reading (Francks et al. 2004)
RFLP	Restriction Fragment Length Polymorphism
Rh	Rhesus Blood Group
RNA	Ribonucleic Acid
RO	Rhyme Oddity
rpm	Revoluations Per Minute
RPN	Rapid Picture Naming
RP1	Retinitis Pigmentosa 1 (gene – human)
RP1L1	Retinitis Pigmentosa 1-like 1
RT	Reverse Transcription
S	Serine
SAP	Shrimp Alkaline Phosphate
SD	Standard Deviation
SDS	Sodium Dodecyl Sulfate
SH2-B	Beta Signalling Protein (gene – human)
SLI	Specific Language Impairment
SNP	Single Nucleotide Polymorphism
SOX4	Sex-Determining Region Y –BOX (SRY-BOX) member 4 (gene – human)
SPSS	Statistical Package for the Social Sciences
SR	Sulforhadamine
SRR	Serine Rich Region
SRY	Sex Determining Region Y
SSADH	Succinate-Semialdehyde Dehydrogenase (gene - human)
SSLD	Severe Speech and Language Disorder
STM	Short Term Memory
T	Thymine
TBE	Tris-Borate EDTA
TDT	Transmission Disequilibrium Test
TFII-I	Transcription Factor II-I
THEM2	Thioesterase Superfamily Member 2 (gene - human)

TNF	Tumour Necrosis Factor
TNFB	Tumour Necrosis Factor Beta
TPR	Tetratricopeptide Repeat
TRAF	Tumour Necrosis Factor Receptor-Associated Factor
TTRAP	TRAF and TNF Receptor Associated Protein (gene - human)
TTRAP	TRAF and TNF Receptor Associated Protein (protein - human)
U	Units
UHW	University Hospital of Wales
UK	United Kingdom
US	United States
USA	United States of America
UTR	Untranslated Region
UV	Ultra Violet
V	Valine
VMP	Vesicular Membrane Protein p24 (gene - human)
VMP	Vesicular Membrane Protein p24 (protein - human)
VNTR	Variable Number Tandem Repeat
WAIS	Wechsler Adult Intelligence Scale
WRAT	Wide Range Achievement Test
WRMT	Woodcock Reading Mastery Test
X	STOP Codon
α	Alpha
θ	Recombination Fraction
λ	Wavelength
χ^2	Chi-Square Test of Statistical Significance

SUMMARY

Developmental dyslexia (DD) is a complex, cognitive disorder, characterised by an impairment in reading despite adequate educational, motivational and intellectual opportunities and in the absence of any sensory or neurological disability. Family and twin studies have shown that genes make a substantial contribution to individual variation in risk of DD.

Genetic linkage and association studies have implicated a number of chromosomal regions that may harbour susceptibility genes for DD, including regions on chromosomes 6p and 15q. The aims of this thesis were to identify novel susceptibility gene(s) for DD on chromosome 6p and to replicate the association reported between DD and *EKN1* on chromosome 15q.

Eleven genes on chromosome 6p were tested for association with DD using data derived from DNA pooling assays of 168 SNPs. Nineteen associations were observed and a minimum set of 13 SNPs were chosen for individual genotyping in a case-control and family-based sample. Nine SNPs revealed association with DD ($p \leq 0.03$) located in *PRL* (1 SNP), *MRS2L* (1 SNP), *KIAA0319* (4 SNPs), *THEM2* (2 SNP) and 1 intergenic SNP. A haplotype comprising rs4504469/rs6935076 (*KIAA0319*) revealed strong evidence for association with DD ($p=0.0001$). This combined with the results of logistic regression analyses suggests that variation within *KIAA0319* increases susceptibility to DD.

Component phenotype analysis of the rs4504469/rs6935076 haplotype suggested that variation on the haplotype may influence a number of components of reading. It may also influence single word reading across the normal ability spectrum, but for other component phenotypes, variation on rs4504469/rs6935076 may influence affection status only.

Association between DD and *EKN1* was tested in a large family-based sample. No association was observed between SNPs previously reported to show association with DD ($p \geq 0.20$). No significant associations were observed between *EKN1* and component phenotypes of DD.

This study identifies *KIAA0319* as a susceptibility gene for DD and suggests that *EKN1* is unlikely to increase vulnerability to DD.

CHAPTER ONE

AN INTRODUCTION TO DEVELOPMENTAL DYSLEXIA

CHAPTER ONE

1. AN INTRODUCTION TO DEVELOPMENTAL DYSLEXIA

'Dyslexia...explains why some very smart people have trouble learning to read'
Sally E Shaywitz, 1996

'...I am not ashamed to declare that I cannot recite the alphabet nor spell.... As I grew older and achieved success...I glossed over it and compensated for my inabilities...I am striving for success and still have ambitions that drive me to reach the highest standards of which I am capable'
Jackie Stewart *OBE*, 1997

1.1 A HISTORY OF DEVELOPMENTAL DYSLEXIA

Developmental dyslexia (DD) was first reported in 1896 by the general doctor, W. Pringle Morgan who described a 14-year-old boy, 'Percy F' (Morgan 1896). He said of Percy, 'His great difficulty has been – and is now – his inability to learn to read... I have no doubt it is due to some congenital defect... he would be the smartest lad in the school if instruction were entirely oral'.

James Hinshelwood wrote extensively on 'word blindness' (now referred to as developmental dyslexia) (Hinshelwood 1900; Hinshelwood 1907; Hinshelwood 1911; Hinshelwood 1917), noting that 'word blindness' could be hereditary and seemed to be more common in boys than girls (Hinshelwood 1917). He suggested that 'word blindness' was a pathological condition caused by damage to a 'visual word-centre' in the brain (Hinshelwood 1917).

In contrast, Samuel Orton believed that DD was an abnormality of physiological development (Miles and Miles 1999). He also suggested that the name, 'congenital word blindness', overstressed the inherent difficulty and underemphasized environmental factors involved in the disorder and recommended the disorder be classified as 'developmental' since it emphasised both hereditary that

tendency and environmental forces influencing the individual. He also pointed out that there is no blindness for words *per se* in the disorder. As a result 'strephosymbolia' ('twisting of symbols') was suggested as a more informative description of the disorder (Orton 1937 / 1989). Hallgren (Hallgren 1950) discarded the names of 'word blindness' and 'strephosymbolia' and penned the term 'dyslexia'. Critchley (Critchley 1981) argued that whilst labels are unimportant if the correct recommendations are made to the patient, the term 'dyslexia' implies not only a delay in learning to read, but expresses a difficulty in the use of words, how they are identified and handled and how they are pronounced and spelt.

1.2 DEFINING DEVELOPMENTAL DYSLEXIA

Dyslexia is derived from the Greek '*dys*' meaning 'difficulty' or 'malfunction' and the root-word '*lexis*' meaning language, literally translating to 'difficulty with words' and therefore suggests not simply a problem with reading, but also with spelling, writing and other aspects of language (Critchley 1981; Thomson 1990). DD (or specific reading disability (RD)) refers to dyslexia where individuals fail to develop competent reading skills rather than having lost their ability to read competently through the result of brain damage ('acquired dyslexia').

Many definitions of DD have been described however no definitive definition has been accepted. The World Federation of Neurology in 1968 described DD as:

> '*A disorder in children who, despite conventional classroom experience, fail to attain the language skills of reading, writing and spelling commensurate with their intellectual abilities.*' (Waites 1968)

Both the World Health Organisation (WHO, 2003) and the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV, 2000) define dyslexia as:

'*... an unexpected, specific and persistent failure to acquire efficient reading skills despite conventional instruction, adequate intelligence, and sociocultural opportunity.*'

1.3 CONSEQUENCES OF DEVELOPMENTAL DYSLEXIA

The ability to read is fundamental for success in society (Shapiro 2001). Poor reading ability has been linked with failure to graduate high school, unemployment, welfare dependency, criminal behaviour and mental disorders (see Schonhaut and Satz 1983). That said, many dyslexics have become successful in business, politics, literature, science, sport and entertainment.

1.4 DEVELOPMENTAL DYSLEXIA

Studies undertaken to determine whether DD is a pathological condition or whether it lies at the extreme end of a normal reading continuum take a number of forms. The first compares the two groups of poor readers (discrepant and non-discrepant) on reading related tasks. Results from these studies tend to yield inconsistent results due to the effect of age, DD definition and choice of outcome measures (Shapiro 2001). Other studies compare older children with DD to younger non-DD children who read at the same age as them. If the two groups do not differ in reading related skills, it is concluded that poor readers are acquiring reading normally but at a slower rate than most, supporting the developmental lag hypothesis. A third type of study compares children with DD to children of the same age chronologically. These studies can identify deviations from normal reading

ability in DD individuals at a given age, but do not indicate whether a developmental lag is present unless a comparison to a group matched on reading ability (i.e. of younger aged children) is undertaken.

Using the Connecticut Longitudinal Study (CLS) sample, Shaywitz and colleagues (Shaywitz et al. 1992) have suggested that reading ability and reading disability occur along a continuum, where dyslexic individuals form the lower tail end of the normal distribution. No 'cut point' could be used to distinguish children with DD from typically normal readers. Later, Shaywitz and colleagues (Shaywitz et al. 2001) observed that over time, poor readers and good readers tend to maintain their relative positions along the reading ability spectrum.

Badian (Badian 1996) did not observe a difference in reading ability between DD children and reading-level matched controls, suggesting DD does not form a distinct pathological condition in children aged 6-7 years, however in older children (8-10 years) support for a phonological deficit in DD was observed, where the reading-discrepant group could be distinguished from the non-reading-discrepant group, a finding supported by Meyers and colleagues (Meyer et al. 1998).

Olson and colleagues (Olson et al. 1989) found evidence consistent with DD forming a pathological condition, although 20% of individuals suggested that DD was a developmental lag in reading. It was posed that developmental deficits rather than a lag may be present depending on the remediation of DD, the presence of a normal IQ and whether normal reading instruction was experienced (Olson et al. 1989; Olson et al. 1990; Snowling et al. 2003; Carroll and Snowling 2004). Remediation was shown to improve reading related measures (such as phonological coding) (Olson et al. 1990) and alter results to suggest a developmental lag in reading. Remediation was also found to improve reading in dyslexic readers by

Hatcher and colleagues (Hatcher et al. 2004).

1.5 EPIDEMIOLOGY OF DEVELOPMENTAL DYSLEXIA

1.5.1 Prevalence of Developmental Dyslexia

DD is the most common of the learning disabilities, accounting for 80% of learning disabled diagnoses (Lerner 1989). In clinic and school identified samples, DD has a prevalence of 5-10%. In unselected population-based samples of children the prevalence is more akin to 17.5% (Shaywitz et al. 1994; Shaywitz 1998) depending on DD criteria and sample size.

Yule and colleagues (Yule et al. 1974) reported DD to have a prevalence of 3.1% in children aged 10 years from the Isle of Wight and 6.3% in 10-year-old children from London. Rodgers (Rodgers 1983) estimated the prevalence of DD to be slightly less at 2.29% using a British sample. In a US sample, Shaywitz and colleagues (Shaywitz et al. 1990) estimated the prevalence of DD to be 5.6% in 1st grade children, 7% in 3rd grade children and 5.4% in 5th grade children.

Difference in prevalence rates could be due to the tests used to calculate IQ and reading ability in children, the geographical location of the children and the definition used to classify children as affected. Although many of the investigations have been undertaken using English as the first language of children, the use of other languages, particularly transparent languages (such as Welsh, Italian and Spanish) rather than the non-transparent languages (such as English and French) may also influence the prevalence of DD in a population.

1.5.2 Gender Differences in Developmental Dyslexia

Hinshelwood (Hinshelwood 1917) suggested DD was more common in males. Supporting this, Bakwin (Bakwin 1973) reported that DD was nearly twice as common in males than females, a view supported by Finucci and Childs (Finucci and Childs 1981) who estimated the prevalence of DD to be 2-4 times higher in males than females in a clinic-referred sample. However, even before this study, doubt existed over the gender differences with Naiden (Naiden 1976) suggesting the increased prevalence of DD in males was the result of sample ascertainment bias.

Summarising gender ratios in five independent studies (DeFries and Decker 1982; Smith et al. 1983; Stevenson et al. 1984; DeFries et al. 1987; Gilger et al. 1991), Wadsworth and colleagues (Wadsworth et al. 1992) found that only in studies with referred or clinical populations (DeFries and Decker 1982; Gilger et al. 1991) did the gender ratios differ from 1:1. Research-identified samples (Stevenson et al. 1984; DeFries et al. 1987; Gilger et al. 1992b) revealed no gender differences.

Supporting Naiden (Naiden 1976), research-identified samples of children with DD, show male to female ratios of approximately 1.4:1 (Hallgren 1950; Sladen 1970; Shaywitz et al. 1990; DeFries et al. 1991; Wadsworth et al. 1992; DeFries and Gillis 1993; Flynn and Rahbar 1994; Marlow et al. 2001). Indeed Shaywitz (Shaywitz et al. 1990) found no significant difference in the prevalence of DD between males and females at either 2nd or 3rd grade in a research-identified sample. The identification of significantly more boys with DD (approximately 2-4 times more than females) in a school-identified sample in either grade, supported earlier observations (Finucci and Childs 1981) and those of a more recent study (Sauver et al. 2001). However, given that teachers rate boys significantly more active, inattentive and as having more behaviour, language and academic problems than

females, resulting in more males being referred for help (Shaywitz et al. 1990; Vogel 1990), evidence suggests that the increased frequency of DD in males may be the result of ascertainment bias. Indeed further support for sample ascertainment bias comes from family studies of DD. Family studies, for example, Hallgren (Hallgren 1950) report a ratio of 3.3:1 (males: females) in index cases but only 1.3:1 in affected siblings, a finding supported by Vogler and colleagues (Vogler et al. 1985). If DD was more frequent in males the ratio of males to females in siblings of index case would likely be higher than that observed.

In addition to ascertainment bias, a higher frequency of males with DD than females could be the result of a gender bias in the prediction of reading from IQ. Share and Silva (Share and Silva 2003) suggested predicted reading scores for males were overestimated based on a regression analysis, thus inflating IQ-reading discrepancies; for females the opposite was true. When defined separately by gender, underachievement in reading was equally prevalent in males and females. They suggest that the bias arose from different reading score distribution in males and females (males having a lower mean and greater variance than females), thus when reading score cut-offs utilised in the definition of DD, based on data pooled from males and females, performance results in over-identification of males with DD and under-identification of females with DD.

Lykken and colleagues (Lykken et al. 1978) hypothesised that differential volunteer rates for males and females in studies may lead to over-representation of female monozygotic (MZ) twins in studies such as that by DeFries and colleagues (DeFries et al. 1997a).

Symmes and Rapoport (Symmes and Rapoport 1972) proposed that a recessive allele on the X chromosome, might explain why more males have DD.

Although family studies have shown no supporting evidence for this (DeFries and Decker 1982; Lubs et al. 1993), recent linkage evidence has shown some evidence of a susceptibility locus on the X chromosome (Fisher et al. 2002; Kovel et al. 2004).

Based on the hypothesis of Geschwind and Behan (Geschwind and Behan 1982), gender difference could be explained by an excess of, or sensitivity to, androgens such as testosterone, consequently delaying left-hemisphere maturation which could result in abnormalities of neuronal migration, and/or connections during gestation and ultimately DD. Males have been shown to be more prone to androgenic deviance (Alarcon et al. 1995) and migration abnormalities have been observed in a number of dyslexics (Galaburda et al. 1985). The differential rate of maturity in males and females could alter interhemispheric connections, which may exaggerate learning disabilities in males (Nass 1993).

1.6 DEVELOPMENTAL DYSLEXIA AND COMORBID DISORDERS

A number of disorders have been found to exist in individuals with DD more often than in matched controls.

1.6.1 Attention Deficit Hyperactivity Disorder

A high rate of attention deficit hyperactivity disorder (ADHD) is observed in children with DD, with ADHD occurring in around 30-50% of DD cases (Gilger et al. 1992a; Shaywitz et al. 1995; Willcutt and Pennington 2000a). Comorbidity seems higher for the attentive rather than the hyperactive-impulsive form of ADHD (Hynd et al. 1991). It is important to note, disruptive behaviour in DD children is attributable to ADHD rather than academic frustration (Willcutt and Pennington 2000a). Twin and family studies on the comorbidity of DD and ADHD have

suggested that the link is, in part, due to shared genetic underpinnings. Heritability of DD and inattention symptoms has been estimated at 0.39, but only 0.05 between DD and hyperactivity-impulsivity symptoms (Willcutt et al. 2000). Ninety-five percent of the covariance between DD and inattention symptoms had common genetic factors whilst only 21% of the overlap between DD and hyperactivity-impulsivity symptoms was due to the same genetic factors (Willcutt et al. 2000; Friedman et al. 2003).

Willcutt and colleagues (Willcutt et al. 2002) have suggested that the comorbidity between DD and ADHD may occur, at least in part to the effects of a QTL on chromosome 6p, whilst Loo and colleagues (Loo et al. 2004) have shown linkage to chromosomes 16p, 17q and 10q using a sample of children with DD and ADHD.

1.6.2 Internalising Symptomatology (Anxiety, Depression and Social Withdrawal), External Psychopathology (Aggressive Behaviour and Delinquent Behaviour) and Schizotypy

Internalising symptomatology, such as anxiety, depression or social withdrawal are associated with DD (Boetsch et al. 1996; Beitchman and Young 1997). Willcutt and Pennington (Willcutt and Pennington 2000b) observed a higher rate of internalising symptoms in DD individuals than those without DD, and to a lesser extent, externalising symptoms. Further, the relationship between DD and internalising symptoms was largely restricted to females with externalising psychopathology being stronger in males.

Positive schizotypal traits (psychotic-like traits in healthy individuals) such as unusual cognitive and perceptual experiences (magical ideation and perceptual aberration respectively), cognitive disorganisation (for example attentional

difficulties) and social anxiety have been observed in dyslexic individuals (Richardson 1994). It has been suggested that these characteristics may indicate a predisposition to psychotic disorders (Claridge et al. 1997).

1.6.3 Developmental Coordination Disorder (Dyspraxia)

A high degree of overlap (around 30-50%; Richardson and Ross 2000) has been suggested between DD and dyspraxia (Kadesjo and Gillberg 1999), an impairment in the development of motor coordination, not attributable to a general medical condition or mental retardation. Although a relationship between lower motor ability, such as hand motor skill and DD has been observed, the genetic effects on motor skill are largely or wholly distinct from DD (Francks et al. 2003).

1.6.4 Specific Language Impairment (SLI)

Half the children diagnosed with SLI have comorbid DD. In a recent paper, Stein and colleagues (Stein et al. 2004) found linkage of SLI to a region of chromosome 3, a region implicated in linkage studies of DD (see Chapter two, Section 2.9.4), suggesting a shared genetic aetiology may account for the comorbidity between DD and SLI.

Other deficits found in dyslexics other than those associated with DD itself include: oral language acquisition (dysphasia), writing abilities (dysgraphia and misspelling), mathematical abilities (dyscalculia) (Lewis et al. 1994; Light and DeFries 1995; Knopik et al. 1997), postural stability and dexterity, temporal orientation (dyschronia) and visuospatial abilities (developmental right-hemisphere syndrome). Specific disorders often found comorbid with DD include speech-sound disorder (SSD) and auditory processing disorder (APD).

1.7 THEORIES OF READING DEVELOPMENT AND THE READING SYSTEM

The exact nature of how individuals learn to read and the processes involved are unknown. Subsequently theories have been developed to try and describe the development of reading.

1.7.1 Stage Theories of Reading Development

Stage theories are defined by the belief that children go through a series of phases in learning to read (Marsh et al. 1981; Seymour and MacGregor 1984; Frith 1985). Frith (Frith 1985) proposed a theory consisting of three phases. The logographic phase involves children learning to recognise sight vocabulary cues for words (for example the two 'tall sticks' in 'yellow'). The second, alphabetic phase requires the child to acquire phonic knowledge and start to read words using letter-sound correspondences. The final, orthographic phase requires children to read words as orthographic units. No phonological conversion is used and recognition using lexical procedures of letter-letter strings are utilised. DD results from getting 'stuck' in the alphabetic phase (where words from a restricted sight vocabulary can be read but new (unfamiliar) and nonsense words can not be read) or orthographic phase (where words obeyed by rules learnt by the child can be read but words in which the spelling-to-sound relation is irregular are not, unless the word was included in the sight vocabulary learnt in the logographic phase) (Frith 1985).

Ehri (Ehri 1989; Ehri 1992) proposed a theory focused on integrative routes of word recognition. Phonology and sight and a well functioning, rapid system is required to access and retrieve words. Poor phonological recoding skills form the basis of the failure to achieve full orthographic representations.

Perfetti (Perfetti 1985; Perfetti and Bell 1991; Perfetti 1992; Perfetti 1994) described learning to read as the acquisition of orthographically addressable words through the alteration of individual representations using specificity and redundancy. Specificity increases the number of position-correct letters in a lexical representation whereas redundancy increases the establishment of redundant phonemic representations.

In general, the stage theories suggest that as reading skills develop, phonological decoding strategies are disregarded in favour of more rapid, orthographic representations of lexicons. DD may result from the failure to progress through a stage (Frith 1985) or failure to master the strategy of word recognition at a particular stage (Ehri 1992).

Berninger and Abbot (Berninger and Abbot 1994) suggested that the term 'stage' may be misleading in development as in adult reading, aspects of numerous stages are utilised in order to read, for example, phonological decoding may be a more useful strategy than orthographic coding in the recognition of unfamiliar words. Further, only simplistic views of the interactions between phonological and orthographic coding are described. Integrated models, such as the connectionist models of reading development attempt to overcome this problem. No acknowledgement is made of the transition between stages and the importance of phonological awareness development before reading training. The orthographic stage may also be more complex than implied, for example, using context to work out how to read a word: 'the *tear* in the eye', 'the *tear* in the cloth'. Finally, Ehri (Ehri 1992) suggested that skilled readers recognise words as a whole (orthographically), which should be classified as a further developmental stage.

1.7.2 The Dual-Route Theory of Reading

The dual-route theories (Coltheart 1978; Morton and Patterson 1980) suggest two separate routes, an indirect route and direct route, in the recognition of printed words (see figure 1.1). The direct route refers to the identification of words by activation of connections between the orthographic representations of words and their meaning as a result of extensive practice (Baron 1979). Difficulty in the direct route is thought to result in surface dyslexia. The indirect route involves the use of phonemes. Letters are translated into sounds, following letter-sound rules, allowing the identification of words and thereby giving access to word meaning (Barron 1986). Difficulty in the indirect route is thought to result in phonological dyslexia. When both routes have been mastered, one is considered a normally skilled reader (Coltheart 1987; Castles and Coltheart 1993). Manis and colleagues (Manis et al. 1996) noted that whilst the profiles of surface dyslexics were similar to younger readers, phonological dyslexic profiles were not, concluding that surface dyslexia may be a general developmental delay in word recognition, whereas phonological dyslexia is a specific deficit in phonological processing and represents a deviant group (see Section 1.4).

Support for the theory originates from individuals with brain injuries. Surface dyslexics show impairments in the direct route, whereas phonological dyslexics show impairments in the indirect route, leading to difficulties in reading words where no previous form-meaning association had been established.

The theory can explain DD in two ways:

1) Individuals with DD often show impairments in reading nonsense and irregular words suggesting that the indirect route is less efficient than the direct route.

2) The two routes may fail to develop at the same rate as in 'normal' readers (Frith 1985). Since children differ in the manner in which they learn to read (Bryant and Impey 1986; Stuart and Coltheart 1988), the two routes may develop at different rates in some and result in symptoms of DD.

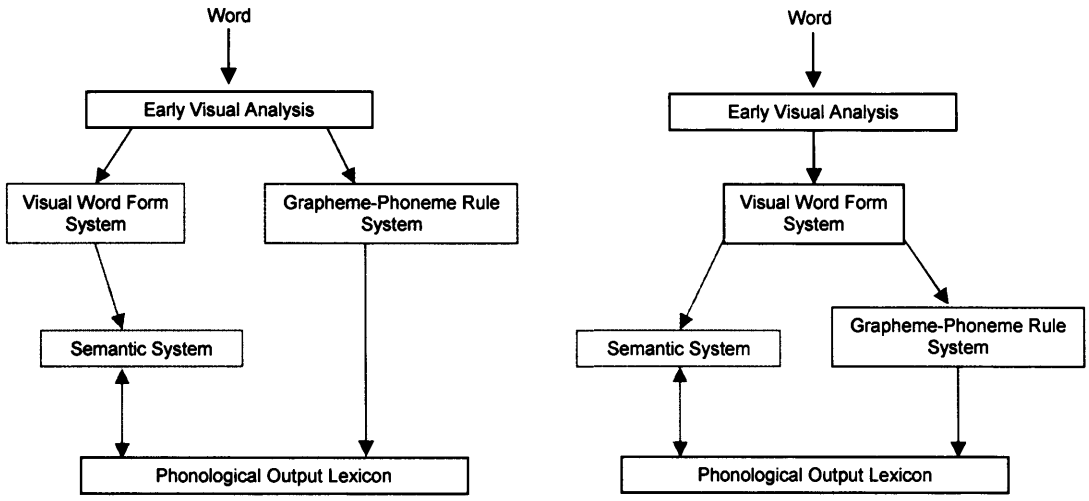


Figure 1.1 Models of the Dual-Route Theory. Word reading is achieved through the indirect route, that requires words to be transformed into their auditory counterparts through a grapheme-to-phoneme correspondence then to its meaning, or through a direct route, corresponding to the direct association between the visual form of the word and its meaning. It has been argued that the visual word-form system is dedicated to the direct route only (left), whilst others assume it is common to both routes (right). Adapted from Jobard and colleagues (Jobard et al. 2003).

Early versions of these theories suggest that the two routes are independent. Later theories however, acknowledge the two routes are dependant on each other in terms of knowledge structure and processes, but retain distinction between lexical and nonlexical processing.

Humphreys and Evett (Humphreys and Evett 1985) and others (Goswami 1988; Manis et al. 1990; Gough and Walsh 1991; Stanovich et al. 1997) criticised the theory, suggesting the two separate routes were artificial and that skilled reading depends on the integration of phonological and orthographic knowledge (Juel et al.

1986; Stanovich and West 1989; Manis et al. 1990; Stanovich 1991). Others have argued that there are more than two ways to read words, for example words can be read by making analogies to known sight words by detecting and pronouncing orthographic patterns (Baron 1979; Glushko 1979; Glushko 1981; Goswami 1986; Brown 1987). The theory also fails to explain why most children with DD have difficulty reading regular non-words and irregular words (Pennington 1999). The Dual-Route theory makes no reference to learning processes and is not developmental.

It is unlikely that the Dual-Route theory is represented in our brains in the form of hard-wired neural units (Ellis 1985; Castles and Coltheart 1993) since it would not explain why children can learn to read with alphabetic, syllabic or logographic writing systems depending on the culture into which they were born.

1.7.3 Connectionist Theory of Reading

The connectionist theory of reading (or parallel distributed processing theory; PDP) emphasises a single, interconnected system, in which words of all types are recognised (Seidenberg and McClelland 1989; Plaut et al. 1996). No representation of letter-sound rules is required, although rule-like (i.e. indirect) phonological coding is a property as well as whole word recognition. Processing of information occurs through the interaction of large numbers of simple processing units (Schneider and Graham 1992) and three interconnected layers (Seidenberg and McClelland 1989) representing visual word pronunciation. The input layer corresponds to graphemes, the middle layer to the correspondence between letters and sounds and the third layer to phonemes. The model works by training the network, by repeated trials, to result in automatic recognition of words, as a consequence of patterns of phonologic and orthographic input features. Weights (how likely a connection is to fire) between

connections are adjusted until there is near optimal performance in the pronunciation of single printed words. Initial connections are made between input and output by intermediate units ('hidden units'). The learning process then optimises the connections in a process called backpropagation. Connections are strengthened when they are deemed to be correct, and weakened if incorrect (i.e. made less likely to fire). When new words are observed the network generalises from its previous experiences.

The theory does not take into account the prior knowledge and skills children have acquired before learning to read and assumes that the phonological store is unstructured below the level of the phoneme when reading acquisition begins. Evidence against the assumption that the phonological store is unstructured below the level of the phoneme comes from dyslexic individuals who show deficits in pure tone perception (Ben-Yehudah and Ahissar 2004; Ben-Yehudah et al. 2004; Lachmann et al. 2005).

1.8 THEORIES OF DEVELOPMENTAL DYSLEXIA

Although DD is a neurological disorder with a complex genetic aetiology, the underlying biological and cognitive causes are unknown. A number of theories of DD have attempted to explain DD and its origins. There are two main ideas within the theories of DD; the first that there is a phonological deficit; the second is that there is a perceptual/timing deficit. In addition, recent work indicates a possible attentional deficit in individuals with DD.

1.8.1 The Phonological Core Deficit Theory of Developmental Dyslexia

The phonological core deficit theory (see figure 1.2) claims that individuals with DD have a specific impairment in the representation, storage and/or retrieval of

speech sounds from long-term memory. It is based on the belief that learning to read requires learning grapheme-phoneme correspondence and that reading comprises of two processes, decoding and comprehension. A deficit in the phonologic elements impairs the ability to segment a word into its phonologic units and results in difficulty to decode and identify words. The phonologic deficit is believed to be independent of non-phonologic abilities and so higher-order cognitive and linguistic functions (IQ, reasoning, vocabulary) remain intact, and are only blocked by lower-order linguistic deficits. Until the lower-order processes have been achieved, meaning of a word/text cannot be drawn from the text.

The neurological basis of DD is assumed to be a congenital dysfunction of the left-hemisphere perisylvian brain areas which underlie phonological representations or connections between phonological and orthographic representations (Galaburda et al. 1985; Geschwind and Galaburda 1985; Paulesu et al. 1996; Shaywitz et al. 1998; Brunswick et al. 1999; McCrory et al. 2000; Pugh et al. 2000; Paulesu et al. 2001; Temple et al. 2001; Shaywitz et al. 2002). Support is provided by the observation that dyslexic individuals perform badly on tasks requiring phonological awareness. Evidence from verbal short term memory (STM) and slow automatic naming in dyslexics suggests that a basic phonological deficit exists, possibly in relation to phonological representations or access and retrieval of phonological representations (Snowling 2000).

The main problem of the theory is that it does not explain the occurrence of sensory and motor disturbances in dyslexic individuals. However it has been suggested that such disorders are not part of the core DD deficit and act only as a marker of DD (Snowling 2000). Castles and Coltheart (Castles and Coltheart 2004) have also criticised the phonological core deficit theory since no study has been able

to show unequivocal evidence that there is a causal link from ability in phonological awareness and success in reading and spelling acquisition.

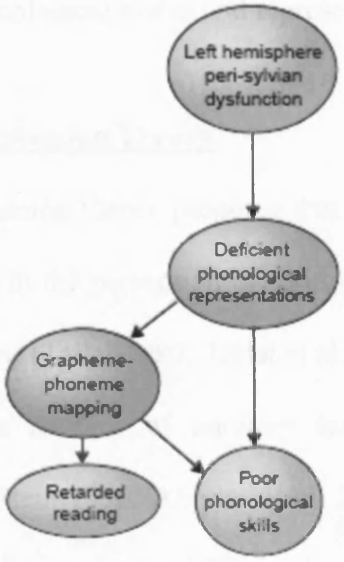


Figure 1.2 The Phonological Core Deficit Theory of Developmental Dyslexia. Bubbles indicate impairments at the neurological (red), cognitive (green) and behavioural (blue) levels. Arrows represent causal connections. Taken from Ramus and colleagues (Ramus 2003)

1.8.2 Double Deficit Hypothesis of Developmental Dyslexia

The double deficit hypothesis proposes that dyslexics have two core deficits, one in phonological processing and one in rapid naming of stimuli (deficits in the processes underlying the rapid recognition and retrieval of visually presented linguistic stimuli), believed to make independent contributions to reading (Wolf et al. 1994; Wolf and Bowers 1999). Indeed studies have shown naming-speed deficits contribute variance to reading, independent of variance contributed by phonological awareness measures (Blachman 1984; Mann 1984; Bowers and Swanson 1991; Olson et al. 1994).

Wolf and Bowers (Wolf and Bowers 1999) have discussed the potential effects of naming speed deficits on reading. Firstly, deficits in the processes underlying naming-speed hinder lower level perceptual requirements that result in

non-fluency of word identification and hinder comprehension. Secondly, the deficits in naming-speed represent a broader system of rate or efficiency-based difficulties that affect orthographic and phonological routes and representations.

1.8.3 The Rapid Auditory Processing Theory

The rapid auditory processing theory proposes that phonological deficits are secondary to an auditory deficit in the perception of short or rapidly varying sounds (not necessarily language sounds) (Tallal 1980; Tallal et al. 1993). Indeed dyslexics show poor performance on a number of auditory tasks including frequency discrimination (McAnally and Stein 1996; Ahissar et al. 2000) and temporal order judgement (Tallal 1980; Nagarajan et al. 1999). Abnormal neurophysiological responses to various auditory stimuli have also been observed (McAnally and Stein 1996; Nagarajan et al. 1999; Kujala et al. 2000; Temple et al. 2000; Ruff et al. 2002).

The failure to correctly identify sounds and fast transitions can cause difficulties when the acoustic events are cues to phonemic contrasts, for example, /ba/ versus /da/. Poor categorical perception of certain sound contrasts has been shown in dyslexics (Mody et al. 1997; Adlard and Hazan 1998; Sernoclaes et al. 2001) indicating they could cause the phonological deficits resulting in the failure to read.

Although the theory had no biological basis, the magnocellular theory (see Section 1.8.4) provides some suggestion of underlying biological mechanisms. Like the visual system (which contains magno- and parvo cells), there is some evidence that two types of neurones, fast and slow activating neurones, akin to the magno- and parvo cells in the visual system exist in the auditory system (Stein 2000).

1.8.4 The Magnocellular Theory of Developmental Dyslexia

Stemming from the observation that dyslexic individuals show poor thresholds for stimuli of low contrasts, low spatial or high temporal frequencies (Lovegrove et al. 1980; Talcott et al. 1998) and poor sensitivity to visual motion (Talcott et al. 2000a), the magnocellular theory attempts to integrate these observations and other deficits observed in DD. Studies haven suggested that children with DD are less sensitive to visual motion and auditory stimuli (Talcott et al. 2003) and both vision and audition influence word recognition (Talcott et al. 2000b).

Stein and Walsh (Stein and Walsh 1997) suggest that magnocellular dysfunction is not restricted to the visual pathways but to all sensory modalities. The cerebellum receives input from magnocellular systems in the brain and has been predicted to be affected by magnocellular defects (Stein et al. 2001). This would account for all aspects of DD, including visual, auditory, motor, tactile and phonological difficulties (Ramus et al. 2003b). Magnocellular abnormalities have been observed in the medial and lateral geniculate nuclei of dyslexic brains (Livingstone et al. 1991; Galaburda et al. 1994), although these studies use very few brains (five in total).

The magnocellular theory (see figure 1.3) fails to describe why not all deficits are observed in all dyslexic individuals (Tallal 1980; Reed 1989; Manis et al. 1997; Mody et al. 1997; Adlard and Hazan 1998; Lorenzi et al. 2000; Marshall et al. 2001; Rosen and Manganari 2001). The inconsistency of an auditory deficit in 'rapid' auditory processing, such that 'rapid' auditory processing is intact in some tasks whilst 'slow' auditory processing is impaired in other tasks (Reed 1989; McAnally and Stein 1996; Adlard and Hazan 1998; Schulte-Körne et al. 1998b; Witton et al.

1998; Nittrouer 1999; Lorenzi et al. 2000; Rosen and Manganari 2001) raises problems with the theory. It has also been argued that auditory deficits do not predict phonological deficits (Mody et al. 1997; Schulte-Körne et al. 1998a; Bishop et al. 1999; Marshall et al. 2001; Rosen and Manganari 2001; Share et al. 2002). Visual impairments observed in dyslexics, tend to be across a whole range of stimuli rather than restricted to the magnocellular system (Skottun 2000; Amitay et al. 2002; Farrag et al. 2002). Although Talcott and colleagues (Talcott et al. 2000b) suggest both vision and audition influence a childs ability in word recognition, the results suggest that vision and audition may separately affect ability to extract phonological and orthographic information in reading (Talcott et al. 2000b). Further, Kronbichler and colleagues (Kronbichler et al. 2002) observed significant differences between DD cases and controls in phonological tests but not in visual, auditory or motor tasks, supporting the idea of a phonological deficit but not a deficit in the magnocellular system.

The basis of the magnocellular theory, that dyslexic individuals have visual motion deficits, is in itself a contentious issue. There is debate into whether the visual deficits result from a magnocellular deficit or whether they result from cortical dysfunction (Cornelissen et al. 1995).

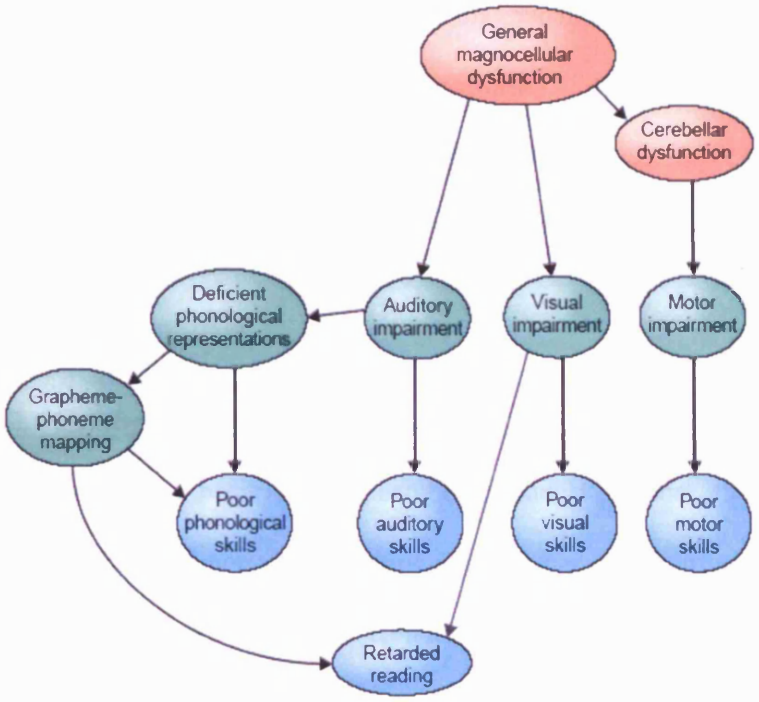


Figure 1.3 The Magnocellular Theory of Developmental Dyslexia. A general magnocellular dysfunction is hypothesised to encompass auditory, visual and cerebellar/motor deficits. The auditory deficit in turn causes a phonological deficit that triggers the events predicted by the phonological core deficit theory of developmental dyslexia. The visual magnocellular deficit is seen as a second direct cause of reading problems. Bubbles indicate impairments at the neurological (red), cognitive (green) and behavioural (blue) levels. Arrows represent causal connections. Taken from Ramus and colleagues (Ramus 2003)

1.8.5 The Visual Theory of Developmental Dyslexia

The visual theory of DD suggests a visual impairment gives rise to problems in the processing of letters and words in text, in particular, deficits in the magnocellular visual pathway lead to deficiencies in visual processing and, via the posterior parietal cortex, to abnormal binocular control and visuospatial attention (Stein and Walsh 1997; Hari et al. 2001). Anatomical studies have shown abnormalities in the magnocellular layers of the lateral geniculate nucleus (Livingstone et al. 1991; Flowers 1993) and psychophysical studies have shown decreased sensitivity in the magnocellular range (i.e. low frequencies and high temporal frequencies) in dyslexics. However, Skottun has raised questions in the

validity of tests used to assess evidence for a magnocellular deficit, particularly those involving contrast sensitivity (Skottun 2000; Skottun 2001; Skottun 2005).

Although proponents of the theory suggest there is also a phonological deficit in DD, no association between the visual deficits and phonological problems are made.

1.8.6 The Cerebellar Theory of Developmental Dyslexia

The biological basis of the cerebellar theory is that the cerebellum in dyslexics is mildly dysfunctional, resulting in a number of cognitive difficulties. The cerebellum plays a role in dexterity, automaticity and motor control and as a consequence in speech articulation. Retarded or dysfunctional articulation will lead to deficiencies in phonological representations via impaired articulatory skills. The cerebellum also plays a role in the automation of tasks; therefore a weak capacity to automatise could affect the learning of grapheme-phoneme correspondences.

The observation that dyslexics perform poorly in a number of motor tasks (Fawcett et al. 1996), in dual tasks that demonstrate impaired automatisation of balance (Nicholson and Fawcett 1990) and in time estimation, a non-motor cerebellar task (Nicholson et al. 1995) provide support for the theory. Brain imaging studies have shown anatomical, metabolic and activation differences in the cerebellum of dyslexics (Rae et al. 1998b; Nicholson et al. 1999; Brown et al. 2001; Leonard et al. 2001).

Although the theory fails to account for sensory disorders, supporters suggest the presence of two distinct subtypes of DD, one involving the cerebellum, the other the magnocellular pathways (Fawcett and Nicholson 2001). The theory also relies on outdated ideas of the motor theory of speech in order to explain the link between articulation and phonology. The idea that phonological representations relay on

speech articulation has been abandoned since several cases of normal phonological development have been observed despite the presence of severe dysarthria or apraxia of speech. It is uncertain how many dyslexics have motor problems since some studies have failed to find such problems (Wimmer et al. 1998; Daal and Leij 1999; Kronbichler et al. 2002) and others find motor problems subgroups of dyslexics (Yap and Leij 1994; Ramus et al. 2003a). Finally, the architecture of the memory system subserving procedural learning is not restricted to the cerebellum and involves interplay between cortical, subcortical and some cerebellar areas during skill learning (Doyon et al. 2003). As a result, deficits of procedural learning could result from dysfunction of a number of regions of the brain.

It is worthy of note, that much of the evidence supporting cerebellar dysfunction in DD has come from investigations undertaken on the same sample of individuals from the UK. Very little replication has been completed on other UK and/or non-UK samples.

1.8.7 Attentional Deficit Theory of Developmental Dyslexia

The neural processes which allow the processing of stimuli relevant to a particular task whilst inhibiting irrelevant stimuli is referred to as attention. In the reading of text, processing of both crowded and small printed letters is required. In order for this to be undertaken, shifts of attention are required between individual letters and letter groups as well as rapid and accurate integration of visual and auditory (if reading aloud) cues. Accurate reading also requires the suppression of interfering information from the periphery, resulting in a narrowing of focus of attention. Deficits in the attending to centrally located stimuli rather than peripherally located stimuli, suppressing interfering information and narrowing the focus of attention have been observed in dyslexic individuals (Geiger et al. 1994;

Steinman et al. 1998; Facoetti et al. 2000). Although research into attentional deficits tends to focus on the visual system, other sensory dysfunctions are being recognised (Hairston et al. 2005).

Further evidence that an attentional deficit in DD may exist is that dyslexic children are often easily distracted and often have problems maintaining attention on one task for prolonged periods (Keogh and Margolis 1976). Further, ADHD is often found to be co-morbid with DD (see Section 1.6.1).

Ramus and colleagues (Ramus et al. 2003b), suggest that it could be possible that all the theories of DD are true – in different individuals with partially overlapping subtypes of DD. Alternatively they suggest that one theory may explain DD and different manifestations of DD reflect differing markers of the disease (i.e. they are associated with DD but not a cause of DD). Ramus and colleagues (Ramus et al. 2003b) studied 16 DD cases and 16 controls (university students) using a range of psychometric, phonological, auditory, visual and cerebellar tasks. All 16 DD cases suffered from a phonological deficit whilst only 10 had an auditory deficit, 4 a motor deficit and 2 a visual magnocellular deficit. The phonological deficit was found to be present in the absence of any other sensory or motor disorder and was aggravated by auditory disorders. However the auditory problems were not solely in rapid auditory processing as predicted by the magnocellular theory. No evidence was observed to suggest that motor impairments were cerebellar in origin, nor that they reflect an automaticity deficit. In summary the results support the phonological theory, however the presence of additional sensory and motor disorders in certain individuals were are not explained by the theory.

1.9 NEUROBIOLOGY OF DEVELOPMENTAL DYSLEXIA

The organisation of behaviour and cognitive functions in the brain involves a complex network of systems involving many brain areas (Taylor 1958; Luria 1973). 'Normal' cognition and behaviour depends on normal anatomic, physiologic and biochemical brain development, taking place throughout infancy, childhood and adolescence (Lyon and Rumsey 1996). The development of the brain systems involved in cognitive processes such as reading starts with the development of the brainstem, motor and sensory areas and proceeds with the development of tertiary areas (including the prefrontal cortex and its connections). In these areas neurons have the ability to respond to sensory inputs.

1.9.1 Anatomical Studies of Developmental Dyslexia

The planum temporale is located within the sylvian fissure on the superior-posterior surface of the temporal lobe. It functions in auditory comprehension and possibly in phonologic processing (Frank and Pavlakis 2001). Generally the planum temporale is larger on the left side than the right side, with asymmetry forming as early as 33 weeks gestation, suggesting a biologic prenatal mechanism. An absence of asymmetry has been shown in dyslexic brains, which has been suggested to result from enlargement of the right side (Galaburda et al. 1985).

Undertaking microscopic analysis of brain tissue, Galaburda and colleagues (Galaburda and Kemper 1979; Galaburda et al. 1985; Humphreys et al. 1990) have shown the presence of ectopia, dysplasia and vascular micro-malformations in cortical parietal regions of dyslexic brains. However, several individuals showing these abnormalities reported oral language delay as well as DD (Cohen et al. 1989). Habib (Habib 2000) suggests that the observation of abnormalities in dyslexic brains suggests abnormal cortical development.

There is increasing evidence that cortical abnormalities may have an impact on cognitive function. Mice and rats with ectopia and microgyri exhibit a number of learning deficits (Denenberg et al. 1991; Schrott et al. 1992; Rosen et al. 1995; Balogh et al. 1998) with the specific location of the cortical disruption influencing the type of learning difficulty exhibited by the rat/mouse (Hyde et al. 2001; Hyde et al. 2002). The presence of an ectopia in the prefrontal cortex impaired learning of the Morris maze and Lashley reversal learning, whilst ectopia in the motor cortex impaired learning of the Lashley maze. Since the ectopia observed are structurally similar to those of dyslexic individuals, the results may suggest, that since ectopic mice learn differently than non-ectopic mice, so to do dyslexic individuals compared to those individuals without the learning disability. Also, since the location of the ectopia is related to learning differences in mice this may reflect the variability in the extent of the learning disability in dyslexic individuals. The authors suggest that the source of the variability may be the related to the location of the ectopia within the cortex of the dyslexic individual (Hyde et al. 2001). Limited support for the presence of cortical abnormalities underlying DD also comes from the observation of association between DD and *EKN1* (see Chapter six), a gene shown to be involved in neuronal migration.

1.9.2 Structural Brain Imaging of Developmental Dyslexia

Structural imaging studies of DD have suggested that 70% of dyslexic patients have a symmetric planum temporale compared to only 30% of control patients, supporting anatomic studies. Although it has been suggested that the planum temporale and anterior temporal asymmetry patterns are related to measures of language processing, phonologic coding and reading development (Rumsey et al.

1986; Hynd et al. 1990; Kushch et al. 1993; Semrud-Clikeman 1997), when age adjustments are made, no significant differences are observed (Schultz et al. 1994).

The corpus callosum, a communication pathway between the left and right hemispheres of the brain, has been studied in relation to DD. In learning and reading, transfer of information between the two hemispheres is required. Some studies have reported a larger splenium of the corpus callosum in dyslexics, however other studies fail to observe such a difference (Larson et al. 1992; Hynd et al. 1995). In one study of children, dyslexics (and dysphasics) had a thicker corpus callosum than controls (Njiokiktjien et al. 1994).

1.9.3 Functional Brain Imaging in Developmental Dyslexia

In poor readers, deficient patterns of cerebral blood flow activation, often in the left hemisphere have been observed (Semrud-Clikeman 1997). During language tasks on poor readers, children have shown activation of the temporoparietal region located posterior to the left supratemporal region (corresponding to Wernickes' area) (Flowers et al. 1991).

Males undertaking an oral reading task have shown differences in normalised regional metabolic activity in the prefrontal cortex and the lingual region of the occipital lobe. Whilst controls had an asymmetry in these areas, dyslexic individuals showed a more symmetrical pattern (Gross-Glenn et al. 1991). In a series of studies (Rumsey et al. 1992; Rumsey et al. 1994; Rumsey et al. 1997), dyslexic individuals showed reduced blood flow in the left parietal region near to the angular/supramarginal gyri (a region important in the reading process) during rest. Reduced activation was also shown in the mid-to-posterior temporal cortex bilaterally and in the inferior parietal cortex, mostly on the left side during pronunciation and decision-making. In syntactic processing, dyslexic and control

individuals revealed similar activation. These results support the hypothesis that posterior temporal language areas controlling phonologic processing are impaired in DD, but areas associated with syntactic processing are typical. Shaywitz and colleagues (Shaywitz et al. 1998) extend the study of 1997 (Rumsey et al. 1997) by using functional magnetic resonance imaging (fMRI) and reported decreased activity in the temporoparietal regions (superior temporal gyrus and angular gyrus) during phonological processing tasks.

Paulesu and colleagues (Paulesu et al. 1996) have shown left temporoparietal dysfunction in adults undertaking a phonological processing task, results remarkably similar to Rumsey and colleagues (Rumsey et al. 1992) and Brunswick (Brunswick et al. 1999), particularly since different tasks, subjects (i.e. severely dyslexic versus compensated) and analysis techniques were utilised. Positron emission tomography (PET) scans have also revealed evidence of a disconnection abnormality between the anterior and posterior language regions, possibly because of a dysfunctional left insula which, in normal individuals acts as an anatomic bridge between Broca's area and the superior temporal and inferior parietal cortex (Frank and Pavlakis 2001). Horwitz and colleagues (Horwitz et al. 1998) also reported an abnormal or absence of functional connectivity in adult dyslexics since no correlation was observed between regional blood flow in the angular gyrus and in extra-striate occipital and temporal lobe regions. In later quantitative magnetic resonance imaging (MRI) studies this has been supported by the observation of a smaller insula and anterior superior neocortex in individuals with DD (Pennington et al. 1999) (see Section 1.9).

Reduced temporoparietal activity in dyslexics undertaking phonological processing tasks was reported by Temple and colleagues (Temple et al. 2001), a result similar to those found in adults (Poldrack et al. 1999).

Magnetoencephalography (MEG) studies of dyslexic children (Poldrack et al. 1999; Simos et al. 2000) suggest the adult findings are not due to compensation effects and rather reflect a brain dysfunction fundamental to DD.

1.9.4 Tissue Metabolites in Developmental Dyslexia

Magnetic resonance spectroscopy (MRS) is an imaging method allowing the study of the concentration of tissue metabolites (Guze 1991). Rae and colleagues (Rae et al. 1998a) reported a decreased ratio of choline to *N*-acetylaspartate in the left temporoparietal region of dyslexic individuals, indicative of a developmental abnormality in the region.

Elevated levels of magnetic resonance lactate have been observed in the left anterior quadrant and the left frontal regions of the brain when phonologic processing tasks are being undertaken (Richards et al. 1999; Richards et al. 2000; Richards et al. 2002).

1.9.5 Electrophysiological Studies of Developmental Dyslexia

Schulte-Körne and colleagues (Schulte-Körne et al. 1998a) yielded results that were suggestive of a deficit in the pre-attentive mechanism required for language processes leading to difficulties in learning to read. In a measurement of the response to an auditory contrast such as /da/-/ga/, Bradlow and colleagues (Bradlow et al. 1999) showed diminished responses in dyslexic children compared to control children. When the transition duration of the stimuli was lengthened, the response of the dyslexics was similar to that of the controls. Kujala and colleagues (Kujala et al. 2000) concluded that problems discriminating temporal sound features when they are surrounded by other stimuli could be the case in DD with regards to phonemes in words.

In conclusion, the loss of left/right asymmetry in the planum temporale and temporal areas involved in the processing of language seem to be consistent across neuroradiologic studies and are supported by post-mortem brain studies.

Together, results suggest that reading requires a number of brain areas, in both hemispheres and including posterior (phonological processes) and anterior brain regions (syntactic processing). Abnormalities in dyslexic individuals tend to lie in the posterior regions of the brain including the temporal and inferior parietal cortex, angular gyrus and the striate and extrastriate cortex. Functional imaging studies have supported these findings.

It is important to note that many neurobiological studies have small numbers of participants, often include only adults and controls are often not well matched in terms of sex, handedness, intelligence or educational experience. In order to correct these problems, studies on larger samples, with homogenous patient populations are required. An important aspect in terms of DD is that the criteria/definition of what constitutes DD needs to be clarified and made uniform across studies. The study of children rather than adults is also of importance, particularly in light of a study by Schlaug (Schlaug et al. 1995) who suggested intensive training in language skills can modify the symmetry observed in dyslexic individuals, resulting in 'normalisation' of the brain.

1.10 WHITE MATTER DISRUPTION IN DEVELOPMENTAL DYSLEXIA

The atypical processing in the frontal and posterior brain networks has been suggested by some to indicate that DD may be a 'disconnection' syndrome, an idea supported by neuroimaging studies (Paulesu et al. 1996; Horwitz et al. 1998; Pugh et al. 2000). Connection deficits imply that there is a disruption of the white matter in

the brain as axon projections of white matter function to connect brain regions. Visualisation of white matter organisation has failed to show disturbances in DD in many studies although Klingberg and colleagues (Klingberg et al. 2000) have reported disruption in temporoparietal white matter, a region connecting left hemisphere language areas to frontal and posterior brain areas, in dyslexics using diffusion tensor imaging (DTI). The degree of white matter disorganisation was shown to correlate with reading ability, suggesting white matter may play a role in the ability to read.

CHAPTER TWO

THE GENETIC BASIS OF DEVELOPMENTAL DYSLEXIA

CHAPTER TWO

2. THE GENETIC BASIS OF DEVELOPMENTAL DYSLEXIA

Given the early observations of multiple DD-affected individuals within the same families, research has been undertaken to investigate whether there is a genetic influence on the presence of DD. Family studies have suggested that DD is familial and twin studies have revealed a genetic basis to DD. Although the mode of transmission of DD genes has not been elucidated, it is likely to be a complex disorder. Linkage and association studies have implicated a number of regions across the genome, which may harbour genes involved in DD susceptibility. Studies are now being undertaken to refine these regions in order to identify DD susceptibility genes.

2.1 FAMILY STUDIES OF DEVELOPMENTAL DYSLEXIA

Familiality, whether a disorder runs in families, is measured by comparing the rate of a disorder in relatives of an affected person to the baseline rate found in the general population. It is the first step in determining whether a disorder has a genetic aetiology but is not itself sufficient to suggest genes are important in the disorder since families also share common environment.

Early studies noted familial aggregation of DD (Morgan 1896; Kerr 1897) and large family studies have subsequently confirmed DD familiality (Orton 1930; Orton 1937; Eustis 1947; Hallgren 1950; Walker and Cole 1965; Owen et al. 1971; Yule and Rutter 1975; Lewitter et al. 1980; Pennington 1990; Pennington et al. 1991; Lubs et al. 1993; Schulte-Körne et al. 1996).

The first calculation of the magnitude of familial risk (Hallgren 1950) revealed a risk of recurrence to siblings of affected probands of 40.8% and 42.4% for parents, both considerably higher than the general population risk (5-10%). The main problem of the study was that affected individuals were not classified based on direct testing and ascertainment biases led to the selection of families with a higher than normal proportion of affected relatives. A number of studies have suggested self-report data are not reliable (Scarborough 1989; Schulte-Körne et al. 1996). Although no control group was included in the first study (Hallgren 1950), similar estimates of familial risk (range 34 – 45%) have since been obtained where control groups and direct tests have often been used to determine affection status (Walker and Cole 1965; Klasen 1968; Rutter et al. 1970; Owen et al. 1971; Naidoo 1972; Zahalkova et al. 1972; Finucci et al. 1976; Omenn and Weber 1978; Vogler et al. 1985; Gilger et al. 1991).

It has been noted that probands collected based on positive DD affection status, have parents affected by DD more often than would be expected given the population risk of 5-10%. Finucci and colleagues (Finucci et al. 1976) observed that 18.8% of DD probands had two affected parents (based on IQ and reading tests) and 62.5% had one affected parent. Vogler and colleagues (Vogler et al. 1985) noted male probands reported 29% of fathers being affected and 17% of mothers. Slightly higher observations were obtained for females, with 36% reporting an affected father and 25% reporting an affected mother. Gilger and colleagues (Gilger et al. 1991) noted that 50% of probands had one affected parent and only 3% having two affected parents. More recently, Wolff and Melngailis (Wolff and Melngailis 1994) found 80% of their DD cases came from families with some evidence of familial transmission.

Finucci and colleagues (Finucci et al. 1976) calculated the risk of recurrence to siblings of affected probands at 42.5%, whist for parents it was 47.2%, supporting early studies (Hallgren 1950) but in contrast to a later study by Gilger and colleagues (Gilger et al. 1991) who analysed a bigger sample (39 families) to report recurrence rates of 27% in parents and 38.5% in siblings.

Vogler and colleagues (Vogler et al. 1985) asked a slightly different question regarding familiality of DD. Using the Colorado Family Reading Study (CFRS) sample, they asked what was the risk of having DD given at least one affected parent. The relative risk to the children of affected parents for male probands varied between 4.9 and 3.9 depending on the parent affected by DD (father and mother respectively) and for female probands, between 10.2 and 8.5 (father and mother respectively), both indicating that the risk of a child developing DD if a parent reports difficulty in learning to read is significantly increased compared to children with no affected parents.

The studies by Finucci, Vogler, Gilger and colleagues (Finucci et al. 1976; Vogler et al. 1985; Gilger et al. 1991) were the first family studies to have adequately large sample sizes, appropriate sampling schemes and a systematic study of all ascertained families.

Scarborough (Scarborough 1989) overcame the problems arising from self-report data used in early family studies by using reading tests and a discrepancy criterion to diagnose DD consistently. However, the advertisement used to recruit participants to the study (which asked for "families in which anyone has experienced a severe childhood reading problem") may have inflated results and could indeed explain the higher prevalence of DD in the families of this study (67% of DD individuals reported a parent or sibling with DD). Further, DD risk may also be

associated with the severity of DD in relatives since children of dyslexics were more likely to be affected if the parent was more severely affected.

2.2 TWIN STUDIES OF DEVELOPMENTAL DYSLEXIA

2.2.1 Twin Concordance Rates

Early twin studies of DD, typically employed a comparison of MZ and dizygotic (DZ) concordance rates as a test for genetic aetiology, where the concordance rate (based on a categorical definition of DD) is calculated by taking the ratio of the number of concordant pairs to the total number of twin pairs. Twin analyses such as these rely on a number of assumptions (Harris 1986):

1) That MZ twins share 100% of genes identical by descent (IBD) since they result by definition from the splitting of one egg (zygote) and DZ twins share 50% of genes IBD as they result from the uniting of two ova and sperm
2) Twins, both DZ and MZ, share environment to a similar degree
3) The population risk is the same for both MZ and DZ twins.

Since MZ twins have identical genes and are assumed to share environment to the same extent as DZ twins, which share only half their genes, differences in concordance between MZ twins and DZ twins is a reflection of genetic influences.

Hallgren (Hallgren 1950) reported 100% DD concordance in MZ twins and only 33.3% in DZ twins. Although the results of Hallgren's study seem to be replicated in later studies (Hermann 1956; Hermann and Norrie 1958; note some sample overlap), no attempt to distinguish general retardation from specific deficiencies in reading was undertaken and sample size was very small. In a review of previous twin-studies, Zerbin-Rüdin (Zerbin-Rüdin 1967) reported concordance rates of 100% in MZ twins but 52% in DZ twins. Since many of the samples had

been referred to clinics and concordant twin pairs are more likely to be reported in research than discordant twins (Harris 1986), ascertainment bias resulting in inflation of concordance rates are possible. Indeed, lower concordance rates have been reported (Bakwin 1973; MZ = 84%, DZ = 29%), although DD definition was very vague.

Stevenson and colleagues (Stevenson et al. 1987) and Pennington (Pennington 1989) argued that many twin studies suffered from:

1) A lack of independent assessment of reading problems, producing bias in the direction of increased concordance rates
2) Lack of an adequate or systematic definition of DD
3) Reliance on volunteers, clinical referrals or retrospective searches through reported cases resulting in bias such as over-inclusion of 'interesting' i.e. concordant cases.
4) Failure to determine zygosity adequately
5) Failure to limit DZ twin samples to same-sex twins

A study on 285 twin pairs from the London Twin Study sample by Stevenson and colleagues (Stevenson et al. 1984; Stevenson et al. 1987) tried to overcome these flaws. Concordance rates of 33-59% for MZ twins and of 29-54% for DZ twins were reported (depending on the reading or spelling measure analysed). Stevenson and colleagues (Stevenson et al. 1987) hypothesised a 'developmental dissociation' whereby genetic factors are important in causing reading problems in young readers but by 13 years of age, spelling becomes more genetically influenced. The lower concordance rates obtained by Stevenson and colleagues (Stevenson et al. 1987) from earlier studies (Hallgren 1950; Hermann 1956; Hermann and Norrie 1958; Bakwin 1973) may be due to differences in ascertainment, DD definition or zygosity

determination (Stevenson et al. 1987). Subsequent studies have found similar results (Decker and Vandenberg 1985, MZ = 85%, DZ = 55%; DeFries and Gillis 1991, MZ = 71%, DZ = 49%; DeFries et al. 1997a, MZ = 67%, DZ = 37%; DeFries et al. 1997b, MZ = 67%, DZ = 38%).

2.2.2 Twin Concordance Rates for Componential Measures of the Developmental Dyslexia Phenotype

Componential measures of DD, like DD defined categorically, have yielded higher concordance estimates in MZ twins compared to DZ twins (word recognition, MZ = 71%, DZ = 32%; phonological decoding, MZ = 66%, DZ = 29%; orthographic coding, MZ = 37%, DZ = 14% and phoneme awareness, MZ = 57%, DZ = 26%; Knopik et al. 2002) estimates supported by previous studies (Stevenson et al. 1984; Stevenson et al. 1987; DeFries and Gillis 1991; DeFries et al. 1997b).

2.2.3 Heritability Estimation

In the analysis of complex disorders, such as DD, where quantitative measures are often used to diagnose reading problems, concordance rate results can lead to inaccuracies since data is lost in the categorisation process (Stevenson et al. 1987; DeFries et al. 1997b). In order to overcome the problems of categorisation, a regression technique was developed (DeFries and Fulker 1985) using continuous measures of reading and/or componential measures of DD. DD affected probands are selected from twin pairs. The extent that the co-twin of the proband regresses towards the population mean is examined as a function of the genetic relatedness of the twin pairs (i.e. MZ versus DZ). The method assess whether DZ co-twins are more similar to the control population than MZ twins. Since the genetic relatedness of the DZ twins is 0.5 (compared to 1.0 for MZ twins), the greater the regression of

the DZ co-twin towards the population mean, the more heritable the disorder, since if genetics play a role in the aetiology of DD/componential phenotypes, the MZ twins should show significantly less regression towards the population mean. The measure of heritability (h^2_g), the proportion of variability that can be attributed to genetic factors, is therefore based on the differential regression towards the population mean of MZ and DZ twins and is specific for the trait being measured, the population under study and the point in time at which it is estimated. Changes in environment, genetic diversity or measurement error can alter estimates.

2.2.4 Heritability of Reading Disability

DeFries and Fulker (DeFries et al. 1987) suggested DD had a heritability (h^2_g) of 0.29 indicating approximately 29% of the reading deficit in probands arises from heritable factors. In a similar sized sample, LaBuda and DeFries (LaBuda and DeFries 1988) reported h^2_g at 0.39 for a discriminant function score (based on the results of six tests administered to each individual) of reading disability.

Two studies using samples from the Colorado reading project (CRP) have suggested that half reading performance deficit, based on performance scores of a number of tests, is due to heritable factors (DeFries and Gillis 1991; Pennington and Gilger 1996). Alarcón and DeFries (Alarcón and DeFries 1997) suggested more than half the deficit could be influenced by genetic factors ($h^2_g = 0.82$).

Hohnen and Stevenson (Hohnen and Stevenson 1999) have shown approximately 60% of variance in literacy is attributable to heritable factors, with 40-50% of the variance in language and 50-60% in phonological awareness.

2.2.5 Heritability Estimates Across Age Ranges

Wadsworth and colleagues (Wadsworth et al. 1989) analysed composite reading scores by grouping their sample in to young (<11.5 years) and old (11.5 – 20.2 years) participants. Heritability estimates of 0.60 and 0.38 were obtained respectively, suggesting heritable influences may influence reading more in younger children, conclusions consistent with those of Stevenson and colleagues' (Stevenson et al. 1987) dissociation theory. However, given the heritability for componential measures of reading in the same study (word recognition = 0.57 vs 0.36, young and old respectively; reading comprehension = 0.68 vs 0.31, young and old respectively and spelling = 0.63 vs 0.52, young and old respectively; Wadsworth et al. 1989), the lack of a higher heritability for spelling in older individuals does not provide support for the dissociation theory.

DeFries and colleagues (DeFries et al. 1997a) also found h^2_g for reading ability was higher in younger participants (h^2_g = 0.61) than older participants (h^2_g = 0.49). Heritability estimates for word recognition, reading comprehension and spelling in the younger sample (0.64, 0.40, 0.52 respectively) and the older sample (0.47, 0.39, 0.68 respectively) also suggested higher h^2_g in the younger participants for reading recognition. Spelling revealed opposite results, providing support for the dissociation hypothesis (Stevenson et al. 1987).

2.2.6 Heritability Estimates for Componential Measures of the Developmental Dyslexia Phenotype

2.2.6.1 Phonological Coding

Phonological coding skill has been shown to have a heritable nature, with h^2_g estimated at ~4.6 in early studies (Olson et al. 1989; Stevenson 1991a). Further

analyses by Olson and colleagues suggested the heritable precursor to the phonological coding deficit was in phoneme segmentation skills (Olson et al. 1989). More recently, Olson and colleagues have estimated h^2_g of phonological coding to be 0.57 (Olson et al. 1991).

2.2.6.2 Orthographic Coding

Early studies suggested orthographic coding ability was not heritable (Olson et al. 1989; Stevenson 1991a). However, subsequent studies have suggested that orthographic coding is in part due to genetic influences (h^2_g = 0.29 (Olson et al. 1991), h^2_g = 0.58 (Gayán and Olson 2001), h^2_g = 0.64 (Knopik et al. 2002)).

2.2.6.3 Phonological Awareness

The heritability of phonological awareness has been estimated to be 0.56 by Gayán and Olson (Gayán and Olson 2001) and 0.64 by Knopik and colleagues (Knopik et al. 2002).

2.2.6.4 Phonological Decoding

Gayán and Olson (Gayán and Olson 2001) noted the h^2_g of phonological decoding as 0.61.

2.2.6.5 Word Recognition

Olson and colleagues estimated h^2_g at 0.62 for word recognition (Olson et al. 1991). Analyses of word recognition based on deficit severity, estimated h^2_g at 0.51 and 0.80 for the worst and best scorers respectively, implying the more severe the deficit, the less heritable the trait (Olson et al. 1991). More recently Gayán and Olson (Gayán and Olson 2001) estimated h^2_g of word recognition between 0.54 –

0.59 depending on the test utilised, whilst Knopik and colleagues estimated h^2_g at 0.60 (Knopik et al. 2002).

2.2.6.6 Spelling

Stevenson and colleagues (Stevenson et al. 1987) estimated 73% of variation in spelling ability to be due to genetic factors. Further support for a genetic influence in spelling ability has come from DeFries and colleagues (h^2_g = 0.62, DeFries et al. 1991), Stevenson and colleagues (h^2_g = 0.47, Stevenson 1991b) and Gayán and Olson (h^2_g = 0.55, Gayán and Olson 2001).

2.3 GENETIC INFLUENCES ON DIFFERENT COMPONENTIAL MEASURES OF DEVELOPMENTAL DYSLEXIA AND READING

Although it has been suggested that the more severe the word recognition deficit, the less heritable the trait (Olson et al. 1991), the same study also found word recognition to be more heritable in individuals performing poorly in phonological coding skills (low = 0.74, high = 0.54) and orthographic coding skills (low = 0.74, high = 0.54) (Olson et al. 1991).

Genetic correlation analyses, based on predicting a co-twins score on one test from a probands score on a different test, revealed that there is a strong genetic relationship between phonological coding and word recognition (r = 0.62) and a weaker relationship between orthographic coding and word recognition (r = 0.22). Phonological coding and word recognition deficits may therefore be due to the same genetic factors but orthographic coding due to environmental factors (DeFries and Gillis 1991) or different genetic factors. Indeed h^2_g estimates of 0.38 and 0.36 (Olson et al. 1999) obtained for phonological coding and orthographic coding respectively suggest that the genetic influences in the two processes of reading may

not entirely due to the same genetic factors given the slight difference in h^2_g estimates.

Twin studies using twin concordance rates and heritability estimates clearly suggest a genetic aetiology to DD. The exact extent of the genetic influence, the number of genes and to what degree different aspects of cognitive processes important in DD are influenced by genetic factors remains uncertain.

2.4 GENDER DIFFERENCES AND DEVELOPMENTAL DYSLEXIA

2.4.1 Twin Concordance Rates

As discussed in Chapter one, Section 1.5.2, the prevalence of DD may differ in males and females. The mechanism behind such a difference and the effect of any genetic influences on gender remains unknown.

Bakwin (Bakwin 1973) analysed concordance rates across twins according to sex and observed a concordance rate of 84% for males and 83% for females (MZ twins). In DZ twins the concordance rates were higher for males (42%) than females (8%). A major problem was the very small number of individuals included thus raising the question, of whether the study was sufficiently powered to detect differences.

2.4.2 Heritability Estimates

Alarcón and colleagues (Alarcon et al. 1995) reported higher MZ correlations than DZ correlations, suggesting heritable factors are involved in reading performance. Male correlations tended to be higher than female correlations, although this was not significant. LaBuda and DeFries and LaBuda and colleagues (LaBuda and DeFries 1990; LaBuda et al. 1990) tested the

interaction of gender with h^2_g of a number of psychometric measures and observed significant effects of gender, although it should be noted that the sample size of twins used in the former study was small (LaBuda and DeFries 1990).

2.4.2.1 Word Recognition

Olson and colleagues (Olson et al. 1991) found no evidence for a difference in h^2_g between males and females for word recognition (males $h^2_g = 0.66$; females $h^2_g = 0.58$; $p = 0.419$).

Although opposite to the findings of Olson and colleagues (Olson et al. 1991), consistent with the findings of Finucci and Childs (Finucci and Childs 1981), at the most extreme end of the word recognition there were more males with word recognition problems than females in a study by LaBuda and colleagues (LaBuda et al. 1990). The disproportionate number of males to females at this extreme end may indicate a differential genetic aetiology in males and females, however this needs to be confirmed in a larger sample.

2.4.2.2 Spelling

The h^2_g of spelling was analysed as a function of gender by DeFries and colleagues (DeFries et al. 1991) who observed a non significant difference in males and females ($h^2_g = 0.66$, $h^2_g = 0.56$ respectively), a result consistent with the finding observed by Olson and colleagues (Olson et al. 1991).

2.5 MODE OF TRANSMISSION

Twin studies have shown that there is a genetic influence on DD and DD component processes, however these studies do not address the mode(s) of genetic transmission. There have been a number of hypotheses about the genetic

transmission of DD including autosomal dominant transmission (Hallgren 1950; Zahalkova et al. 1972; Childs and Finucci 1983), autosomal dominant with incomplete penetrance (Sladen 1970), autosomal recessive (Stephenson 1907; Sladen 1970) and sex-linked recessive (Symmes and Rapoport 1972). Due to the high and similar recurrence rates in parents and siblings, an additive or dominant major locus effect has also been postulated (Pennington and Gilger 1996).

Hallgren (Hallgren 1950) concluded that DD had autosomal dominant transmission. However no formal tests of diagnosis were utilised and genetic heterogeneity was not considered.

In a larger sample, Lewitter and colleagues (Lewitter et al. 1980) found no support for a single major locus (autosomal dominant, autosomal recessive or co-dominant transmission). In families with a female proband, autosomal recessive transmission could not be rejected. The authors suggested that the results were indicative of genetic heterogeneity, a view supported by Finucci and colleagues (Finucci et al. 1976). A major problem of the study by Lewitter (Lewitter et al. 1980) is that adults with DD who used compensation strategies to help them overcome the problems of DD (compensated adults) were not classified as dyslexic.

Evidence of major gene (dominant or additive) transmission with sex-dependent penetrance was observed by Pennington and colleagues (Pennington et al. 1991).

Pennington and Gilger (Pennington and Gilger 1996) looked for sex effects on transmission of DD through families (including compensated dyslexics). Evidence for transmission from fathers to son was observed, allowing the exclusion of a single major X-linked locus and of mitochondrial transmission as the sole cause. The average rates of transmission in male and female offspring were equal,

providing little evidence for sex-influenced transmission. However, the slight excess of affected males reported suggests at least some sex-influenced transmission (although not necessarily a sex-chromosome). Indeed, majority of the sample analysed showed sex-influenced autosomal dominant transmission.

Observations by Wolff and Melngailis (Wolff and Melngailis 1994) suggested that a child with two affected parents was more likely to be affected (and more severely) than children with only one affected parent, indicating a possible additive genetic effect.

2.6 IDENTIFICATION OF DISEASE SUSCEPTIBILITY GENES

In order for the gene(s) for DD to be identified, linkage and association studies have been undertaken to elucidate linkage regions and refine them using association studies.

2.7 LINKAGE STUDIES

Linkage studies involve the study of co-segregation of genetic markers and a phenotype through families, with the aim being to detect departure from independent assortment. The genetic distance between a genetic marker and a disease locus is estimated by observing the segregation of the marker locus with the disease/phenotype status.

Parametric linkage analysis results are often shown in the form of a LOD score function (Morton 1955) that is the logarithm of the odds that the locus is linked to the trait compared to the odds that the locus is not linked to the trait. For a simple genetic disorder a LOD score >3 is considered evidence for linkage, whilst a LOD score <-2 rejects linkage to a region (Morton 1955; Lander and Kruglyak

1995). When LOD scores are calculated with the assumption of heterogeneity they are known as HLOD. Model-free LOD scores (MFLOD) do not require the specification of disease model parameters and allow data to be analysed parametrically (Sham et al. 2000).

The main advantage of parametric LOD-score analysis is that if approximately correct models are applied, they have high power to detect linkage (Fisher and Smith 2001).

In the case of complex disorders such as DD it often difficult to apply the correct model. Consequently non-parametric linkage analysis is commonly used (Molenaar et al. 1993). Non-parametric analyses assess allele-sharing IBD between relative pairs, most commonly sibling pairs. Allele-sharing IBD increases above the chance level when a marker is linked to a variant influencing susceptibility to the phenotype/trait.

2.8 ASSOCIATION STUDIES

Association is a statistical statement referring to the co-occurrence of a particular allele(s) and a phenotype above a level expected by chance (Edwards 1965). Fundamentally association and linkage rely on similar principles and assumptions (Borecki and Suarez 2001). Both rely on the co-inheritance of DNA variants adjacent to each other. However, whilst linkage looks at the information by identifying haplotypes that are inherited intact over only a few generations (where little recombination has occurred), association looks at the retention of adjacent variants over many generations (where lots of recombination may have occurred). Generally, regions of association are smaller than linked regions (Hartl and Clark 1997).

Association analyses are less sensitive to misdiagnoses, decreased penetrance and genetic heterogeneity than linkage analyses (Crowe 1993; Hodge 1993; Spielman et al. 1993) and have been able to detect loci that linkage has failed to identify (Julier et al. 1991). Genes of minor effect are often better investigated using association methods as the number of pedigrees required for linkage analysis would be too large (Risch and Merikangas 1996).

2.8.1 LINKAGE DISEQUILIBRIUM AND ASSOCIATION STUDIES

If loci do not segregate independently in a population they are said to be in linkage disequilibrium (LD). During meiosis, reshuffling of genes reduces the level of LD between pairs of loci from one generation to the next. The extent of LD varies across the genome and is affected by numerous factors including:

2.8.1.1 Meiotic Recombination

For single nucleotide polymorphisms (SNPs), recombination in prophase I, is the primary drive in the breakdown of LD (Tsunoda et al. 2004). Recombination rate varies from zero to values as high as 5cM/Mb across the genome (Kong et al. 2002). On the same chromosome, interspersed regions of several megabases displaying unusually low recombination rates (<0.3cM/Mb) or unusually high recombination rates (>3.0cM/Mb) have been observed (Kong et al. 2002). An inverse correlation between the length of the chromosome and the average recombination rate also exists (Consortium 2001; Kong et al. 2002).

Sex-averaged recombination rates have been reported between 1.33cM/Mb and 1.1cM/Mb (Yu et al. 2001; Kong et al. 2002) with the average rate for males being 0.9cM/Mb (higher nearer the telomeres), and for females 1.7cM/Mb (higher

nearer the centromeres) (Broman et al. 1998; Kong et al. 2002). Since autosomes are not known to contain sex-specific differences in sequence, the difference in recombination rate suggests that other features may contribute to recombination rate (Kong et al. 2002).

A new mutation arising on a chromosome will be in complete LD with all of the other polymorphisms carried by that chromosome. As recombination occurs, LD between variations is broken down. As a result, LD between alleles in part reflects the age of the mutation (Golding 1984; Pritchard and Przeworski 2001; Weiss and Clark 2002).

2.8.1.2 Assortative Mating / Inbreeding

Inbreeding, often considered a type of assortative mating, violates the assumption that two genes segregate independently. When two people who are closely related breed, their genes resemble each other, increasing the likelihood that their children will be homozygous, resulting in the presence of rare, recessive conditions. Inbreeding can produce high levels of association with a reduction in the levels of variation (Nordberg and al 2002).

2.8.1.3 Population Bottlenecks

Population bottlenecks, a temporary and often dramatic reduction in population size, reduce the number of haplotypes in a population so that they are not represented in a population in the proportions expected by the frequencies of alleles. If, a bottleneck has occurred recently (in terms of generations) there may be extensive LD because of the removal of some haplotypes.

2.8.1.4 Selection

Selection inflates LD especially in cases where a single gene has an influence on the risk for a disease. For example the hitchhiking effect, where an entire haplotype that flanks a particular variant is swept rapidly into high frequency (Parsch et al. 2001; Verrelli and Eanes 2001; Wang et al. 2002). Alleles on the same chromosomal segment as an advantageous allele are spread as a result of being 'dragged' along with the advantageous allele (Hurst 1999). Selection against deleterious variants can also increase LD as the haplotypes containing deleterious variants are removed from the population (Charlesworth et al. 1983). The loss of deleterious mutations can 'drag' alleles in LD out of the population resulting in the reduction of the effective population size of the chromosomal region with the disadvantageous allele (Hurst 1999).

2.8.1.5 Genetic Drift

Genetic drift (allele frequency changes occurring by chance) is accentuated in small populations where transmitted gametes are unlikely to represent the gametes within a population (of sperm or ova). When there is increased drift in small populations, LD increases since haplotypes are lost from the population.

2.8.1.6 Population Growth

In populations where there has not been a recent demographic expansion there may be greater levels of LD than in exponentially expanded populations (Terwilliger and Weiss 1998; Terwilliger et al. 1998). Laan and Pääbo examined several populations, which varied in size and structure and showed that LD between

microsatellite markers was greater in a non-expanded (Saami) population than in a rapidly expanded (Finnish) population (Laan and Pääbo 1997).

2.8.1.7 Admixture or Migration

Admixture or migration, the introduction of genes from one distinct population into another, can create LD. When individuals from two genetically different populations mate, the next generation will have association between both linked and unlinked pairs of loci. Over time the LD decays, but much more rapidly for the unlinked loci pairs than the linked pairs.

2.8.1.8 Inversion Polymorphisms

Inversion polymorphisms suppress recombination (Roberts 1976; Martin 1999) (Pritchard and Przeworski 2001), resulting in slow LD break down. In individuals heterozygous for an inversion, little or no recombination occurs between the 'normal' chromosome and the chromosome carrying the inversion polymorphism (Brown et al. 1998). As the mutations accumulate on both genetic backgrounds, the two arrangements diverge, leading to a potential build-up of LD. The extent of this depends on the history and frequency of the inversion, including any natural selection acting on the inversion (Andolfatto et al. 2001).

A region of chromosome 17q21, showing linkage to frontotemporal dementia (FTD; between markers D17S1787 and D17S958) is characterised by an inversion common in the population, which results in high LD across the region. Consequently the genetic defect underlying FTD has been difficult to identify.

2.8.1.9 Gene Conversion

During meiosis, gene conversion (a nonreciprocal exchange of short tracts of DNA between homologous chromosomes) can result in the breakdown of LD similar to recombination or recurrent mutation since the conversion acts like two closely spaced recombination events. Conversion rates in humans are high and become relatively important for very tightly linked markers (Collins et al. 1999; Frisse et al. 2001; Quintana et al. 2001). Gene conversions disrupt LD between closely linked markers, particularly if the conversion spans one of the markers, faster than may be predicted by the recombination rate but will have a smaller role in disrupting LD between distant markers because, as the distance between markers increases, recombination events become more common than gene conversion events.

2.8.2 MEASURING LINKAGE DISEQUILIBRIUM

Early theoretical work, led to the prediction that LD should extend only a few kilobases from common SNPs in the genome and it has been shown that, under a uniform recombination rate of 1cM/Mb and a simple model of human demography, LD in the human genome is unlikely to extend further than about 3kb (Kruglyak 1999). However, if the assumptions of the models are relaxed, LD may extend further. Indeed Gabriel and colleagues (Gabriel et al. 2002) have shown that haplotype blocks (sizeable regions of the genome which show little evidence for historical recombination) exist across the genome. These haplotype blocks may extend upto 22kb in African and African-American populations and 44kb in European and Asian populations (Gabriel et al. 2002). Within these blocks a small number of haplotypes (typically three to five) will capture 90% of the chromosomes

in a population (Gabriel et al. 2002). Although smaller than estimates by Gabriel and colleagues (Gabriel et al. 2002), Hinds and colleagues (Hinds et al. 2005) have shown that haplotype blocks extend approximately 8.8kb in African-American populations, 20.7kb in European-American populations and 25.2kb in Hans Chinese populations.

Two statistical measures, D' and r^2, summarise the LD between two markers and are based on the basic pairwise-disequilibrium coefficient, D, the difference between the probability of observing two markers alleles on the same haplotype and observing them independently in the population. When D' is 0, independence of the two loci is implied. A value of 1 implies all copies at one locus occur exclusively with one of the two possible alleles at the second marker. Any D' value <1 indicates historical recombination has taken place (Hudson and Kaplan 1985). If r^2 (the square of the correlation coefficient that measures the association between alleles) equals 0, independence is implied. It equals 1 when the occurrence of an allele at one marker perfectly predicts the allele at a second locus, in contrast to D', which can reach 1 when allele frequencies differ widely since it reflects the correlation only since the most recent mutation occurred (Zondervan and Cardon 2004). D' is particularly biased in small population sizes and for low allele frequencies.

2.8.3 ASSOCIATION STUDY DESIGN

The basic aim of any genetic association study design is to assess correlations between genotypes and disease phenotypes. Many factors have been cited as contributing to the lack of success in finding complex disease genes through association, including genetic and phenotypic complexity, environmental influences,

sub-optimal sampling and data over-interpretation (Weiss and Terwilliger 2000; Cardon and Bell 2001).

2.8.3.1 Indirect Association Studies

Indirect association studies assume that the actual susceptibility allele may not be genotyped, but that it is located near to a genotyped marker so any association detected is due to LD. Since LD varies across the genome, LD must be considered locally for informative studies to be undertaken.

2.8.3.2 Direct Association Studies

Direct association studies assume that if they exist in a gene, susceptibility or causal alleles themselves are evaluated. In practice however, all direct association studies can capture information indirectly as well. It can be difficult to distinguish between the two.

2.8.4 NON REPLICATION IN GENETIC ASSOCIATION STUDIES

There are a number of reasons why associations observed in one study may not be observed in subsequent studies. These include:

i) Power: Studies often lack power to generate reliable and reproducible results. The power of a study can be influenced by sample size, degree of LD between genotyped polymorphisms and the causal variation(s) and the size of the genetic effect.

ii) Population Stratification: Population stratification can arise when two samples being compared differ because they consist of different sub-

populations (Kidd 1993). Stratifying samples according to ancestry and demographic location can help control for population stratification. If population stratification is not considered in an association study, type I and type II errors may be observed.

iii) Phenotype Difference: Complex diseases can typically vary in severity and type of symptoms resulting in difficulties defining the global phenotype. For each component of a complex phenotype, the processes involved may have unique or shared genetic determinants. As a result, using slightly different phenotypes to study the same gene can result in different observations across studies. Even highly correlated phenotypes can share little genetic variance (Deng et al. 2002; Recker and Deng 2002).

iv) Quality Control: Errors in genotyping and phenotyping can result in the lowering of the power of a study and the creation of type I and type II errors. Problems with quality of phenotyping can be an issue, particularly when a number of individuals are employed to phenotype and different measures are used across different samples.

v) Multiple Testing: By testing multiple genetic markers and componential phenotypes within a study, the number of tests undertaken increases the likelihood of a type I error occurring. As a consequence the nominal significance level of 0.05 is not appropriate for studies. Methods of correcting for multiple testing have been developed including the use of permutations or the Bonferroni correction.

vi) Complex Aetiology: Inconsistent results from different populations may be the result of real biological differences. Even if the causal variant is

under investigation, the variant could be more or less important in different populations, especially if the variant has low relative risk, variable penetrance and different allele frequency in the different populations.

vii) Interactions: Differences in gene-gene and gene-environment interactions within samples or differences in allele frequencies in the disease population could account for non-replication of associations. Analysis of gene-gene and gene-environment interactions could help reduce the non-replication of associations.

2.8.5 CASE-CONTROL SAMPLES AND FAMILY BASED SAMPLES

In the analysis of association of complex genetic traits, there are a number of sample designs, which can be utilised in such a study. These include:

i) Cross-Sectional Samples: A random samples of the general population is genotyped and phenotyped. This is inexpensive and provides estimates of disease prevalence however it can result in few affected individuals, especially if a disease is rare.

ii) Cohort Samples: A subsection of the population is genotyped and disease incidence is followed for a specified period of time. This method provides an estimate of disease prevalence although it can be expensive to follow up and has issues with drop-out rates.

iii) Case-Control Samples: A number of affected individuals (cases) and unaffected individuals (controls) are genotyped. There is no need for follow-up in this sample design and can provide an estimate of exposure effects. Less genotyping is required to generate the same power as a

family based sample. Controls must be selected carefully due to the potential of population stratification.

iv) Family Based Samples: Affected individuals are genotyped along with their parents. This sample collection is robust to population stratification and can estimate maternal and imprinting effects (Cordell and Clayton 2005). However, these samples tend to be less powerful than case-control samples, it can be difficult to collect parents and a large number of individuals has to be genotyped to achieve the same power as case-control samples (Cordell and Clayton 2005). In these samples, alleles or genotypes transmitted to affected individuals are compared to untransmitted alleles or genotypes, thus providing a control sample matched to the case sample with regard to population structure.

2.9 LINKAGE AND ASSOCIATION STUDIES OF DEVELOPMENTAL DYSLEXIA

In order to identify the genes that underlie the genetic predisposition to DD, studies have generally been driven by positional genetics rather than functional candidate gene analyses given the absence of a solid understanding of the likely biological mechanisms that underpin DD. Linkage studies have identified a number of regions, which may harbour DD susceptibility gene(s). Regions showing replicated evidence of linkage to DD have been named DYX1–DYX9 by the Human Gene Nomenclature Committee (www.gene.ucl.ac.uk/cgi-bin/nomenclature/search genes.pl).

2.9.1 CHROMOSOME 15 – DYX1

A summary of the linkage findings on chromosome 15 is shown in Figure 2.1.

Smith and colleagues (Smith et al. 1983) achieved a LOD score of 3.24 at the centromere of chromosome 15. In an extension to this study (Smith et al. 1990), the LOD score fell to 1.7 suggesting evidence for linkage was no longer significant under criteria set out by Lander and Kruglyak (Lander and Kruglyak 1995).

Two small samples (Bisgaard et al. 1987; Lubs et al. 1991b) also failed to replicate the original linkage to chromosome 15. However, it should also be noted that there was limited assessment of the DD phenotype by Bisgaard and colleagues (Bisgaard et al. 1987) (LOD = -3.42) and the marker set utilised by Lubs and colleagues was very sparse (Lubs et al. 1991b).

Smith and colleagues (Smith et al. 1991) observed a linkage region distal to that identified previously on chromosome 15q (Smith et al. 1983). Using a qualitative phenotypic measure (individuals were classified affected or unaffected), marker YNZ90 yielded evidence of linkage ($p = 0.009$) whilst marker JU201 was linked ($p = 0.03$) with a quantitative phenotypic measure. Linkage to a third marker was observed when the sample was cut to include only sibling pairs with at least one severely affected sibling (THH114, $p = <0.05$), suggesting that DYX1 may have more influence on those individuals falling at the tail end of the DD distribution (Smith et al. 1991).

Grigorenko and colleagues (Grigorenko et al. 1997) used five subphenotypes including phonological awareness, phonological decoding, rapid automated naming, single-word reading and discrepancy between intelligence and reading performance to test for linkage. Parametric linkage analysis gave a LOD score of 3.15

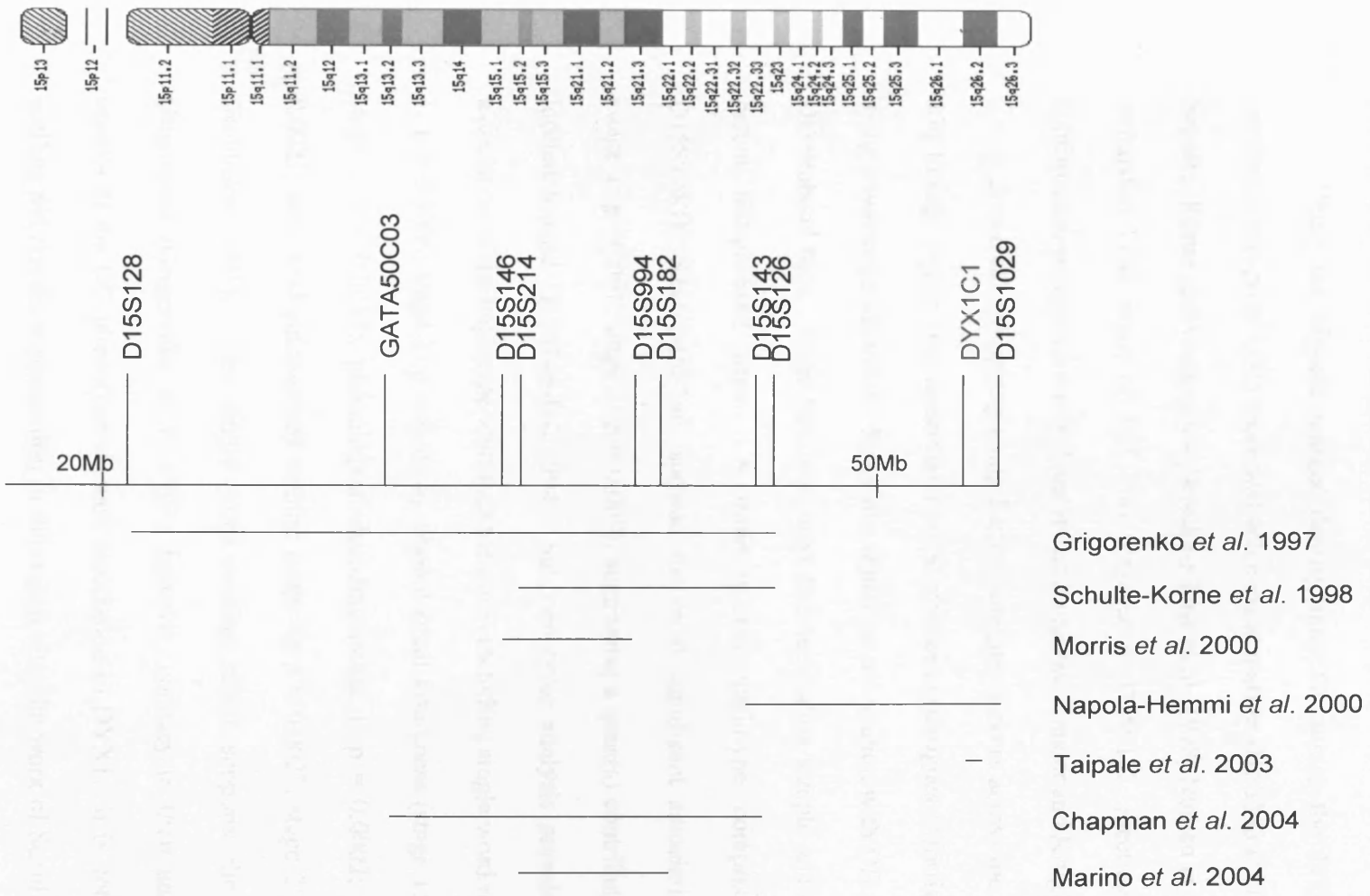


Figure 2.1 Linkage and association regions for DD on chromosome 15. Studies show linkage and/or association of DD to a region spanning approximately 30Mb between markers D15S128 and D15S1029. Most replicated evidence lies around the region D15S146-D15S126. Located distal to this region is the *DYX1C1* (*EKN1*) gene identified as a susceptibility gene for DD by Taipale and colleagues (Taipale et al. 2003; Raskind et al. 2005).
Note, the study by Napola-Hemmi and colleagues (Nopola-Hemmi et al. 2000) is discussed in Chapter six.

(D15S143) with single-word reading and a marker within the region implicated previously (Smith et al. 1983).

Using the highest marker density thus far across the DYX1 locus, a maximum two-point LOD score was achieved at marker D15S143 (LOD = 1.26) by Schulte- Körne and Colleagues (Schulte-Körne et al. 1998; Nöthen et al. 1999). A multipoint LOD score of 1.78 was achieved at D15S132, meeting the linkage confirmation criteria set out by Lander and Kruglyak (Lander and Kruglyak 1995).

Association mapping using 8 microsatellite markers across the chromosome 15q linkage region was undertaken by Morris and colleagues (Morris et al. 2000) using a two-stage approach. Stage one tested for association with DD in 101 parent DD-proband trios. Stage two, was used as a replication sample and contained 88 parent DD-proband trios. A three marker haplotype comprising markers D15S994/D15S214/D15S146, showed the most significant association with DD (stage 1: $p < 0.001$; stage 2: $p = 0.009$), suggesting a gene(s) contributing to DD is located around D15S146-D15S994. Subphenotypic analysis revealed replicated association of the haplotype with four measures including single word reading (stage 1: $p = 0.0006$; stage 2: $p = 0.0014$), phonological awareness (stage 1: $p = 0.0001$; stage 2: $p = 0.0058$), phonological decoding (stage 1: $p = 0.0002$; stage 2: $p = 0.0021$) and rapid automatised naming (stage 1: $p = 0.0023$; stage 2: $p = 0.0339$) (Robinson 2001). The single word reading result supports the findings of Grigorenko (Grigorenko et al. 1997), however, contrary to their analyses, other aspects of the DD phenotype showed association to DYX1. It is noteworthy that spelling did not show association to the region after the work of Schulte-Körne and Nöthen (Schulte-Körne et al. 1998; Nöthen et al. 1999).

Extending this work, Morris and colleagues (Morris et al. 2000) selected two candidate genes (Phospholipase C β 2 (*PLCB2*) and phospholipase A_2, group IVB (cytosolic; *PLA2G4B*)) from the region and tested for association to DD in a case-control and family-based sample (Morris et al. 2004). Although evidence of association was observed in the case-control sample, no association was observed in the family-based sample. The differences in results between the case-control and family-based sample were probably attributable to sampling variance and multiple testing and so are likely to represent a false positive.

Chapman and colleagues showed evidence of linkage between single word reading (D15S143, LOD = 1.72) and DYX1 and weak evidence of linkage for phonological decoding (D15S659, LOD = 0.39; D15S143, LOD = 0.63). Multipoint parametric analysis generated a maximum LOD score of 2.47 in the interval GATA50C03 to D15S143.

Using 6 microsatellite markers across 9Mb of chromosome 15q, Marino and colleagues tested for association with DD in a sample of 121 parent DD-proband trios of Italian ancestry (Marino et al. 2004). Although some evidence for association was observed, after correction for multiple testing, none of the results remained significant. It is particularly interesting that D15S994, shown to be associated to DD previously (Morris et al. 2000) was not associated to DD in their sample (Marino et al. 2004). Non-replication of association studies has been discussed in Section 2.8.4.

2.9.2 CHROMOSOME 6 – DYX2 (6p) AND DYX4 (6q)

Smith and colleagues (Smith et al. 1991) showed evidence of linkage between glyoxylase 1 (GLO) and a qualitative definition of DD ($p = 0.015$) and also

between properdin factor (BF) and a quantitative definition of DD ($p = 0.0$). The significance achieved was equivalent to a LOD score of greater than 3, suggesting linkage between BF and GLO.

Cardon and colleagues (Cardon et al. 1994) reported a 2cM region of linkage between markers D6S105 and tumour necrosis factor beta (TNFB) ($p = 0.04$). This was replicated in a DZ twin sample ($p = 0.0003$) and appeared strongest in individuals representing the most severe deficits in reading (twin sample $p< 0.00001$; sibling sample $p = 0.066$). Subsequently, Gayán and colleagues (Gayán et al. 1995) used the same sample of DZ twins to test for linkage with word recognition, reading comprehension, spelling, non-word reading, orthographic coding, pig latin and a time-limited word recognition test. Word recognition and the time-limited word recognition measure ($p<0.001$) yielded most evidence of linkage, although linkage was also shown with spelling, reading comprehension and non-word reading ($p < 0.05$). The results are consistent with those shown by Cardon and colleagues (Cardon et al. 1994; Cardon et al. 1995) (see Figure 2.2 for a graphical representation of the linkage and association results between DD and chromosome 6).

Grigorenko and colleagues (Grigorenko et al. 1997) dissected the composite phenotype of DD into five components for linkage analysis, including phonological awareness, phonological decoding, rapid automated naming, single word reading and discrepancy between intelligence and reading performance. Multipoint affected pedigree member (APM) linkage results showed significant linkage of phonological awareness ($p < 10^{-6}$) to five adjacent markers between D6S109 and D6S306. By using subphenotypes of DD, Grigorenko and colleagues (Grigorenko et al. 1997) also suggested the region may influence phonological awareness.

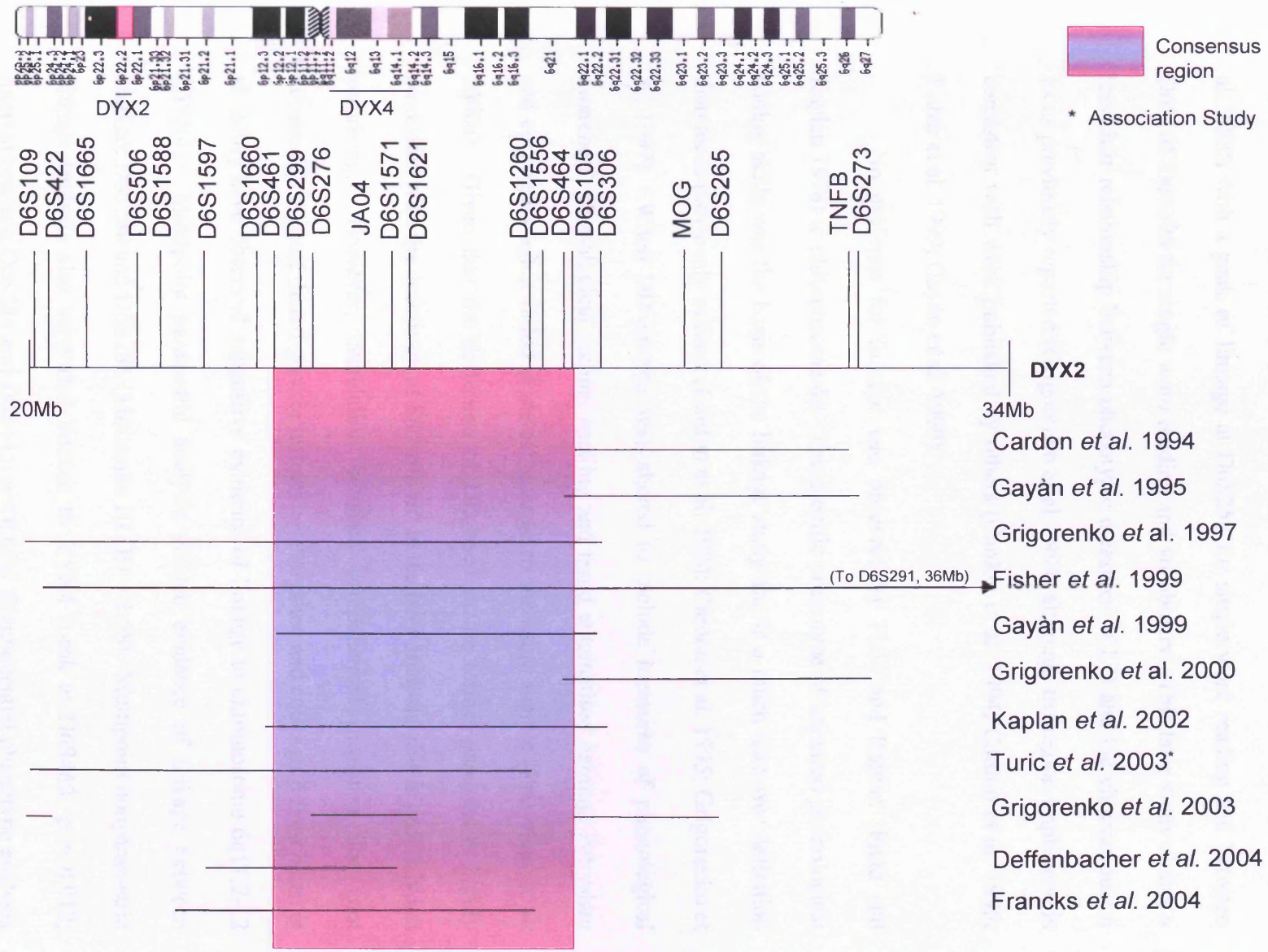


Figure 2.2 Linkage and association regions between chromosome 6 and DD. Studies show linkage and/or association of DD to a region spanning approximately 16Mb between markers D6S109 and D6S291. The region D6S461-D6S105 (4.2Mb) shows most evidence for linkage and/or association across studies (see consensus region) and is the most likely location of a QTL influencing DD.

A later analysis also by Grigorenko and colleagues (Grigorenko et al. 2000), reported the linkage region localised between D6S464 and D6S306 (Grigorenko et al. 2000) with a peak of linkage at D6S258 for single word reading and between D6S105-D6S306 for single word reading and vocabulary. This later study shows a less clear relationship between phenotypic measures of DD and the chromosome 6 locus previously reported (Grigorenko et al. 1997) although the region implicated is consistent with work published by others (Cardon et al. 1994; Cardon et al. 1995; Fisher et al. 1999; Gayán et al. 1999).

No evidence for linkage was observed by Field and Kaplan (Field and Kaplan 1998) to chromosome 6p. The specific phenotype of impaired phonological coding skills was the basis of the linkage study and is a much narrower definition than those previously utilised (Cardon et al. 1994; Cardon et al. 1995; Grigorenko et al. 1997). When DD criteria was altered to include measures of phonological awareness, phonological coding, spelling and rapid automatised naming, Petryshen and colleagues again failed to detect linkage in the same sample (Petryshen et al. 2000a). Given that the definition of DD used in the former analyses is highly correlated with the definition of DD utilised in the later analysis, this is probably not surprising. However, correlation between componential measures does not necessarily suggest shared genetic influence. Petryshen and colleagues (Petryshen et al. 2001) have observed suggestive evidence of linkage to chromosome 6q11.2-12 (DYX4). Multipoint parametric analysis yielded evidence of linkage between markers D6S280 and D6S286 (Maximum HLOD = 1.58). Multipoint nonparametric linkage analysis also supported linkage to DYX4 (peak at D6S460, $p = 0.012$; interval between D6S286 and D6S445, $p < 0.05$). Componential phenotype analysis suggested the QTL influenced, spelling (D6S865, peak LOD = 3.34), phonological

coding (D6S965, peak LOD = 2.08) and phonological awareness (sib-pair analyses, D6S455, p = 0.026). As yet, no other samples have replicated the linkage reported in this region.

No evidence for linkage of spelling to chromosome 6p was observed (two-point LOD < 0.24) by Schulte-Körne and Nöthen and colleagues (Schulte-Körne et al. 1998; Nöthen et al. 1999). However, if a gene on chromosome 6 has only a minor effect on spelling, the size of the sample may have been too small to detect an effect.

Measures including IQ/reading discrepancy scores, word recognition, irregular word reading and non-word reading, have placed the linkage region between markers D6S422 and D6S291, with maximal evidence between D6S276 and D6S105 by Fisher and colleagues (Fisher et al. 1999). In that study the locus affected both phonological and orthographic skills and was not specific to phoneme awareness (phonological decoding, p = 0.00035; orthographic coding, p = 0.0035; composite score, p = 0.0003). These results are relatively consistent with those found by several others (Cardon et al. 1994; Cardon et al. 1995; Grigorenko et al. 1997; Gayán et al. 1999) although the marker order used in this study differs from that used by Grigorenko and colleagues (Grigorenko et al. 1997).

Gayán and colleagues (Gayán et al. 1999) tested for linkage between chromosome 6p and componential measures of word recognition, orthographic coding, phonological decoding and phoneme awareness. Evidence was reported of a locus in the vicinity D6S461 to D6S306-D6S258 that influenced orthographic and phonological skills (orthographic choice LOD = 3.10; phonological decoding LOD = 2.42) (Gayán et al. 1999). More specifically the position of the locus lies between markers D6S276 and D6S105, overlapping the region observed by Grigorenko and

colleagues (Grigorenko et al. 1997). It is of note that the marker order does not correspond exactly between the two studies. Markers D6S276 and D6S105 positioned proximally to D6S306 by Grigorenko (Grigorenko et al. 1997) are located distal to D6S306 later on, making a more compact region of linkage (Gayán et al. 1999).

Kaplan and colleagues tested 11 componential phenotypes for linkage to the 6p region (Kaplan et al. 2002). All 11 component phenotypes yielded evidence for association to 6p ($p<0.05$) however the significance and location of the linkage varied amongst traits and which method of analysis was undertaken. The most significant results were obtained near D6S461 with measures of orthographic choice, orthographic composite score and timed word recognition. A peak of association was also found with orthographic choice around JA04 ($p = 0.0021$), 1Mb from D6S461. Kaplan and colleagues suggested that the most likely location for a quantitative trait locus (QTL) is 4Mb around JA04. Londin and colleagues (Londin et al. 2003) followed on from this study and identified 19 genes and 2 pseudogenes around JA04. Reverse-transcription polymerase chain reactions (RT-PCR) testing the tissue specific expression of the genes, identified five possible candidate genes based on high expression of the genes in brain tissue (*P24*, *SSADH*, *GPLD1*, *KIAA0386* and *KIAA0319*).

Multipoint linkage analysis by Fisher and colleagues (Fisher et al. 2002) suggested linkage between D6S1610 and phonological decoding (variance-components, $p = 0.00001$), suggesting phonological awareness is influenced by a gene(s) in the DYX2 region.

Grigorenko and colleagues (Grigorenko et al. 2003) observed linkage most prominently with the single markers D6S299 (LOD = 2.01, phonemic

awareness/decoding/single word reading pathway) and D6S2222 (LOD = 2.57, phonological awareness). In all, three interesting linkage regions were found: D6S109-JA01 (relating to a report by Turic and colleagues (Turic et al. 2003)), D6S299-D6S1621 (relating to a report by Kaplan and colleagues (Kaplan et al. 2002)) and D6S105-D6S265 (relating to reports by Cardon, Fisher and Grigorenko (Cardon et al. 1994; Cardon et al. 1995; Fisher et al. 1999; Grigorenko et al. 2000; Grigorenko et al. 2003).

Turic and colleagues (Turic et al. 2003) used 21 microsatellites to cover an 18cM region of chromosome 6 to test association to DD using the 2 stage approach employed by Morris and colleagues (Morris et al. 2000). One two-marker haplotype (D6S422-D6S1665) showed replicated evidence of association (stage 1 p = 0.01; stage 2 p = 0.04) but most association was found with the haplotypes D6S109/D6S422/D6S1665,D6S422/D6S1665/D6S506, D6S506/D6S1029/D6S1660 and D6S1281/D6S1558/D6S1260. Haplotypes D6S109/D6S422/D6S1665 and D6S506/D6S1029/D6S1660 revealed most evidence of association across the two samples. Association between the most significant haplotype D6S109/D6S422/D6S1665 and subphenotypic measures showed haplotype association with single-word reading, spelling, phonological awareness, phonological decoding, orthographic accuracy and rapid automatised naming. The results suggest a broad region of association spanning markers D6S109 to D6S1260. The region D6S461-D6S1260 of the narrow region of association lies within the consensus region identified from other linkage and association studies of DD on chromosome 6p (see Figure 2.1).

A region, 10cM distal to other regions of linkage on chromosome 6p has shown weak evidence of linkage to DD and no evidence for linkage with

componential measures of DD (Chapman et al. 2004). Scoring of the trait and ascertainment criteria meant probands were less impaired than in other studies with similar componential measure scores. The failure to replicate specific linkages may be related to the differences in the measures rather than observation of a 'false negative'.

It is unclear why studies differ over the exact nature of the linked phenotype. One explanation is that significance levels vary as a consequence of variation of methodological factors such as the sample size for each component of the phenotype and the reliability of phenotypic testing (Fisher et al. 1999). However, the evidence summarised above suggests the QTL on chromosome 6p may affect both phonological and orthographic components of reading. It is noteworthy that many studies show a high correlation between measures of phonological decoding and orthographic coding (Gayán et al. 1999).

After starting this thesis, work has been published narrowing the chromosome 6p linkage region further. Deffenbacher and colleagues (Deffenbacher et al. 2004) tested five component phenotypes of DD for linkage with 22 markers spanning 22Mb of 6p21.3. Significant linkage was observed with all five component phenotypes, with the five converging over a region spanning ~3.24Mb between markers D6S1597 and D6S1571. Ten genes were tested for association with components of DD. Association was detected in five of the genes tested, *VMP*, *DCDC2*, *KIAA0319*, *TTRAP* and *THEM2*.

Francks and colleagues (Francks et al. 2004) used association analysis in 223 siblings (UK Caucasian) to refine the QTL on 6p. A 77kb region revealed association with DD spanning *TTRAP* and the first four exons of *KIAA0319*. This was replicated in a sample of 159 families from the Colorado twin study of reading

(American Caucasian). Association was observed with a number of component phenotypes including orthographic coding, phonological decoding, phonological awareness, spelling and single word reading.

The results of Deffenbacher and colleagues (Deffenbacher et al. 2004) and Francks and colleagues (Francks et al. 2004) are discussed further in Chapters four and five.

2.9.3 CHROMOSOME 2 – DYX3

A LOD score of 3.53 was obtained at D2S2183 (assuming sex dependant penetrance and specific phenocopy values and differential weight factors; model 1), 2.92 at D2S393 (included all normal readers and only the affected family members with a history and positive test for DD; model 2) and 4.32 at D2S378 (excluded all individuals under the age of 20; model 3) by Fagerheim and colleagues (Fagerheim et al. 1999). Lander and Kruglyak (Lander and Kruglyak 1995) criteria for genomewide significance was exceeded in models 1 and 3. A gene was suggested to lie between D2S2352 and D2S1337. Further analysis of this family (including extra family members and new microsatellite markers) (Fagerheim et al. 2002), defined a new region of ~4.3Mb between D2S2153 and D2S444 (Fagerheim et al. 2002) (see Figure 2.3 for a graphical representation of the regions of linkage observed between DD and chromosome 2).

Petryshen and colleagues found weak evidence for linkage of phonological decoding dyslexia to the region DYX3 (D2S1352, maximum two-point LOD = 0.77, $\theta = 0.3$; D2S1352 multipoint peak HLOD = 0.07) (Petryshen et al. 2002). However non-parametric linkage analysis showed more evidence of linkage (D2S1352, $Z_{all} = 2.33$, $p = 0.0087$; and D2S1352 to D2S2352, $p < 0.05$). The discrepancy between

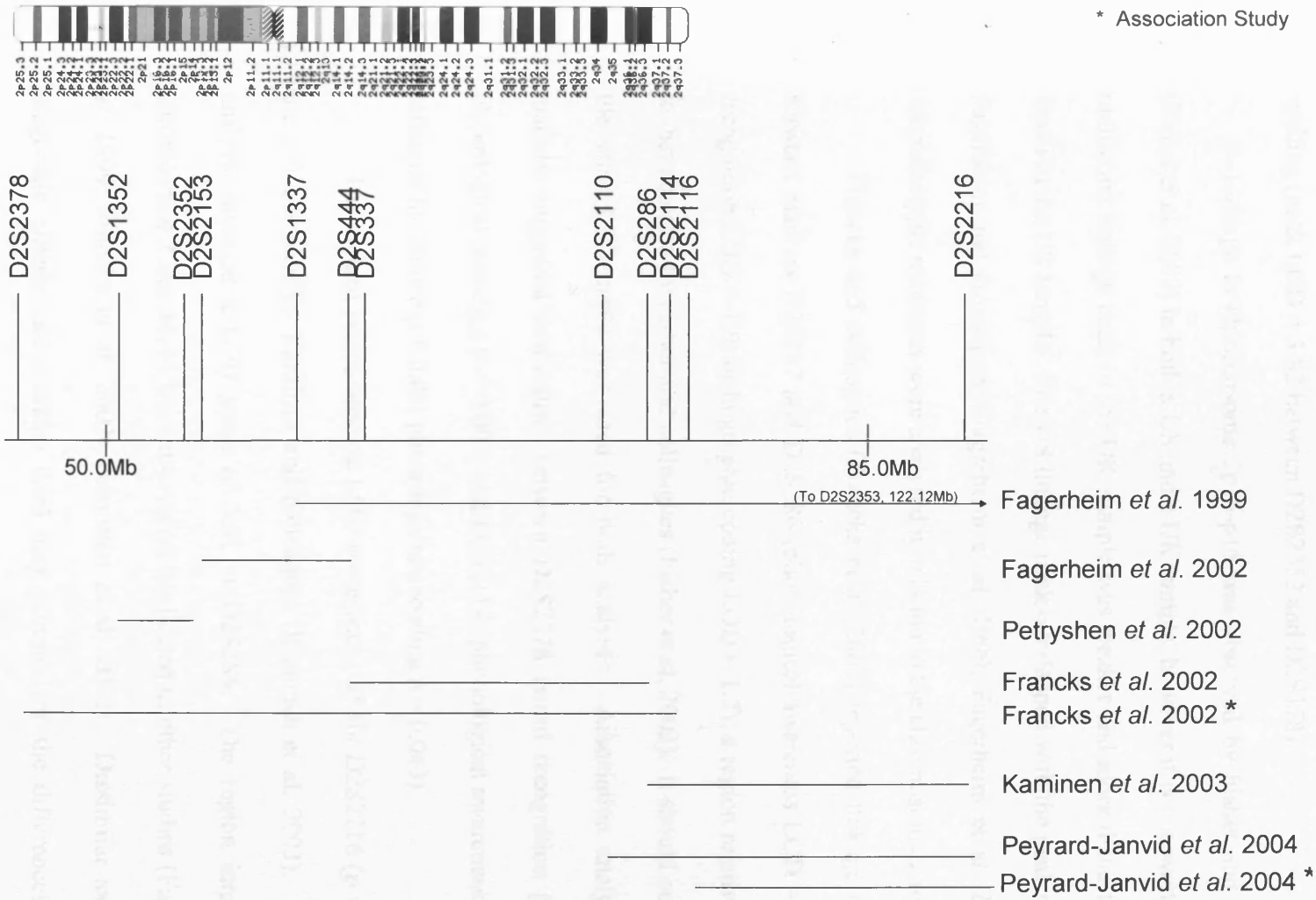


Figure 2.3 Linkage and association regions between chromosome 2 and DD. Studies show linkage and/or association of DD to a region spanning approximately 72Mb between markers D2S2378 and D2S2999. The region D2S1352-D2S2216 (35Mb) shows most evidence for linkage and/or association across studies.

parametric and nonparametric linkage results could be due to mis-specification of the model. Multipoint analyses found evidence for linkage of phonological coding (D2S378, LOD = 1.13), phonological awareness (D2S378, LOD = 1.01) and spelling (peak LOD = 3.82 between D2S2352 and D2S378).

Linkage to chromosome 2p15-p16 was observed by Fisher and colleagues (Fisher et al. 2002) in both a US and a UK sample, however it is noteworthy that the multipoint linkage peak in the UK sample was weaker and more distal to the peak found in the US sample. The US linkage peak overlapped with the peak reported by Fagerheim and colleagues (Fagerheim et al. 1999; Fagerheim et al. 2002). No subphenotypic measures were analysed in relation to the chromosome 2 results.

Francks and colleagues (Francks et al. 2002) reported linkage to a region between markers D2S337 and D2S286 (phonological awareness LOD = 2.3; word recognition LOD = 1.9; orthographic coding LOD = 1.7); a region relatively similar to that reported by Fisher and colleagues (Fisher et al. 2002). It should be noted that the same US sample was used for both analyses. Association analysis of the markers suggested association between D2S2378 (word recognition p = 0.004; phonological decoding p = 0.004) and D2S2114 (phonological awareness p = 0.015; orthographic choice p = 0.03; phonological decoding p = 0.043).

The highest nonparametric LOD score was 2.55 for D2S2216 (p = 0.004) in the genome scan by Kaminen and colleagues (Kaminen et al. 2003). Parametric analysis revealed a LOD score of 3.01 at D2S286. The region implicated on chromosome 2 lies 34cM from the region implicated in other studies (Fagerheim et al. 1999; Francks et al. 2002; Petryshen et al. 2002). Dissimilar sample sets, diagnostic criteria and markers used may account for the differences observed (Kaminen et al. 2003) or more than one candidate gene could be located on

chromosome 2. Extending these findings (with a slightly larger, but overlapping sample), Peyrard-Janvid (Peyrard-Janvid et al. 2004) used fine mapping (21 markers, average distance 1.8cM) across the 40cM region between markers D2S391 and D2S2181 and six SNP markers spanning ~670kb around D2S286/rs3220265. Nonparametric linkage analysis revealed the highest LOD score for marker D2S2216 (LOD = 3.0; p = 0.001). All markers spanning D2S2110 to D2S2181 showed LOD scores above 2.0 ($p < 0.01$) a region spanning ~15cM. Association studies narrowed the region to ~12Mb between D2S2116 and D2S2181, supporting the findings of others (Fagerheim et al. 1999; Petryshen et al. 2000b; Fagerheim et al. 2002; Petryshen et al. 2002).

In a genome-screen of phonemic decoding efficiency and reading accuracy, Raskind and colleagues (Raskind et al. 2005) showed evidence of linkage to chromosome 2q (~q22.3) (phonemic decoding efficiency, LOD = 2.65). Additional markers on chromosome 2 showed linkage of phonological decoding efficiency to D2S1399 (LOD = 3.0; multipoint $LOD_{max} = 2.89$). This signal lies on the q arm of chromosome 2, in contrast to previous studies (Fagerheim et al. 1999; Petryshen et al. 2000b; Fagerheim et al. 2002; Francks et al. 2002; Petryshen et al. 2002; Peyrard-Janvid et al. 2004).

2.9.4 CHROMOSOME 3 – DYX5

A genome-scan by Nopola-Hemmi and colleagues (Nopola-Hemmi et al. 2001) revealed non-parametric linkage to the region 3p12-q13 (Z_{all} 5.8, p = 0.0017). A second family, genotyped for seven microsatellites spanning 60cM of the linked region implicated in the first family and 11 additional markers (genotyped in both families; average intermarker distance ~2cM), revealed that 90% (n = 19) of dyslexic subjects shared a haplotype on chromosome 3 (Nopola-Hemmi et al. 2001).

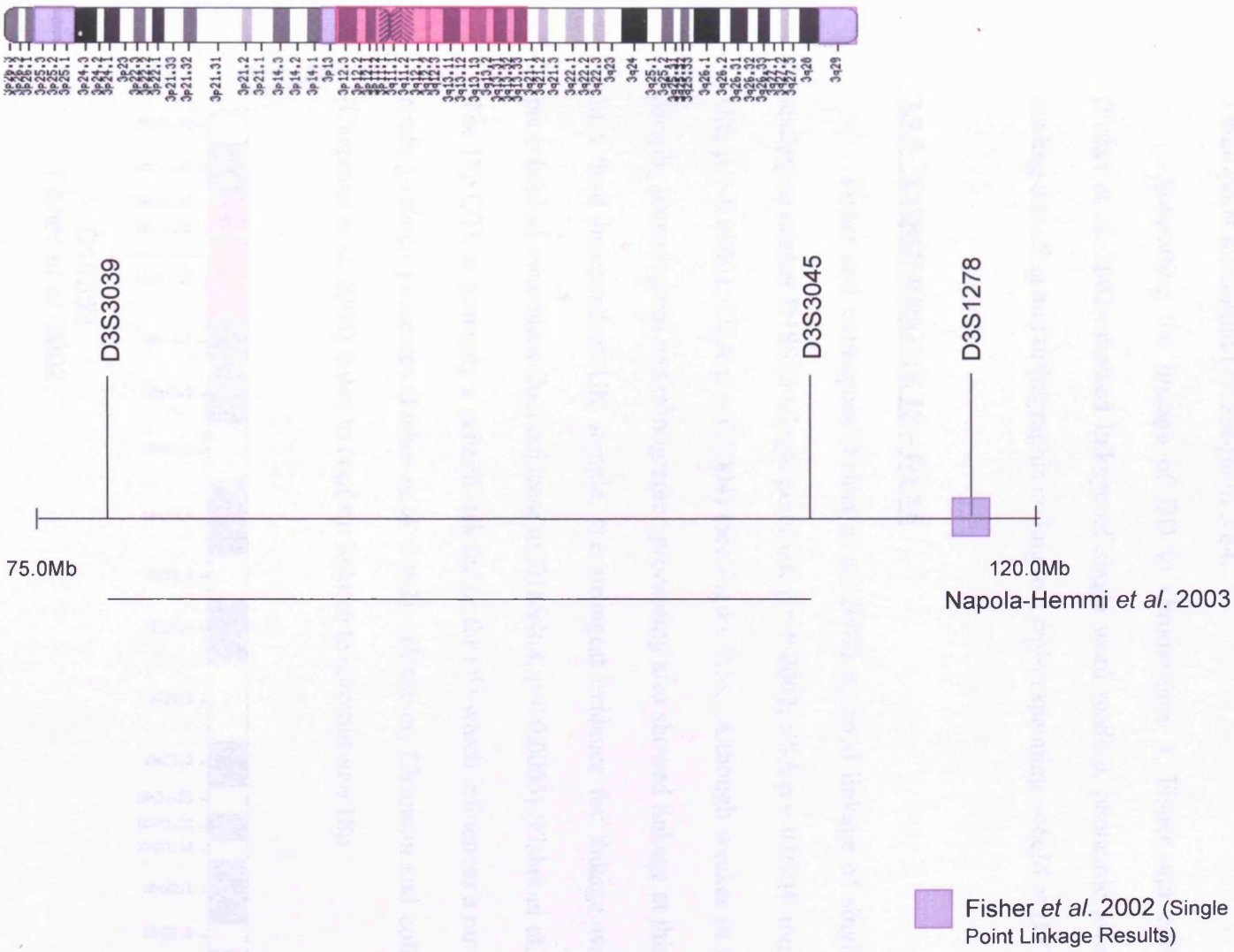


Figure 2.4 Linkage between DD and chromosome 3. A region of linkage between markers D3S3039 and D3S3045 has been identified by Napola-Hemmi and colleagues (Nopola-Hemmi et al. 2001). Linkage has also been observed between D3S1278 and DD by Fisher and colleagues (Fisher et al. 2002).

The absence of the haplotype in 2 individuals may represent the complexity of the DD phenotype or phenocopies. One individual limited the haplotype to 20cM between D3S3039 and D3S3045 (see Figure 2.4). Parametric linkage analysis using the four most informative markers (D3S2454, D3S3039, D3S1595, D3S3655) gave a multipoint maximum LOD score of 3.84.

Supporting the linkage of DD to chromosome 3, Fisher and colleagues (Fisher et al. 2002) showed linkage of single word reading, phonemic awareness, reading decoding and orthographic coding to a region spanning ~46cM at DYX5.

2.9.5 CHROMOSOME 18 – DYX6

Fisher and colleagues (Fisher et al. 2002) reported linkage of single word reading to marker D18S53 (single point UK p = 0.0002; USA p = 0.0004; multipoint UK, p = 0.00001; USA p = 0.0004) (see Figure 2.5). Although weaker in the UK sample, phonological and orthographic processing also showed linkage at this locus. In a third independent UK sample, the strongest evidence for linkage was with phonological awareness (loci adjacent to D18S464, p < 0.0005) (Fisher et al. 2002). The 18p QTL is probably a general risk factor for DD which influences a number of reading-related processes (Fisher et al. 2002). However, Chapman and colleagues (Chapman et al. 2004) failed to confirm linkage to chromosome 18p.

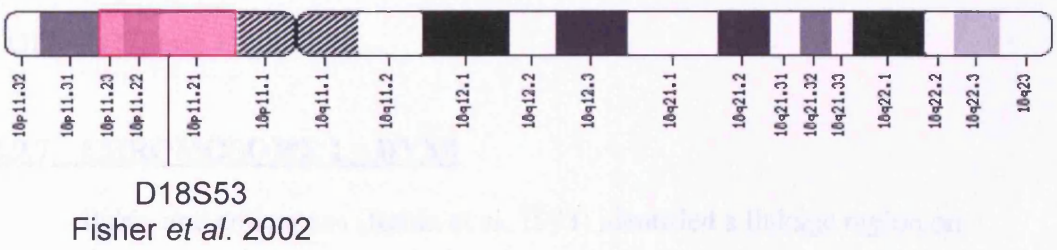


Figure 2.5 Linkage between DD and chromosome 18. A region around D18S53 has been identified by (Fisher et al. 2002).

2.9.6 CHROMOSOME 11 – DYX7

Given the comorbidity of DD and ADHD and the possibility of shared genetic influences (August and Garfinkel 1990; Purvis and Tannock 1997; Shaywitz 1998; Kaplan et al. 2001) genes showing association to ADHD are reasonable candidate genes for DD. The candidate gene dopamine D4 receptor (*DRD4*) located on chromosome 11p15.5 and particularly the 7-repeat allele of a 48bp variable number tandem repeat (VNTR) in exon 3, has been shown to contribute to ADHD susceptibility (see meta-analyses by Faraone et al. 2001; Maher et al. 2002). Hsiung and colleagues (Hsiung et al. 2004) subsequently tested for linkage of DD to the region of chromosome 11 surrounding *DRD4*. Two-point linkage analysis results suggested linkage to the *DRD4* variable number tandem repeat (VNTR; MFLOD = 2.27) and several nearby markers including D11S1984 (MFLOD = 2.32), D11S1363 (MFLOD = 2.13) and HRAS (MFLOD = 2.68). Three-point analysis also suggested linkage between the VNTR and HRAS (MFLOD = 3.57, p = 0.00005) however association analysis was not significant (p = 0.30). The linkage results suggest that a gene influencing DD susceptibility may lie on chromosome 11 in the region of *DRD4*. Supporting the evidence for linkage to this region, Fisher and colleagues (Fisher et al. 2002) have observed linkage of DD to D11S1338 (UK, phoneme awareness p = 0.001) and a weak signal has been reported by Raskind and colleagues (Raskind et al. 2005) with phonological decoding efficiency (D11S2002, LOD = 1.27) (see Figure 2.6).

2.9.7 CHROMOSOME 1 – DYX8

Rabin and colleagues (Rabin et al. 1993) identified a linkage region on chromosome 1p using 9 three-generation families, in particular with markers Rh (Z_{max} = 1.95) and D1S165 (Z_{max} = 2.33). Lubs and colleagues (Lubs et al. 1991a)

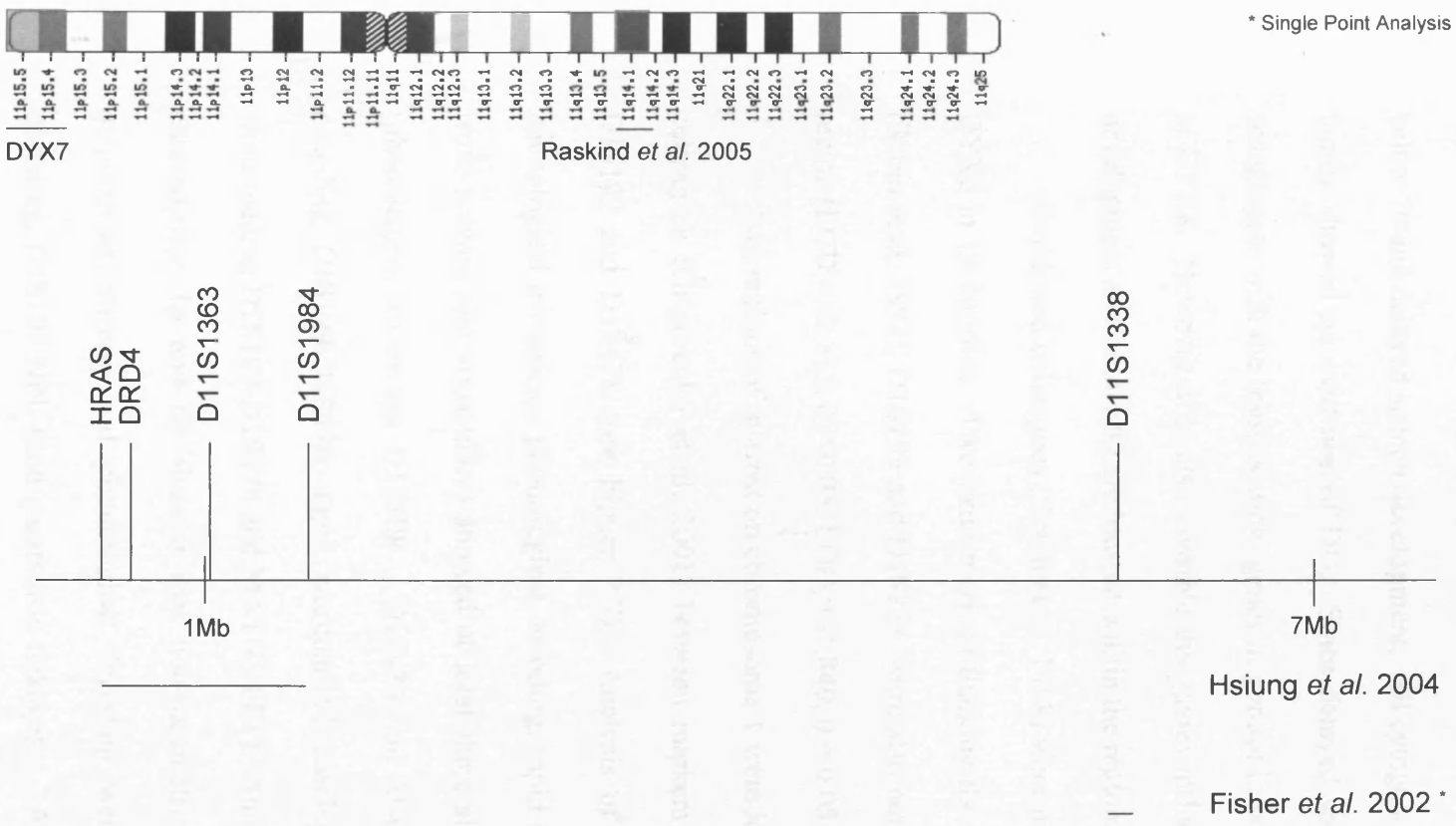


Figure 2.6 Linkage observed between DD and chromosome 11. Chromosome 11 was tested for linkage to DD based on evidence of association between chromosome 11 and ADHD and the comorbidity of DD and ADHD. Fisher and colleagues (Fisher et al. 2002) have observed some evidence for linkage between phonological awareness and chromosome 11. This linkage signal lies approximately 5Mb from D11S1984 at the proximal end of the linkage region identified by Hsiung and colleagues (Fisher 1905; Thomas 1905; Hinshelwood 1907; Stephenson 1907; Hinshelwood 1911; Hsiung et al. 2004). Raskind and colleagues observed a weak linkage signal with D11S2002, which lies on the q arm of chromosome 11.

found linkage to 1p36 including Rh (LOD = 1.5).

A balanced translocation (t[1;2][1p22;2q31]) has also been found to segregate in a family with DD (Froster et al. 1993) using an age discrepancy definition of DD. It is noteworthy however that all translocation carriers had an IQ below 70 and delayed speech development. All cytogenetically normal people in the family showed no evidence of DD. Since delayed speech development and DD cosegregate with the translocation, genes important in both processes may be linked to DYX8. However, it is also possible that genes influencing only delayed speech development and/or low IQ are located within the region.

Smith and colleagues (Smith et al. 1998) were unable to replicate linkage to DYX8 in 19 families. Two extra markers flanking the region implicated by Rabin (Rabin et al. 1993), D1S199 and D1S234, were also not indicative of linkage to the region (LOD = -2.315, $\theta = 0.05$; LOD = -1.840, $\theta = 0.05$ respectively).

Two regions of interest on chromosome 1 were identified by Grigorenko and colleagues (Grigorenko et al. 2001) between markers D1S253 and D1S436 and D1S199 and D1S478 (see Figure 2.7). Analysis of all phenotypes (including, phonological awareness, phonological decoding, rapid automatised naming, single word reading and vocabulary) showed at least three significant peaks for linkage (phonological awareness D1S508 – D1S436 and MATN1 – PPT; phonological decoding D1S199-D1S478; rapid automatised naming D1S253-D1S507; single word reading D1S199-D1S478 and MATN1-PPT). An interaction analysis between chromosome 1p and 6p showed that non-parametric linkage scores for rapid automatised naming and phonological decoding were increased (phonological decoding, D1S199 NPL (non-parametric linkage) = 2.62; rapid automated naming, D1S470 NPL = 5.74) suggesting an interaction between genes on chromosome 1p

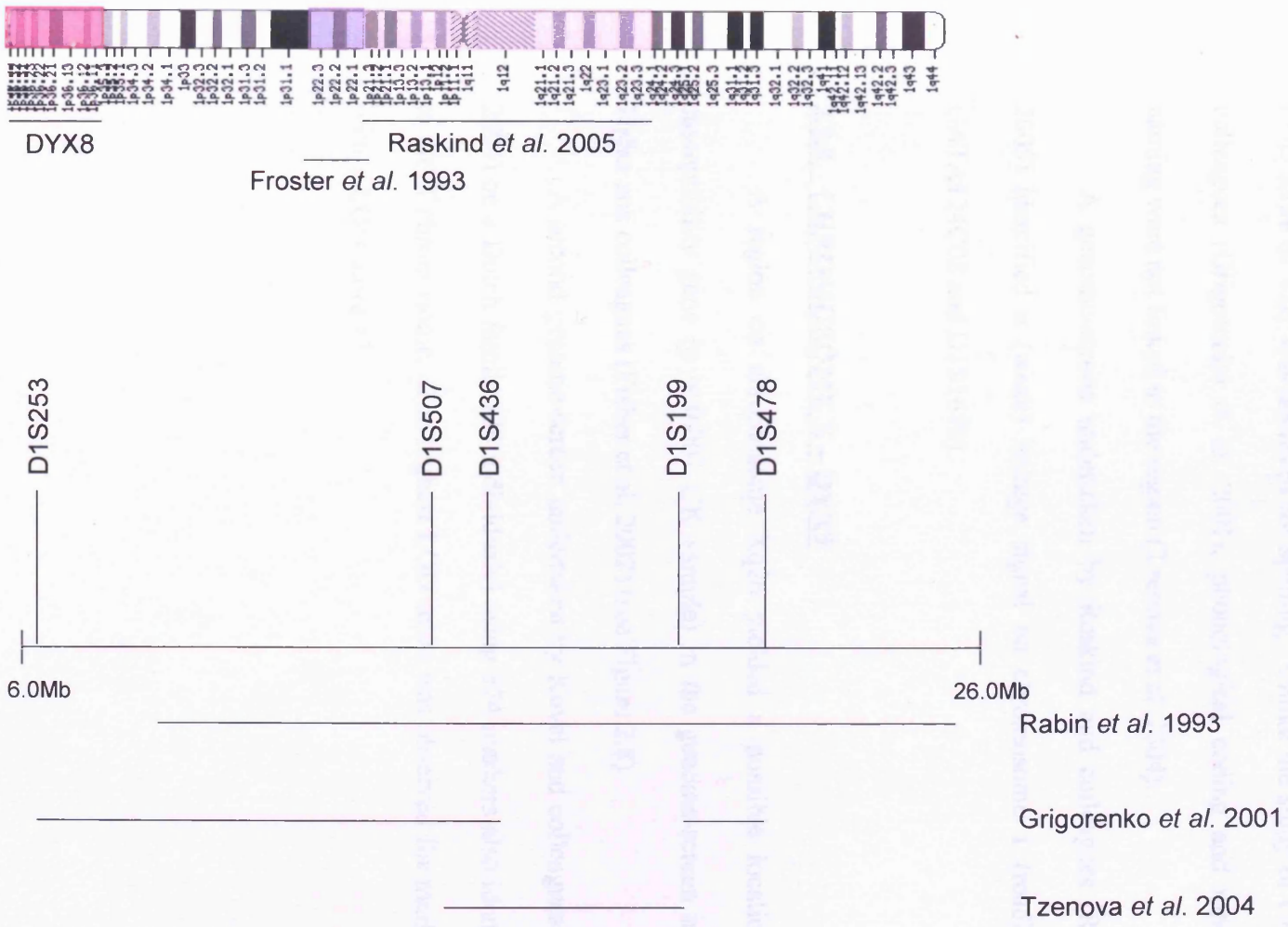


Figure 2.7 Linkage regions observed between chromosome 1 and DD. A region spanning approximately 20Mb has shown some evidence for linkage to DD. In addition linkage has also been observed towards the centromere of chromosome 1 (Froster et al. 1993).

and 6p. It is possible that the two regions contain genes producing related proteins.

Using a categorical definition of DD, the strongest evidence of linkage was observed at D1S507 (multipoint maximum LOD score = 3.65) by Tzenova and colleagues (Tzenova et al. 2004). Using quantitative measures of DD a maximum LOD score of 4.01 was achieved for spelling. Unlike the study of Grigorenko and colleagues (Grigorenko et al. 2001), phonological coding and rapid automated naming were not linked to the region (Tzenova et al. 2004).

A genome-screen undertaken by Raskind and colleagues (Raskind et al. 2005) identified a (weak) linkage signal on chromosome 1 (reading accuracy: GATA124C08 and D1S1679).

2.9.8 CHROMOSOME X – DYX9

A region on chromosome Xq26 yielded a possible location for a DD susceptibility gene (p = 0.001 UK sample) in the genome-screen undertaken by Fisher and colleagues (Fisher et al. 2002) (see Figure 2.8).

A second genome-screen undertaken by Kovel and colleagues (Kovel et al. 2004) on a Dutch family (29 individuals) using 374 markers also identified linkage to the X chromosome. The highest LOD score was observed for marker DXS8043 with a LOD score >3.

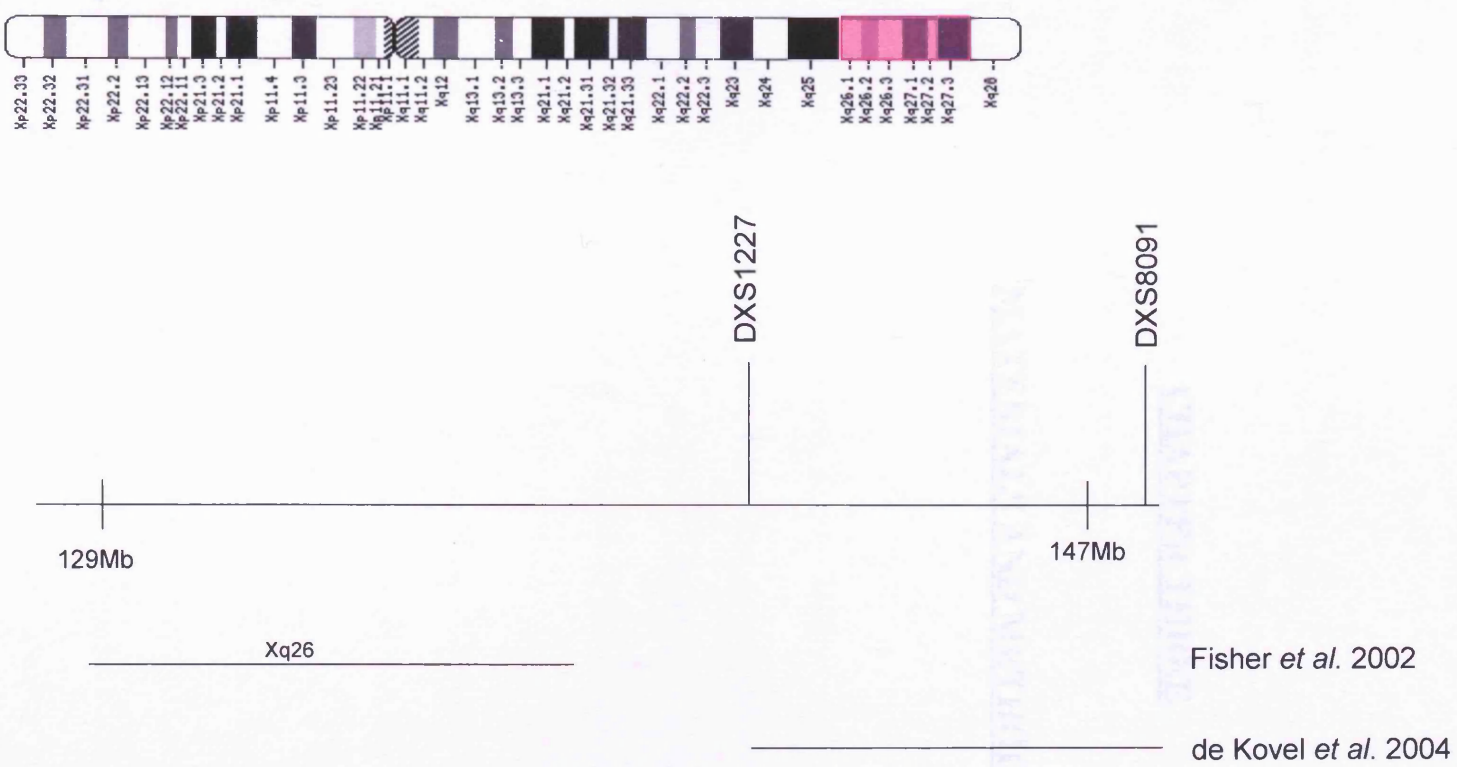


Figure 2.8 Linkage regions observed between chromosome X and DD. A region spanning 20cM has been identified on chromosome X showing linkage to DD.

CHAPTER THREE

MATERIALS AND METHODS

CHAPTER THREE

3. MATERIALS AND METHODS

3.1 COLLECTION OF THE SAMPLE

The collection of DNA samples and phenotypic information for this study was untaken by Dr Lucie Robinson and Dr Gary Hill and colleagues under the supervision of Prof. Julie Williams. Initial collection of the sample began in 1997 and more recent sample collection commenced in 2003. The study has ethical approval obtained from local ethics committees in the UK and for all participants of the study, informed written consent was obtained. Written consent for children under the age of 18 years was obtained from parental guardians.

DD-probands and their families were ascertained through contacts with Local Education Authorities (LEA) in South Wales and schools specialising in the education of children with reading difficulties in England. All English schools, with the exception of one, were members of Crested (the Council for the Registration of Schools Teaching Dyslexic Pupils). In total, 1326 families were contacted to take part in the study. Four hundred and thirty-seven families replied. Three hundred and ninety children were tested for DD and 254 met the criteria for our categorical definition of DD, 65% of the children tested. The categorical definition of DD required probands to have an IQ of 85 or above and a reading age 2.5 years or more behind their chronological age. This criterion represents a severe degree of reading disability and is likely to represent the lower 5th centile of children. English was the first language of all participants. No reward or feedback was given for taking part in the study.

A pro-rated full-scale IQ score was calculated using four subtests from the WISC III UK, including vocabulary, similarities, block design and picture completion (Wechsler 1992). Reading disability was calculated using either the Neale Analysis of Reading Disability (NARA) (Neale 1989) or the British Ability Scale (BAS) single word reading test (Elliot 1983) depending on the age and ability of the proband. Whilst NARA is based on prose reading, the BAS single word reading is based on reading a list of words (correlation coefficient between NARA and BAS tests, r = 0.89). The measure used to calculate reading disability is comparable to other definitions of DD used in molecular genetic studies including those of Grigorenko and colleagues (Grigorenko et al. 1997) and others (Fisher et al. 1999; Gayán et al. 1999).

3.1.1 CHOICE OF TESTS MEASURING COMPONENT PROCESSES OF READING

There are many tests available that allow the measurement of components of reading. Since this study is a genetic study, reading components were tested using tests with known heritable outcome and which show reliability in a UK population (Hohnen and Stevenson 1999).

3.1.2 THE TEST BATTERY

To refine the linkage regions for DD to more specific reading and DD related processes, all DD cases and 101 control children were given a full battery of tests that measured specific reading and language related processes. The component phenotypes and tests used in this thesis are shown in Sections 3.1.2.1-3.1.2.10.1.

3.1.2.1 Phonological Awareness

Phonological awareness is the ability to reflect explicitly on and manipulate the units of spoken language. In the English language, words are made up of combinations of 44 speech sounds (phonemes). For example, the word cat is made up of three phonemes, /kuh/, /aah/, /tuh/. Phonological awareness refers to the ability to be aware of and manipulate these sounds.

Castles and Coltheart (Castles and Coltheart 2004) suggested that although the mechanism is unknown, phonological skills are associated with development of word recognition skills (ultimately therefore with reading ability). Early studies (Savin 1972; Liberman et al. 1974) argued that letters represent phonemes in spoken words and in order for a child to read effectively, the child has to become aware of the phonemes in speech. Tasks involving manipulation of phonemes are complex tests requiring more than one operation on a task. Whilst information is processed, data must be held in working memory for the task to be completed successfully. Later studies have suggested that rhyme ability, rather than the ability to recognise phonemes is a better predictor of subsequent progress in learning to read (Bradley and Bryant 1983). Rhyming tasks are especially difficult for dyslexics who compensate by using visual-orthographic mechanisms, for example 'shoe' and 'toe' are orthographically similar but do not rhyme whereas 'head' and 'said' rhyme but are orthographically dissimilar. In this study, three measures of phonological awareness are used, a rhyming task – the rhyme oddity task and two phoneme awareness tasks including a phoneme deletion task and a task of auditory analysis.

3.1.2.1.1 Rhyme Oddity Task

The rhyme oddity task was adapted from a study by Bowey and Patel

(Bowey and Patel 1988). The test has a retest reliability of 0.66 and a heritability of 0.62 (Hohnen and Stevenson 1999). In essence the task tests auditory discrimination. Three or four words are presented to the child who is told that two (or three) of the words sound the same and one is the odd one out. The aim of the task is for the child to determine which word is the odd one out. For example, the three words 'made', 'hide', and 'fade' are read out, the child has to reply that 'hide' is the odd one out as it does not rhyme with the other two words.

3.1.2.1.2 Phoneme Deletion Task

For this task children are required to say a word after removing one of the sounds. The words vary in length from one syllable (e.g. cat) to multi-syllabic (e.g. driver). For example, children are asked to say 'cup' without the /k/ sound. Three trials are given so that the interviewer knows the child understands the task. The task is discontinued if the child fails the first five items consecutively; otherwise, the task is completed to the end. A shortened list of words was used for this study, adapted from a list by Wagner (Wagner et al. 1993; Wagner et al. 1994). The task has a retest reliability of 0.93 and a heritability of 0.62 (Hohnen and Stevenson 1999).

3.1.2.1.3 Task of Auditory Analysis

The task of auditory analysis is similar to the phoneme deletion task and involves removing phonemes from words. For example, the interviewer asks the child to say 'sunshine' without the 'shine'. In essence this test is very similar to the phoneme deletion task, but is often found more difficult by dyslexic individuals.

3.1.2.2 Phonological Decoding

Phonological decoding involves parsing written text into phonetic units. This is usually measured by the reading of nonsense words, for example twamket, which have not been seen by a participant before. Correct pronunciation of a word can only be achieved by converting graphemes to phonemes. Phonological decoding involves a number of steps in the correct pronunciation of a nonsense word including breaking down the word into the relevant phonetic units and then translating the phonetic units into a string of speech sounds so that the word can be pronounced according to a single set of language-specific rules. The use of unfamiliar words prevents the use of sight word strategies, forcing the use of a phonological decoding strategy.

The strategy involved in phonological decoding allows individuals to tackle new and unfamiliar words and may be especially important in early reading (Castles and Coltheart 1993). The measurements of this test can often reveal lifelong deficits in individuals with DD (Francks et al. 2002).

3.1.2.2.1 Non-Word Reading Task

A list of unfamiliar words is given to the child who then has to read them. It prevents the use of a sight word strategy and enforces a phonological decoding strategy. The list of words in this study is taken from Spring and Davis (Spring and Davis 1988). The retest reliability of this task is 0.93, whilst the heritability is estimated at 0.61 (Hohnen and Stevenson 1999).

The word list comprises 40 pronounceable letter strings. Words contain from one to five syllables. At the start of the task three trails are attempted so that the child understands the task. The words are not graded so no discontinue rule is

applied. Examples of words include 'fik', 'rond', and 'elgrund'.

3.1.2.3 Orthographic Coding

The process involved in recognising words from previous exposure to the word, without the use of phonological processes, is orthographic coding. Orthographic coding is usually tested using irregular words such as 'yacht' that violate standard letter-sound conventions such that phonological processing can not be utilised in the identification of words.

As readers gain experience, orthographic coding may become a more efficient mechanism for reading familiar words than phonological decoding (Francks et al. 2002) (see Chapter one, Section 1.7.1). Some dyslexic individuals show more striking deficits in orthographic coding tasks than they do on some measures relating to phonology (Castles and Coltheart 1993; Francks et al. 2002).

3.1.2.3.1 Pseudohomophone Judgement Task

The pseudohomophone judgement task requires the identification of the 'real' word from a pair of 'words' that sound the same when recoded phonologically, but only one of which has the correct orthographic pattern (i.e. spelt correctly). For example, children will be presented with pairs of words such as 'take' and 'taik', 'believe' and 'beleave' and 'engine' and 'enjine' and asked to identify the correctly spelt word. The task used in this study was adapted from a test devised by Olson and colleagues (Olson et al. 1994) and contains 70 pairs of 'words'. Five practice items are given to the child at the start of the task to check they understand the test. No discontinue rule is employed. The retest reliability is 0.74 and the heritability has been estimated at 0.48 (Hohnen and Stevenson 1999).

3.1.2.4 Word Recognition / Single Word Reading

Word recognition is a multicomponent process involving a range of processes from phoneme processing to the comprehension of text and has been described as a broad and definitive indicator of reading skill (Grigorenko et al. 1997). It can be described as a lexical retrieval process involving the visual recognition of letters (which make up familiar words) and the retrieval of the name and meaning of the word from memory. Vellutino suggests that the basic deficit in DD is the inability to decode the printed word (Vellutino et al. 2004). Such deficits can be related to inadequate skills related to word recognition such as spelling or phonological decoding.

Tests evaluating processes such as phonological decoding, spelling and verbal memory are good predictors of word identification ability (Vellutino et al. 2004).

3.1.2.4.1 British Ability Scales (BAS) Single Word Reading

The British Ability Scales (BAS) single word reading (Elliot 1983) is a reading test where single words are read from a list rather than prose text. The test is divided into 9 blocks of 10 words and is discontinued when one block is failed. Each block contains two lines of five words. Individuals read the words across each line before proceeding to the next. Words get progressively more difficult. Failure of a block occurs when five words on one line have been read incorrectly. The test has a retest reliability of 0.96 and a heritability of 0.44 (Hohnen and Stevenson 1999).

3.1.2.5 Rapid Automatic Naming

Many dyslexics show subtle impairments in perception and/or articulation of speech that is believed by some (Habib 2000) to be the most important mechanism of reading problems. Oral language problems in dyslexia can be measured using tests of rapid automated naming (RAN). Such tests assess the ability to rapidly call the names of simple visually presented stimuli (for example digits, letters, colours or objects). A measure of speed of processing, RAN may be important in the success of reading outside the group of processes relating more closely to language (Francks et al. 2002). Deficits in RAN have been shown to persist into late childhood and adulthood (Wolf 1986; Korhonen 1995; Snyder and Downey 1995) and the double deficit hypothesis (see Chapter one, Section 1.7.6) proposes that, along with phonological deficits, rapid automated naming forms the basis of DD.

3.1.2.5.1 Rapid Digit Naming Task

In this task, 50 digits, ranging from 1 to 9 and arranged in 10 lines of 5 digits, are presented to the individual. The proband has to recite the numbers as fast as possible. Probands repeat the task twice and the average time for the two trials recorded. Individuals are told to try and not make mistakes as correcting wrong answers will add to the time taken to do the task. The retest reliability of the rapid digit-naming task is 0.72, whilst the heritability is 0.54 (Hohnen and Stevenson 1999).

3.1.2.5.2 Rapid Picture Naming Task

In this task individuals are presented with a series of fifty pictures (pictures include: a box, a desk, a ball and a hat) in 5 rows of 10 of pictures. Like the digit

naming task participants are required to name the objects in the pictures in order, as fast as possible. An average of the time taken for two attempts is used as the measure. The retest reliability of the rapid picture naming naming task is 0.53, whilst the heritability is 0.34 (Hohnen and Stevenson 1999).

The rapid picture-naming task requires extra processing than the rapid digit-naming task since it is more semantic based, requiring information to be processed from a picture to a word.

3.1.2.6 Spelling

Spelling is measured as the ability to spell real words of various types and difficulties correctly. Deficits in spelling are commonly used in the diagnosis of dyslexia in both clinical and educational settings. Spelling ability is thought to comprise both orthographic and phonological components.

One of the greatest genetic effects in literacy is that of spelling ability and whilst the genetic effects involved in literacy decline with age, effects on spelling increase or are maintained (Olson et al. 1989; Wadsworth et al. 1989) (see Chapter two). Based on three pieces of information, (1) that spelling improves less over time, (2) spelling is more constrained because there are fewer contextual cues and (3) the genetic effects differ more as a function of age for reading than spelling, Stevenson and colleagues (Stevenson et al. 1984) and later, DeFries and colleagues (DeFries et al. 1991) suggested that the genetic influences on literacy problems are more appropriately studied through their impact on spelling than on measures of word recognition or reading comprehension.

3.1.2.6.1 British Ability Scales (BAS) Spelling

Spelling test D of the BAS Spelling tests is used to assess the spelling age of children. During the test, the interviewer reads single words to the child and then puts the word into a sentence before repeating the word again. The child then has to write down the correct spelling of the word. Retest reliability has been estimated at 0.89 and heritability at 0.51 (Hohnen and Stevenson 1999). The test consists of 20 words, given in a set order and is discontinued when five consecutive words have been spelt incorrectly. From the spelling test result, a spelling age is calculated and a discrepancy measured between this age and a probands chronological age.

3.1.2.7 Phonological Working Memory

Working memory focuses on the function of memory in cognitive tasks such as reading, speech comprehension and mental arithmetic. Working memory corresponds to short term memory (STM) which acts as a working storage system in everyday cognitive tasks. Baddeley (Baddeley 1986) has described the working memory system suggesting it has three components, a central executive system and two passive storage systems used to store information which interact with it. The articulatory (phonological) loop temporarily holds verbal information, whilst the visuo-spatial scratch pad is responsible for the storage of visual-spatial information. The central executive primarily coordinates activity within the cognitive system, but also increases the capacity that can be held by the articulatory loop and visuo-spatial scratch pad (Swanson 1999).

Reading comprehension deficits in DD have been attributed to impairments in working memory (Swanson and Berninger 1995; Swanson and Alexander 1997), in particular with the utilisation and/or operation of the articulatory loop (see Hulme

and Snowling 1992 for review). Since the articulatory loop is controlled by the central executive, deficits in reading comprehension may be due to deficiencies in controlling functions of the central executive (Baddeley 1992; Gathercole and Baddeley 1993).

3.1.2.7.1 Non-Word Repetition

The non-word repetition test (Gathercole et al. 1994) involves the child hearing a single novel word and being required to repeat it back immediately. Forty items are used in this task. Non-words range from 2 syllables to 5 syllables and are arranged in blocks of 10 words (ten 2 syllable words, ten 3 syllable words, ten 4 syllable words and ten 5 syllable words). Words are presented from a tape recoding in order that accents and dialects don't interfere with the pronunciation of words. The retest reliability of the test is 0.91 and the heritability is 0.71 (Hohnen and Stevenson 1999). Examples of words used in this task are 'dopelate', 'confrantually', 'sladding' and 'skiticult'.

3.1.2.8 Reading Accuracy, Comprehension and Rate

Reading accuracy, the ability to read a paragraph of text without making mistakes is often used as a measure of reading ability. It is often believed to be a more accurate measure of reading ability than single word reading since contextual clues can be obtained from text, which reflects more accurately reading in general.

Reading comprehension is multifaceted and requires the use and synchrony of a number of reading related processes in order to allow the derivation of meaning from text. Dyslexic children show normal listening comprehension but often have poor reading comprehension. Inadequate reading comprehension is believed to stem

from poor phonological processing and reading fluency (measured by rapid naming of objects etc) and deficits in both these processes have been noted in DD. Gough and Tunmer (Gough and Tunmer 1986) have suggested that word recognition is necessary although not sufficient to allow reading comprehension and that grammatical skills and vocabulary knowledge are important in the ability to comprehend text. More recently, Muter and colleagues predicted vocabulary knowledge and grammatical ability would predict reading comprehension ability, a prediction confirmed following tests on 90 children (Muter et al. 2004).

Reading rate, i.e. how fast a child can read a paragraph of prose text, has not been extensively researched in regards to DD. However, slow rapid automatised naming has been reported in dyslexic individuals.

3.1.2.8.1 Neale Analysis of Reading Ability (NARA)

The NARA reading test consists of a set of graded prose passages that allow the testing of rate, comprehension and accuracy of oral reading. Test material is presented in the form of a book, which consists of short, graded narratives, each constructed with a limited number of words and with a central theme, action and resolution. Pictures accompany the narrative however these set the scene rather than tell the story.

Within the book there are six passages of increasing difficulty. Comprehension questions are available after the oral reading of the passage, which tap into the child's use of contextual cues, pictures and prompts. They also test the immediate recall of the main idea of the narrative, the sequence of events and other details. In order to answer some questions inference is required.

The NARA tests are only suitable for children up to the age of 12 years. As

a result, analysis undertaken using measures from this test are only done on children aged 12 years and under.

3.1.2.8.1.1 Reading Accuracy

Children start reading the passages at a level below expected by their age (since they are poor readers), but which is not too low that they lose interest. Children move to the next level of passage until 16 mistakes have been made (20 on level 6). At this point the reading test is discontinued. The accuracy score is based on the number of words correct out of the number read. The accuracy score is converted to a reading age based on population norms and for this thesis a discrepancy score was calculated between reading accuracy age and the child's chronological age.

3.1.2.8.1.2 Reading Comprehension Task

Reading comprehension was measured on the number of correct questions answered by the child based on the number of passages they read and at what level. Like reading accuracy a reading comprehension age is calculated and a discrepancy measure calculated between this and the child's chronological age.

3.1.2.8.1.3 Reading Rate Task

Whilst undertaking the prose reading task, the child is timed. Question answering is not included in the timing of individuals. A time is calculated and used in the following equation:

$$\text{Words Per Minute} = \frac{\text{Total Number of Words}}{\text{Total Time (seconds)}} \times 60$$

From this equation a reading rate age is calculated based on child norms. An age discrepancy is calculated between the reading rate age and the child's chronological age.

3.1.2.9 Attention Deficit Hyperactivity Disorder (ADHD)

The comorbidity between DD and ADHD is well known and it is likely that genes contribute to both disorders (August and Garfinkel 1990; Purvis and Tannock 1997; Shaywitz 1998; Kaplan et al. 2001). The shared genetic effects between DD and ADHD are particularly interesting given reports of a gene influencing DD and ADHD susceptibility located on chromosome 6p (Willcutt et al. 2002).

The Du Paul parent reported symptoms of ADHD questionnaire has been shown to accurately assess ADHD symptoms, however it does not allow the segregation of ADHD symptoms into inattention, hyperactivity and impulsivity symptoms. Consequently the Connors' questionnaire was utilised to allow the segregation of ADHD symptoms.

3.1.2.9.1 Connors' Abbreviated Parent Reported Symptoms of ADHD Questionnaire

The Connors' Abbreviated Parent Reported Symptoms of ADHD Questionnaire is a valid and reliable measure for evaluating symptoms of ADHD (Connors 1973; Rosenberg et al. 1989). The questionnaire consists of 27 questions, which relate to inattention, hyperactivity and impulsivity in the child and are answered by parents. To each question there is four possible answers given marks 0, 1, 2 or 3. The investigator adds up these scores to give a total score. A score above 15 is suggested to indicate the presence of ADHD in categorical analyses.

Continuous measures of symptoms can also be utilised. The questionnaire was also used to calculate a continuous measure of inattention symptoms in probands.

3.1.2.9.2 Du Paul Parent Reported Symptoms of ADHD Questionnaire (Shortened Version)

The Du Paul parent reported symptoms of ADHD is similar to the Connors questionnaire and is filled in the same way. There are 26 questions which have four possible answers graded 0, 1, 2 and 3. Again these are added together to give a total score. A total score above 35 is suggestive of ADHD presence in categorical analyses. Continuous measures of symptoms can also be utilised.

3.1.2.10 IQ

The definition of DD differs between genetic studies. Whilst some studies, including this study define DD as a reading lag despite normal intelligence, other studies define DD as a lag in reading relative to what would be expected given their intelligence. The latter definition would include individuals, which in the former study would be excluded due to low IQ, so long as the individuals read below the reading age predicted by their low IQ. As IQ plays a role in defining DD, it is important to determine that QTLs influencing DD, do not reflect an underlying association with IQ, especially given the reports of linkage and association for IQ overlapping and within DYX2 (Plomin et al. 2004; Posthuma et al. 2005).

3.1.2.10.1 Weschler Intelligence Scale for Children (WISC III UK)

A pro-rated full scale IQ was determined using 4 subtests of the WISC III UK (Wechsler 1992). Vocabulary, similarities, block design and picture completion were utilised to form a pro-rated IQ score.

3.2 DNA AND SAMPLES

3.2.1 DNA EXTRACTION

Venous blood was taken from participants willing to give blood and DNA extracted from lymphocytes using standard procedures. If blood samples could not be obtained, 25ml saline mouthwashes were obtained. DNA was extracted from buccal cavity epithelial cells by centrifugation at 2000*g* for 10 minutes, followed by incubation with proteinase K (Sigma, USA), SET buffer (Qiagen, UK) and SDS (Invitrogen, UK) at 50°C for 12 hours. The DNA was then isolated by standard phenol-chloroform extraction (Morris et al. 2000). DNA extraction was undertaken by other members of the laboratory team.

Following DNA extraction, DNA was quantified in two ways.

1) Quantification using a spectrophotometer (Beckmann Instruments, UK). Absorbance of UV light at 260nm and 280nm wavelengths (λ) were measured and the ratio between A_{260nm} and A_{280nm} calculated. A ratio of above 1.8 indicated a suitable level of clean DNA and the absence of contaminating protein.
2) The PicoGreen dsDNA Quantification kit (Molecular Probes, USA) and a Fluoroskan Ascent fluorometer (LifeSciences International, UK) with Ascent Software (Labsystems, UK) were used to calculate more accurately the concentration of DNA.

All DNA stocks were kept at -20°C in individual eppendorf tubes. Working dilutions of DNA, diluted to 8ng/μl, were kept in 96-deep-well plates at 4°C. During the study a number of sample sets were utilised.

3.2.2 SAMPLES

3.2.2.1 Dyslexia Sample For *De Novo* Polymorphism Discovery

A sample of 16 dyslexic individuals diagnosed with a reading age ≥2.5 years behind their chronological age and an IQ of ≥85 was employed when screening for polymorphisms. The descriptive statistics of the 3 females and 13 males comprising the sample is shown below (Table 3.1).

Variable	Mean (years)	SD (years)	Minimum (years)	Maximum (years)
Age	14.28	2.12	9.92	16.58
IQ	102	10.99	85	131
RD	-5.58	2.16	-2.91	-9.16

Table 3.1 Demographics for the DD sample used for *de novo* polymorphism discovery.

3.2.2.2 Pooled DNA Sample 1

Two sets of DNA pools were used in this study. These included in the first instance, a case pool containing 140 DD individuals (116 males, 24 females) and control pools containing 550 blood donor controls (split between 3 pools) containing 391 males and 159 females. The sample is described in Table 3.2.

	Cases				Controls	
Variable	Mean (years)	SD (years)	Min (years)	Max (years)	Mean (years)	SD (years)
Age	13.22	2.3	7.92	17.08	41.39	12.5
IQ*	100.13	11.03	85	134	-	-
RD*	-5.07	1.76	-2.5	-9.76	-	-

Table 3.2 Demographics of DNA pooled sample 1. * No IQ and RD data was available for blood donor controls.

3.2.2.3 Pooled DNA Sample 2

Later in the study, three pools of 80 DD individuals (totalling 189 males and 51 females) and four pools of 78 age-matched control children (totalling 178 males and 134 females) were utilised. Descriptive statistics of the sample are shown in Table 3.3.

	Cases				Controls			
Variable	**Mean (years)**	**SD (years)**	**Min (years)**	**Max (years)**	**Mean (years)**	**SD (years)**	**Min (years)**	**Max (years)**
Age	13.17	2.18	7.67	17.56	11.98	2.39	5.5	16.67
IQ*	98.88	18.38	85	136	103.35	11.97	85	137
RD	-4.93	1.87	-2.5	-13	+1.14	1.45	-0.5	3.83

Table 3.3 Demographics for DNA pooled sample 2. * Not all controls were IQ tested therefore demographics are based on the 153 participants tested for IQ.

3.2.2.4 Case Control Sample

A case-control sample containing 223 cases and 273 controls was created for individual genotyping in this study (Table 3.4) and represents a subset of the cases and controls used in pooled DNA sample 2. The case sample comprised 173 males and 46 females and the control sample, 137 males and 134 females. DNA was plated into 96-well plates containing 42 DD cases and 48 controls. Six wells were left empty to allow negative (water) controls to be run.

	Cases				Controls			
Variable	**Mean (years)**	**SD (years)**	**Min (years)**	**Max (years)**	**Mean (years)**	**SD (years)**	**Min (years)**	**Max (years)**
Age	13.20	2.16	7.67	17.58	11.98	2.39	5.5	16.67
IQ*	102.93	13.42	85	136	103.25	11.95	85	137
RD	-5.19	1.79	-2.5	-13	+1.14	1.45	-0.5	6.92

Table 3.4 Demographics for the case-control sample. * Not all controls were IQ tested therefore demographics are based on the 153 participants tested for IQ.

3.2.2.5 Parent DD-Proband Trios: Sample 1

Sample 1 comprised of 254 parent-DD proband trios (probands comprised 205 males and 49 females; see Table 3.5). DNA was plated into 96-well plates containing parents and probands. Three wells were left empty to allow negative (water) controls to be run. The majority of the cases utilised in this sample were also included in the case-control sample.

Variable	Mean (years)	SD (years)	Minimum (years)	Maximum (years)
Age	13.24	2.14	7.92	17.17
IQ	102.55	11.44	85	136
RD	-5.06	1.72	-2.5	-10.25

Table 3.5 Demographics of the parent DD-proband trio sample 1. Demographics are shown for DD probands only.

3.2.2.6 Parent DD-Proband Trios: Sample 2

Later in the study, 143 parent DD-proband trios were used. This sample represents a subset of sample 1 for whom there was adequate DNA available to carry out large-scale association studies (Table 3.6). The sample contains 110 male probands and 33 female probands.

Variable	Mean (years)	SD (years)	Minimum (years)	Maximum (years)
Age	13.17	2.08	7.92	17.08
IQ	104.01	11.88	85	136
RD	-5.06	1.76	-2.5	-10.25

Table 3.6 Demographics of the parent DD-proband trio sample 2. Demographics are shown for DD probands only.

3.2.2.7 Intermediate Phenotype Sample

A sample of individuals was set up to allow quantitative analysis of DD component phenotypes. This sample represents individuals of normal IQ (IQ $\geq$ 85) with a reading disability between 6 months and 2.5 years behind chronological age. Table 3.7 shows the demographics of this sample.

Variable	Mean (years)	SD (years)	Minimum (years)	Maximum (years)
Age	12.19	2.62	5	16.58
IQ	107.22	13.51	85	138
RD	-1.59	1.26	0.42	-2.42

Table 3.7 Demographics of the intermediate phenotype group. Note IQ calculations are based on those individuals for whom data was available.

3.2.2.8 ADHD Sample

A sample of 144 parent ADHD-proband trios and 115 single parent ADHD proband duos were made available to me by Professor Anita Thaper. Probands were identified and referred to the study by child and adolescent psychiatrists or paediatricians from the Greater Manchester, South Wales or Avon areas of the UK. ADHD was defined using criteria from ICD-10 hyperkinetic disorder criteria or ADHD criteria from DSM-IV or DSM-III-R. All probands were required to have an IQ of 70 or above and were of UK Caucasian origin. In order to prevent retrospective recall bias from the parents reporting ADHD symptoms in the child, probands had not been on medication for ADHD for more than one year. Probands with epilepsy, fragile X syndrome, pervasive developmental disorders or Tourette's syndrome were excluded.

Variable	Mean	Standard Deviation	Minimum Value	Maximum Value
Age	9 years 2 months	2 years 1 month	6 years	16 years
IQ	90	12.1	70	134
DSM-IV and ICD-10 ADHD Symptoms	14.74	2.45	7	18

Table 3.8 Demographics of the ADHD sample. Demographics are based on the ADHD-proband only from parent ADHD-proband trios and duos. Ninety percent of the participants were male and 10% female.

3.3 MATERIALS AND METHODS

3.3.1 POLYMERASE CHAIN REACTION (PCR)

Primers for PCR reactions were designed using the software Primer 3.0, freely available at http://www-genome.wi.mit.edu/cgi-bin/primer/primer 3www.cgi. To find the optimum temperature for primer annealing, optimisation reactions were undertaken for each primer pair on control DNA using a temperature gradient on MJ Tetrad thermocyclers (MJ Research, UK). Each primer set was then tested at 6 different temperatures simultaneously (see Table 3.9).

Temperatures Used on Temperature Gradient					
52	53	56	60	64	66

Table 3.9 Temperatures used in a temperature gradient to allow the optimal primer annealing temperature to be determined.

PCR reactions were carried out using the following mix (note: the amount of each reagent is given per single PCR reaction):

Reagent	Company	Volume
Buffer (10X containing 15mM $MgCl_2$)	Qiagen, UK	1.2µl
dNTPs (2.5mM)	Amersham, UK	1.0µl
Primer (10pmol/µl) – Forward	Sigma-Genosys, UK	0.28µl
Primer (10 pmol/µl) – Reverse	Sigma-Genosys, UK	0.28µl
ddH_2O	UHW, UK	3.2µl
Hot Start Taq Polymerase (5U/µl)	Qiagen, UK	0.06µl
Genomic DNA (6ng/µl)		6µl

Table 3.10 Reagents required for PCR reactions.

For each assay, one of the cycling parameter sets shown in Tables' 3.11 and 3.12 was used. Details of individual PCR reactions undertaken can be found in results Chapters four and six and Appendix 1.

PCR Steps	Programme
Initial Denaturation	94°C 15 minutes
Denaturation	94°C 30 seconds
Annealing	See Individual SNPs 30 seconds
Extension	72°C 45 seconds
Number of Cycles	34
Final Extension	72°C 10 minutes
Taq	Hot Star

Table 3.11 PCR assay conditions. The correct annealing temperature for each assay is determined as described below.

The optimal temperature for primer annealing is determined by analysis of PCR products on an agarose gel (see Section 3.3.2). Products should show clear bands at the correct size to be fully optimised.

PCR Steps	Programme - Touchdown
Initial Denaturation	95°C 15 minutes
Denaturation	94°C 5 seconds
Annealing	65°C 5 seconds (-1°C per cycle)
Extension	72°C 10 seconds
Number of Cycles	11
Denaturation	94°C 5 seconds
Annealing	55°C 5 seconds
Extension	72°C 10 seconds
Number of Cycles	30
Final Extension	72°C 5 minutes

Table 3.12 Touchdown PCR assay conditions. Touchdown PCR was attempted if annealing temperatures could not be determined.

3.3.2 AGAROSE GEL ELECTROPHORESIS

To check the PCR process had worked, horizontal gel electrophoresis was undertaken. Since DNA carries a negative charge, when placed in an electrical field, DNA will migrate towards the positive pole, with migration dependant on:

1) Size of fragments
2) Conformation of fragments.

Therefore fragments of differing sizes and shapes can be separated.

Gels were constructed using the reagents shown below (Table 3.13) and heated until reagents were mixed appropriately. After the molten gel was allowed to cool, ethidium bromide stain was added and the gel poured into a gel-casting tray and allowed to set at room temperature.

Reagent	Source	Amount
0.5X TBE Buffer	National Diagnostics, UK	50ml
dH_2O	UHW, UK	50ml
Agarose	Sigma, UK	3g*
Ethidium Bromide (10mg/ml)	Sigma, UK	2.5µl

Table 3.13 Reagents required to make an agarose gel. *Makes a 3% Agarose gel.

Three microlitres of PCR amplified gDNA was mixed with 3µl of loading buffer (see Table 3.14). This was aliquoted into the wells of the gel and immersed in 0.5X TBE buffer in the gel tank (Thermo Hybaid, UK). Gel electrophoresis was undertaken at 100 volts for approximately 75 minutes. At the same time, a digested DNA fragment (100bp DNA ladder, New England Biolabs, USA) containing DNA fragments of known fragment sizes was used to identify the size of DNA fragments amplified by PCR and check the primers had amplified the correct sequence. DNA products were visualised using an ultra violet (UV) transilluminator fitted with a Polaroid camera (UVP. Inc, USA).

Reagent	Company	Amount
dH_2O	UHW, UK	20ml
EDTA (0.5M)	Sigma, UK	4ml
Ficoll	BDH, USA	3g
Orange G (0.05g)	Sigma, UK	0.01g

Table 3.14 Reagents required for loading buffer for gel electrophoresis. Loading buffer allows the DNA sample to 'fall' into the wells formed in the gel and also allows the tracking of DNA migration through the agarose gel.

3.3.3 DENATURING HIGH PERFORMANCE LIQUID CHROMATOGRAPHY (dHPLC)

Denaturing high performance liquid chromatography (dHPLC), is based on the detection of heteroduplexes in PCR products by ion-pair reverse-phase high-performance liquid chromatography which allows the detection of single base substitutions, insertions and deletions. After detection, polymorphisms are characterised (location, nature of variation) by DNA sequencing.

Genes analysed by dHPLC, were PCR amplified in a number of fragments in a screening sample of 16 dyslexic individuals. To form heteroduplexes, PCR products were heated to 94°C and gradually reannealed by cooling at a rate of 1°C/min for 40 minutes.

Optimal temperatures for each fragment and corresponding elutent gradients were determined using the dHPLC melt programme available at http://insertion.stanford.edu/melt.html. In order to achieve optimal mutation detection, based on work previously done in the lab, fragments were analysed at the recommended temperature (n°C) and at n+2°C to allow greater than 96% sensitivity.

DHPLC analysis was undertaken on a Transgenomic WAVE™ DNA Fragment Analysis System (Transgenomic, USA). 5µl of heteroduplexed PCR product was injected onto a DNASep™ column. Hetero- and homoduplexes were eluted with a linear ACN gradient formed by mixing buffer A (0.1M TEAA, pH 7.0)

with buffer B (0.1M TEAA, pH 7.0, containing 25% ACN) (Transgenomic, USA) at a constant flow rate of 0.9ml/min. DNA was detected at 260nm. The analytical gradient was 4 minutes long and buffer B was increased at 2%/min. For each fragment, the initial and final concentrations of buffer B were adjusted to obtain a retention time between 3 and 5 minutes. Between samples the column was cleaned with 100% buffer B for 30 seconds and equilibrated at starting conditions for 2 minutes. Resultant chromatograms were compared, and any differences in the chromatograph trace were indicative of a heteroduplex.

3.3.4 DNA SEQUENCING

DNA sequencing was undertaken to confirm and characterise polymorphisms found during dHPLC analysis. Samples were selected for DNA sequencing analysis if they showed a difference in the chromatogram indicative of a heteroduplex in dHPLC analysis. All samples showing such differences were sequenced. The sequencing undertaken was based on fluorescent dye-terminator cycle sequencing, based on the chain-termination dideoxynucleotide sequencing method of Sanger and co-workers (Sanger et al. 1977).

In order to undertake the sequencing reaction, DNA was amplified in a PCR reaction (Section 3.3.1) to yield 32µl of PCR product. Amplified DNA was then purified using the QIAquick purification kit (Qiagen, UK)- 5 volumes of buffer PB (160µl) (Qiagen, UK) were added to 1 volume of PCR product (32µl). After centrifugation (60 seconds at 13000rpm), 0.75ml of buffer PE (Qiagen, UK) was added to the DNA and centrifuged (60 seconds at 13000rpm). DNA was eluted with 30µl of buffer EB (elution buffer) (Qiagen, UK), left to stand for 1 minute. Centrifugation allows the DNA to be collected in a microcentrifuge tube.

Sequencing reactions were performed using the Big Dye Terminator (v2.0) Cycle Sequencing kit (Applied Biosystems, UK) in a total volume of 10µl. For each PCR reaction, two sequencing reactions were performed; one using the forward PCR primer and one using the reverse PCR primer. The following reaction mix and thermocycling conditions were used:

Reaction Mix:	Terminator Mix	4µl
	Primer (3.2µM)	1µl
	Purified PCR Product	5µl

Thermocycling Conditions:	96°C	2 minutes	
	96°C	30 seconds	50 cycles
	55°C	15 seconds	
	60°C	4 minutes	

In order to generate high quality DNA sequence data, unincorporated dye terminators (see Table 3.15), salt and any other small molecule contaminants must be removed from the sequencing reaction prior to electrophoresis. DNA was removed from these components using gel filtration with Sephadex G-50 Fine Resin (Sigma, UK) loaded MultiScreen HV filtration plates (Millipore, UK). Dry Sephadex G-50 was loaded into a 96 well filtration plate using a Column Loader (Millipore, UK). 300µl of ddH_2O was added to each well to allow the resin to swell (4 hours). To pack the columns and remove excess water from the filtration plate, the plate was centrifuged at 2500rpm for 5 minutes.

ddNTP	Dye Label	Colour of Analysed Data
A	dR6G	Green
C	dROX	Blue
G	dR110	Black
T	dTAMRA	Red

Table 3.15 ddNTPs and their corresponding dye.

The individual sequencing products were diluted with 10µl of ddH_2O and added to the centre of individual columns on the filtration plate, which was centrifuged at 2500rpm for 5 minutes. Eluted sequencing products collected in a 96-well plate were subsequently dried at 65°C for 45min in a speed vacuum.

The purified sequencing products were resuspended in 10µl of Hi-Di Formamide (Applied Biosystems, UK) and electrophoresed on an ABI PRISM® 3100 Genetic Analyzer using a 36cm capillary array and POP-6™ polymer (Applied Biosystems, UK).

The identification of polymorphisms in sequence data was undertaken by comparing the forward reaction sequences for the heterozygous and homozygous samples. Allelic variants should be observable in both forward and reverse reactions. However, there are a number of reasons why polymorphisms may only be identified in one reaction, these include:

1) If the polymorphism is within 20 bases of the sequencing primer, the identification of polymorphisms can be difficult as the first 30-40 bases of a sequencing reaction can be of poor quality.
2) Fragments that are over 500 bases in length can show poor quality of the sequencing towards the end of the fragment, therefore polymorphisms located here can be difficult to detect.
3) The sequencing of a fragment can be disrupted if there is a long stretch of As or Ts in a sequence. Thus if there is a run of As 100 bases from the forward

primer, sequence may only be readable as far as the run of As. As a result such polymorphisms may only be identifiable in the reverse sequencing reaction where it will be present in the sequence upstream to the run of Ts.

3.3.5 DNA POOLING

DNA pooling was used in order to allow the estimation of allele frequency differences between cases and controls using fewer PCR and genotyping reactions than would be necessary if individual samples were individually assessed. DNA concentration (for all steps) was determined using the PicoGreen dsDNA Quantitation Reagent (Molecular Probes, USA) and quality was assessed by PCR amplification of microsatellite markers under standard conditions. DNA samples not amplified in the PCR reaction of microsatellites were not included in the DNA pools. To produce pools, water was added to stock DNA to produce a target DNA concentration of 40ng/µl. Samples were allowed to equilibrate at 4°C for 48 hours before quantification using the PicoGreen dsDNA Quantification kit. Based upon re-estimation of concentration of every step, each sample was further diluted to 10ng/µl and then 4ng/µl. Samples were allowed to equilibrate for 48 hours at 4°C between each dilution. Samples diluted to 4ng/µl (±0.5ng/µl) were accepted for pooling. DNA above 4ng/µl on final quantification were rediluted to an exceptable concentration. Samples below 4ng/µl were rediluted from stocks. Equal volumes of each sample were combined to form DNA pools.

3.3.6 ABI PRISM® SNaPshot™ MULTIPLEX KIT REACTIONS

The ABI Prism® SNaPshot™ Multiplex Kit (Applied Biosytems, UK) allows the examination of SNPs at known locations on DNA templates. The reaction is

based on the dideoxy single-base extension of an unlabelled oligonucleotide primer.

PCR reactions were carried out in 12µl reactions as described previously (see Section 3.3.1). Preparation for SNaPshot™ required incubation of 12µl of PCR product with 0.25µl of exonuclease I (1U) (Amersham, UK) to remove residual single-stranded primers, 1µl shrimp alkaline phosphatase (1U, SAP) (Amersham, UK) to dephosphorylate remaining dNTPs and render them inactive for the primer extension step and 1.75µl of water for 1 hour at 37°C and 15 minutes at 80°C. The SNaPshot™ primer extension method was performed by combining 2µl of treated PCR product with 2.5µl of SNaPshot™ solution (AmpliTaq DNA polymerase, fluorescently labelled ddNTPs and reaction buffer), 2.5µl of buffer (8ml Tris-HCL [1 molar, pH 8], 8ml $MgCl_2$ [25mM], 34ml water), 1µl of 0.5pmol extension primer (Sigma-Genosys, UK) (designed using a programme available at http://m034.pc.uwcm.ac.uk/FP_Primer.html; Ivanov et al.) and 2.0µl of water. The reaction was then incubated at 94°C for 2 minutes and subject to 25 cycles of 95°C for 5 seconds, 43°C for 5 seconds and 60°C for 5 seconds. To prevent the unincorporated fluorescent ddNTPs ([F]ddNTP) obscuring the primer extension products during electrophoresis, the reactions were treated with 1U SAP (2.5µl of reaction mix added to: 0.5µl (1U) SAP and 2µl water) and incubated at 37°C for 1 hour and 80°C for 15 minutes to denature the SAP enzyme. 2µl of reaction product was then added to 8µl of Hi-Di Formamide (Applied Biosystems, UK) in a 96-well 3100 optical microamp plate (Applied Biosystems, UK) and loaded onto a 3100 Genetic Analyzer (Applied Biosystems, UK). The 3100 Genetic Analyzer was set up with a 36cm capillary array and POP-4 polymer. For each SNP tested for association to DD using DNA pools, a negative control containing water was used alongside each of the sets of DNA pools to check for contamination.

The fluorescence signals corresponding to the allele-specific extended primers were determined using Genotyper version 2.5 (PE Biosystems, UK). The peaks of fluorescence on the resultant electropherographs correspond to the specific labelled-base present from the extension of the allele-specific primers. Below is a table summarising the dyes assigned to each of the bases nucleotides and the colour used in the data analysis.

ddNTP	Dye Label	Colour of Analysed Data
A	dR6G	Green
C	dTAMRA	Black
G	dR110	Blue
T	dROX	Red

Table 3.16 Dye labels attached to bases involved in the dideoxy single-base extension of the unlabelled extension primer in SNaPshot™ and the subsequent colour of chromatographs analysed using Genotyper software.

Genescan Analysis is unable to quantify accurately fluorescent signals above 8000 fluorescence units since the detector becomes saturated at this point. Consequently reactions were only analysed when peak heights were less than 8000 fluorescence units. In reactions where peak heights exceeded 8000 fluorescence units the reactions were repeated, lowering the concentration of extension primer using the following equation:

$$\text{Required Peak Height} = \frac{Y'}{(Y/X)}$$

where Y' is the required peak height (*e.g.* 3000), Y is the initial peak height and X is the initial primer concentration (*e.g.* 0.5μM) (Norton et al. 2002).

In the analysis of bi-allelic markers in DNA pools, the primer extension products for each allele may not be represented with equal efficiency, thus not

providing an accurate basis for the calculation of allele frequencies. There are several reasons why alleles may not be equally represented:

a) Differential PCR amplification of each of the alleles (Liu et al. 1997)

b) Differential efficiencies of the incorporation of the ddNTPs for each allele-specific primer extension reaction (Haff and Smirnov 1997; Barnard et al. 1998)

c) Unequal emission energies of the different fluorescent dyes (Norton et al. 2002)

d) Different efficiency of primer annealing based upon nearest neighbour effects (Moskvina et al. 2005)

To allow for the unequal representation of alleles, the estimated allele frequencies from pools were corrected by using the mean of the ratios obtained from measurements taken from heterozygous samples. Since heterozygous individuals contain one of each allele at a known polymorphism, fluorescence corresponding to each allele of a SNP should be the same (under perfect conditions) resulting in a 1:1 ratio of fluorescence units for each allele. Any deviation away from 1:1 fluorescence ratio can be determined and any pooled assays can be corrected for any unequal representation of the alleles from the known heterozygote ratio. The correction could be made using the following equation:

$$f(\mathrm{a}) = A/(A+kB)$$

Where A and B are the peak heights of the primer extension products representing alleles A and B in a pool and k is the mean of the replicates of A/B ratios observed in a heterozygote (Hoogendoorn et al. 2000). f(a) is the frequency of

allele A. The frequency of allele B (f(b)) was then calculated from the formula:

$$f(b) = 1 - f(a)$$

Frequencies were converted to allele number by multiplying the estimated allele frequency of A or B by twice the number of individuals represented in the pools. From the estimated allele count, a chi-squared test of association could be performed to test for differences in allele distribution between cases and controls (Hoogendoorn et al. 2000).

Where there were no known heterozygotes for an assay (since SNPs were chosen from databases and often individual genotypes were not available for CEPH individuals in the International HapMap Project; www.hapmap.org), correction could not be made based on the correction factor k in the equation above. In these cases values of k were utilised which had been empirically derived from 152 SNPs based on all possible pairs of bases (Moskvina et al. 2005). Values used for the correction factor k for pooled DNA analysis included 0.25, 0.32, 0.47, 0.88, 1.00, 1.14, 2.13, 3.10 and 4.00.

For each SNP assayed negative controls were used, where instead of DNA template, water was added. In the case of DNA pooling, one negative control was run, in the case of individual genotyping, at least three negative controls were run per 96 individuals assayed. If fluorescence peaks were observed in the negative controls, likely due to extension primer hairpin extension or primer dimer extension, a new extension primer was designed and the assays repeated. Since in pooled DNA analysis, the allele frequency in each pool was determined as the mean of two replicate assays, if a large discrepancy was observed between the peak height ratios of the two replicates, the entire experiment was repeated. However, replicates gave

consistent results throughout, with, in general, no more than 1-2% difference between the two samples.

The SNaPshot™ reaction can be used to genotype individual samples or used to compare allele frequencies between case and control DNA pools. In later investigations, SNaPshot™ was also used in allele specific expression assays (see Chapter five).

3.3.7 AMPLIFLUOR® SNPs GENOTYPING SYSTEM

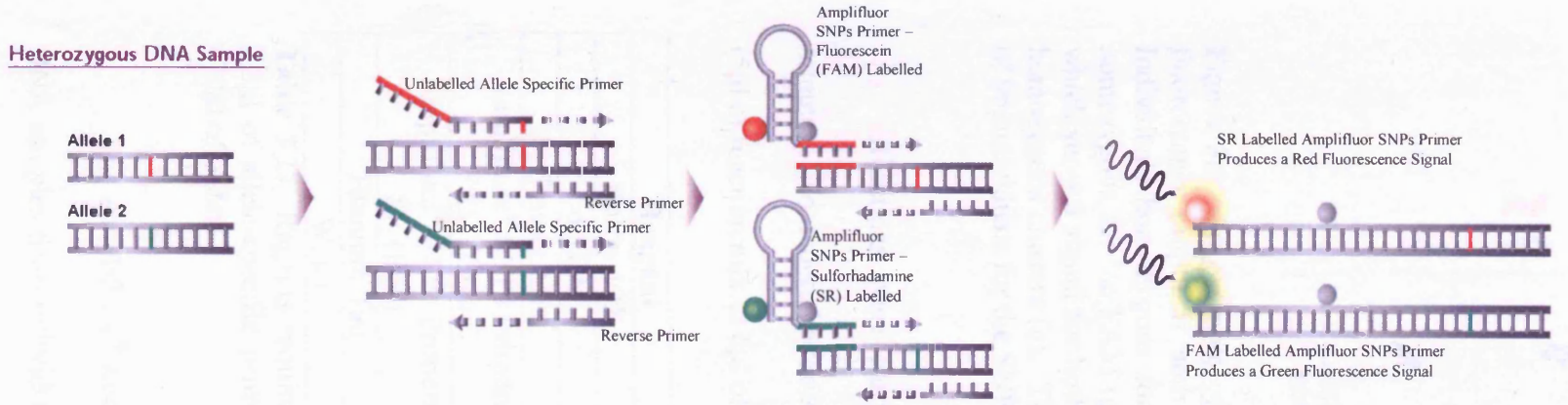
The Amplifluor® SNPs Genotyping System (Amplifluor®) (Chemicon International, USA) is a method of individual genotyping individual DNA samples in a single-step PCR reaction. The PCR reaction involves five primers, two Amplifluor® SNPs primers and three unlabeled primers.

The Amplifluor® SNPs primers are labelled with either fluorescein (FAM) or sulforhadamine (SR). A single reverse primer and two allele specific primers (one for each allele in the SNP) are designed (using a primer design programme available at www.assayarchitect.com) to amplify across the SNP of interest. Each of the allele specific primers have a 5' tail that corresponds to one of the two Amplifluor® SNPs primers.

As SNP-specific PCR products are generated, the primer sequences of the Amplifluor® SNPs primers hybridise to the complement of the tail sequence of the generated PCR product, and with the reverse primer, amplify the product further. During the PCR reaction, the hairpin structure of the Amplifluor® SNPs primer is unfolded, releasing the fluorphore (either FAM or SR) from the quencher (Dabsyl). As a result, depending on the allele present in the target sequence (resulting in the specific annealing of either a FAM or SR labelled Amplifluor® SNPs primer), a

green (FAM) or red (SR) fluorescent signal is produced. Heterozygotes have signals from both fluorophores. Fluorescence is detected on an LJL Biosystems Analyst platform. Figure 3.1 shows the Amplifluor® process.

The fluorescence for each individual is plotted on a graph with other individuals. Along the x-axis of the graph is plotted the FAM (green) fluorescence value and on the y-axis is plotted the SR (red) fluorescence value for each individual. Graphs reveal three clusters of individuals corresponding to genotypes of each homozygote and heterozygote (see Figure 3.2). Individuals cluster depending on their fluorescence signal values.



Allele specific primers anneal to the target sequence and are elongated using *Taq Polymerase*. Each primer has a unique 5' tail sequence, complementary to the priming domain of one of the Amplifluor SNPs primers. The reverse primer anneals to its target sequence and is also elongated with *Taq Polymerase*. This results in the synthesis of the tail sequence complement.

Amplifluor SNPs primers anneal specifically to the products of the reverse reaction and are elongated by *Taq Polymerase*. During the PCR reaction, the hairpin structure of the Amplifluor SNPs primers is unfolded and a fluorescent signal is generated.

Depending on which base is present in the target SNP, either a green or a red signal is generated. Heterozygotes produce a signal combined for both fluorophores. The fluorescent signal can then be measured on a fluorescent plate reader.

Figure 3.1 Amplifluor® SNPs Genotyping. The diagram shows the genotyping of a heterozygous individual. Homozygous individuals generate only one coloured signal depending on the allele present at the SNP. Adapted from Amplifluor® SNPs Genotyping System for Assay Development Handbook (www.chemicon.com).

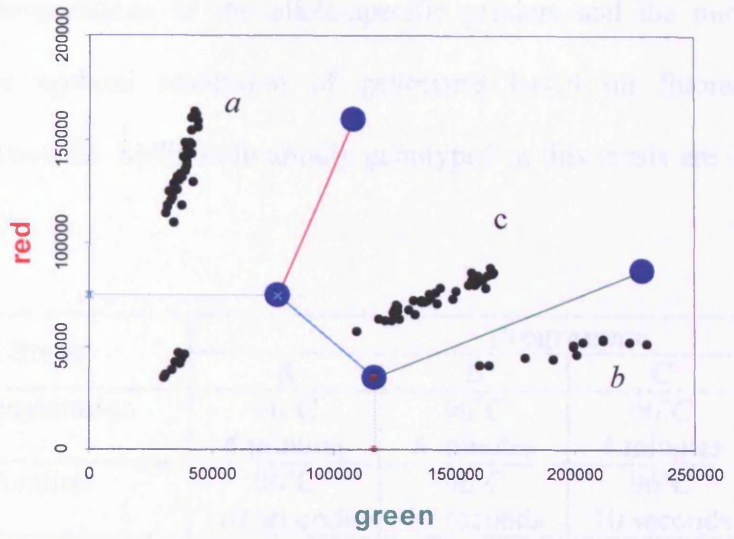


Figure 3.2 An example of an Amplifluor® output. For each DNA sample, the fluorescence output of both FAM and SR is determined and plotted on a graph. Individuals homozygous for the SR (red) signal form one cluster (*a*) and those homozygous for the FAM (green) signal form a second cluster (*b*). Heterozygotes which show a signal for both SR and FAM form a third group lying in between the homozygous clusters (*c*). The graph shows the results obtained for the genotyping of 96 individuals for the SNP rs2179515 (see Chapter five).

Reactions were carried out on 5µl of 8ng/µl dried DNA (90°C for 10 minutes). To undertake the Amplifluor™ assay the following reaction mix was used (5µl of reaction mix to 6µl of DNA):

Reagent	Company	Amount
Buffer 10X	BD Biosciences, USA	0.5µl
dNTPs	Amersham, UK	0.4µl
Primer Mix*		0.07µl
Amplifluor® SNPs Primers – FAM (Green)	Intergen, UK	0.07µl
Amplifluor® SNPs Primers – SR (Red)	Intergen, UK	0.07µl
Titanium Taq	BD Biosciences, USA	0.05µl
Water	UHW, UK	3.84µl

Table 3.17 Regents required for Amplifluor® genotyping. *Primer mix contains 2.5µl of allele-specific primer (100pMol), 25µl of reverse primer (100pMol) and 470µl of water.

The Amplifluor® assay was optimised for PCR reaction condition using 90 DNA samples from individuals from CEPH DNA. Optimisation is required for the

annealing temperatures of the allele-specific primers and the number of cycles required for optimal resolution of genotypes based on fluorescence output. Conditions used for SNPs individually genotyped in this thesis are shown in Table 3.18.

PCR Steps	Programme			
	A	B	C	D
Initial Denaturation	96°C 4 minutes	96°C 4 minutes	96°C 4 minutes	96°C 4 minutes
Denaturation	96°C 10 seconds	96°C 10 seconds	96°C 10 seconds	96°C 10 seconds
Annealing	58°C 5 seconds	60°C 5 seconds	60°C 5 seconds	58°C 5 seconds
Extension	72°C 10 seconds	72°C 10 seconds	72°C 10 seconds	72°C 10 seconds
Number of Cycles	19	19	19	19
Denaturation	96°C 10 seconds	96°C 10 seconds	96°C 10 seconds	96°C 10 seconds
Annealing	55°C 20 seconds	55°C 20 seconds	55°C 20 seconds	55°C 20 seconds
Extension	72°C 40 seconds	72°C 40 seconds	72°C 40 seconds	72°C 40 seconds
Number of Cycles	22 cycles	22 cycles	27 cycles	20 cycles
Final Extension	72°C 3 minutes	72°C 3 minutes	72°C 3 minutes	72°C 3 minutes

Table 3.18 Amplifluor® genotyping assay conditions.

3.3.8 RESTRICTION FRAGMENT LENGTH POLYMORPHISMS (RFLP)

Restriction fragment length polymorphisms were used to genotype individual samples of DNA when the Amplifluor® method of genotyping failed to optimise or local sequence made primers too difficult to design.

In the current context, RFLP genotyping is a PCR based assay where DNA is amplified by PCR and then fragments digested using appropriate restriction endonuclease enzymes, which cut DNA at specific nucleotides. Restriction enzymes recognise specific sequences of bases and cut the sequence at a specific location. For example, in the case of a C→T SNP, located in the sequence 5'....TCGA....3',

the restriction enzyme (TaqI – see Appendix 1) will recognise the sequence TCGA and cut the DNA after the C nucleotide. When a T nucleotide is in place of the C nucleotide, resulting in the sequence TTGA, the restriction enzyme does not recognise the sequence and consequently does not cut the DNA sequence. Genotypes can be identified by the different fragment sizes after digestion using agarose gel electrophoresis.

PCR assays were performed as described in Section 3.3.1. PCR products were then digested for 16 hours with 5 units of the appropriate restriction enzyme (New England Biolabs, USA) in the presence of the appropriate restriction buffer, 1X BSA (bovine serum albumin) (if appropriate) and at the recommended temperature (all followed from manufacturers guidelines per enzyme). Digested PCR products were electrophoresed on 3% agarose gels and genotypes read manually. To check no mistakes were made in the manual reading of genotypes a second laboratory member blindly read all samples.

Restriction enzymes were chosen using the program Sequencher™, which analyses sequence patterns and highlights restriction endonuclease digestion sites.

CHAPTER FOUR

POSITIONAL CANDIDATE GENE ASSOCIATION STUDIES BETWEEN DEVELOPMENTAL DYSLEXIA AND CHROMOSOME 6p: THE IDENTIFICATION OF *KIAA0319* AS A SUSCEPTIBILITY GENE FOR DEVELOPMENTAL DYSLEXIA

CHAPTER FOUR

4. POSITIONAL CANDIDATE GENE ASSOCIATION STUDIES BETWEEN DEVELOPMENTAL DYSLEXIA AND CHROMOSOME 6p: THE IDENTIFICATION OF *KIAA0319* AS A SUSCEPTIBILITY GENE FOR DEVELOPMENTAL DYSLEXIA

Few linkage studies into complex disorders replicate a region of the genome as robustly as the chromosome 6p locus (DYX2) in DD. To date, over eleven studies, including both linkage and association, have found evidence for a gene in the region D6S109-D6S291. Complex disorders are often hard to replicate in terms of linkage and association due to variable phenotype definition, phenocopy, heterogeneity, low penetrance and oligogenicity.

Only two studies (Field and Kaplan 1998; Petryshen et al. 2000) (both using the same sample) have failed to find evidence of linkage or association to the region *DYX2*.

In Figure 2.2 (Chapter two), the location of signals emerging from linkage and association studies of chromosome 6p are shown. The region, spanning approximately 16Mb shows particular evidence of linkage where samples are selected for lower scores ($\leq$ 2 SD below the population mean) (Cardon et al. 1994; Cardon et al. 1995; Gayán et al. 1999; Francks et al. 2004). Combining all the linkage data from studies between DD and chromosome 6p reveals a consensus region. This region is a 4.2Mb region between markers D6S461 and D6S105 and may be the location of a QTL influencing DD (Figure 2.2, Chapter two).

4.1 THE SEARCH FOR SUSCEPTIBILITY GENES ON CHROMOSOME 6p

Given the evidence for a susceptibility gene on 6p and given the relatively consistent evidence across 6p22.2-21.3, candidate genes were selected for analysis. Since relatively little is known about the neurobiology of DD, candidate genes selection based on function is challenging. The broad linkage region contains 447 known genes from which candidates were selected. Previous studies (Turic et al. 2003) across DYX2 yielded significant evidence for a susceptibility gene for DD between markers D6S109 and D6S1260 (for detailed description see Chapter two). Genes in this region were prioritised for analysis in this study, along with genes with known expression in the brain and functions relevant to cognitive disorders such as IQ and lissencephaly.

Between D6S109 and D6S1260, there are 95 genes, of which 43 encode histone proteins. Histone proteins are not obvious functional candidate genes for specific cognitive disorders and were not examined in this study. Of the 52 remaining genes, 11 genes were selected for association analysis based on expression in brain, location between D6S109 and D6S1260 and in some cases, previously reported associations (see Table 4.1).

Gene	Genomic Size (kb)	Number of Exons	Amino Acid Sequence Length	Protein Size (kDa)
ID4	3.3	3	162	16.6
SOX4	4.8	1	474	47.3
PRL	10.3	5	227	25.9
VMP	21.3	44	195	21.5
DCDC2	211.1	10	476	52.8
KAAG1	1.4	1	84	9.0
MRS2L	21.3	11	443	50.3
KIAA0319	102.1	21	1071	117.6
TTRAP	16.9	7	362	40.9
THEM2	34.7	3	140	15.0
C6orf62	14.2	5	229	27.1

Table 4.1 Genomic and proteomic information for genes selected for association analysis within the region D6S109 to D6S1260.

4.1.1 DNA-BINDING INHIBITOR PROTEIN 4 (*ID4*)

Figure 4.1 shows a schematic representation of the structure of *ID4*. *ID4* is a member of a family of genes that encode helix-loop-helix (HLH) proteins that mediate processes involved in development, including cellular differentiation, proliferation and apoptosis (Andres-Barquin et al. 2000), cell growth and cell cycle progression and embryonic development (Bounpheng et al. 1999). Studies of ID4 indicate that the protein may contribute to mammalian nervous system development (Riechmann and Sablitzky 1995; Andres-Barquin et al. 1998; Andres-Barquin et al. 2000; Yokota 2001; Yun et al. 2004).

ID4 is expressed as four major transcripts generated by differential use of polyadenylation sites (Cruchten et al. 1998). The abundance of the four transcripts varies across tissues, suggesting a tissue-specific regulation of polyadenylation and/or post-transcriptional regulation of *ID4* expression (Cruchten et al. 1998). Expression occurs in neuronal tissue, the ventral portion of the epithelium of the developing stomach (Jen et al. 1996), adult brain, kidney, testis (Cruchten et al. 1998), thyroid, foetal tissue and some nervous system tumour cell lines (Rigolet et al. 1998). The ID4 protein is also detectable in the cytoplasm of type A1 spermatogonia, as well as in late pachytene and in diplotene spermatocytes (Sablitzky et al. 1998). By binding to basic HLH transcription factors, the ID proteins regulate gene expression.

ID4 is required for normal brain size and regulates lateral expansion of the proliferative zone in the developing cortex and hippocampus (Yun et al. 2004). In the absence of ID4, proliferation of stem cells in the ventricular zone is compromised. In early cortical progenitor cells, ID4 is required for the normal progression from G1 to S phase of the cell cycle. It has also been suggested that ID4

is a regulator of neural stem cell proliferation and fate determination (Yokota 2001).

Schematic Representation of DNA-Binding Inhibitor Protein 4 (*ID4*)

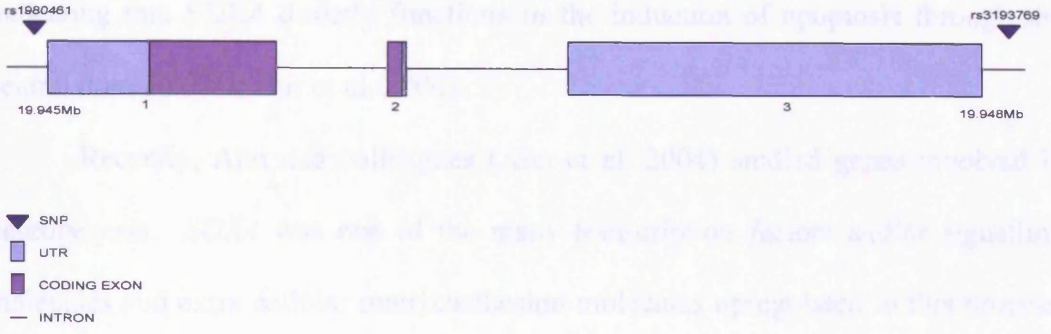


Figure 4.1 Schematic representation of the gene *ID4*. SNPs selected for association analysis in this study are indicated by triangles. An indirect association analysis was undertaken for *ID4* using SNPs selected from public databases.

4.1.2 SEX-DETERMINING REGION Y –BOX (SRY-BOX) MEMBER 4 (*SOX4*)

The family of *SOX* genes encodes transcription factors defined by a conserved high mobility group (HMG) domain that mediates their binding to a short target DNA sequence. The HMG box consists of approximately 80 amino acids, a motif that is also found in the chromatin-associated proteins HMG-1 and HMG-2. This motif suggests that as well as being transcription factors, SOX proteins may bend DNA, altering chromosome structure and leading to recruitment of various regulators and the binding and forming of biologically active complexes (Reppe et al. 2000; Lee et al. 2002).

The HMG region spans amino acids 57-135 of SOX4 and lies just upstream from a glycine rich region (GRR) between amino acids 152-227. A third serine rich region (SRR) is located between amino acids 333-397. This SRR acts as a

transactivation domain towards the C-terminus of SOX4. The HMG and SRR regions regulate transcription, whilst the function of the GRR region is unknown. A central domain (CD) of SOX4 has a role in regulating apoptosis. Deletions which destroy the HMG and SRR have no effect on its ability to induce apoptosis, indicating that SOX4 directly functions in the induction of apoptosis through the central domain (CD; Hur et al. 2004).

Recently, Ahn and colleagues (Ahn et al. 2004) studied genes involved in neurogenesis. *SOX4* was one of the many transcription factors and/or signalling molecules and extra cellular matrix/adhesion molecules upregulated in this process, suggesting *SOX4* has a role in the development of neurones.

A schematic representation of *SOX4* is shown in Figure 4.2.

Schematic Representation of Sex-Determining Region Y –BOX (SRY-BOX) Member 4 (*SOX4*)

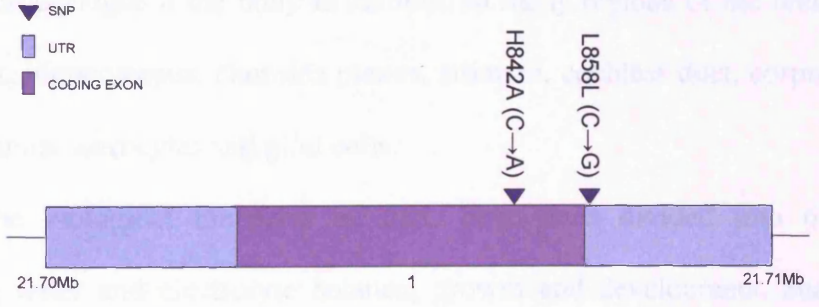


Figure 4.2 A schematic representation of *SOX4*. An indirect association analysis was undertaken for *SOX4*. SNPs identified by dHPLC and sequencing are indicated by triangles. *SOX4* contains no introns.

SOX family proteins have been shown to play a role in development, sex determination, testis formation, neuronal development, lymphocyte differentiation and chondrogenesis.

4.1.3 PROLACTIN (*PRL*)

Prolactin (PRL) is one of a family of related hormones including growth hormone (GH) and placental lactogen (PL) (Niall et al. 1971). A schematic representation of the gene encoding PRL is shown in Figure 4.3.

PRL is primarily thought of as a hormone synthesised and secreted by the lactotrophic cells of the anterior pituitary gland. However, it has also been shown to be synthesised by the brain, deciduas, myometrium, lacrimal gland, thymus, spleen, circulatory lymphocytes, lymphoid cells of the bone marrow, mammary epithelium cells and tumours, skin fibroblasts and sweat glands (Ben-Jonathan et al. 1996). As well as being present in serum, PRL is also found in cerebrospinal fluid, amniotic fluid, tears, milk, follicular fluid and sweat.

PRL, secreted by the pituitary (pPRL), is transported by the circulatory system and acts on target cells at peripheral sites that contain PRL receptors (PRLR) in their plasma membranes. PRL has multiple functions including acting as a growth factor, neurotransmitter and immunomodulator. PRL binding sites have been found throughout the body in addition to many regions of the brain including the cortex, hippocampus, choroids plexus, striatum, cochlear duct, corpus callosum, hypothalamus, astrocytes and glial cells.

The biological functions of PRL have been divided into 6 categories including water and electrolyte balance, growth and development, endocrinology and metabolism, brain and behaviour, reproduction and immunoregulation and protection. Within the brain, PRL has been shown to have effects on stress responses, increased dopamine turnover and maturation of the neonatal neuroendocrine system.

Schematic Representation of Prolactin (*PRL*)

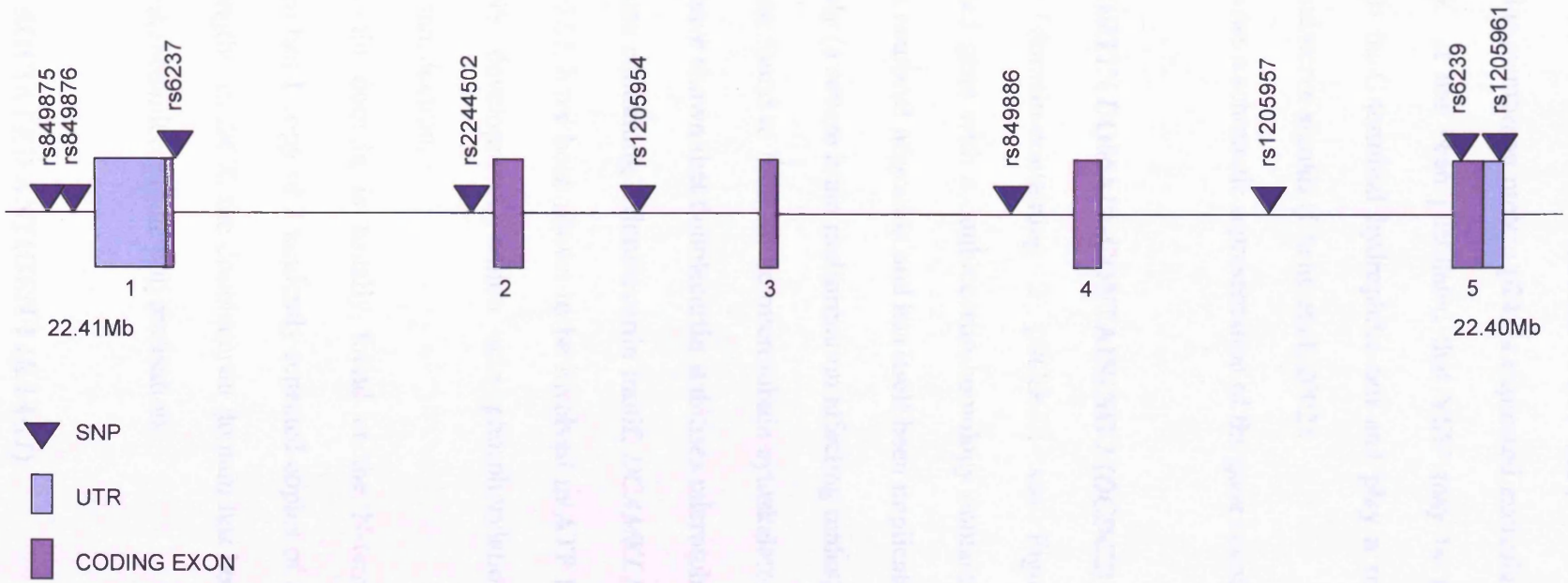


Figure 4.3 A schematic representation of *PRL*. Triangles indicate SNPs tested for association with DD in this study. An indirect association analysis was undertaken for *PRL* using SNPs selected from public databases.

4.1.4 VESICULAR MEMBRANE PROTEIN P24 (*VMP*)

VMP, vesicular membrane protein p24, is expressed exclusively in the brain (Cheng et al. 2002). It has been postulated that VMP may be associated with microtubules through its C-terminal hydrophilic tail and play a role in vesicular organelle transport and nerve signals (Cheng et al. 2002).

Figure 4.4 shows a schematic representation of the gene encoding VMP.

4.1.5 DOUBLECORTIN DOMAIN-CONTAINING 2 (*DCDC2*)

Doublecortin domain-containing 2 (*DCDC2*; see Figure 4.5) is a ubiquitously expressed gene with a doublecortin-homology domain. Doublecortin (*DCX*) is involved in neuronal migration and has itself been implicated as a cause of X-linked lissencephaly (a severe brain malformation affecting males) (Gleeson et al. 1998). DCX has been found to bind to the microtubule cytoskeleton. Assays both in vivo and in vitro have shown that doublecortin stabilises microtubules and causes bundling. Other genes containing a doublecortin motif, *DCAMKL1*, *DCX*, *DCX.1*, *DCX.3*, *RP1* and *RP1L1*, have been shown to be involved in ATP binding, protein kinase activity, CNS development, amino acid phosphorylation, intracellular signalling and phototransduction.

The doublecortin domain is usually found at the N-terminus of DCX proteins. The domain has 1 copy or 2 tandemly repeated copies of a region around 80 amino acids in length. In *DCX*, the doublecortin domain has been suggested to bind to tubulin and enhance microtubule polymerisation.

4.1.6 KIDNEY-ASSOCIATED ANTIGEN 1 (*KAAG1*)

In a study by Van den Eynde and colleagues (Eynde et al. 1999), cytolytic T

lymphocytes (CTLs) were obtained from lymphocytes stimulated from renal cell carcinoma patients with autologous tumour cells. The antigenic peptide recognised by the CTLs has the sequence LPRWPPPQL. This sequence is encoded by *KAAG1*. In short term cultures of epithelial cells from the renal proximal tubule, *KAAG1* was expressed at significant levels and was recognised by CTLs. They therefore concluded that the antigen is not tumour specific but corresponds to a self-antigen with restricted tissue distribution.

KAAG1 starts at a promoter on the reverse strand of the first intron of *DCDC2* and ends on the reverse strand of the *DCDC2* promotor (see Figure 4.5), which contains a polyadenylation signal. *KAAG1*, is expressed in a high proportion of tumours of various histological origins. In normal tissue it seems only to be expressed in testis and kidney and to a lower extent in the bladder and liver. *KAAG1* was included in this study as SNPs selected across *DCDC2* also covered this gene. *KAAG1* would not otherwise have been selected a good candidate gene for DD.

4.1.7 MAGNESIUM HOMEOSTASIS FACTOR (*MRS2L*)

The product of this gene is important in magnesium homeostasis (Zsurka et al. 2001). The protein contains two predicted transmembrane domains in its carboxyl-terminus. The middle of the sequence is leucine rich (amino acids 157-270) and it has been suggested that this region may form a helix-turn-helix structure, which may be involved in protein-protein interactions (Schmidt et al. 1998). *MRS2L* is expressed ubiquitously (Londin et al. 2003). This gene had been included in a previous association study of DD, however analysis was not comprehensive (Deffenbacher et al. 2004). A schematic representation of *MRS2L* is shown in Figure 4.6.

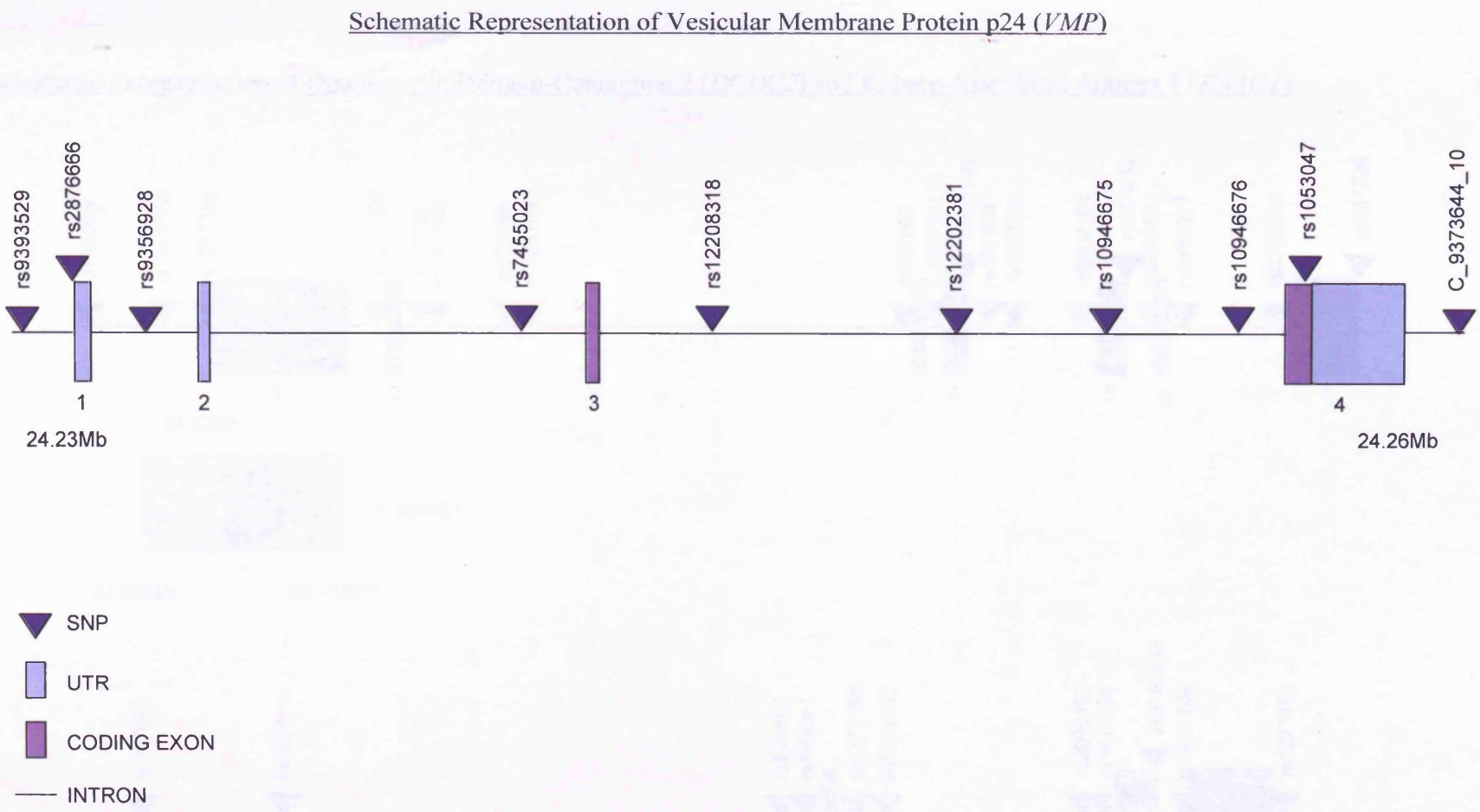


Figure 4.4 A schematic representation of *VMP*. SNPs tested for association with DD in this study are highlighted with triangles. Previous studies have revealed association between DD and SNPs within *VMP* (Deffenbacher et al. 2004). An indirect association analysis was undertaken for *VMP* using SNPs selected from public databases.

Schematic Representation of Doublecortin Domain-Containing 2 (*DCDC2*) and Kidney-Associated Antigen 1 (*KAAG1*)

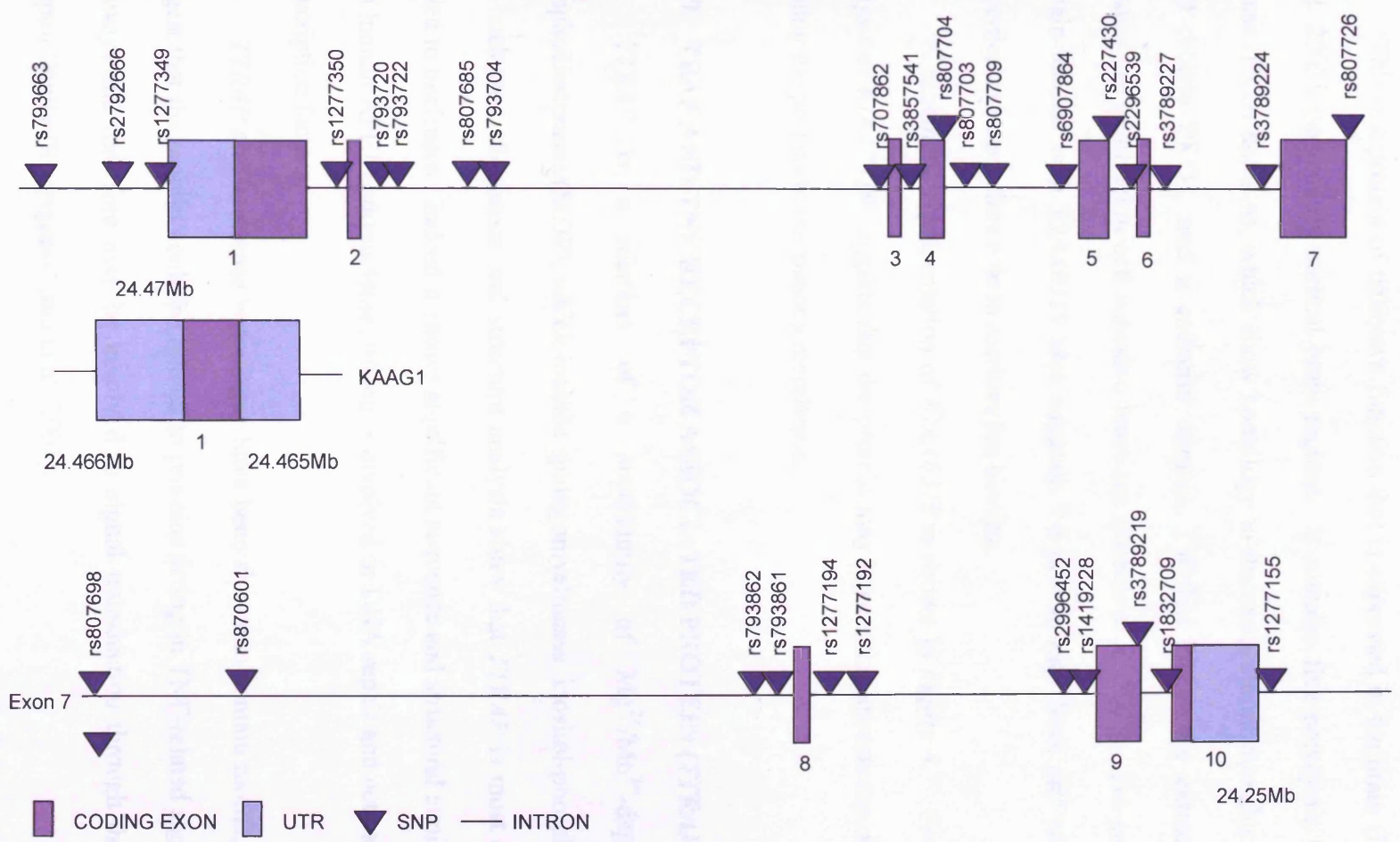


Figure 4.5 Schematic representation of *DCDC2*. SNPs indicated with a triangle were tested for association with DD in this study. Association had been observed previously between *DCDC2* and DD (Deffenbacher et al. 2004). An indirect association analysis was undertaken for *DCDC2* using SNPs selected from public databases.

4.1.8 *KIAA0319*

This is a protein of unknown function that is expressed in the brain (Londin et al. 2003), particularly cortical brain regions. It contains five polycystic kidney disease (PKD) domains, which show homology to the extracellular domains of the PKD protein PKD1, and a cadherin domain. Within PKD1 the extracellular domains are involved in cell-adhesive functions (Streets et al. 2003). The cadherin domain located with *KIAA0319* also suggests the protein may have cell adhesion properties and a possible role in calcium ion binding.

A schematic representation of *KIAA0319* is shown in Figure 4.7. Sequence analysis of *KIAA0319* suggests that the protein may have a transmembrane domain, locating the protein to the plasma membrane.

4.1.9 TRAF AND TNF RECEPTOR ASSOCIATED PROTEIN (*TTRAP*)

TTRAP is a member of a superfamily of Mg^{2+}/Mn^{2+}-dependant phosphodiesterases (MDP), which include sphingomyelinases, inositol-phosphatases and nucleases. Sequence and structure analysis show that *TTRAP* is most closely related to nucleases. Indeed it shows significant sequence and structural similarities with human APE1 endonuclease, which is involved in DNA repair and activation of transcription factors.

TTRAP and its mouse homologue have been shown to contain motifs, which suggest that the protein could be related to proteins acting in TNF-related signalling pathways and therefore may be involved in signal transduction through the TNF receptor family (Rodrigues-Lima et al. 2001).

TTRAP is also a CD40 (a tumour necrosis factor, TNF) and TRAF-interacting protein. The protein interacts with cytoplasmic TNF receptor-associated factors (TRAFs), as well as with the cytoplasmic domains of specific members of

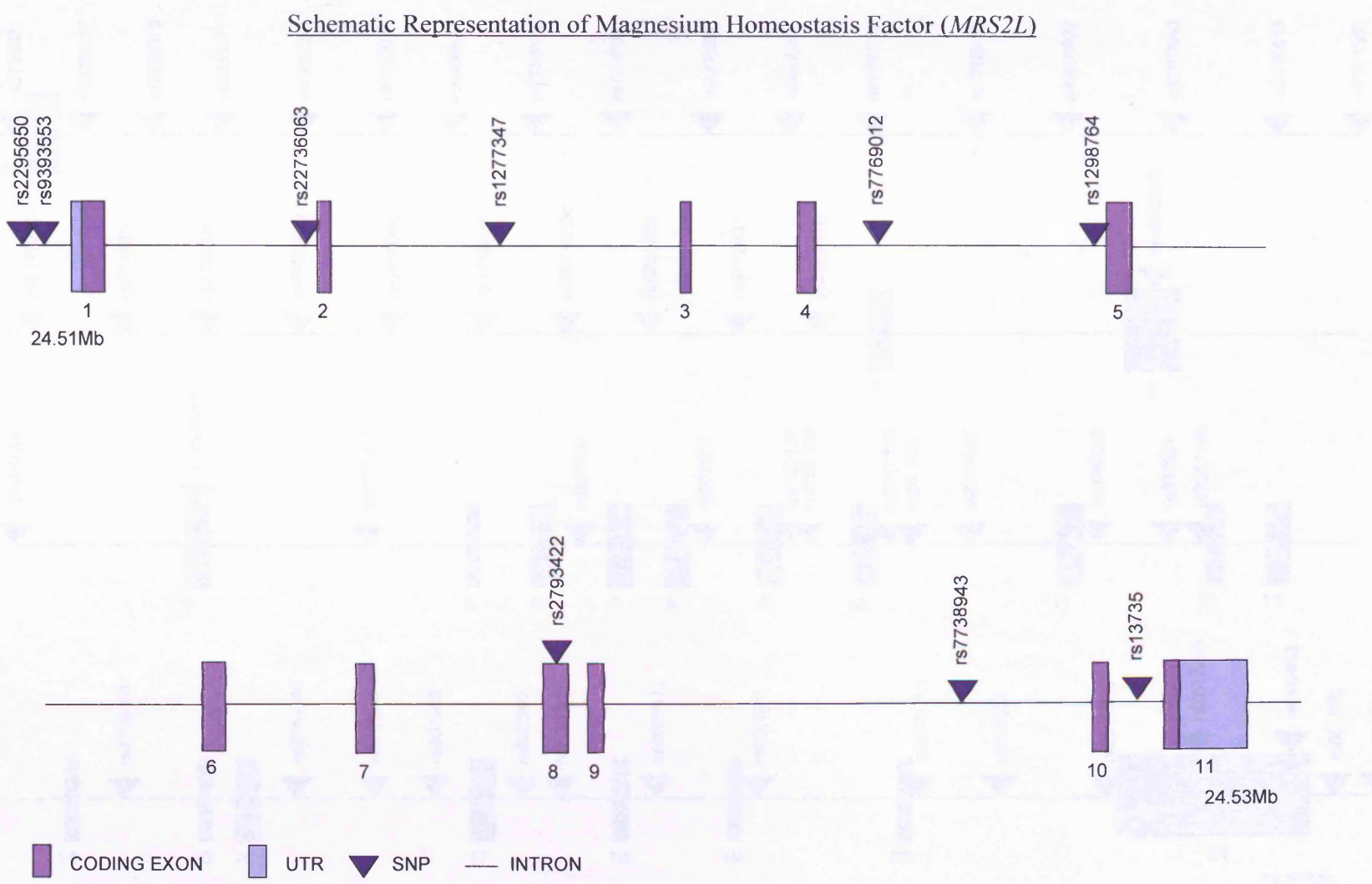


Figure 4.6 A schematic representation of *MRS2L*. SNPs tested for association with DD in this study are indicated by triangles. An indirect association analysis was undertaken for *MRS2L* using SNPs selected from public databases.

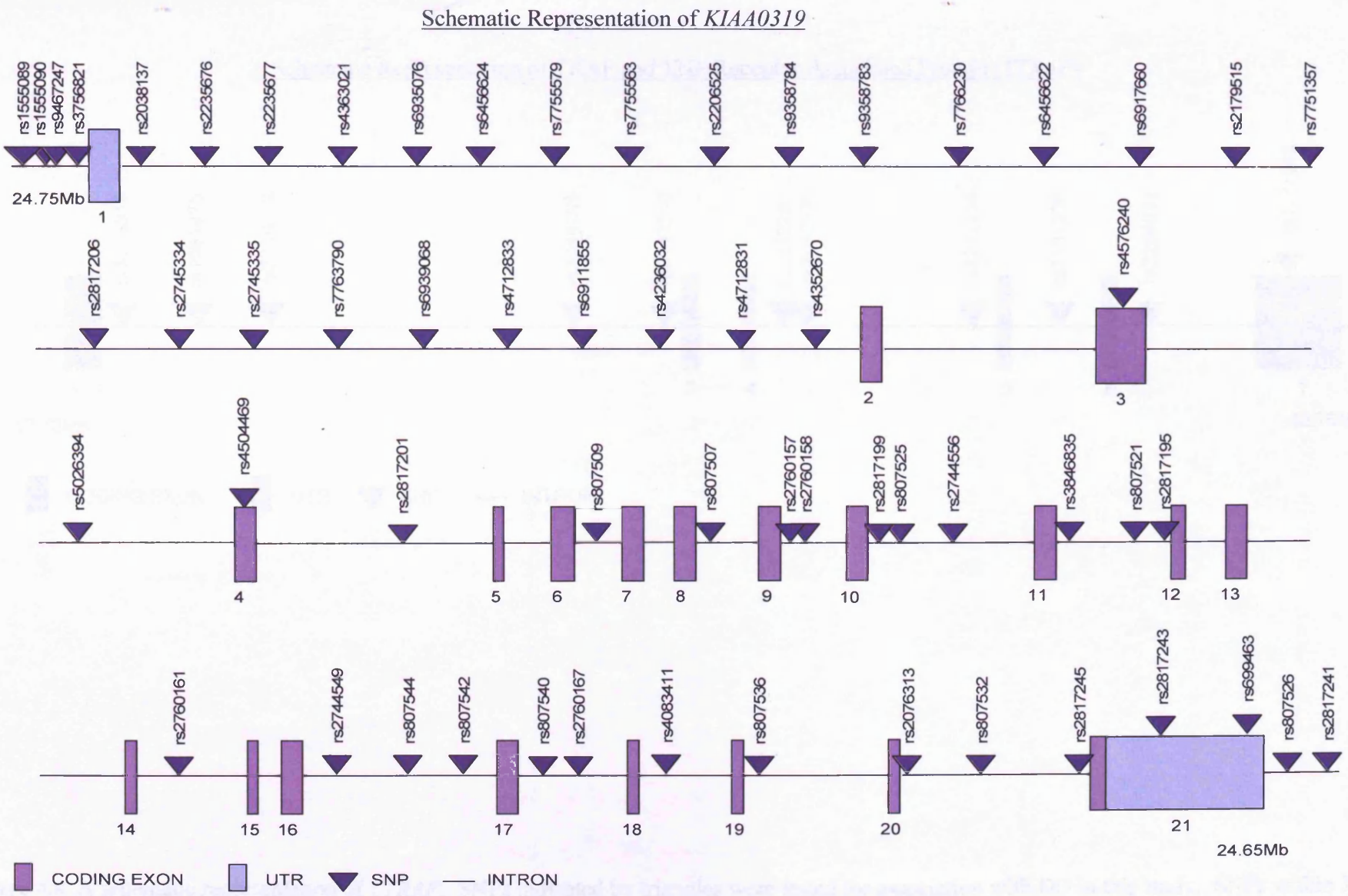


Figure 4.7 A schematic representation of *KIAA0319*. SNPs tested for association with DD in this study are indicated with triangles. Previously two studies had shown association between DD and *KIAA0319* (Deffenbacher et al. 2004; Francks et al. 2004). In particular the first four exons revealed evidence for association with DD (Francks et al. 2004). An indirect association analysis was undertaken for *KIAA0319* using SNPs selected from public databases.

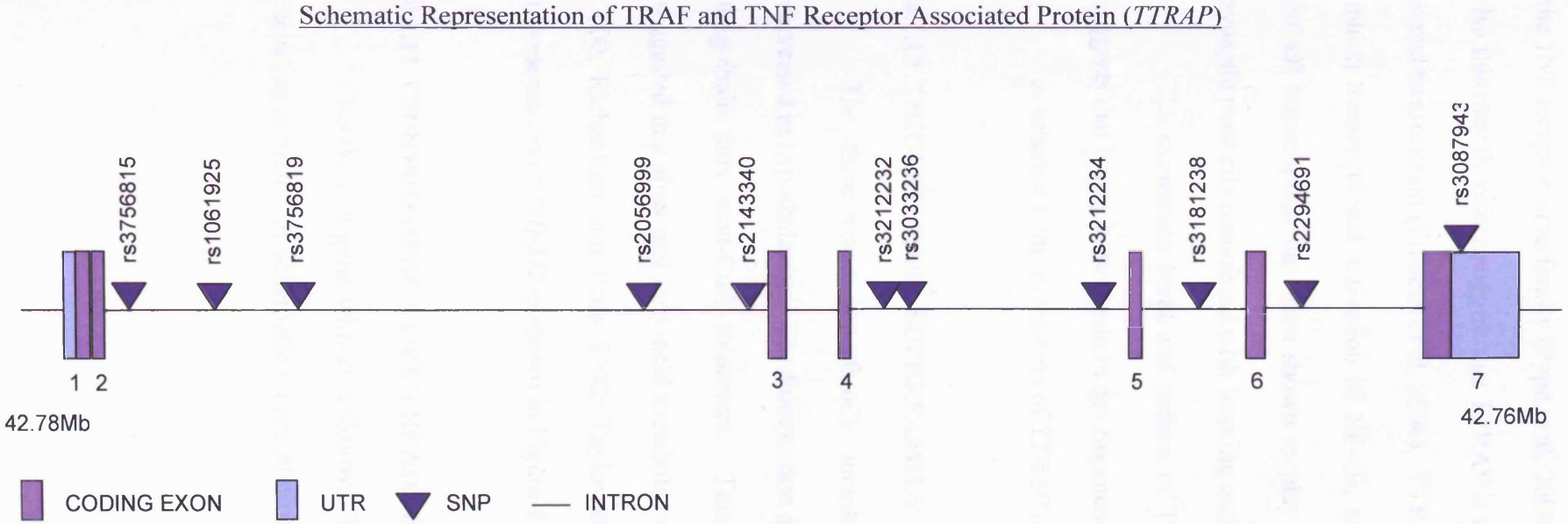


Figure 4.8 A schematic representation of *TTRAP*. SNPs indicated by triangles were tested for association with DD in this study. SNPs within *TTRAP* have previously been reported to show association with DD (Deffenbacher et al. 2004; Francks et al. 2004). An indirect association analysis was undertaken for *TTRAP* using SNPs selected from public databases.

the TNF receptor superfamily (Pype et al. 2000). Francks and colleagues suggested that this may therefore suggest that TTRAP is involved as a regulatory factor in TNF signal transduction (Francks et al. 2004). TTRAP overexpression has been shown to inhibit transcriptional activation of NF-κB, a TNF-responsive transcription factor. NF-κB transcription has been shown to play a role in long-term potentiation and synaptic plasticity associated with learning and memory (Meffert et al. 2003).

The expression level and pattern of TTRAP mRNA during embryogenesis suggests that it may have a role in development (Pype et al. 2000).

A schematic representation of *TTRAP* is shown in Figure 4.8.

4.1.10 THIOESTERASE SUPERFAMILY MEMBER 2 (*THEM2*)

The thioesterase superfamily member 2, *THEM2*, encodes a protein expressed in hypothalamus. The thioesterase superfamily catalyses the hydrolysis of long-chain fatty acyl-CoA thioesters. This gene was selected as it has been suggested that abnormal fatty-acid metabolism plays a role in DD (Richardson et al. 2000; Richardson and Ross 2000; Taylor and Richardson 2000). A schematic representation of *THEM2* is shown in Figure 4.9.

4.1.11 CHROMOSOME 6 OPEN READING FRAME 62 (*C6orf62*)

C6orf62 is a gene with an unknown function. It is expressed ubiquitously, including in brain. A schematic representation of *C6orf62* is shown in Figure 4.10.

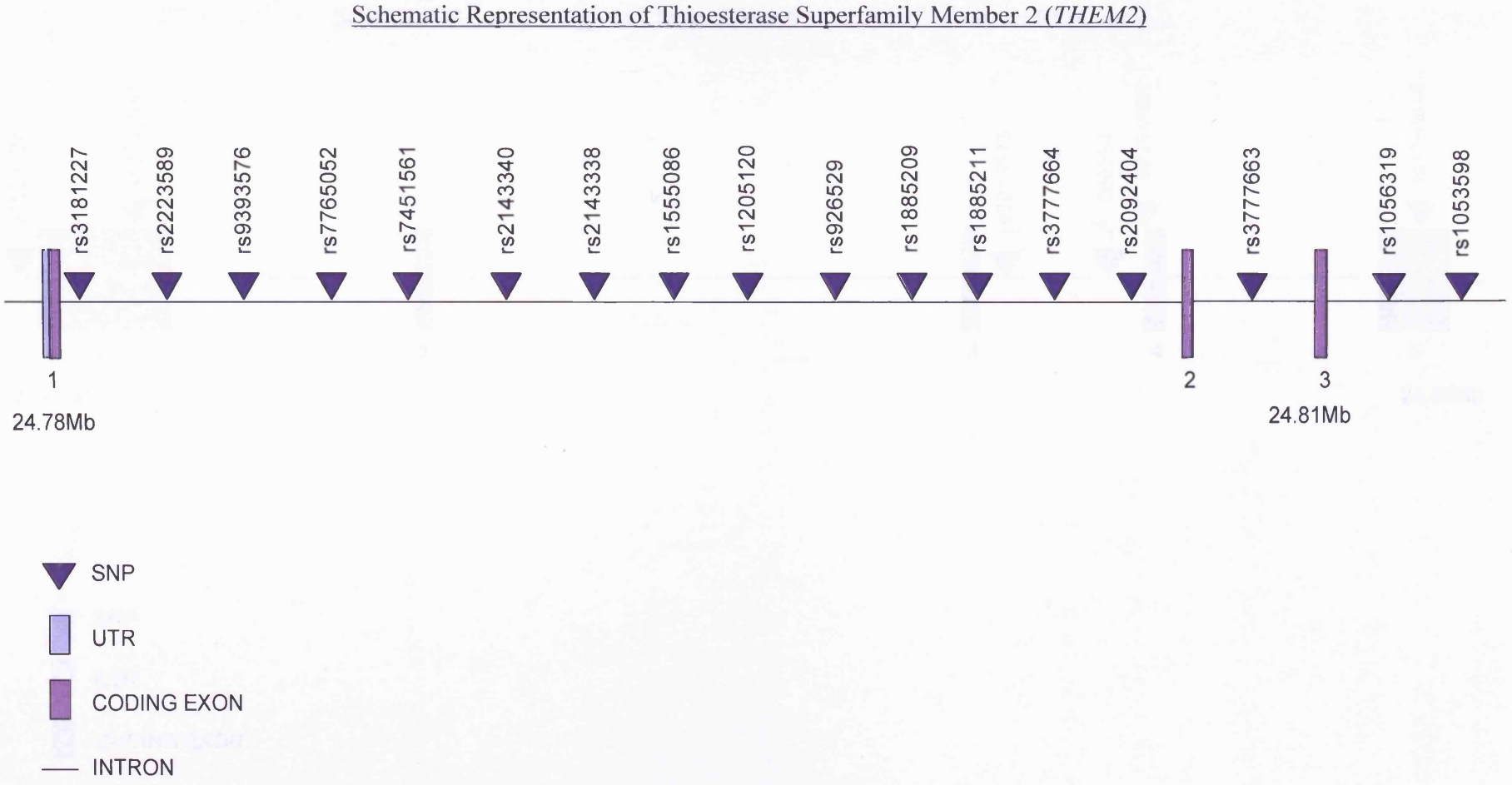


Figure 4.9 A schematic representation of *THEM2*. Triangles indicate SNPs tested for association with DD in this study. Some evidence of association between DD and *THEM2* has been reported previously (Francks et al. 2004). An indirect association analysis was undertaken for *THEM2* using SNPs selected from public databases.

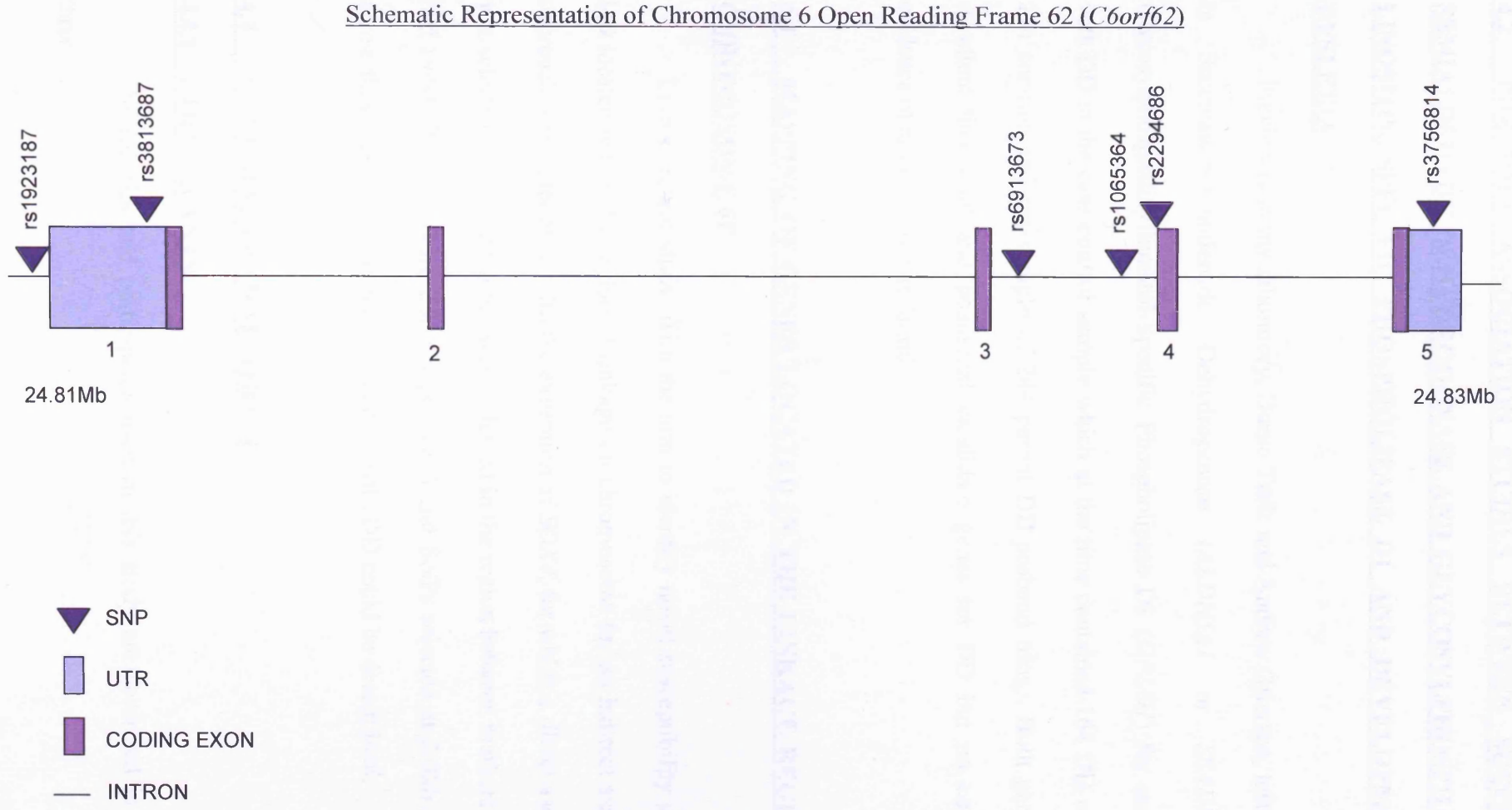


Figure 4.10 A schematic representation of *C6orf62*. Triangles indicate SNPs tested for association with DD in this study. An indirect association analysis was undertaken for *C6orf62* using SNPs selected from public databases.

4.2 PREVIOUS ASSOCIATION STUDIES BETWEEN SUCCINATE-SEMIALDEHYDE DEHYDROGENASE AND GLYCOSYLPHOSPHATIDY-LINOSITOL-SPECIFIC PHOSPHOLIPASE D1 AND DEVELOPMENTAL DYSLEXIA

Previously in my laboratory, Darko Turic and Andrew Grierson, tested SNPs in Succinate-Semialdehyde Dehydrogenase (*ALDH5A1* or *SSADH*) and Glycosylphosphatidylinositol-Specific Phospholipase D1 (*GPLD1*) for association with DD in the case-control sample which at the time contained 164 DD cases and 174 controls and the sample of 244 parent DD-proband trios. Both genes were excellent functional and positional candidate genes for DD but no convincing evidence of association was found

4.3 MAPPING OF GENES LOCATED IN THE LINKAGE REGION ON CHROMOSOME 6P

In this present study, with the aim to identify novel susceptibility genes for DD located within the region of linkage on chromosome 6p, an indirect association approach was undertaken with the exception of *SOX4*, for which a direct association was selected. Candidate genes were selected in the region between markers D6S109 and D6S1260 (see Figure 2.2, Chapter two) and SNPs selected at 2-3kb intervals across these genes in order that association with DD could be determined.

4.4 MATERIALS AND METHODS

4.4.1 PARTICIPANTS

The samples and participants used in this study are described in Chapter three.

4.4.2 ANALYSIS OF CANDIDATE GENES FOR DIRECT ASSOCIATION ANALYSIS WITH DEVELOPMENTAL DYSLEXIA

The gene *SOX4* was tested for association with DD using a direct association analysis approach. The gene was examined for polymorphisms using dHPLC and sequencing. Since there is only one exon in *SOX4*, the coding sequence, 5' and 3' UTR and 1kb of putative promotor sequence was examined for polymorphisms. The region, spanning approximately 6kb was divided into 15 fragments (average size 367bp) (see Appendix 1, Table 1).

SNPs discovered were assessed for association with DD using the DNA pooling method (see Figure 4.11). Since heterozygotes were known for SNPs, an empirically derived heterozygote ratio for the correction factor *k* could be used.

4.4.3 CHOICE OF SNPS FOR INDIRECT ASSOCIATION ANALYSIS OF CANDIDATE GENES ON CHROMOSOME 6p

SNPs were selected from Ensembl and CHIP Bioinformatics Tools (available at www.ensembl.org and http://snpper.chip.org/ respectively). SNPs with known allele frequencies were only selected if the minor allele frequency was equal to or greater than 5%. The distance between SNPs is shown in Table 4.7. SNPs were chosen to cover all introns, exons, and 3kb each of 5' and 3' flanking sequence of each gene. Introns 2, 7 and 8 of *DCDC2*, would have required 60-90 SNPs for SNP coverage every 2-3kb, therefore, for pragmatic reasons, analysis of these introns was restricted to 3kb of flanking sequence on either side of each of the exons.

During this study, two studies were published (Deffenbacher et al. 2004; Francks et al. 2004) which had assessed evidence of association across the

chromosome 6p region. Based on the results of these studies extra SNPs were selected for analysis in our sample. From the paper by Deffenbacher and colleagues (Deffenbacher et al. 2004) eight SNPs were chosen and from the Francks and colleagues paper (Francks et al. 2004), 12 extra SNPs were chosen. The additional SNPs were chosen based upon significant association with DD and absence of LD ($r^2 \leq 0.8$) with other SNPs genotyped based upon data available from the HapMap project and the two published studies (Deffenbacher et al. 2004; Francks et al. 2004).

4.4.4 SAMPLE STRATEGY EMPLOYED

Following identification of SNPs, the sample strategy shown in Figure 4.11 was utilised.

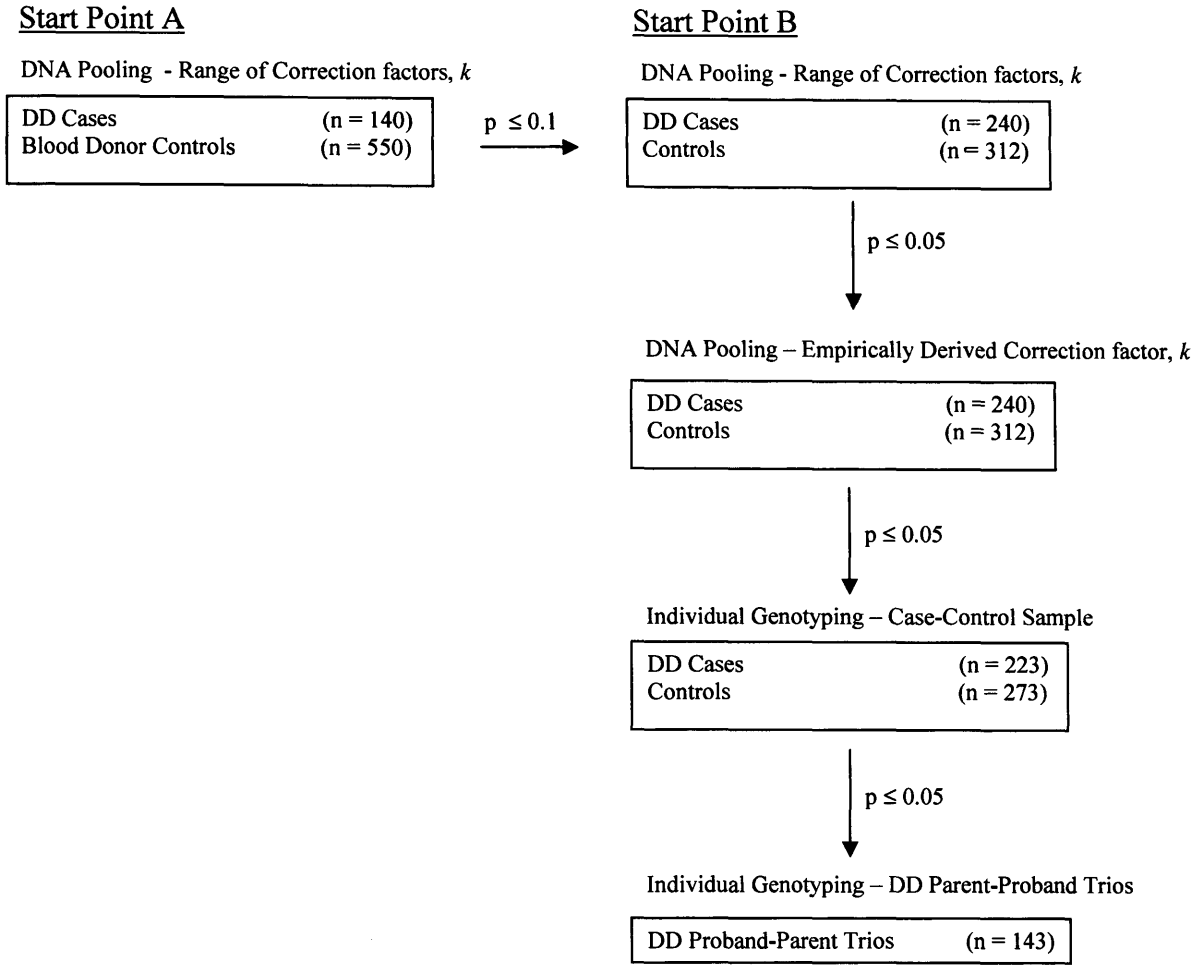


Figure 4.11 Overview of the strategy employed in the analysis of candidate genes on chromosome 6p.

At the start of the study, DNA pooling was undertaken using pooled DNA sample 1 (see Chapter three) and the DNA pooling approach was begun at start point A (see Figure 4.11). As samples became available, DNA pooling was undertaken using pooled DNA sample 2 (see Chapter three) and the DNA pooling approach was begun at start point B (see Figure 4.11). Since blood donor controls had not been tested for DD, SNPs assessed from start point A (see Figure 4.11) were tested for association, with a relaxed alpha level of 0.1. Since SNaPshot™, used to determine allele frequencies in the DNA pools, needs to be corrected for unequal representation of the specific bases involved in the SNP, a range of values of the correction factor *k* were used to correct the pools. If any of the SNPs showed

association ($p < 0.1$) with DD using any of the nine correction factors employed, DNA pools were later reanalysed in pools using DNA pooled sample 2 (if previously started at start point A, Figure 4.11). An alpha level of 0.05 was set as evidence for association DNA pools run with sample 2.

Following analysis in DNA pools sample 2, 16 individual DNA samples were genotyped by SNaPshot™ to find heterozygotes for those SNPs which had shown some evidence for association with any one of the range of correction factors used in DNA pooled sample 2. These pools were then corrected using the known heterozygote ratios. In total 168 SNPs were analysed in DNA pools.

SNPs showing association with DD in the DNA pools were then individually genotyped in 42 DD cases and 48 controls. LD was calculated between SNPs. If two SNPs showed evidence of LD ($r^2 \geq 0.8$), only one SNP from the pair was individually genotyped.

All non-redundant SNPs showing association with a p value of ≤ 0.05 in DNA pools were individually genotyped in a sample of 223 cases and 273 controls. In order to check that any observed associations between SNPs and DD were not due to population stratification, SNPs were genotyped in a semi-independent sample of 143 parent DD-proband trios.

4.4.5 VARIANTS AND GENOTYPING

The PCR conditions and assay parameters for assays involved in the analysis of *SOX4* and DD are shown in Tables 1 and 2 in Appendix 1. Following the direct detection of variants in *SOX4*, variants were analysed for association with DD using DNA pooling with sample 2.

For the SNPs analysed in the direct association analysis, PCR conditions, PCR primers and extension primers for SNaPshot™ reactions are listed in Tables 3 and 4 in Appendix 1. It should be noted that the DNA pooling for the 8 SNPs in *VMP* and 9 SNPs located in *MRS2L* was undertaken by Dr Denise Harold. For these SNPs, an empirically derived value for *k* was used in the analysis of DNA pools.

The primers and conditions for each assay undertaken for individual genotyping are shown in Tables 5 and 6 in Appendix 1.

4.4.6 STATISTICS

All Genotypes were tested for Hardy-Weinberg Equilibrium using an in-house chi square goodness of fit test in the HW program written by P. McGuffin and J. Williams (modified by Marian Hamshere). HAPLOVIEW, available for download at the website: http://www.genome.wi.mit.edu/personal/jcbarret /haploview, was used to analyse LD between markers.

Genotypic and allelic associations between SNPs and DD in the case-control sample were determined using standard contingency tables in a χ^2 test. Odds ratios and 95% confidence intervals were calculated using statistical packages available at http://home.clara.net/sisa/. Association in parent DD-proband trios was determined using the TDTphase program in the graphical interface (GLUE) of the statistical package UNPHASED (Dudbridge 2003). Haplotype analyses were undertaken in the case-control sample using COCAphase and in parent DD-proband trios using TDTphase, both in the statistical package UNPHASED (Dudbridge 2003).

Logistic and conditional logistic regression models used included both additive and dominance coding in order to assess both genotypic and allelic effects (see Table 4.2) (Cordell 2002).

Genotype	Allelic (Additive) Coding	Genotypic (Dominance) Coding
1 1	-1	-0.5
1 2	0	0.5
2 2	1	-0.5

Table 4.2 Additive and dominance coding used for stepwise logistic and conditional logistic regression analyses. Additive coding reflect allelic effects, dominance coding reflect genotypic effects.

Logistic and conditional logistic regression analyses were performed on case-control and trio data respectively using the Statistical Package for the Social Sciences (SPSS) software, version 12.0.1 for windows.

4.4.7 ALLELE SPECIFIC EXPRESSION

In order to investigate potential *cis*-acting influences on gene expression, the relative expression of paternal and maternal copies of genes was examined using RNA samples extracted from human brain tissue. SNP rs4504469 was used as a marker polymorphism. Sixty-six samples of RNA from brains were available to investigate allelic expression differences. Samples were *post-mortem* brain tissue samples derived from frontal, parietal and temporal cortex. The samples were from anonymous, unrelated adult individuals, principally from psychiatric control groups. Brain samples were obtained from the MRC London Neurodegenerative Diseases Brain Bank (United Kingdom), the Stanley Medical Research Institute Brain Bank (Bethesda, USA) and the Karolinska Institute (Stockholm, Sweden). For each individual, approximately 500mg of tissue was processed using standard phenol-chloroform procedures to extract gDNA. Approximately 300-500mg of tissue was processed for total RNA using RNAwiz isolation reagent (Ambion, Huntingdon, UK). Total RNA was treated with DNAase prior to reverse transcription with the RETROscript kit (Ambion, Huntingdon, UK). These steps were performed by Dr Nick Bray.

Samples derived from the brain samples were subjected to PCR amplification using primers (see Table 4.3) based on a single exonic sequence and capable of amplifying both gDNA and cDNA.

SNP	Forward PCR Primer (5' → 3')	Reverse PCR Primer (5' → 3')	Extension Primer (5' → 3')	Size (bp)	Annealing Temperature
rs4504469	catagtcttcctccgg caag	tgtcctgggagcagtgg tag	tcccaacacctc ccactagc	177	60

Table 4.3 Primers and PCR conditions for the PCR and SNaPshot reaction™ for analysis of allele specific expression. rs4504469 was used as the tag SNP for this assay.

Initially gDNA samples from all subjects were genotyped to identify heterozygotes for rs4504469. For the allele specific expression assay, utilising SNaPshot™ technology (see Chapter three), cDNA prepared from RNA samples taken from people heterozygous for the marker were assayed twice as two separate reverse transcription- (RT-) PCR reactions alongside the corresponding gDNA samples. In total, three samples were assayed per individual, a gDNA sample and two cDNA samples each reverse transcribed from RNA separately. Assays were not performed in duplicate. Assay conditions were the same for both cDNA and gDNA to enable the use of the average of the ratios observed from the gDNA to correct allelic ratios obtained from the cDNA analyses for any unequal representation of alleles specific to each assay (Bray et al. 2003). The relative expression of the alleles at rs4504469 in each sample was calculated by the ratio of the peak heights corrected by the use of the average gDNA ratio from the heterozygote samples as described for SNaPshot™ in Chapter three.

Due to the results of allele specific expression experiments by another research group (Paracchini 2005), the gDNA taken from the brain samples was

genotyped for the Francks and colleagues (Francks et al. 2004) risk haplotype using SNaPshot™.

In order to ensure that data did not derive from gDNA contaminating the cDNA, the RNA samples (used to create the cDNA) were amplified using PCR. Primers were designed to amplify exon 4 of the gDNA and cDNA. Since RNA is single stranded, the reverse primer will not bind and so no amplification should occur. Any contamination of the RNA with gDNA would result in the presence of a PCR product on an agarose gel. The absence of any PCR product in the RNA samples suggests no contaminating gDNA in the RNA samples. RNA derived from four brain samples was used to test for contamination, two samples with the Francks and colleagues (Francks et al. 2004) risk haplotype and two without. In addition, for each RNA the corresponding gDNA was amplified with the same reaction mix, parameters and conditions to check that the absence of contaminating gDNA of the RNA was not due to the failure of the PCR assay.

4.5 RESULTS

4.5.1 DIRECT ASSOCIATION ANALYSIS RESULTS: *SOX4*

Fifteen DNA fragments corresponding to *SOX4* were examined for the presence of variants using dHPLC and sequencing. One fragment showed a chromatograph consistent with heterozygosity in at least one PCR product for the fragments from the individuals. These were sequenced to characterise the variant. Two SNPs were found in the gene (see Table 4.4).

Location of SNP	Base Pair Position (Relative to ATG)	Base Pair Change	Amino Acid Change
Exon 1	842	C → A	Histidine → Asparagine
3' UTR	858	C → G	Leucine → Leucine

Table 4.4 SNPs identified in SOX4 using dHPLC and characterisation by sequencing.

Following sequencing of the DNA fragments containing the variants, the PCR amplification step failed to amplify the DNA. New primers were designed, different PCR cycling parameters, different temperature gradients for optimal primer annealing and the use of a GC-rich kit to overcome the problems of sequences rich in GC bases were attempted to reoptimise the PCR reaction. Despite attempts, the attempts reoptimise the PCR reaction failed.

4.5.2 DNA POOLING RESULTS

Allele frequency differences calculated in pools were validated by individually genotyping the individuals in the pools to allow actual allele frequency differences to be determined. Comparison of the differences showed a very small error rate in the determination of allele frequency differences in DNA pools (see Table 4.5). It should be noted that this step forms a validation step of the DNA pools.

rs4504469	Allele Frequency in Cases			Allele Frequency in Controls		
Genotyping	**Individual**	**Pooled**	**Difference**	**Individual**	**Pooled**	**Difference**
Allele 1	0.66	0.64	0.02	0.57	0.57	0.00
Allele 2	0.34	0.36		0.43	0.43	

	Individual	Pooled	Pooling Error
Difference Between Allele 1 in Cases and Controls	0.09	0.07	0.02

Table 4.5 Calculation of the accuracy of DNA pool allele frequency estimation compared to individual genotyping. It should be noted that the estimation of error is inexact because some DNA used in the pools was unavailable for individual genotyping in the case of DD cases, resulting in fewer genotypes in this sample. Estimates also assume individual genotyping is 100% accurate.

One hundred and twenty-six SNPs were selected across the positional candidate genes and tested for association with DD using a range of correction factors for *k*. Of these SNPs, 80 were assayed using an age-matched control sample (start point B, Figure 4.11) and 46 were assayed using a sample of blood donor controls (start point A, Figure 4.11). Seventeen of the SNPs showing association to DD using the blood donor control sample revealed some evidence of association with DD (with empirically derived values of correction factor *k*). Consequently these were assayed using the pooling strategy with age-matched controls (start point B, Figure 4.11).

Twenty-five SNPs assayed using the age-matched controls showed some evidence of association with DD. These SNPs were then re-assayed using the pooling strategy, with pools being corrected for unequal allelic representation using a known heterozygote allelic ratio. In addition, 42 extra SNPs were assayed and tested for association with DD from this point, having been chosen from reports of association with DD that were emerging as this work was being undertaken. In total 168 SNPs were tested for association with DD using the DNA pooling strategy. Of these SNPs, 19 revealed association with DD after correction using a known heterozygote ratio. The results from the DNA pooling assays are shown in Tables 4.6 and 4.7. The tables show the SNPs tested for association in column two. The gene in which the SNP lies, the base change, position (Mb) and the position within each gene, are shown in columns one, three, four and six respectively. The distance between each SNP is shown in column five. For each SNP the frequency of the major allele is shown for the cases and controls. The major allele corresponds to the first allele in the column containing the base change of each SNP (column two). For each test of association the allelic p-value and chi-square (χ^2) are shown. Indicated

in the final column is the control sample utilised (see Figure 4.11). For the 17 SNPs showing association with DD using the blood-donor control sample and that were re-analysed for association with DD using the age-matched controls, the p-value for the association test using the age-matched controls is shown.

Gene	rs Number	Base Change (Major Allele /Minor Allele)	Position	Distance Between SNPs	Position in Gene	Case Frequency Allele 1	Control Frequency Allele 1	χ^2	p	Control Sample
ID4	1980461	G/C	19942988		5' Flank	0.71	0.71	0	0.99	Age
ID4	3193769	G/C	19947313	4325	Intron	0.71	0.71	0	0.95	Age
PRL	1205961	T/C	22393991	2446678	3'UTR	0.5	0.51	0.05	0.82	Age
PRL	6239	C/T	22395724	1733	Exon	0.82	0.86	2.96	0.09	Age
PRL	1205957	G/A	22397119	1395	Intron	0.5	0.46	0.89	0.35	Donor
PRL	849886	C/T	22399346	2227	Intron	0.54	0.59	3.05	0.08	Age
PRL	1205954	G/A	22401833	2487	Intron	0.55	0.54	0.03	0.85	Age
PRL*	2244502	T/A	22402966	1133	Intron	0.55	0.46	9.06	**0.003**	Age
PRL	6237	C/T	22405134	2168	Intron	0.86	0.86	0.08	0.78	Age
PRL	849876	T/C	22407617	2483	5' Flank	0.76	0.73	1.46	0.23	Age
PRL	849875	G/A	22407979	362	5' Flank	0.59	0.59	0.02	0.88	Donor
VMP	12208318	G/C	24243823	1835844	Intron	0.73	0.75	0.59	0.44	Age
VMP	10946676	C/T	24252846	9023	Intron	0.87	0.86	0.04	0.85	Age
DCDC2	1277155	C/T	24276123	23277	3' UTR	0.97	0.96	0.5	0.48	Age
DCDC2*	1832709	A/C	24280941	4818	3' UTR	0.6	0.53	5.88	**0.02**	Age
DCDC2	3789219	T/C	24283103	2162	Intron	0.54	0.55	0.2	0.65	Age
DCDC2	1419228	T/C	24286285	3182	Intron	0.66	0.62	1.53	0.22	Donor
DCDC2	2996452	C/T	24288345	2060	Intron	0.61	0.61	0.001	0.97	Donor
DCDC2	1277192	C/A	24308388	20043	Intron	0.6	0.62	0.41	0.52	Age
DCDC2	1277194	C/A	24311703	3315	Intron	0.56	0.57	0.21	0.65	Age
DCDC2	793861	A/T	24314595	2892	Intron	0.53	0.54	0.25	0.62	Age
DCDC2	793862	T/C	24315179	584	Intron	0.56	0.57	0.07	0.79	Age
DCDC2	870601	C/T	24369039	53860	Intron	0.66	0.65	0.07	0.79	Age
DCDC2	807698	C/T	24384232	15193	Intron	0.51	0.56	2.16	0.14	Donor
DCDC2	807726	T/C	24385979	1747	Intron	0.5	0.53	1.08	0.2	Donor

Table 4.6 Continued …

Gene	rs Number	Base Change (Major Allele /Minor Allele)	Position	Distance Between SNPs	Position in Gene	Case Frequency Allele 1	Control Frequency Allele 1	χ^2	p	Control Sample
DCDC2*	3789224	C/T	24388080	2101	Intron	0.62	0.7	8.12	**0.004**	Age
DCDC2	3789227	C/T	24395508	7428	Intron	0.57	0.54	0.83	0.36	Donor
DCDC2	2296539	G/A	24397431	1923	Intron	0.65	0.62	1.25	0.26	Donor
DCDC2	2274305	C/T	24399182	1751	Exon	0.6	0.59	0.25	0.62	Age
DCDC2*	6907864	C/T	24401303	2121	Intron	0.6	0.53	4.37	**0.04**	Age
DCDC2	807709	T/C	24405282	3979	Intron	0.59	0.58	0.13	0.72	Age
DCDC2	807704	A/G	24408825	3543	Intron	0.71	0.73	0.27	0.61	Age
DCDC2	807703	T/C	24410066	1241	Intron	0.97	0.99	2.69	0.1	Age
DCDC2	3857541	T/C	24410512	446	Intron	0.84	0.82	0.55	0.46	Age
DCDC2	707862	G/A	24412730	2218	Intron	0.78	0.79	0.35	0.55	Age
DCDC2	807685	T/A	24418623	5893	Intron	0.8	0.82	0.82	0.37	Age
DCDC2	793704	G/C	24444623	26000	Intron	0.56	0.57	0.11	0.74	Age
DCDC2*	793722	T/C	24460342	15719	Intron	0.55	0.49	4.67	**0.03**	Age
DCDC2	793720	T/C	24461259	917	Intron	0.5	0.45	2.51	0.11	Donor
DCDC2	1277350	G/A	24463129	1870	Intron	0.5	0.44	3.67	0.06	Age
DCDC2*	1277349	C/G	24466462	3333	5'Flank	0.76	0.65	17.38	**0.00003**	Age
DCDC2	2792666	C/G	24469168	2706	5'Flank	0.57	0.54	0.64	0.43	Donor
DCDC2	793663	C/T	24477494	8326	5'Flank	0.68	0.64	1.9	0.17	Age
MRS2L	9393553	A/G	24509633	32139	5'Flank	0.72	0.75	1.05	0.3	Age
KIAA0319	2817241	C/T	24650242	140609	3' Flank	0.54	0.55	0.02	0.88	Donor
KIAA0319	807526	T/C	24651699	1457	3' UTR	0.63	0.62	0.14	0.71	Donor
KIAA0319*	2817243	A/G	24653575	1876	3' UTR	0.53	0.61	7.98	**0.005**	Age
KIAA0319	2817245	A/G	24655585	2010	Intron	0.54	0.58	1.45	0.23	Donor
KIAA0319	807532	G/A	24657647	2062	Intron	0.87	0.87	0.006	0.94	Donor
KIAA0319	2076313	G/A	24659603	1956	Intron	0.54	0.5	1.53	0.22	Age

Table 4.6 Continued ...

Gene	rs Number	Base Change (Major Allele /Minor Allele)	Position	Distance Between SNPs	Position in Gene	Case Frequency Allele 1	Control Frequency Allele 1	χ^2	p	Control Sample
KIAA0319	807536	A/C	24662253	2650	Intron	0.59	0.57	0.53	0.47	Age
KIAA0319	4083411	A/G	24664281	2028	Intron	0.58	0.55	1.18	0.28	Donor
KIAA0319	2760167	T/A	24666498	2217	Intron	0.53	0.49	2.14	0.14	Age
KIAA0319	807542	A/G	24668312	1814	Intron	0.53	0.55	0.2	0.66	Donor
KIAA0319	807544	A/C	24670281	1969	Intron	0.7	0.7	0.02	0.88	Donor
KIAA0319	2744549	A/G	24671531	1250	Intron	0.85	0.84	0.005	0.95	Donor
KIAA0319	2760161	A/G	24673389	1858	Intron	0.56	0.51	1.87	0.17	Donor
KIAA0319	2817195	T/C	24678323	4934	Intron	0.97	0.97	0.1	0.75	Donor
KIAA0319	807521	G/T	24678410	87	Intron	0.78	0.81	0.89	0.35	Donor
KIAA0319	3846835	A/C	24680310	1900	Intron	0.8	0.83	1.53	0.22	Age
KIAA0319	2744556	C/G	24682096	1786	Intron	0.57	0.56	0.08	0.77	Donor
KIAA0319	807525	C/T	24684016	1920	Intron	0.51	0.52	0.003	0.96	Donor
KIAA0319	2817199	A/G	24684227	211	Intron	0.53	0.5	0.79	0.38	Donor
KIAA0319	2760158	T/A	24686016	1789	Intron	0.52	0.52	0	1	Donor
KIAA0319	2760157	T/C	24686251	235	Intron	0.58	0.55	0.83	0.36	Donor
KIAA0319	807507	C/G	24687846	1595	Intron	0.57	0.52	2.16	0.14	Donor
KIAA0319	807509	A/G	24690011	2165	Intron	0.53	0.54	0.26	0.61	Donor
KIAA0319	2817201	A/C	24693193	3182	Intron	0.54	0.59	2.18	0.14	Donor
KIAA0319*	4504469	G/A	24696863	3670	Exon	0.52	0.42	11.68	**0.0006**	Age
KIAA0319	5026394	T/G	24698526	1663	Intron	0.53	0.57	2.1	0.15	Age
KIAA0319*	4576240	G/T	24704457	5931	Exon	0.72	0.78	5.73	**0.02**	Age
KIAA0319	4352670	A/C	24712047	7590	Intron	0.75	0.79	1.45	0.23	Age
KIAA0319	4712831	G/A	24713548	1501	Intron	0.68	0.72	2.09	0.15	Age
KIAA0319	4236032	T/G	24714706	1158	Intron	0.52	0.57	2.5	0.11	Age
KIAA0319*	6911855	C/T	24715290	584	Intron	0.7	0.81	16.5	**0.00005**	Age
KIAA0319	4712833	T/C	24719122	3832	Intron	0.58	0.62	2.16	0.14	Age

Table 4.6 Continued ...

Gene	rs Number	Base Change (Major Allele /Minor Allele)	Position	Distance Between SNPs	Position in Gene	Case Frequency Allele 1	Control Frequency Allele 1	χ^2	p	Control Sample
KIAA0319*	6939068	T/A	24721274	2152	Intron	0.69	0.79	16.29	**0.00005**	Age
KIAA0319	7763790	A/G	24723042	1768	Intron	0.51	0.46	2.57	0.11	Age
KIAA0319	2745335	T/C	24725393	2351	Intron	0.8	0.76	1.79	0.18	Age
KIAA0319	2745334	C/T	24725509	116	Intron	0.6	0.65	3.23	0.07	Age
KIAA0319	2817206	A/T	24728381	2872	Intron	0.74	0.7	1.86	0.17	Age
KIAA0319*	7751357	G/C	24733377	4996	Intron	0.56	0.49	6.47	**0.01**	Age
KIAA0319*	6917660	G/A	24737085	3708	Intron	0.56	0.48	7.84	**0.005**	Age
KIAA0319*	6456622	G/A	24739537	2452	Intron	0.57	0.51	4.67	**0.03**	Age
KIAA0319	7766230	G/A	24741408	1871	Intron	0.72	0.75	1.16	0.28	Age
KIAA0319*	2206525	T/C	24744882	3474	Intron	0.56	0.62	3.85	**0.05**	Age
KIAA0319	7755563	A/G	24746690	1808	Intron	0.54	0.49	22.68	0.1	Age
KIAA0319*	7755579	A/C	24746719	29	Intron	0.55	0.48	5.95	**0.01**	Age
KIAA0319*	6935076	A/G	24752301	5582	Intron	0.53	0.45	6.15	**0.01**	Age
KIAA0319	4363021	C/T	24753130	829	Intron	0.89	0.88	0.08	0.78	Age
KIAA0319*	2038137	G/T	24753922	792	Intron	0.57	0.47	10.1	**0.001**	Age
KIAA0319	3756821	C/T	24754800	878	5' Flank	0.68	0.67	0.01	0.92	Age
KIAA0319	1555089	A/G	24756385	1585	5' Flank	0.54	0.49	2.47	0.12	Age
TTRAP*	3212236	T/C	24756434	49	3' UTR	0.55	0.61	5.04	**0.03**	Age
TTRAP	3087943	T/C	24758740	2306	Exon	0.53	0.59	3.68	0.06	Age
TTRAP*	2294691	A/G	24760822	2082	Intron	0.78	0.73	4.06	**0.04**	Age
TTRAP	3181238	C/T	24762422	1600	Intron	0.63	0.63	0.005	0.94	Donor
TTRAP	3212234	C/G	24763930	1508	Intron	0.51	0.56	2.74	0.09	Age
TTRAP*	3212232	T/C	24765876	1946	Intron	0.53	0.47	3.94	**0.05**	Age
TTRAP	2143340	A/G	24767050	1174	Intron	0.55	0.54	0.17	0.68	Age
TTRAP	2056999	G/A	24769490	2440	Intron	0.6	0.57	0.6	0.44	Donor
TTRAP	3756819	G/T	24773319	3829	Intron	0.59	0.56	0.97	0.33	Donor

Table 4.6 Continued ...

Gene	rs Number	Base Change (Major Allele /Minor Allele)	Position	Distance Between SNPs	Position in Gene	Case Frequency Allele 1	Control Frequency Allele 1	χ^2	p	Control Sample
TTRAP	1061925	A/C	24773986	667	Intron	0.64	0.67	0.54	0.46	Donor
TTRAP	3756815	G/C	24774173	187	Intron	0.87	0.91	2.79	0.1	Donor
THEM2*	3181227	T/C	24775778	1605	Intron	0.58	0.52	4.33	**0.04**	Age
THEM2	2223589	C/G	24777451	1673	5' UTR	0.63	0.62	0.07	0.79	Donor
THEM2	9393576	C/T	24777924	473	Intron	0.63	0.62	0.12	0.73	Donor
THEM2	7765052	G/A	24781599	3675	Intron	0.71	0.74	1.15	0.28	Age
THEM2	7451561	T/C	24785967	4368	Intron	0.51	0.47	1.53	0.22	Age
THEM2	2143338	A/G	24791260	5293	Intron	0.52	0.48	1.85	0.17	Donor
THEM2	1555086	C/T	24793000	1740	Intron	0.53	0.5	0.98	0.32	Donor
THEM2	1205120	T/G	24793537	537	Intron	0.76	0.79	1.27	0.26	Donor
THEM2*	926529	G/A	24795744	2207	Intron	0.59	0.47	17.04	**0.00004**	Age
THEM2	1885209	G/A	24796937	1193	Intron	0.71	0.67	1.36	0.24	Donor
THEM2	1885211	G/A	24800158	3221	Intron	0.74	0.74	0.004	0.95	Donor
THEM2*	3777664	A/G	24801825	1667	Intron	0.57	0.51	4.42	**0.04**	Age
THEM2	2092404	A/C	24805367	3542	Intron	0.57	0.56	0.12	0.72	Donor
THEM2	3777663	T/C	24808214	2847	Intron	0.52	0.49	0.99	0.32	Age
THEM2	1056319	A/C	24811025	2811	3'Flank	0.89	0.88	0.13	0.72	Donor
THEM2*	1053598	C/T	24812517	1492	3'Flank	0.58	0.49	9.42	**0.002**	Age
C6orf62	3756814	T/C	24813814	1297	3'UTR	0.5	0.48	0.36	0.55	Donor
C6orf62	2294686	A/G	24817181	3367	Intron	0.79	0.83	2.73	0.09	Age
C6orf62	1065364	A/G	24818379	1198	Intron	0.94	0.94	0.004	0.95	Donor
C6orf62	6913673	A/C	24821123	2744	Intron	0.86	0.87	0.03	0.86	Age
C6orf62	3813687	G/A	24827957	6834	5'Flank	0.57	0.52	3.3	0.07	Age
C6orf62	1923187	A/T	24829646	1689	5'Flank	0.54	0.53	0.09	0.76	Donor

Table 4.6 Results of DNA pooling assays. All p-values represent the best possible p-value obtained using a range of correction factors, *k*.
* These pools were later repeated in DNA pools using a known heterozygote ratio for correction factor *k*

Gene	rs Number	Allele 1	Allele 2	Case Frequency Allele 1	Case Frequency Allele 2	Control Frequency Allele 1	Control Frequency Allele 2	χ^2	p
PRL	2244502#	A	T	0.26	0.74	0.32	0.68	5.58	**0.02**
VMP	9393529	G	A	0.9	0.1	0.91	0.09	0.69	0.41
VMP	2876666	A	G	0.65	0.35	0.69	0.31	2.22	0.14
VMP	9356928	G	A	0.49	0.51	0.53	0.47	1.79	0.18
VMP	7455023	A	G	0.54	0.46	0.58	0.42	1.58	0.21
VMP	12202381	A	G	0.88	0.12	0.88	0.12	0.06	0.81
VMP	10946675	G	A	0.53	0.47	0.49	0.51	1.66	0.20
VMP	1053047	G	A	0.49	0.51	0.52	0.48	1.29	0.26
VMP	C_9373644_10*	G	A	0.57	0.43	0.52	0.48	2.03	0.15
DCDC2	2792666	G	C	0.43	0.57	0.47	0.53	1.63	0.20
DCDC2	1832709#	A	C	0.15	0.85	0.13	0.87	1.62	0.20
DCDC2	1277349#	G	C	0.04	0.96	0.05	0.95	1.59	0.21
DCDC2	793720	T	C	0.69	0.31	0.63	0.37	3.21	0.07
DCDC2	793722#	T	C	0.69	0.31	0.63	0.37	5.00	**0.03**
DCDC2	707862	G	A	0.93	0.07	0.93	0.07	0.01	0.93
DCDC2	807704	G	A	0.08	0.92	0.07	0.93	0.34	0.56
DCDC2	6907864#	T	C	0.09	0.91	0.13	0.87	3.37	0.07
DCDC2	3789224#	T	C	0.15	0.85	0.13	0.87	0.76	0.38
DCDC2	793862	T	C	0.26	0.74	0.29	0.71	0.81	0.37
DCDC2	870601	C	T	0.66	0.34	0.65	0.35	0.15	0.7
DCDC2	793704	G	C	0.44	0.56	0.43	0.57	0.09	0.8
DCDC2	793663	G	A	0.68	0.32	0.64	0.36	1.85	0.17
MRS2L	2273606	A	G	0.85	0.15	0.84	0.16	0.33	0.57
MRS2L	7769012	C	A	0.66	0.34	0.64	0.36	0.34	0.56

Table 4.7 Continued ...

Gene	rs Number	Allele 1	Allele 2	Case Frequency Allele 1	Case Frequency Allele 2	Control Frequency Allele 1	Control Frequency Allele 2	χ^2	p
MRS2L	2793422	G	A	0.67	0.33	0.61	0.39	4.10	**0.04**
MRS2L	7738943	G	C	0.88	0.12	0.92	0.08	3.07	0.08
MRS2L	13735	C	T	0.61	0.39	0.65	0.35	1.78	0.18
MRS2L	2295650	RARE							
MRS2L	2273606	C	T	0.15	0.85	0.16	0.84	0.33	0.57
MRS2L	1277347	RARE							
MRS2L	1298764	RARE							
ALDH5A1	2817220	C	T	0.95	0.05	0.95	0.05	0.02	0.91
KIAA0319	4504469#	G	A	0.64	0.36	0.57	0.43	6.92	**0.009**
KIAA0319	6939068#	A	T	0.81	0.19	0.71	0.29	12.45	**0.0004**
KIAA0319	2817206	A	T	0.92	0.08	0.88	0.12	3.70	0.06
KIAA0319	7751357#	G	C	0.69	0.31	0.62	0.38	6.30	**0.012**
KIAA0319	6917660#	G	A	0.68	0.32	0.63	0.37	3.03	0.08
KIAA0319	6456622#	G	A	0.68	0.32	0.61	0.39	5.78	**0.02**
KIAA0319	2206525#	T	C	0.31	0.69	0.36	0.64	3.17	0.08
KIAA0319	7755579#	A	C	0.67	0.33	0.58	0.42	11.34	**0.001**
KIAA0319	6935076#	G	A	0.56	0.44	0.64	0.36	7.47	**0.006**
KIAA0319	4576240#	G	T	0.87	0.13	0.9	0.1	1.54	0.21
KIAA0319	6911855#	T	C	0.78	0.22	0.68	0.32	12.19	**0.0005**
KIAA0319	2817243#	G	A	0.11	0.89	0.08	0.92	3.11	0.08
KIAA0319	2817200	T	C	0.65	0.35	0.61	0.39	2.32	0.13
KIAA0319	2038137#	G	T	0.61	0.39	0.54	0.46	5.89	**0.02**
KIAA0319	699463	C	T	0.73	0.27	0.72	0.28	0.004	0.95
KIAA0319	807540	G	A	0.16	0.84	0.13	0.87	3.16	0.08
KIAA0319	2179515	C	T	0.66	0.34	0.58	0.42	7.52	**0.006**
KIAA0319	9358783	T	G	0.66	0.34	0.6	0.4	3.88	**0.05**

Table 4.7 Continued ...

Gene	rs Number	Allele 1	Allele 2	Case Frequency Allele 1	Case Frequency Allele 2	Control Frequency Allele 1	Control Frequency Allele 2	χ^2	p
KIAA0319	9358784	G	A	0.33	0.67	0.39	0.61	4.22	**0.04**
KIAA0319	6456624	C	T	0.66	0.34	0.6	0.4	4.96	**0.03**
KIAA0319	2235677	G	A	0.69	0.31	0.65	0.35	2.1	0.15
KIAA0319	2235676	G	T	0.82	0.18	0.84	0.16	0.34	0.56
KIAA0319	9467247	A	C	0.2	0.8	0.19	0.81	0.33	0.57
KIAA0319	1555090	G	C	0.33	0.67	0.4	0.6	6.53	**0.01**
TTRAP	2143340	G	A	0.17	0.83	0.17	0.83	0.03	0.86
TTRAP	3212232#	T	C	0.75	0.25	0.76	0.24	0.38	0.54
TTRAP	2294691#	G	A	0.12	0.88	0.09	0.91	2.50	0.11
TTRAP	3212236#	T	C	0.81	0.19	0.83	0.17	0.44	0.51
TTRAP	3033236	G	T	0.16	0.84	0.17	0.83	0.11	0.74
THEM2	926529#	G	A	0.73	0.27	0.66	0.34	7.14	**0.008**
THEM2	3777664#	G	A	0.28	0.72	0.35	0.65	5.73	**0.02**
THEM2	3181227#	T	C	0.75	0.25	0.77	0.23	1.00	0.32
INTERGENIC	1053598#	T	C	0.27	0.73	0.32	0.68	3.97	**0.05**
C6orf62	6456632	C	T	0.72	0.28	0.75	0.25	1.06	0.3
INTERGENIC	1419229	G	A	0.49	0.51	0.52	0.48	0.79	0.38

Table 4.7 DNA pooling results for assays with a known heterozygote ratio for correction factor *k*.

* Assay-on-Demand ID. # SNPs showing association in previous DNA pooling analysis (using a range of correction factors)

Within the pooled DNA SNP analysis, 27 SNPs could not be genotyped in a SNaPshot assay or were found not to be polymorphic in our sample. In each case, a new SNP was chosen from the databases in order to fill the gaps left by these SNPs and are included in pooled DNA results tables.

4.5.3 MARKER-MARKER LINKAGE DISEQUILIBRIUM ANALYSES

The 19 SNPs showing association to DD in the DNA pools were genotyped in 42 DD cases and 48 controls in order that LD between the SNPs could be calculated (see Tables 4.8 and 4.9). SNP rs2143340 was also included in analysis. Although it did not show association in the DNA pool analysis, it had previously been shown to contribute to a significant haplotype (Francks et al. 2004). SNPs were chosen for individual genotyping if they did not show LD with any other SNP ($r^2 < 0.8$). Where two SNPs were in LD only one of the SNPs was individually genotyped. A set of 13 SNPs was chosen for individual genotyping in the case-control sample (see Table 4.10).

4.5.4 INDIVIDUAL GENOTYPING RESULTS

Individual genotyping results can be seen in Table 4.10 and 4.11. The most significant results showing association to DD with a p-value ≤ 0.01 in the case-control sample were rs2793422 (*MRS2L*), rs4504469, rs2179515, rs6935076, rs2038137 (*KIAA0319*), rs926529 and rs3777664 (*THEM2*) (see Table 4.10). Other SNPs showing association with a p-value between 0.01 and 0.05 were in the gene *PRL* and an intergenic SNP between genes *THEM2* and *TTRAP*. In order to check that the results were not due to population stratification, 146 parent DD-proband trios were genotyped (see Table 4.11).

Markers	rs2244502	rs793722	rs2793422	rs4504469	rs6911855	rs6939068	rs7751357	rs2179515	rs6456622	rs9358783	rs9358784	rs7755579	rs6456624	rs6935076	rs2038137	rs1555090	rs2143340	rs926529	rs3777664	rs1053598
rs2244502	1.00	0.29	0.04	0.17	0.21	0.43	0.10	0.11	0.15	0.22	0.16	0.13	0.21	0.03	0.14	0.16	0.06	0.06	0.06	0.03
rs793722	0.37	1.00	0.05	0.01	0.03	1.00	0.12	0.08	0.08	0.12	0.07	0.03	0.04	0.06	0.05	0.07	0.16	0.00	0.46	0.10
rs2793422	0.10	0.24	1.00	0.18	1.00	0.35	0.19	0.15	0.19	0.27	0.17	0.06	0.03	0.05	0.05	0.17	0.08	0.03	0.16	0.14
rs4504469	0.20	0.09	0.28	1.00	1.00	0.17	0.84	0.85	0.88	0.88	0.89	0.68	0.70	0.62	0.71	0.89	0.58	0.42	0.49	0.58
rs6911855	1.00	1.00	0.22	0.49	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
rs6939068	1.00	1.00	0.54	1.00	1.00	1.00	0.06	0.00	0.16	0.30	0.08	0.25	0.18	0.95	0.09	0.08	1.00	0.05	0.27	0.45
rs7751357	0.13	0.40	0.39	0.78	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.29	0.72	0.78	0.78
rs2179515	0.20	0.40	0.39	0.78	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.20	0.69	0.79	0.79
rs6456622	0.20	0.40	0.43	0.79	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.29	0.67	0.76	0.78
rs9358783	0.18	0.38	0.39	0.77	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.35	0.66	0.74	0.74
rs9358784	0.13	0.43	0.39	0.78	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.27	0.68	0.77	0.78
rs7755579	0.08	0.22	0.17	0.71	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.06	0.90	0.63	0.81
rs6456624	0.08	0.21	0.17	0.72	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.19	0.85	0.61	0.75
rs6935076	0.36	0.42	0.36	0.50	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.71	1.00
rs2038137	0.08	0.31	0.17	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.05	0.76	0.69	0.77
rs1555090	0.13	0.37	0.36	0.77	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.27	0.68	0.77	0.78
rs2143340	1.00	0.14	0.05	0.01	0.14	0.28	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.73	1.00	1.00	1.00	1.00	1.00	0.00
rs926529	0.08	0.09	0.04	0.44	1.00	1.00	0.65	0.57	0.57	0.59	0.57	0.78	0.78	1.00	0.68	0.74	1.00	1.00	0.76	0.95
rs3777664	0.13	0.14	0.30	0.48	1.00	1.00	0.50	0.50	0.57	0.53	0.50	0.58	0.58	0.49	0.60	0.55	1.00	0.90	1.00	1.00
rs1053598	0.13	0.19	0.07	0.52	0.30	0.12	0.44	0.44	0.50	0.51	0.44	0.51	0.51	0.77	0.53	0.48	1.00	0.90	0.91	1.00

Table 4.8 LD based on D' between SNPs individually genotyped in 42 cases and 48 controls. LD was assessed separately for cases and controls.

LD between cases: below the diagonal

LD between controls: above the diagonal

Markers	rs2244502	rs793722	rs2793422	rs4504469	rs6911855	rs6939068	rs7751357	rs2179515	rs6456622	rs9358783	rs9358784	rs7755579	rs6456624	rs6935076	rs2038137	rs1555090	rs2143340	rs926529	rs3777664	rs1053598
rs2244502	1.00	0.08	0.00	0.01	0.00	0.02	0.01	0.01	0.02	0.03	0.02	0.01	0.02	0.00	0.01	0.02	0.00	0.00	0.00	0.00
rs793722	0.03	1.00	0.00	0.00	0.00	0.03	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.01
rs2793422	0.00	0.02	1.00	0.02	0.02	0.01	0.02	0.01	0.02	0.03	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.02	0.01
rs4504469	0.03	0.00	0.02	1.00	0.02	0.00	0.61	0.63	0.63	0.54	0.65	0.38	0.44	0.16	0.50	0.65	0.05	0.18	0.24	0.28
rs6911855	0.01	0.09	0.00	0.01	1.00	0.38	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.05	0.02	0.02	0.00	0.02	0.02	0.02
rs6939068	0.02	0.11	0.04	0.03	0.65	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.01	0.00	0.01	0.02
rs7751357	0.02	0.07	0.03	0.36	0.01	0.02	1.00	1.00	1.00	0.79	1.00	0.73	0.80	0.35	0.87	1.00	0.02	0.44	0.50	0.58
rs2179515	0.03	0.07	0.03	0.36	0.01	0.02	1.00	1.00	1.00	0.80	1.00	0.75	0.81	0.37	0.88	1.00	0.01	0.44	0.52	0.57
rs6456622	0.03	0.07	0.03	0.40	0.01	0.02	1.00	1.00	1.00	1.00	1.00	0.75	0.83	0.36	0.91	1.00	0.02	0.43	0.47	0.58
rs9358783	0.03	0.06	0.03	0.32	0.01	0.02	0.92	0.92	0.92	1.00	0.96	0.60	0.65	0.29	0.70	0.96	0.02	0.34	0.37	0.53
rs9358784	0.02	0.09	0.03	0.36	0.01	0.02	1.00	1.00	1.00	0.92	1.00	0.73	0.80	0.36	0.87	1.00	0.01	0.42	0.49	0.55
rs7755579	0.01	0.03	0.01	0.35	0.01	0.02	0.92	0.85	0.85	0.78	0.85	1.00	0.92	0.50	0.88	0.73	0.00	0.65	0.35	0.46
rs6456624	0.01	0.02	0.01	0.36	0.01	0.02	0.92	0.85	0.85	0.78	0.85	1.00	1.00	0.45	0.92	0.80	0.00	0.64	0.35	0.42
rs6935076	0.04	0.11	0.08	0.12	0.05	0.00	0.31	0.27	0.27	0.26	0.27	0.32	0.33	1.00	0.43	0.36	0.06	0.41	0.22	0.35
rs2038137	0.01	0.05	0.01	0.41	0.01	0.02	0.92	0.92	0.92	0.85	0.92	0.92	0.92	0.30	1.00	0.87	0.00	0.58	0.44	0.47
rs1555090	0.02	0.07	0.02	0.37	0.01	0.02	1.00	1.00	1.00	0.91	1.00	0.92	0.92	0.34	0.92	1.00	0.01	0.42	0.49	0.55
rs2143340	0.07	0.00	0.00	0.00	0.00	0.03	0.06	0.06	0.06	0.05	0.06	0.07	0.06	0.12	0.06	0.07	1.00	0.08	0.09	0.00
rs926529	0.00	0.00	0.00	0.10	0.01	0.02	0.33	0.28	0.28	0.32	0.28	0.43	0.43	0.23	0.35	0.37	0.04	1.00	0.55	0.71
rs3777664	0.01	0.01	0.01	0.13	0.01	0.02	0.23	0.23	0.28	0.28	0.23	0.26	0.27	0.06	0.30	0.25	0.05	0.74	1.00	0.76
rs1053598	0.01	0.02	0.00	0.16	0.00	0.00	0.20	0.20	0.23	0.24	0.20	0.22	0.22	0.17	0.26	0.21	0.05	0.68	0.75	1.00

Table 4.9 LD based on r^2 between SNPs individually genotyped in 42 cases and 48 controls. LD was assessed separately for cases and controls.

LD between cases: below the diagonal

LD between controls: above the diagonal

Two of the SNPs did not show association after analysis in the parent DD-proband trios, these included rs2244502 (*PRL*) and rs2038137 (*KIAA0319*) (see Table 4.11). Since the control (nontransmitted) alleles are independent of the controls in the case-control study, the analysis of SNPs using the parent DD-proband trios provides semi-independent replication (Plomin et al. 2004). Seven SNPs showed evidence of association to DD in the parent DD-proband trios. These are located in *MRS2L* (1 SNP), *KIAA0319* (3 SNPs) and within or flanking *THEM2* (3 SNPs).

4.5.5 HARDY-WEINBERG EQUILIBRIUM ANALYSES

With the exception of rs2038137 and rs6939076, all genotypes were in Hardy-Weinberg equilibrium in probands and in parents. SNP rs2038137, showed slight distortion in the controls ($p = 0.03$), consequently rs2038137 was regenotyped both in the case-control sample and the CEPH trios. All genotypes matched previous calls and the CEPH genotypes in our study matched those available from the HapMap project suggesting the genotyping assay was accurate. Parent proband trios revealed that all genotypes Mendelised as did the parent DD-probands used in our sample. LD analysis showed that rs2038137 was in strong LD with rs2179515 ($r^2 = 0.89$) suggesting that association would be picked up by rs2179515.

The cases of the case-control sample showed distortion from Hardy-Weinberg equilibrium at rs6939076. This may reflect genotypic association to DD.

Gene	rs Number	Sample	No. 11 Genotype	No. 12 Genotype	No. 22 Genotype	χ^2	p	No. Allele 1	% Allele 1	χ^2	p	OR	LOWER 95% CI	UPPER 95% CI
PRL	2244502*	CASE	14	75	112	5.71	0.06	103	26	4.87	0.03	1.39	1.04	1.86
		CONTROL	23	117	112			163	32					
DCDC2	793722	CASE	22	93	97	3.39	0.18	137	32	1.96	0.16	1.21	0.93	1.59
		CONTROL	42	106	111			190	37					
MRS2L	2793422*	CASE	99	94	22	6.89	0.03	304	69	8.65	0.003	1.5	1.14	1.96
		CONTROL	91	117	43			299	60					
KIAA0319	4504469*	CASE	101	117	22	16.33	0.002	319	66	9.86	0.002	1.51	1.17	1.95
		CONTROL	88	124	2			300	57					
KIAA0319	6911855	CASE	200	17	1	3.52	0.17	417	96	3.38	0.07	0.51	0.24	1.06
		CONTROL	253	12	0			518	98					
KIAA0319	6939068	CASE	180	19	1	3.70	0.16	379	95	3.48	0.06	0.52	0.26	1.04
		CONTROL	234	14	0			482	97					
KIAA0319	2179515*	CASE	116	100	16	9.56	0.008	332	72	6.03	0.007	1.45	1.11	1.9
		CONTROL	109	108	40			326	63					
KIAA0319	6935076*	CASE	65	131	35	10.16	0.006	261	56	7.55	0.006	0.7	0.54	0.9
		CONTROL	107	118	30			332	65					
KIAA0319	2038137*	CASE	112	104	13	17.07	0.0001	328	72	10.73	0.001	1.57	1.2	2.05
		CONTROL	106	105	46			317	62					
TTRAP	2143340	CASE	140	56	7	2.68	0.26	336	83	1.00	0.32	0.83	0.58	1.19
		CONTROL	179	68	3			426	85					
THEM2	926529*	CASE	120	81	9	11.46	0.0032	321	76	9.22	0.0024	0.64	0.47	0.85
		CONTROL	118	98	32			334	67					
THEM2	3777664*	CASE	119	92	13	7.63	0.02	330	74	6.93	0.008	1.45	1.1	1.92
		CONTROL	112	113	31			337	66					
INTERGENIC	1053598*	CASE	124	92	9	5.86	0.05	340	76	4.57	0.03	1.36	1.03	1.81
		CONTROL	123	112	23			358	69					

Table 4.10 Individual genotyping results in the case-control sample. *These include 25 extra DD probands from the parent DD-proband trios.

Gene	rs Number	Transmission	No. of Allele 1	Percentage of Allele 1	χ^2	p	OR*	LOWER 95% CI*	UPPER 95% CI*
PRL	2244502	TRANSMITTED	63	25	0.82	0.36	0.83	0.56	1.24
		NONTRANSMITTED	72	28					
MRS2L	2793422	TRANSMITTED	191	71	4.4	0.04	1.47	1.02	2.1
		NONTRANSMITTED	168	62					
KIAA0319	4504469	TRANSMITTED	166	68	4.28	0.04	1.48	1.02	2.14
		NONTRANSMITTED	144	59					
KIAA0319	2179515	TRANSMITTED	184	71	4.15	0.04	1.46	1.01	2.1
		NONTRANSMITTED	162	62					
KIAA0319	6935076	TRANSMITTED	154	56	9.55	0.002	0.57	0.41	0.82
		NONTRANSMITTED	189	69					
KIAA0319	2038137	TRANSMITTED	169	69	2.59	0.11	1.36	0.94	1.97
		NONTRANSMITTED	152	62					
THEM2	926529	TRANSMITTED	203	77	5.76	0.02	1.6	1.09	2.35
		NONTRANSMITTED	179	68					
THEM2	3777664	TRANSMITTED	165	74	4.6	0.03	1.55	1.04	2.33
		NONTRANSMITTED	144	64					
INTERGENIC	1053598	TRANSMITTED	186	76	6.65	0.01	1.67	1.13	2.48
		NONTRANSMITTED	160	65					

Table 4.11 Individual genotyping results in the parent DD-proband trios. * ORs and 95% CIs refer to allele 1.

4.5.6 LOGISTIC REGRESSION ANALYSES

To try to determine which minimal set of SNPs could account for the association signal, stepwise logistic regression analyses and conditional logistic regression analyses were performed on the case-control and parent DD-proband trio data respectively.

4.5.6.1 Stepwise Logistic Regression

Stepwise logistic regression was performed using data from all SNPs genotyped in the case-control sample. To allow the best fit to be determined with allelic and genotypic effects added, alleles were coded to assess additive and dominance effects as described by Cordell and colleagues (Cordell 2002). To determine the model best describing the association, diallelic SNPs with alleles denoted 1 and 2 were coded 1 for 11 homozygotes, 0 for 12 heterozygotes and -1 for 22 homozygotes. Genotypic effects were modelled using the codings -0.5 for 11 and 22 homozygotes and 0.5 for 12 heterozygotes. For all SNPs the major allele was classified as 1. All SNPs for which individual genotyping was available in the case-control sample were initially submitted into the logistic regression model and yielded evidence for association (χ^2 = 40.91, p = 0.017, 24 df). A backwards stepwise procedure reduced the number of SNPs to additive (allelic) effects of rs4504469, rs2178515, rs6935076 and rs2143340 and genotypic (dominance) effects of rs6939068, rs2038137 and rs2143340, showing a highly significant fit (χ^2 = 28.97, p = 0.0001; 7df) (see Table 4.12).

rs Number	Effect Size	p	Exp(B)	Lower 95% CI	Upper 95% CI
rs4504469	Additive	0.06	1.59	0.98	2.57
rs6939068	Additive	0.05	0.42	0.18	1.01
rs2179515	Additive	0.097	0.58	0.30	1.11
rs6935076	Additive	0.001	0.42	0.25	0.69
rs2143340	Additive	0.003	0.25	0.10	0.64
rs2038137	Dominance	0.003	2.15	1.30	3.55
rs2143340	Dominance	0.06	0.41	0.16	1.05
Constant		0.007	5.22		

Table 4.12 Backward stepwise logistic regression model suggesting the effects of the SNPs in table best account for the association signal. The model shows a highly significant fit (χ^2 = 28.97, p = 0.0001; 7df). Variables entered on step 1 from the case-control sample: rs2793422a, rs4504469a, rs6911855a, rs6939068a, rs2179515a, rs6935076a, rs2038137a, rs2143340a, rs3777664a, rs1053598a, rs2244502a, rs793722a, rs926529a, rs2793422d, rs4504469d, rs2179515d, rs6935076d, rs2038137d, rs2143340d, rs3777664d, rs1053598d, rs2244502d, rs793722d, rs926529d (where a – additive effects, d = dominance effects).

4.5.6.2 Conditional Logistic Regression

Conditional logistic regression was performed on parent DD-proband trio data where DD-probands were considered as cases and the nontransmitted alleles were employed to create pseudocontrols (Cordell and Clayton 2002). Both additive and dominance coding were used in the analysis to determine allelic and genotypic effects. All the SNPs for which individual genotyping data was available in the parent DD-probands trios were submitted into the conditional logistic regression model (χ^2 = 21.25, p = 0.267, 18 df). Backward stepwise removal of SNPs left an allelic effect of rs6935076 best describing our observed association on chromosome 6p (χ^2 = 6.23, p = 0.013, 1 df; see Table 4.13).

rs Number	Effect Size	p	Exp(B)	Lower 95% CI	Upper 95% CI
rs6935076	Additive	0.015	0.52	0.30	0.88
Constant		0.00007	2.07		

Table 4.13 Backward stepwise logistic regression model suggesting the effects of the SNPs in table best account for the association signal. The model shows a highly significant fit ($\chi^2 = 6.23$, $p = 0.013$, 1 df). Variables entered on step 1 from a sample of parent DD-proband trios: rs2793422a, rs2793422d, rs4504469a, rs4504469d, rs2179515a, rs2179515d, rs6935076a, rs6935076d, rs2038137a, rs2038137d, rs926529a, rs926529d, rs3777664a, rs3777664d, rs1053598a, rs1053598d, rs2244502a, rs2244502d (where a – additive effects, d = dominance effects).

4.5.7 HAPLOTYPE ANALYSIS

Haplotypes were tested for association with DD using case-control data and the most significant association globally, was tested for association in the parent DD-proband trio sample data.

In the case-control, sample haplotype analysis was undertaken using the SNPs (rs4504469, rs6939068, rs2179515, rs6935076, rs2038137 and rs2143340 (*KIAA0319*) and rs926529 (*THEM2*)) that the logistic regression had shown to best describe the association with DD with either additive or dominance effects in the case-control or parent DD-proband trios samples.

All 2-, 3-, 4-, 5-, 6- and 7-marker haplotype combinations were calculated. Global p-value results are presented in Tables 4.14 – 4.19. No 6- or 7-marker haplotypes yielded a p-value ≤ 0.01. Only one 5-marker haplotype and four 4-marker haplotypes yielded global evidence of association ($p \leq 0.01$). Twelve 3-marker and thirteen 2-marker haplotypes revealed global association with DD, the most significant being with the SNPs rs4504469 and rs6935076 (global $p = 0.0003$). A similar p-value with the two markers in the parent DD-proband trios sample, suggests the result is not due to population stratification (global $p = 0.02$).

Marker 2 / Marker 1	rs4504469	rs6939068	rs2179515	rs6935076	rs2038137	rs2143340	rs926529
rs4504469		0.01	**0.007**		**0.007**	0.07	**0.007**
rs6939068			0.04	**0.01**	0.02	0.26	0.01
rs2179515				**0.01**	**0.005**	0.16	**0.007**
rs6935076					**0.0009**	**0.01**	**0.004**
rs2038137						0.02	**0.004**
rs2143340							**0.009**
rs926529							

Table **4.14** Global p-values for each two-marker haplotype. Highlighted in pink is the most significant haplotype observed in the association study. Bold indicates haplotypes showing association ($p \leq 0.01$).

Marker 1	Marker 2	Marker 3	P
rs4504469	rs6939068	rs2179515	0.04
rs6939068	rs2179515	rs6935076	0.02
rs2179515	rs6935076	rs2038137	**0.001**
rs6935076	rs2038137	rs2143340	**0.01**
rs2038137	rs2143340	rs926529	**0.01**
rs4504469	rs6939068	rs6935076	**0.005**
rs4504469	rs6939068	rs2038137	0.05
rs4504469	rs6939068	rs2143340	0.05
rs4504469	rs6939068	rs926529	**0.01**
rs4504469	rs2179515	rs6935076	**0.007**
rs4504469	rs2179515	rs2038137	**0.009**
rs4504469	rs2179515	rs2143340	0.18
rs4504469	rs2179515	rs926529	0.04
rs4504469	rs6935076	rs2038137	**0.005**
rs4504469	rs6935076	rs2143340	0.02
rs4504469	rs6935076	rs926529	**0.005**
rs4504469	rs2038137	rs2143340	0.09
rs4504469	rs2038137	rs926529	0.03
rs4504469	rs2143340	rs926529	0.03
rs6939068	rs2179515	rs2038137	0.04
rs6939068	rs2179515	rs2143340	0.14
rs6939068	rs2179515	rs926529	0.03
rs6939068	rs6935076	rs2038137	0.02
rs6939068	rs6935076	rs2143340	0.03
rs6939068	rs6935076	rs926529	0.03
rs6939068	rs2038137	rs2143340	0.12
rs6939068	rs2038137	rs926529	0.02
rs6939068	rs2143340	rs926529	0.07
rs2179515	rs6935076	rs2143340	0.04
rs2179515	rs6935076	rs926529	**0.01**
rs2179515	rs2038137	rs2143340	0.02
rs2179515	rs2038137	rs926529	**0.007**
rs2179515	rs2143340	rs926529	0.08
rs6935076	rs2038137	rs926529	**0.01**
rs6935076	rs2143340	rs926529	0.03

Table 4.15 Global p-values for each three-marker haplotype. Bold indicates haplotypes showing association ($p \leq 0.01$).

Marker 1	Marker 2	Marker 3	Marker 4	P
rs4504469	rs6939068	rs2179515	rs6935076	0.02
rs6939068	rs2179515	rs6935076	rs2038137	0.02
rs2179515	rs6935076	rs2038137	rs2143340	**0.01**
rs6935076	rs2038137	rs2143340	rs926529	0.03
rs4504469	rs6939068	rs2179515	rs2038137	0.06
rs4504469	rs6939068	rs2179515	rs2143340	0.19
rs4504469	rs6939068	rs2179515	rs926529	0.09
rs4504469	rs6939068	rs6935076	rs2038137	0.03
rs4504469	rs6939068	rs6935076	rs2143340	0.03
rs4504469	rs6939068	rs6935076	rs926529	0.03
rs4504469	rs6939068	rs2038137	rs2143340	0.19
rs4504469	rs6939068	rs2038137	rs926529	0.06
rs4504469	rs6939068	rs2143340	rs926529	0.03
rs4504469	rs2179515	rs6935076	rs2038137	**0.005**
rs4504469	rs2179515	rs6935076	rs2143340	0.13
rs4504469	rs2179515	rs6935076	rs926529	0.06
rs4504469	rs2179515	rs2038137	rs2143340	0.12
rs4504469	rs2179515	rs2038137	rs926529	0.06
rs4504469	rs2179515	rs2143340	rs926529	0.13
rs4504469	rs6935076	rs2038137	rs2143340	0.08
rs4504469	rs6935076	rs2038137	rs926529	0.05
rs4504469	rs6935076	rs2143340	rs926529	0.08
rs4504469	rs2038137	rs2143340	rs926529	0.08
rs6939068	rs2179515	rs6935076	rs2143340	0.03
rs6939068	rs2179515	rs6935076	rs926529	0.04
rs6939068	rs2179515	rs2038137	rs2143340	0.03
rs6939068	rs2179515	rs2038137	rs926529	0.04
rs6939068	rs2179515	rs2143340	rs926529	0.14
rs6939068	rs6935076	rs2038137	rs2143340	0.02
rs6939068	rs6935076	rs2038137	rs926529	0.04
rs6939068	rs6935076	rs2143340	rs926529	0.08
rs6939068	rs2038137	rs2143340	rs926529	0.10
rs2179515	rs6935076	rs2038137	rs926529	**0.004**
rs2179515	rs6935076	rs2143340	rs926529	0.07
rs2179515	rs2038137	rs2143340	rs926529	**0.01**

Table 4.16 Global p-values for each four-marker haplotype. Bold indicates haplotypes showing association ($p \leq 0.01$).

Marker 1	Marker 2	Marker 3	Marker 4	Marker 5	P
rs4504469	rs6939068	rs2179515	rs6935076	rs2038137	0.05
rs6939068	rs2179515	rs6935076	rs2038137	rs2143340	**0.01**
rs2179515	rs6935076	rs2038137	rs2143340	rs926529	0.02
rs4504469	rs6939068	rs2179515	rs6935076	rs2143340	0.04
rs4504469	rs6939068	rs2179515	rs6935076	rs926529	0.08
rs4504469	rs6939068	rs2179515	rs2038137	rs2143340	0.19
rs4504469	rs6939068	rs2179515	rs2038137	rs926529	0.09
rs4504469	rs6939068	rs2179515	rs2143340	rs926529	0.19
rs4504469	rs6939068	rs6935076	rs2038137	rs2143340	0.05
rs4504469	rs6939068	rs6935076	rs2038137	rs926529	0.13
rs4504469	rs6939068	rs6935076	rs2143340	rs926529	0.08
rs4504469	rs6939068	rs2038137	rs2143340	rs926529	0.19
rs4504469	rs2179515	rs6935076	rs2038137	rs2143340	0.07
rs4504469	rs2179515	rs6935076	rs2038137	rs926529	0.06
rs4504469	rs2179515	rs6935076	rs2143340	rs926529	0.18
rs4504469	rs2179515	rs2038137	rs2143340	rs926529	0.07
rs4504469	rs6935076	rs2038137	rs2143340	rs926529	0.10
rs6939068	rs2179515	rs6935076	rs2038137	rs926529	0.03
rs6939068	rs2179515	rs6935076	rs2143340	rs926529	0.07
rs6939068	rs2179515	rs2038137	rs2143340	rs926529	0.03
rs6939068	rs6935076	rs6935076	rs2143340	rs926529	0.04

Table 4.17 Global p-values for each five-marker haplotype. Bold indicates haplotypes showing association ($p \leq 0.01$).

Marker 1	Marker 2	Marker 3	Marker 4	Marker 5	Marker 6	P
rs4504469	rs6939068	rs2179515	rs6935076	rs2038137	rs2143340	0.04
rs6939068	rs2179515	rs6935076	rs2038137	rs2143340	rs926529	0.02
rs4504469	rs6939068	rs2179515	rs6935076	rs2038137	rs926529	0.16
rs4504469	rs6939068	rs2179515	rs6935076	rs2143340	rs926529	0.10
rs4504469	rs6939068	rs2179515	rs2038137	rs2143340	rs926529	0.14
rs4504469	rs6939068	rs6935076	rs2038137	rs2143340	rs926529	0.06
rs4504469	rs2179515	rs6935076	rs2038137	rs2143340	rs926529	0.12

Table 4.18 Global p-values for each six-marker haplotype. Bold indicates haplotypes showing association ($p \leq 0.01$).

Marker 1	Marker 2	Marker 3	Marker 4	Marker 5	Marker 6	Marker 7	P
rs4504469	rs6939068	rs2179515	rs6935076	rs2038137	rs2143340	rs926529	0.07

Table 4.19 Global p-value the seven-marker haplotype.

For the haplotype rs4504469/rs6935076 (see Table 4.14), for each sample the G-A haplotype was associated with DD (case-control p = 0.02; parent DD-proband trios p = 0.03) (see Tables 4.20 and 4.21). However, the most striking observation was the under-representation of the A-G haplotype in the DD cases

based on the case-control sample (p = 0.00003) and parent DD-proband trios sample (p = 0.006).

Haplotype	Frequency in Cases	Frequency in Controls	χ^2	p
G-G	0.31	0.27	1.25	0.26
G-A	0.35	0.30	5.82	0.02
A-G	0.25	0.39	17.38	0.00003
A-A	0.09	0.05	1.87	0.17

Table 4.20 Two-marker (rs4504469/rs6939076) haplotype association with DD in the case-control sample. The G-A haplotype is over-represented in the DD cases, whilst the A-G haplotype is under represented in the DD cases. Globally the two-marker haplotype yielded evidence of association with DD (p = 0.0001).

Haplotype	Frequency of Transmitted Haplotype	Frequency of Non-Transmitted Haplotype	p
G-G	0.32	0.32	0.95
G-A	0.35	0.27	0.03
A-G	0.24	0.36	0.006
A-A	0.09	0.05	0.22

Table 4.21 Two-marker (rs4504469/rs6939076) haplotype association with DD in the parent DD-proband trio sample. The G-A haplotype is over-transmitted to the DD cases, whilst the A-G haplotype is under-transmitted to the DD cases. Globally the two-marker haplotype yielded evidence of association with DD (p = 0.02).

In previous studies of the chromosome 6p region, the best evidence for association was described by a 3-marker haplotype (rs4504469/rs2038137/rs2143340) (Francks et al. 2004). Globally, this haplotype did not yield evidence of association in the case-control sample (p = 0.09) (see Table 4.22). However, two individual haplotypes did show evidence of association with DD (see Table 4.22). The G-G-T haplotype was more common in subjects with DD than in controls (p = 0.03), whilst the A-T-T haplotype, which showed significant association in the Francks study (Francks et al. 2004) with the single word reading (READ) phenotype (combined UK sample), showed evidence of association and in the same direction of effect as that observed before (Francks et al. 2004). In our

sample the A-T-T haplotype was more common in controls, while in the Francks study it was associated with better performance on the READ phenotype (Paracchini 2005). The G-G-C haplotype that was reported to show association with componential measures of DD in a US and UK sample (Francks et al. 2004), was not significantly associated with DD in our sample (p = 0.21) (see Table 4.22).

Haplotype	Frequency in Cases	Frequency in Control	χ^2	p
G-G-T	0.47	0.41	4.58	**0.03**
G-G-C	0.15	0.12	1.59	0.21
G-T-T	0.04	0.05	0.19	0.66
A-G-T	0.07	0.08	0.19	0.67
A-G-C	0.02	0.02	0.18	0.67
A-T-T	0.25	0.33	6.63	**0.01**

Table 4.22 Three-marker haplotype (rs4504469/rs2038137/rs2143340) association in the case-control sample. The G-G-T haplotype shows significant over-representation in the cases, whilst the A-T-T haplotype shows under-representation in the cases. Globally the haplotype was not significant (p = 0.09).

In our sample the A-T-T haplotype is perfectly defined by the first two SNPs, rs4504469 and rs2038137, since the A-T-C haplotype has a frequency of zero (see Table 4.22). As a result, rs2143340 was excluded to reduce the degrees of freedom and the two-marker rs4504469 and rs2038137 (A-T) haplotype was tested in our family based sample. Although the A-T haplotype was undertransmitted to the DD-probands, the result was not significant (p = 0.1).

4.5.8 ALLELE SPECIFIC EXPRESSION

Although evidence of association has been observed with the non-synonymous SNP rs4504469, the lack of association between this SNP and a sample from the USA (Francks et al. 2004) suggests that it is unlikely to be the polymorphism responsible for the association. This, and the absence of any

association with other non-synonymous polymorphisms suggested that polymorphisms influencing gene expression for example may be associated with DD. Consequently, an allele specific expression assay was performed to investigate potential *cis*-acting influences upon *KIAA0319* gene expression. The SNP rs4504469 (located in exon 4 of *KIAA0319*) was used as a tag polymorphism to test for differential allelic expression. Thirty of the 66 individuals for which brain derived RNA was available were heterozygous for rs4504469 and were therefore informative for the assay at this locus. No evidence for polymorphic *cis*-acting effects on *KIAA0319* (see Figure 4.12) was observed.

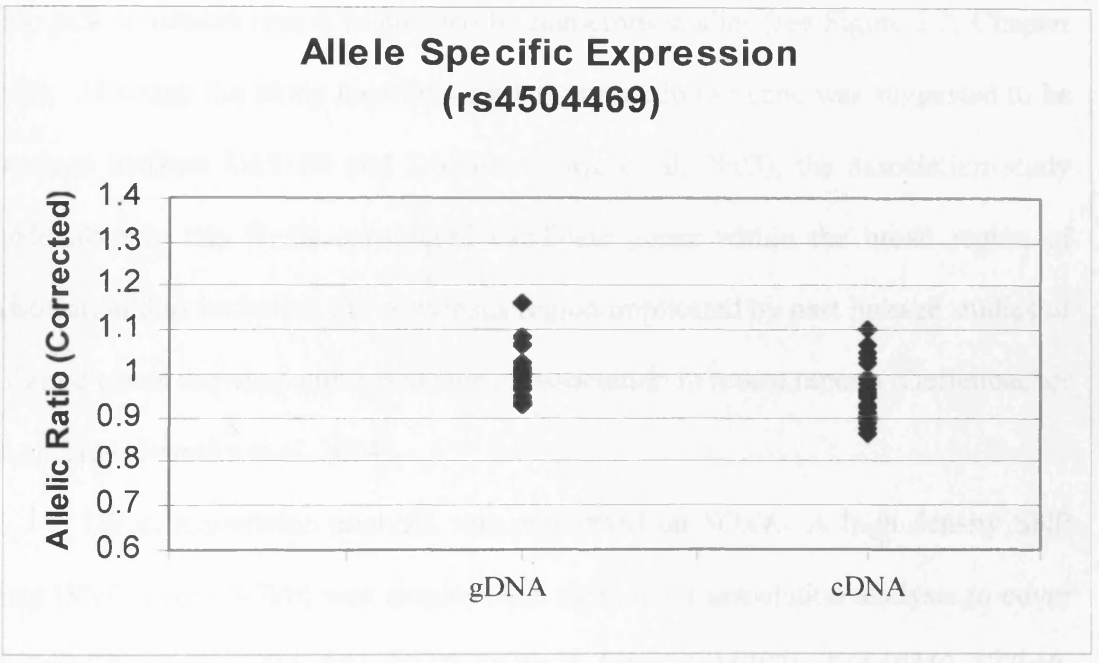


Figure 4.12 The relative expression of alleles at rs4504469. rs4504469 is a G→A polymorphism. Allelic ratio refers to the ratio of G alleles to A alleles. No relative expression differences between alleles at rs4504469 were observed.

4.6 DISCUSSION

There is considerable evidence for a QTL for DD on chromosome 6p (Cardon et al. 1994; Cardon et al. 1995; Grigorenko et al. 1997; Fisher et al. 1999; Gayán et al. 1999; Grigorenko et al. 2000; Smith et al. 2001; Kaplan et al. 2002;

Willcutt et al. 2002; Grigorenko et al. 2003; Turic et al. 2003; Deffenbacher et al. 2004; Francks et al. 2004). Recent evidence from Deffenbacher and colleagues (Deffenbacher et al. 2004) pointed to five genes including *VMP*, *DCDC2*, *KIAA0319*, *TTRAP* and *THEM2*. Francks and colleagues (Francks et al. 2004) further narrowed the region, suggesting the linkage and association to chromosome 6p was most likely due to *KIAA0319*, *TTRAP* or *THEM2*.

Previous association analysis across the chromosome 6p region had yielded evidence for association to DD in the sample used in this thesis (Turic et al. 2003), specifically a broad region spanning markers D6S109 to D6S1260, part of which falls in a consensus region implicated by numerous studies (see Figure 2.2, Chapter two). Although the likely location of a DD susceptibility gene was suggested to be between markers D6S109 and D6S506 (Turic et al. 2003), the association study undertaken in this thesis considered candidate genes within the broad region of association thus including the consensus region implicated by past linkage studies of DD and genes showing some evidence of association in recent reports (Deffenbacher et al. 2004; Francks et al. 2004).

Direct association analysis was performed on *SOX4*. A high density SNP map (SNPs every 2-3kb) was employed in an indirect association analysis to cover the candidate genes *ID4*, *PRL*, *VMP*, *DCDC2*, *KAAG1*, *MRS2L*, *KIAA0319*, *TTRAP*, *THEM2* and *C6orf62*. Genes chosen for analysis were known genes prioritised on previous associations and brain expression. The SNP map included SNPs spanning all introns, coding exon, UTR and flanking sequences of the candidate genes.

Initially DNA pools were genotyped from DD cases and controls. This analysis has been shown to be a cost effective method of genotyping and is highly accurate (Norton et al. 2002). Validation of the pools utilised in this study have

shown a pooling error of just 2% (see Table 4.5). Positive results generated in pooled DNA analysis were followed up by individual genotyping of a case-control sample following removal of redundant markers based upon LD. All SNPs identified as of interest by the DNA pooling experiments, with the exception of rs793722, rs6911855 and rs6939068 yielded significant evidence for association with DD after individual genotyping. In the case of two of the three that did not (rs6911855 and rs6939068), results were very near to significance (allelic p = 0.07 and allelic p = 0.06 respectively). In the case-control sample, association was observed with a marker in *PRL* (rs2244502, allelic p = 0.03), *MRS2L* (rs2793422, allelic p = 0.003), *KIAA0319* (rs4504469, genotypic and allelic p = 0.002; rs2179515, allelic p = 0.007; rs6935076, genotypic and allelic p = 0.006; rs2038137, genotypic p = 0.0001), *THEM2* (rs926529, allelic p = 0.0024; rs3777664, allelic p = 0.008) and an intergenic SNP between *THEM2* and *C6orf62* (rs1053598, allelic p = 0.03). A nested family-based association sample was individually genotyped to ensure that observed associations in the case-control sample were not the result of population stratification. The results indicated that association between rs2244502 and DD detected in the case-control sample could have been due to population stratification (p = 0.36), but given that most of the results were confirmed, it is unlikely to be population stratification and could be due to a lack of power to detect associations in the parent DD-proband sample. The absence of association between DD and rs2038137 in the parent DD-proband trios could also be due to the departure from Hardy-Weinberg Equilibrium in the controls, resulting in a false positive in the case-control sample. Associations with all other SNPs in the case-control sample were confirmed in the parent DD-proband trio sample (rs4504469, p = 0.04; rs2179515, p = 0.04; rs6935076, p = 0.002; rs926529, p = 0.02; rs3777664, p = 0.03;

rs1053598, p = 0.01).

4.6.1 LINKAGE DISEQUILIBRIUM ACROSS THE CHROMOSOME 6P REGION

LD analysis across the chromosome 6p region in my sample with 20 SNPs revealed a large block of LD spanning part of *KIAA0319* and extending into *TTRAP*, specifically between markers rs7751357 and rs2143340, a region spanning 33kb (see Figures 4.13 and 4.14). The strong LD between SNPs within the region, and specifically the region spanning the first four exons of *KIAA0319*, mean that the functionally relevant SNPs for DD may be difficult to identify.

Data available from the HapMap project have shown that there are a number of blocks of LD spanning *KIAA0319*, with two large blocks of LD towards each end of the gene. There is moderate LD spanning the two halves of the gene (based on D'). In future studies these data could be captured by using tag SNPs across *KIAA0319*. Using tagging software available in Haploview and genotypes on CEPH individuals downloaded from the International HapMap Project, based on individual genotyping, a minor allele frequency of 0.001 and an r^2 of 0.8, 43 SNPs would have to be genotyped and tested for evidence of association with DD (individually), along with three 3-marker haplotypes and four 2-marker haplotypes.

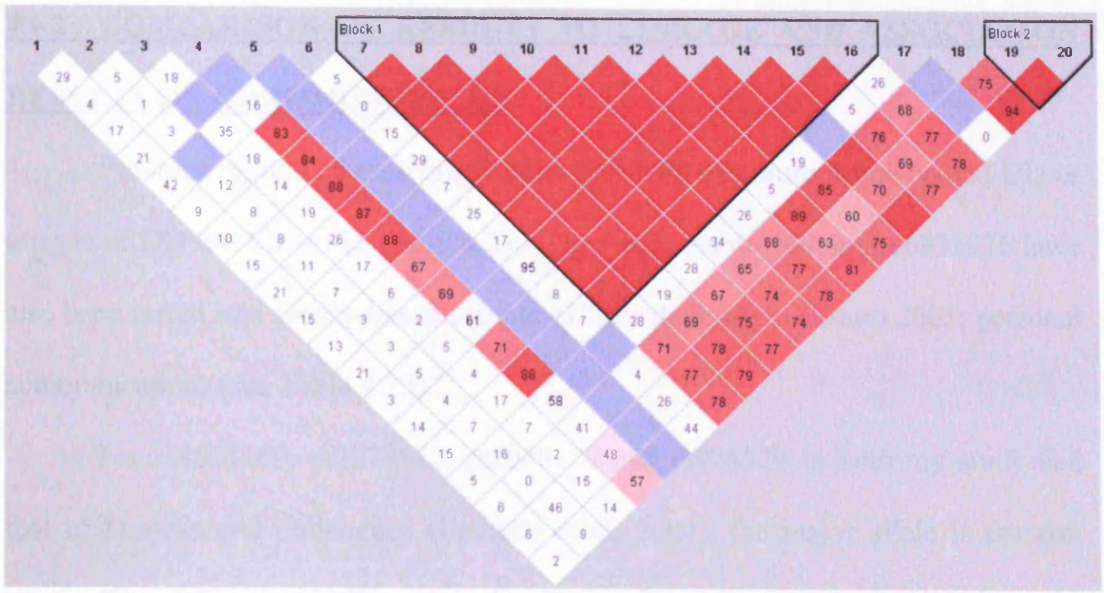


Figure 4.13 LD between SNPs individually genotyped in 48 controls. LD is based on D', darker red illustrating D'= 1, white illustrating D' is low. Specific values of D' are shown in each square. A large block of LD is shown in red, along with a smaller block not observed in the DD cases. 1 = 2244502, 2 = 793722, 3 = 2793422, 4 = 4504469, 5 = 6911855, 6 = 6939068, 7 = 7751357, 8 = 2179515, 9 = 6456622, 10 = 9358783, 11 = 9358784, 12 = 7755579, 13 = 6456624, 14 = 6935076, 15 = 2038137, 16 = 1555090, 17 = 2143340, 18 = 926529, 19 = 3777664, 20 = 1053598.

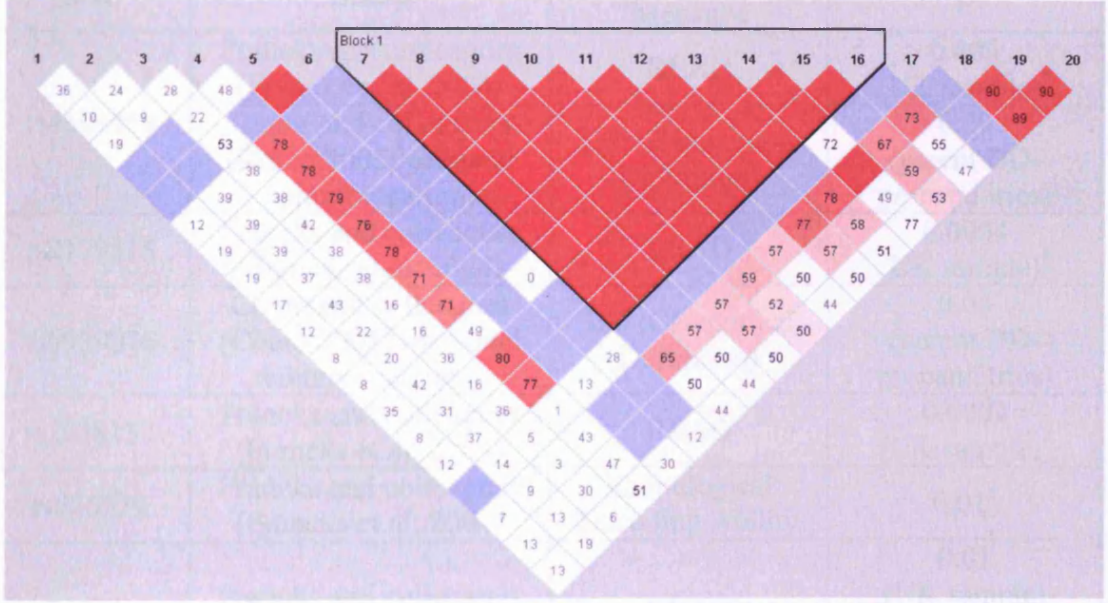


Figure 4.14 LD between SNPs individually genotyped in 42 DD cases. LD is based on D', darker red illustrating D'= 1, white illustrating D' is low. Specific values of D' are shown in each square. A large block of LD is shown in red. A second block of LD, shown in the LD across a control sample is not observed in the case sample. 1 = 2244502, 2 = 793722, 3 = 2793422, 4 = 4504469, 5 = 6911855, 6 = 6939068, 7 = 7751357, 8 = 2179515, 9 = 6456622, 10 = 9358783, 11 = 9358784, 12 = 7755579, 13 = 6456624, 14 = 6935076, 15 = 2038137, 16 = 1555090, 17 = 2143340, 18 = 926529, 19 = 3777664, 20 = 1053598.

4.6.2 COMPARISON OF RESULTS TO LINKAGE AND ASSOCIATION RESULTS ON CHROMOSOME 6P

Previous reports of association between SNPs on chromosome 6p and DD or aspects of DD are shown in Table 4.23. The SNPs rs4504469 and rs6935076 have also been tested in a family-based sample of Canadian origin (Couto 2005; personal communication) (see Table 4.23).

For rs4504469, rs2179515, rs2038137 and rs926529 in both my study and that of Francks and colleagues (Francks et al. 2004), the major allele is the risk allele. The SNP rs2143340, although not significant in our sample did show the same direction of effect to that of Francks and colleagues (Francks et al. 2004), who suggested the minor allele was the risk allele in DD.

SNP	Study	Component Measure	p
rs4504469	Francks and colleagues (Francks et al. 2004)	READ	0.004 (UK sample)*
	Couto and colleagues (Couto 2005; personal communication)	DD (Categorical Definition)	0.29 (parent DD-proband trios)
rs2179515	Francks and colleagues (Francks et al. 2004)	READ	0.0004 (UK sample)*
rs6939076	Couto and colleagues (Couto 2005) (personal communication)	DD (Categorical Definition)	0.03 (parent DD-proband trios)
rs2038137	Francks and colleagues (Francks et al. 2004)	READ	0.0002 (UK sample)*
rs926529	Francks and colleagues (Francks et al. 2004)	Phonological Decoding Ability	0.01#
rs2143340	Francks and colleagues (Francks et al. 2004)	READ	0.01 (UK sample) 0.005 (US sample)

Table 4.23 Associations reported in previous studies with SNPs yielding evidence of association in this thesis. The study in this thesis was unable to replicate the association previously reported between DD and rs2143340 (Francks et al. 2004).
* Associations were not observed in a sample of US origin.
# This measure of phonological decoding ability is not directly comparable to the categorical definition of DD employed in this thesis.

Logistic regression, using a stepwise procedure, revealed rs4504469, rs6939068, rs2178515, rs6935076, rs2038137 (*KIAA0319*) and rs2143340 (*TTRAP*) best capture the association signal to DD (see Table 4.12). Conditional logistic regression using the data from the parent DD-proband trios suggested that allelic effects of rs6935076 (*KIAA0319*) best describe the association (see Table 4.13).

Haplotype analysis in the case-control sample yielded significant evidence for association between the haplotype rs4504469/rs6935076 and DD (see Tables 4.20 and 4.21). This haplotype spans intron 1 to exon 4 of *KIAA0319*. Correcting the p-values for multiple testing is not appropriate at this stage of the study given the large number of markers analysed through the study (including DNA pooling results) which would need to be corrected for.

Jill Couto at the University of Toronto (Couto 2005; personal communication), reported some evidence of association between DD and the rs4504469/rs6935076 haplotype. Although global significance was not observed (p = 0.129), the G-A 'risk' haplotype in our sample also showed significance in their family based sample (p = 0.04). However, this haplotype, although increasing risk to DD in our sample, seemed to be a 'protective' haplotype in their sample, with undertransmission of the haplotype to cases. The 'protective' A-G haplotype in our sample did not yield evidence for association in their sample (p = 0.108).

4.6.3 *KIAA0319* GENE EXPRESSION

At a symposium in Toronto, Canada organised by the Canadian Language and Literacy Research Network, Dr Silvia Paracchini (Paracchini 2005) reported reduced allelic expression of the G-G-C risk haplotype (rs4504469/rs2038137/rs2143340). In my sample of brain derived gDNA and

cDNA, I was able to identify individuals for whom the phased haplotype probability of the G-G-C haplotype was 98% or above or who were most likely to carry the specific haplotype. My results suggest that there is no alteration of the expression of *KIAA0319* in individuals carrying the G-G-C haplotype relative to other haplotype carriers (see Figure 4.15). It should be noted that the expression differences reported by Dr Silvia Paracchini were observed in lymphoblast cells rather than brain samples. The extent to which results on lymphoblast cells can be extrapolated to other human tissue is unknown and it is unlikely that differences in gene expression within lymphoblast cells would alter brain function, which would result in DD. As a result it is more important to look at brain cells directly in order to elucidate any expression differences, which could alter brain function.

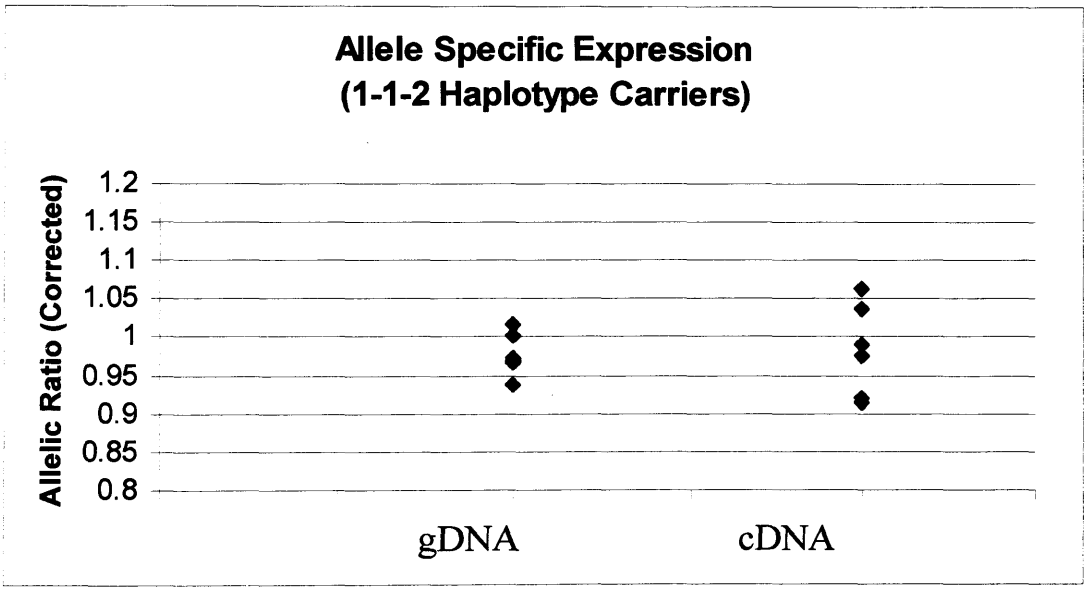


Figure 4.15 Allele specific expression using rs4504469 as a tag SNP. Plotted are individuals who have or are predicted to have the G-G-C risk haplotype (rs4504469/rs2038137/rs2143340) which showed reduced expression in a previous study (Paracchini 2005). The allelic ratio refers to the ratio of G alleles to A alleles at rs4504469.

4.6.4 *KIAA0319*

KIAA0319 (see Figure 4.7) is 102kb in length and contains 21 exons. It encodes 1109 amino acids, which are predicted to form a 121kDa protein. The function of KIAA0319 is unknown at present although sequence analysis suggests it may have a role in cell adhesion or calcium binding ability. It has a transmembrane domain towards the C' terminus of polypeptide, suggesting it is a membrane spanning protein. The protein is predicted to have five PKD domains and a cadherin domain. The PKD (polycystic kidney disease) domains have homology with the extracellular domains of PKD1, which have been shown to be involved in cell-adhesive functions. The presence of an Ig- like fold indicates a possible role in cell-matrix or cell-cell adhesion. Cadherins are glycoproteins involved in calcium-mediated cell-cell adhesion. The cadherin domain is thought to mediate cell-cell adhesion when calcium is binds to it. The domain has been suggested to play a role in cell fate, signalling, proliferation, differentiation and migration.

KIAA0319 also contains a cytoplasmic motif, referred to as the immunoreceptor tyrosine-based inhibitor motif (ITIM). This motif, in other proteins, has been shown to be involved in the modulation of cellular responses and the phosphorylation of ITIM motifs may allow binding of the SH2 domain of several SH2-containing phosphatases. However, the ITIM motif within KIAA0319 is predicted on sequence data, not on experimental evidence. Recently, Nakayama and colleagues (Nakayama et al. 2002), using a yeast-two hybrid system, reported that KIAA0319 interacts with KIAA1299, an SH2-B β-signalling protein, providing some evidence that the ITIM motif may be present since KIAA1299 contains SH2 domains.

KIAA1299 is a SH2-B β-signalling protein containing two motifs, a

pleckstrin-like motif and an SH2 motif. The pleckstrin-like motif has been observed in other proteins that are involved in intracellular signalling or which make up the cytoskeleton of cells. The SH2 motif is also found in intracellular signalling proteins. It functions as a regulatory module of intracellular signalling cascades by interacting with high affinity to phosphotyrosine-containing target peptides.

Although the allele specific expression assay show *KIAA0319* expression in brain from the frontal, temporal and parietal cortex, more specific brain regions showing expression were not identified. Londin and colleagues (Londin et al. 2003) have shown *KIAA0319* expression in total brain using RT-PCR as well as in cerebellum. Expression profiles in humans from the GNF Expression Atlas 1 (based on Affymetrix U95 Chips) have shown high expression of *KIAA0319* in foetal brain and in adult brain, including the cortex, caudate nucleus, amygdala and thalamus (http://genome.ucsc.edu/cgi-bin/hgGene?hgsid=59571891&db=hg17&hgggene=NM_014809&hgg_chrom=chr6&hgg_start=24652310&hgg_end=24754362&hgg_type=knownGene). Although not often the main regions implicated in neurobiological studies of DD, each of the four regions has shown some evidence for a role in DD (Abdullaev and Melnichuk 1997; Raymer et al. 1997; Crosson 1999; Fried et al. 2001; Joseph et al. 2001; Backes et al. 2002; Jaskowski and Rusiak 2005).

KIAA0319 shows a number of regions of sequence conservation from human through chimp, dog, mouse, rat, chicken, fugu and zebrafish. Most of these cover the exons, but there are 8 regions of conservation in intronic sequence (see Figure 4.16). rs4504469, which shows evidence of association with DD in our study, is located in exon 4, which shows conservation across species. The SNP changes an alanine to a threonine at position 311. The alanine residue is only observed in humans. All other SNPs showing association in *KIAA0319* in my study are located

within intron 1. Of these rs2179515 and rs6939076 are located in sequence conserved between human, chimp and dog, whilst rs2038137 is located in a region conserved between the same three species but also mouse and rat. Intronic sequence has been reported to contain a number of regulatory elements that play an important role in gene regulation and splicing (Duan et al. 2003; Fedorova and Fedorov 2003). As a result, associated SNPs in intron 1 of *KIAA0319* may influence such elements.

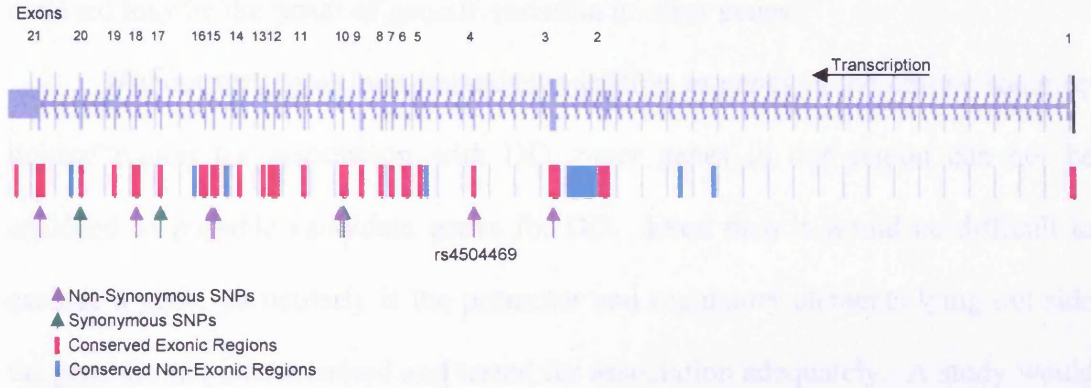


Figure 4.16 The conserved regions of *KIAA0319* across the species of chimp, dog, mouse, rat, chicken, fugu and zebrafish. The pink conserved regions are regions conserved amongst species and contain exonic sequence. Conserved regions containing intronic or 3'UTR sequence are shown in turquoise. SNPs, indicated as non-synonymous and synonymous, are indicated by the arrows.

Given the allele-specific expression results obtained in this study, it is unlikely that there is *cis*-acting variation in the *KIAA0319* gene that affects expression levels. However, this does not exclude the possibility that there are unknown transcripts of *KIAA0319*, which do not include the tag SNP (rs4504469) I used to measure expression. Alternatively, there could be a minor transcript containing rs4504469 where expression is altered but being a minor transcript, it is not sufficient for detection using this assay.

4.6.5 GENES OTHER THAN KIAA0319 IN THE CHROMOSOME 6p REGION

Evidence for association was observed between variations in *PRL*, *MRS2L*, *KIAA0319*, *THEM2* and an intergenic SNP near *THEM2*. Whilst statistical evidence (stepwise logistic regression and conditional logistic regression) and haplotype analysis suggest *KIAA0319* is the susceptibility gene in the chromosome 6p linkage region, the unpredictable nature of LD could mean that the association signal obtained may be the result of genetic variation in other genes.

Without testing all non-redundant variation in genes in the chromosome 6p linkage region for association with DD, other genes in this region can not be excluded as possible candidate genes for DD. Even then it would be difficult to exclude a gene, particularly if the promotor and regulatory elements lying out side the gene are not characterised and tested for association adequately. A study would also need to have a sufficiently large sample with adequate power to detect associations, in particular associations with a small effect size to exclude genes as possible candidate genes.

Association was observed between *THEM2* and DD. However, the association was weaker than the association between the rs4504469/rs6935076 haplotype and DD. Given the moderate LD between markers in *THEM2* and *KIAA0319* the association observed with *THEM2* could reflect the association observed with the two-marker haplotype. rs926529, which shows association in this study, shows only weak evidence of association with phonological decoding efficiency in another study (Francks et al. 2004) and is not replicated in their US sample. Other SNPs in *THEM2* show weak evidence for association with components of DD in a UK sample (Francks et al. 2004) but very little evidence in a

US sample. Unless a variation common in the population is represented more in our DD cases than those of Francks and colleagues (Francks et al. 2004), *THEM2* is less likely, with current knowledge, to increase susceptibility to DD.

Association was also observed between rs2793422 (*MRS2L*) and DD. No association has been reported in a previous study between DD and *MRS2L* (Deffenbacher et al. 2004), however SNP density was not comprehensive. Given the lack of LD between rs2793422 and any other marker analysed in my study, the association observed is not attributable to *KIAA0319*. This could indicate that a second susceptibility gene for DD lies in the chromosome 6p region. Previous studies have suggested that the relatively large region showing linkage to DD on chromosome 6p may harbour more than one susceptibility gene (Grigorenko et al. 2003; Turic et al. 2003). Grigorenko and colleagues (Grigorenko et al. 2003) suggested that the regions may co-act or represent duplicated, functionally similar regions. Interestingly rs2793340 is an exonic SNP, albeit synonymous but could alter expression, mRNA stability, editing and splicing. It is also possible that rs2793422 represents a type I error (false positive).

Like *THEM2*, *MRS2L* can not be excluded as a DD susceptibility gene, however given the weak evidence for association (only one SNP out of twelve showed any evidence of association), the lack of association in a previous study (Deffenbacher et al. 2004) and given the stronger support for association with *KIAA0319*, the further study of *MRS2L* is not a priority at the present time in our laboratory.

4.7 GENERAL CONCLUSION

This study has identified a number of associations between DD and genes on

chromosome 6p. Although haplotype and statistical analysis suggests that variation within *KIAA0319* is responsible for the association, associations in other genes cannot be ignored. Indeed there is the possibility that a second susceptibility gene for DD lies within the same chromosome 6p region as *KIAA0319*. Further work is required to determine whether the association observed in *KIAA0319* is the result of a causal variation within *KIAA0319* itself or the result of unpredictable LD between *KIAA0319* and other genes within the region.

The sequencing of exons and putative promoter of *KIAA0319* in a previous study has failed to reveal any non-synonymous SNPs, with the exception of rs4504469 (Francks et al. 2004). It is unlikely that rs4504469 is a causal variant since the threonine that is present on the 'protective' A-G haplotype is also present on the A-A haplotype which is more common in cases, albeit not significantly more so than controls. This suggests that if rs4504469 can influence risk of DD directly, then its effects can be modified by a second susceptibility allele in the gene, which is a *cis-cis* interaction. The failure to detect association in US (Francks et al. 2004) and Canadian samples (Couto 2005) between rs4504469 and DD also suggests that rs4504469 is unlikely to directly influence DD susceptibility.

Further work is thus required to identify the causal variation(s) within *KIAA0319* or which show LD with it.

CHAPTER FIVE

A SUBPHENOTYPIC ANALYSIS OF THE DEVELOPMENTAL DYSLEXIA SUSCEPTIBILITY LOCUS ON CHROMOSOME 6p

CHAPTER FIVE

5. A SUBPHENOTYPIC ANALYSIS OF THE DEVELOPMENTAL DYSLEXIA SUSCEPTIBILITY LOCUS ON CHROMOSOME 6P

DD is a phenotypically heterogeneous disorder and this may be the result of genetic heterogeneity (Grigorenko et al. 1997; Raskind et al. 2000). In this chapter I examine the different componential phenotypes of DD for a relationship with variation within *KIAA0319*.

5.1 COMPLEX NATURE OF DEVELOPMENTAL DYSLEXIA

Chapter two presented family and twin evidence that component measures of reading and/or DD may be differentially genetically influenced (see Section 2.2-2.4). Although heritability estimates have suggested that there may be a shared genetic aetiology between many of the reading components (see Chapter two), Castles and colleagues have also indicated that there is at least some partial genetic independence between the cognitive processes involved in reading (Castles et al. 1999). It is therefore plausible that genes lying on different chromosomes influence different aspects of the DD phenotype.

5.2 COGNITIVE ASPECTS OF READING: READING RELATED MEASURES

Three main candidate language processes contributing to difficulty in word learning are phonological skills (Liberman et al. 1974; Vellutino 1979; Wagner and Torgesen 1987; Pennington et al. 1990), rapid automated naming (Denckla and

Rudel 1976; Wolf et al. 1986) and orthographic skills (Berninger 1994; Olson et al. 1994; Berninger et al. 2001). Although often highly correlated, component phenotypes do not necessarily show similar support for linkage to the same regions (see Chapter two). If the variation in linkage signals reported index genetic heterogeneity, analysis of association between component measures of reading and the true susceptibility gene might allow the identification of subgroups that are more genetically homogeneous and may lead to an increased understanding of the heterogeneity whilst increasing the reproducibility of findings across studies. This is the rationale for examining the relationship between specific deficits observed in dyslexic individuals and genetic variation in *KIAA0319*.

5.3 CHOICE OF COMPONENTIAL PHENOTYPE ANALYSIS

Typically, molecular genetic studies of component measures of reading have been assessed using categorical definitions of component reading phenotypes (Grigorenko et al. 1997; Fisher et al. 1999; Gayán and Olson 1999; Grigorenko et al. 2000; Grigorenko et al. 2003). Categorical analyses do not assume a normal distribution of the data; an important assumption given some of the reading component scores are not normally distributed.

The associations observed within Chapter four have been observed with DD classified as a disease category. However there are differing opinions to whether DD forms a bi-modal distribution with normal reading ability (and therefore represents a distinct pathological reading group), or whether it represents the tail end of a normal reading distribution (see Chapter one, Section 1.4). As a result it is important to test for associations with continuous measures of component phenotypes of DD regardless of DD affection status. The QTL approach provides a

complementary strategy that can link categorical disorders to continuously distributed traits that more closely underlie the genetic liability in the general population. A QTL analysis will also allow the detection of associations within distinct pathological groups, such as within a DD group or within a normal reading ability group.

As a result two forms of analysis were undertaken in this chapter; a categorical analysis (see analysis one) and a quantitative analysis (see analysis two). In addition, given the findings of analysis one, a third analysis was undertaken, testing association between ADHD and variation in *KIAA0319*. Given the number of SNPs showing association to DD in Chapter four, componential phenotype analysis was undertaken using the haplotype rs4504469/rs6939076 given that it yielded the strongest evidence for association to DD.

5.4 ANALYSIS ONE – A CATEGORICAL COMPONENTIAL PHENOTYPE ANALYSIS OF *KIAA0319*

5.4.1 AIMS AND HYPOTHESES

A categorical analysis was undertaken in order to answer the questions:

1) Does an association exist between componential phenotypes of DD and the haplotype rs4504469/rs6939076?
2) Can specific componential phenotypes of DD explain the association observed between rs4504469/rs6939076 better than a categorical definition of DD (as used in Chapter four)?

5.4.2 MATERIALS AND METHODS

5.4.2.1 PARTICIPANTS AND TEST BATTERY

The sample of 273 control individuals and 241 DD-probands for whom genotypes and phenotypic data were available are described in Chapter three. Two categories were formed (demographics for each sample are shown in Tables 5.1 and 5.2):

1) For each component phenotype, the worst 100 DD cases (based on the original DD classification) for each component measure were selected to form the 'case' category.
2) All controls were utilised to form the 'control' category, regardless of their score on each test.

This categorical analysis represents the comparison between a normal reading population and the DD cases with the most severe deficits in component phenotypes related to DD. An analysis such as this was chosen in order to maximise

the ability to detect differences and therefore association between component phenotypes of DD and variation in *KIAA0319*.

Variable	Mean (years)	SD (years)	Min (years)	Max (years)
Age	13.24	2.12	7.67	17.58
IQ	103.48	11.54	85	136
RD	-5.23	1.97	-2.5	-13.0

Table 5.1 The demographics for the case category.

Variable	Mean (years)	SD (years)	Min (years)	Max (years)
Age	11.98	2.39	5.5	16.67
IQ	103.25	11.95	85	137
RD	+1.14	1.45	-0.5	6.92

Table 5.2 The demographics for the control category.

In total sixteen component phenotypes were analysed (see Chapter three, Section 3.1.2). Three measures of reading including accuracy, comprehension and reading rate were only analysed in children aged 12 years or under since the NARA test only accurately calculates these measures to the age of 12 years.

5.4.2.2 STATISTICAL PROCEDURE

For this analysis, the raw scores for componential phenotype measures of DD, including the rhyme oddity task, phoneme deletion task, task of auditory analysis, non-word reading, pseudohomophone judgement task, rapid picture and digit naming tasks and non-word repetition task were regressed for age to eliminate effects of age. Unstandardised residuals were saved and used as the new scores.

Haplotype data were analysed using COCAphase in the statistical package UNPHASED (Dudbridge 2003).

Correlations between componential phenotype measures in both the cases

and controls were undertaken using Spearman's Rank Correlation Coefficient analyses in SPSS version 12.

5.4.3 RESULTS

The highly correlated nature of many of the componential phenotypes makes correction for multiple testing difficult, particularly using the Bonferonni correction, which is highly conservative. As a result, p values have not been corrected for multiple testing.

The two-marker haplotype rs4504469/rs6935076 was tested for association with each of the reading components (see Tables 5.3 – 5.17). Globally, evidence for association was observed with measures of rapid automated naming, phonological awareness, phonological decoding, phonological working memory, reading rate, comprehension and accuracy, spelling and word recognition. No global evidence of association between orthographic coding and rs4504469/rs6935076 was observed. Association between rs4504469/rs6935076 and ADHD was observed with ADHD symptoms assessed with the Du Paul questionnaire (global p = 0.05) but not globally with ADHD symptoms assessed with the Connors questionnaire. However both measures of ADHD yielded evidence of association with rs4504469A/rs6939076G.

Within the rs4504469/rs6935076 haplotype, the A-G haplotype yielded the most significant evidence for association. For measures of rapid automatised naming, spelling, word recognition and one measure of phonological awareness (the task of auditory analysis), the G-A haplotype was also significantly associated. Phoneme deletion showed an association with the A-A haplotype (p = 0.047). In all instances where the A-G haplotype was significantly associated it was more frequent in controls than cases. In instances of significant association of the G-A haplotype,

the haplotype was more frequent in cases than controls. The A-A haplotype showing association with phoneme deletion, was more frequent in cases than controls.

Word Recognition Global p = 0.009				
Haplotype	**Control Frequency**	**Case Frequency**	**Chi Square**	**p**
G-G	0.26	0.27	0.04	0.84
G-A	0.30	0.38	5.78	0.02
A-G	0.39	0.26	8.81	0.003
A-A	0.05	0.09	2.39	0.12

Table 5.3 rs4504469/rs6939076 haplotype association with word recognition.

Reading Rate Global p = 0.02				
Haplotype	**Control Frequency**	**Case Frequency**	**Chi Square**	**p**
G-G	0.26	0.36	1.37	0.24
G-A	0.30	0.29	0.17	0.68
A-G	0.39	0.23	6.07	0.01
A-A	0.04	0.13	3.18	0.07

Table 5.4 rs4504469/rs6939076 haplotype association with reading rate. Note, this analysis was undertaken on children aged 12 years and under.

Reading Comprehension Global p = 0.02				
Haplotype	**Control Frequency**	**Case Frequency**	**Chi Square**	**p**
G-G	0.26	0.36	1.51	0.22
G-A	0.30	0.29	0.22	0.64
A-G	0.39	0.22	6.53	0.01
A-A	0.04	0.13	3.03	0.08

Table 5.5 rs4504469/rs6939076 haplotype association with reading comprehension. Note, this analysis was undertaken on children aged 12 years and under.

Reading Accuracy Global p = 0.02				
Haplotype	**Control Frequency**	**Case Frequency**	**Chi Square**	**p**
G-G	0.26	0.36	1.51	0.22
G-A	0.30	0.29	0.22	0.64
A-G	0.39	0.22	6.53	0.01
A-A	0.04	0.13	3.03	0.08

Table 5.6 rs4504469/rs6939076 haplotype association with reading accuracy. Note, this analysis was undertaken on children aged 12 years and under.

Spelling Global p = 0.006				
Haplotype	**Control Frequency**	**Case Frequency**	**Chi Square**	**p**
G-G	0.26	0.28	0.02	0.90
G-A	0.30	0.39	7.30	0.007
A-G	0.39	0.25	10.89	0.001
A-A	0.05	0.08	0.92	0.34

Table 5.7 rs4504469/rs6939076 haplotype association with spelling.

Rhyme Oddity Global p = 0.002				
Haplotype	**Control Frequency**	**Case Frequency**	**Chi Square**	**p**
G-G	0.26	0.34	2.80	0.09
G-A	0.30	0.35	3.09	0.08
A-G	0.39	0.23	14.35	0.0002
A-A	0.05	0.08	0.64	0.42

Table 5.8 rs4504469/rs6939076 haplotype association with rhyme oddity.

Phoneme Deletion Global p = 0.01				
Haplotype	**Control Frequency**	**Case Frequency**	**Chi Square**	**p**
G-G	0.26	0.30	0.18	0.67
G-A	0.30	0.33	2.02	0.16
A-G	0.39	0.26	7.84	0.005
A-A	0.05	0.11	3.96	0.047

Table 5.9 rs4504469/rs6939076 haplotype association with phoneme deletion.

Task of Auditory Analysis p = 0.0003				
Haplotype	**Control Frequency**	**Case Frequency**	**Chi Square**	**p**
G-G	0.26	0.32	0.94	0.33
G-A	0.30	0.38	7.26	0.007
A-G	0.39	0.21	17.61	0.00003
A-A	0.05	0.09	0.99	0.32

Table 5.10 rs4504469/rs6939076 haplotype association with the task of auditory analysis.

Non-Word Reading Global p = 0.001				
Haplotype	**Control Frequency**	**Case Frequency**	**Chi Square**	**p**
G-G	0.26	0.33	1.45	0.23
G-A	0.30	0.34	3.39	0.07
A-G	0.39	0.22	13.91	0.0002
A-A	0.05	0.10	2.04	0.15

Table 5.11 rs4504469/rs6939076 haplotype association with non-word reading.

Non-Word Repetition Global p = 0.03				
Haplotype	**Control Frequency**	**Case Frequency**	**Chi Square**	**p**
G-G	0.26	0.30	0.29	0.59
G-A	0.30	0.35	2.86	0.09
A-G	0.39	0.27	7.62	0.006
A-A	0.05	0.09	1.49	0.22

Table 5.12 rs4504469/rs6939076 haplotype association with non-word repetition.

Pseudohomophone Judgement Task Global p = 0.10				
Haplotype	**Control Frequency**	**Case Frequency**	**Chi Square**	**p**
G-G	0.26	0.33	1.64	0.20
G-A	0.30	0.29	0.11	0.74
A-G	0.39	0.29	4.09	0.04
A-A	0.05	0.09	1.17	0.28

Table 5.13 rs4504469/rs6939076 haplotype association with the pseudohomophone judgement task.

Rapid Digit Naming Global p = 0.001				
Haplotype	**Control Frequency**	**Case Frequency**	**Chi Square**	**p**
G-G	0.26	0.31	0.51	0.47
G-A	0.30	0.37	5.90	0.02
A-G	0.39	0.22	14.38	0.0001
A-A	0.05	0.10	1.68	0.19

Table 5.14 rs4504469/rs6939076 haplotype association with rapid digit naming.

Rapid Picture Naming Global p = 0.001				
Haplotype	**Control Frequency**	**Case Frequency**	**Chi Square**	**p**
G-G	0.26	0.31	1.14	0.29
G-A	0.30	0.39	7.37	0.007
A-G	0.39	0.23	15.88	0.00007
A-A	0.05	0.07	0.14	0.71

Table 5.15 rs4504469/rs6939076 haplotype association with rapid picture naming.

ADHD - Connors' Questionnaire Global p = 0.13				
Haplotype	**Control Frequency**	**Case Frequency**	**Chi Square**	**p**
G-G	0.26	0.28	0.01	0.92
G-A	0.30	0.38	2.84	0.09
A-G	0.39	0.26	4.82	0.03
A-A	0.05	0.09	0.79	0.38

Table 5.16 rs4504469/rs6939076 haplotype association with ADHD symptoms using the Connors' questionnaire.

ADHD - Du Paul Questionnaire Global p = 0.05				
Haplotype	**Control Frequency**	**Case Frequency**	**Chi Square**	**p**
G-G	0.26	0.26	0.09	0.77
G-A	0.30	0.38	3.75	0.05
A-G	0.39	0.26	5.57	0.02
A-A	0.05	0.10	2.08	0.15

Table 5.17 rs4504469/rs6939076 haplotype association with ADHD symptoms using the Du Paul questionnaire.

5.4.3.1 NON-PARAMETRIC CORRELATION ANALYSIS

Correlations between the scores for each componential phenotype were assessed using Spearman's rank correlation coefficients (see Tables 5.18 and 5.19). This was deemed the most appropriate method given the non-normal distribution of many of the traits (see Appendix 2 for graphical representations of each component phenotype distribution). Correlations were analysed in cases and controls separately.

Component Phenotype	**Reading Accuracy**	**Reading Comprehension**	**Reading Rate**
Reading Accuracy		0.46	0.42
Reading Comprehension	0.78		0.38
Reading Rate	0.46	0.38	

Table 5.18 Spearman's rank correlation coefficients between reading componential phenotypes. Note these phenotypes were only assessed in children 12 years and under. Correlations above the diagonal are for controls (blue), those below the diagonal represent correlations in the case sample (black).

Component Phenotype	Rapid Picture Naming	Rapid Digit Naming	Task of Auditory Analysis	Phoneme Deletion	Rhyme Oddity	ADHD: Du Paul	Spelling	ADHD: Connors	Non-Word Reading	Pseudohomophone Judgement Task	Non-Word Repetition	IQ	Single Word Reading
Rapid Picture Naming		0.51	0.03	-0.04	0.19	-	0.06	-	-0.09	-0.20	-0.05	-0.05	0.15
Rapid Digit Naming	0.65		0.17	-0.06	0.16	-	0.38	-	-0.19	-0.16	0.03	0.04	0.02
Task of Auditory Analysis	-0.32	-0.29		0.35	0.32	-	0.02	-	0.19	0.20	0.24	0.08	-0.12
Phoneme Deletion	-0.17	-0.21	0.51		0.25	-	-0.19	-	0.19	0.33	0.06	0.07	-0.21
Rhyme Oddity	-0.32	-0.34	0.48	0.38		-	-0.07	-	0.02	0.13	0.18	0.11	-0.05
ADHD: Du Paul	0.01	-0.01	0.11	0.07	-0.04		-	-	-	-	-	-	-
Spelling	0.27	0.28	-0.23	-0.19	-0.14	0.05		-	-0.29	-0.25	0.03	-0.01	0.14
ADHD: Connors	-0.02	-0.04	0.16	0.14	0.09	0.73	0.08		-	-	-	-	-
Non-Word Reading	-0.22	-0.24	0.53	0.46	0.44	-0.07	-0.08	-0.01		0.23	0.15	0.12	-0.25
Pseudohomophone Judgement Task	-0.15	-0.13	0.17	0.18	0.18	0.07	-0.26	0.10	0.22		-0.08	0.16	-0.35
Non-Word Repetition	-0.09	-0.11	0.09	0.24	0.25	-0.02	-0.07	-0.06	0.29	0.10		0.03	-0.05
IQ	-0.03	-0.02	0.11	0.11	0.03	0.21	-0.09	0.11	0.00	0.13	-0.01		-0.12
Single Word Reading	0.21	0.26	-0.24	-0.23	-0.26	-0.06	0.72	0.07	-0.26	-0.24	-0.16	-0.10	

Table 5.19 Spearman's rank correlation coefficients between componential phenotypes.
Cases located below the diagonal, controls above the diagonal. Coefficients in pink are significant at the 0.01 level. (2-tailed). Coefficients in blue are significant at the 0.05 level (2-tailed). Note: ADHD symptoms were only available for cases.

5.4.4 DISCUSSION

The results from the categorical analysis of componential phenotypes of DD and reading revealed that associations do exist between measures of DD and rs4504469/rs6939076. The haplotype rs4504469/rs6939076 is significantly associated with a number of components of reading including measures of rapid automated naming, phonological awareness, phonological decoding, phonological working memory, reading rate, comprehension and accuracy, spelling and word recognition. In addition, although not globally significant, the A-G haplotype was associated with orthographic coding. In children 12 years and under, rs4504469/rs6939076 was associated with reading accuracy, rate and comprehension.

The most significant association observed with rs4504469/rs6939076 remains from Chapter four, defining DD categorically as a lag in reading ability (global p = 0.0001). However, although less significant globally (p = 0.0003), the A-G haplotype and the task of auditory analysis showed the same level of significance as DD as a lag in reading (p = 0.00003). The A-G haplotype yielded some evidence of association with all component phenotypes analysed, although the level of significance varied. This makes it difficult to answer the question of whether component phenotypes of DD better explain the association observed in Chapter four. The results obtained suggest that rs4504469/rs6939076 is associated with a number of component phenotypes but the relative effects on each component phenotype are difficult to measure, making further refinement of the DD association of Chapter four difficult.

The results support the previous studies of Grigorenko, Fisher, Gayán and colleagues (Grigorenko et al. 1997; Fisher et al. 1999; Gayán et al. 1999) who

found linkage and/or association between DYX2 and component measures of phonological decoding ability, rapid automatised naming, single word reading, a discrepancy measure based on vocabulary, orthographic coding, an IQ-reading discrepancy score and phonological awareness. All three studies used a categorical definition for each component phenotype measure.

5.4.4.1 LIMITATIONS OF THE STUDY

In categorical analyses the arbitrary nature of the cut-off scores used to define categories can result in the differing observations of association between studies due to the inclusion or exclusion of cases on either side of the cut-off score. Grigorenko (Grigorenko et al. 1997) suggested that the population-based cut score employed in their analysis meant the number of individuals for each subphenotype varied, possibly influencing power between reading components for linkage to be detected. Our study overcomes this problem since the worst 100 DD cases for each measure are included and therefore maintains sample size to allow some power to detect associations however, the cut-off scores used mean for each reading component different percentages of low performers are used. The cut scores employed by Field and Kaplan (Field and Kaplan 1998) (a deficit of two years in a number of tests) could explain the absence of linkage to DYX2 in their sample as this will include only the most severe cases for each trait and reduce the sample size and therefore the power to detect association.

The categorisation approach of using age-referenced percentiles was not possible in our sample since a large sample representing normal reading ability across all age ranges was not available to calculate percentiles. The use of a set number of standard deviations below the mean was also deemed an inappropriate

method of categorisation in this sample as this relies on the mean as an accurate measure of central tendency. Given the non-normality of many of the component phenotypes, the median would be a better measure of central tendency, yielding the use of standard deviations inappropriate.

Adjustment of the α-level to account for multiple testing assumes that the tests in question are independent. In our data this assumption is violated given the correlation between a number of reading components. Any correction, to avoid being too overly conservative, would need to take into account these factors.

5.5 ANALYSIS TWO – A QUANTITATIVE PHENOTYPIC ANALYSIS OF CHROMOSOME 6p

The use of quantitative data may enhance power to detect associations compared to categorical analysis as information is not lost in the categorisation process by using arbitrary cut-offs.

Kaplan and colleagues (Kaplan et al. 2002) looked at quantitative measures of reading components with markers on chromosome 6. Significant associations were observed with measures of orthographic coding and phonological awareness. Associations were particularly observed with the marker JA04 located within the 5'UTR of *KIAA0319*, suggesting that variation in or around *KIAA0319* may influence normal ability in orthographic coding ability and/or phonological awareness ability.

Both Deffenbacher and colleagues (Deffenbacher et al. 2004) and Francks and colleagues (Francks et al. 2004) have utilised quantitative measures of component processes of reading. Like Kaplan and colleagues (Kaplan et al. 2002), Deffenbacher and colleagues (Deffenbacher et al. 2004) observed association between SNPs on 6p and quantitative measures of orthographic coding and phonological awareness but also with phonological decoding, word recognition and a reading discrepancy score measuring reading ability in general. Francks and colleagues (Francks et al. 2004) also observed association with quantitative measures of orthographic coding and phonological awareness and like Deffenbacher and colleagues (Deffenbacher et al. 2004) observed association with phonological decoding and word recognition and in addition, spelling ability.

Given these replicated observations my aim was to assess whether

continuous measures of DD were associated with the haplotype rs4504469/rs6939076 spanning *KIAA0319*.

5.5.1 AIMS AND HYPOTHESES

The quantitative analysis was undertaken in order to answer the questions:

1) Do the associations observed with categorical components of reading extend to continuous measures of the same phenotypes?
2) Does the haplotype rs4504469/rs6939076 influence reading ability in general (across all reading abilities), suggesting DD lies on a continuum with normal reading ability, or does variation on the haplotype influence cases and/or controls differently, suggesting DD forms a group pathologically distinct from normal readers?
3) Is there evidence of independent replication of the association observed between groups in analysis one using a quantitative within groups design?

5.5.2 MATERIALS AND METHODS

5.5.3.1 PARTICIPANTS AND TEST BATTERY

The 101 control individuals (for whom component phenotype data were available) and 241 DD-probands included in the categorical analysis were included in this quantitative analysis of componential phenotypes. In addition, 207 individuals forming a DD intermediate phenotype (i.e. with an IQ $\geq$ 85 and a reading age 6 months to 2.5 years below that expected from their chronological age) were included in the analysis (see Chapter three) to gain a sample more representative of the general population. The individuals with intermediate phenotype were genotyped at SNPs rs4504469 and rs6939076 as described in Chapter four, Section

.4.5. The demographics of this sample are shown in Chapter three.

5.5.3.2 SUBPHENOTYPIC MEASURES

Componential phenotype measures (regressed for age since many are age dependant – see Section 5.4.2.2) were used as continuous phenotypic variables in order for significant differences between groups to be tested.

5.5.3.3 STATISTICAL PROCEDURE

Raw scores for componential measures of DD were age regressed and unstandardised residuals used as the component phenotype score. These were then tested for normality using measures of skewness and kurtosis. For a phenotype to be classified as 'normally distributed' values of both skewness and kurtosis were required to lie between –1 and +1. When this did not occur, attempts were made to transform the data and skewness and kurtosis tested on transformed data. If distributions could not be transformed to form a normal distribution, further analyses of these phenotypes required non-parametric tests to be used as an alternative to the parametric tests used for analysis of phenotypes with a normal distribution.

For the analysis of the rs4504469/rs6935076 haplotype, the probability of each person having a particular phased diplotype was calculated. In order to determine this probability, the frequency of each haplotype was estimated using EHPlus (Zhao et al. 2000). The probability of each haplotype was calculated from the genotypes of individuals using the formula:

$$\frac{F_1 \times F_4}{(F_1 \times F_4)(F_2 \times F_3)}$$

Where F_1 is the frequency of one possible haplotype from an individuals genotype, F_4 is the frequency of another possible haplotype given the first haplotype is present. F_2 and F_3 are the frequencies of the other two possible haplotypes that could be from the two SNPs. The probability of phased diplotypes for each individual was calculated using the formula:

$$1 - \frac{F_1 \times F_4}{(F_1 \times F_4)(F_2 \times F_3)}$$

As a result in any individual there are four possible phased diplotypes, although some have probabilities equal to zero if a diplotype is not possible for a given set of genotypes. Phased diplotype probabilities were determined in a program written by Dr Valentina Moskvina (HADES program available from http://www.cardiff.ac.uk/medicine/psychological_medicine/index.htm). The phased diplotype probabilities were then used as the independent variable in a linear or binary logistic regression depending on the underlying distribution of the component phenotype.

Association between component phenotypes of DD and rs4504469/rs6939076 was tested using linear regression. Component phenotypes normally distributed (single word reading, spelling, rhyme oddity, task of auditory analysis, non-word repetition and IQ), were tested for association with rs4504469/rs6939076 using linear regression. Continuous measures of the reading related measures were utilised as the dependant variable and the probability of each phased diplotype as the independent variable. Component phenotype scores not normally distributed (phoneme deletion, rapid picture and digit naming, pseudohomophone judgement task and non-word reading) were tested for

association to rs4504469/rs6939076 using binary logistic regression. Individuals were classed as affected if they scored above (rapid digit naming and rapid picture naming) or below (pseudohomophone judgement task, non-word reading and phoneme deletion task) the median, depending on whether a high score was deemed good or bad. All other individuals were classified as normal for that phenotype score. The dependant variable was the 'affected' or 'unaffected' status of the individual and the covariate, the probability of each phased diplotype.

All statistical tests were undertaken using the Statistical Package for the Social Sciences (SPSS) version 12.0.1 for windows.

5.5.4 RESULTS

The reading ability distribution of the sample utilised in this quantitative analysis of DD is shown in Figure 5.1. The distribution is normally distributed (skewness = -0.061, kurtosis = -0.371).

The data obtained from the phoneme deletion task, pseudohomophone judgement task, non-word reading, rapid picture naming and rapid digit naming were not normally distributed (see Table 5.20). Despite attempts to normalise the data using various transformations, normality could not be achieved. Appendix 2 shows the distribution of each component phenotype after regression for age.

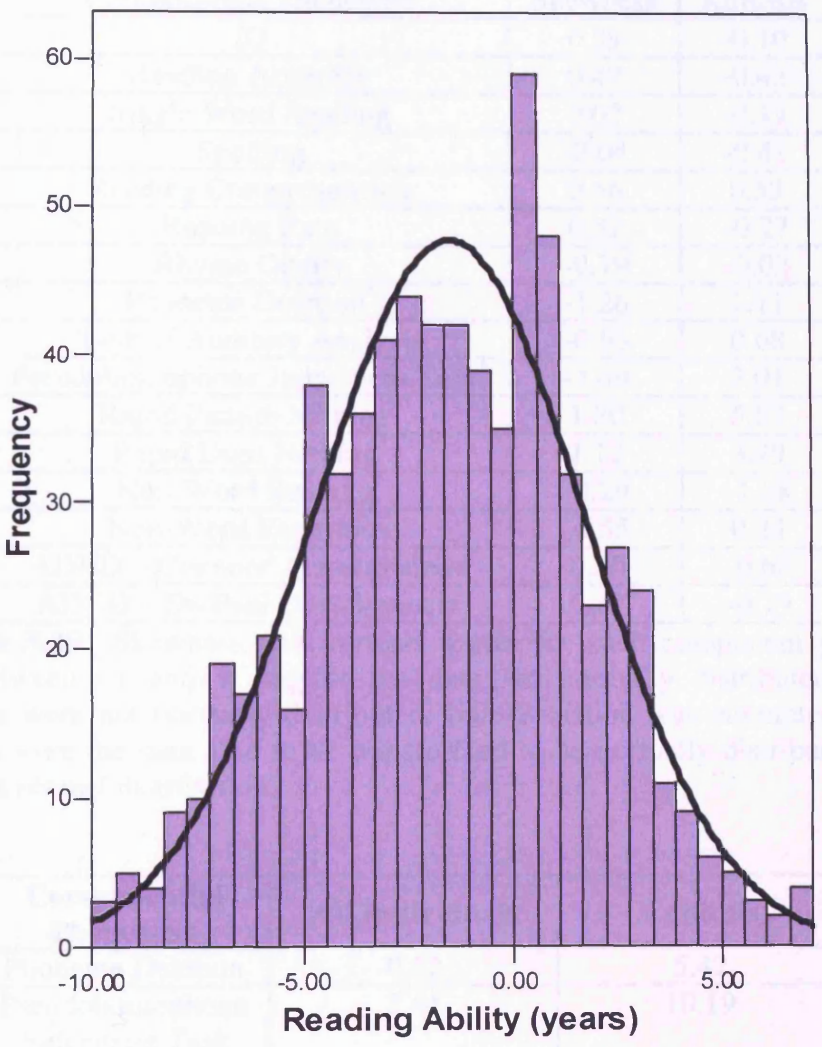


Figure 5.1 The reading disability distribution of the individuals used in the quantitative analysis of component phenotypes of DD. This covers the whole reading ability spectrum. A positive value represents a reading age above that expected from chronological age, a negative value represents a reading age behind chronological age.

For all analyses, association was tested in all individuals grouped together (regardless of DD definition previously i.e. DD case, control or of intermediate phenotype) and in controls and cases separately.

Componential Phenotype	Skewness	Kurtosis
IQ	0.59	-0.10
Reading Accuracy	0.49	-0.43
Single Word Reading	0.02	-0.17
Spelling	-0.04	-0.41
Reading Comprehension	0.66	0.53
Reading Rate	0.32	-0.27
Rhyme Oddity	-0.79	-0.02
Phoneme Deletion*	-1.26	2.11
Task of Auditory Analysis	-0.95	0.68
Pseudohomophone Judgement Task*	-1.49	3.01
Rapid Picture Naming*	1.80	5.57
Rapid Digit Naming*	1.12	3.49
Non-Word Reading*	-0.29	-1.08
Non-Word Repetition	-0.55	0.24
ADHD – Connors' Questionnaire	0.50	-0.60
ADHD – Du Paul Questionnaire	0.39	-0.79

Table 5.20 Skewness and kurtosis scores for each component phenotype. Values between −1 and 1 suggest the data are normally distributed. Where phenotypes were not normally distributed, transformation was attempted. For no raw scores were the data able to be transformed to be normally distributed. * Not following a normal distribution.

Componential Phenotype	**All Individuals**	**Controls**
Phoneme Deletion	0.32	5.42
Pseudohomophone Judgement Task	2.41	10.19
Rapid Digit Naming	-0.99	-7.63
Rapid Picture Naming	-1.43	-7.63
Non-Word Reading	-0.44	12.45

Table 5.21 Medians for componential phenotypes not following a normal distribution

Global evidence of association was observed between rs4504469/rs6935076 and single word reading, phoneme deletion, non-word reading, pseudohomophone judgement task and rapid digit naming (see Table 5.22). In children 12 years and under, association between rs4504469/rs6935076 and reading rate and reading accuracy was observed (see Table 5.22).

Regarding the specific haplotypes, reading rate (under 12 years only),

phoneme deletion, non-word reading, pseudohomophone judgement task and rapid digit naming showed association with all haplotypes. For each of these component phenotypes the most significant haplotype was the A-G haplotype. Reading accuracy in children less than 12 years showed association with only the A-A haplotype. Single word reading was specifically associated with the A-G and A-A haplotypes although the A-G haplotype was the most significant.

Analysis of controls (see Table 5.23) separately did not reveal any evidence for global association between continuous measures of reading related processes and rs4504469/rs693076. Single word reading yielded evidence of association with the A-G haplotype in control individuals. In particular the A-G haplotype was associated with better reading.

No evidence for association was observed between rs4504469/rs6935076 and component measures of DD in cases analysed separately (see Table 5.24).

Component Phenotype	Global p	Reference Haplotype	Haplotype	B	p
Single Word Reading	**0.000004**	G-A	G-G	0.16	0.78
			A-G	-1.91	0.0001
			A-A	2.56	0.01
Reading Rate*	**0.0001**	G-G	G-A	-1.70	0.02
			A-G	-2.64	0.0004
			A-A	3.16	0.006
Reading Comprehension*	0.20	G-G	G-A	-0.42	0.49
			A-G	-0.86	0.17
			A-A	1.50	0.10
Reading Accuracy*	**0.01**	G-A	G-G	1.09	0.10
			A-G	-0.73	0.22
			A-A	2.95	0.01
Spelling	0.36	G-A	G-G	-0.08	0.90
			A-G	-0.87	0.10
			A-A	0.48	0.65
Rhyme Oddity	0.06	G-A	G-G	-1.64	0.16
			A-G	1.51	0.13
			A-A	-2.68	0.18
Phoneme Deletion	**0.03**	A-A	G-G	-1.41	0.05
			G-A	-2.05	0.01
			A-G	-2.42	0.005
Task of Auditory Analysis	0.07	G-G	G-A	-1.05	0.41
			A-G	1.78	0.10
			A-A	-3.72	0.10
Non-Word Reading	**0.02**	A-A	G-G	-1.53	0.04
			G-A	-2.20	0.01
			A-G	-2.57	0.003
Non-Word Repetition	0.61	G-A	G-G	-1.20	0.42
			A-G	-0.35	0.79
			A-A	-3.43	0.18
Pseudohomophone Judgement Task	**0.02**	A-A	G-G	-1.46	0.06
			G-A	-2.28	0.008
			A-G	-2.68	0.004
Rapid Digit Naming	**0.03**	A-A	G-G	-1.78	0.02
			G-A	-1.82	0.03
			A-G	-2.47	0.005
Rapid Picture Naming	0.08	A-A	G-G	-0.31	0.65
			G-A	-0.26	0.74
			A-G	-1.22	0.15
IQ	0.79	G-A	G-G	-0.55	0.83
			A-G	-2.16	0.34
			A-A	-1.39	0.76

Table 5.22 A quantitative analysis of each component phenotype including all individuals covering the whole reading ability spectrum. *Includes children 12 years and under only.

Component Phenotype	Global p	Reference Haplotype	Haplotype	B	p
Single Word Reading	0.18	G-A	G-G	0.50	0.31
			A-G	0.95	0.03
			A-A	0.14	0.89
Reading Rate*	0.33	G-A	G-G	1.12	0.26
			A-G	-0.37	0.65
			A-A	3.15	0.18
Reading Comprehension*	0.52	G-A	G-G	0.94	0.18
			A-G	0.64	0.28
			A-A	0.30	0.83
Reading Accuracy*	0.74	A-A	G-G	0.09	0.83
			A-G	-0.15	0.67
			G-A	0.85	0.33
Spelling	0.59	G-A	G-G	0.46	0.57
			A-G	0.66	0.28
			A-A	-0.91	0.53
Rhyme Oddity	0.70	G-A	G-G	0.01	0.99
			A-G	0.38	0.63
			A-A	-1.85	0.32
Phoneme Deletion	0.55	A-A	G-G	1.45	0.37
			G-A	0.68	0.69
			A-G	1.52	0.38
Task of Auditory Analysis	0.84	G-G	G-A	0.00	1.00
			A-G	-0.64	0.42
			A-A	0.21	0.91
Non-Word Reading	0.46	A-A	G-G	-2.49	0.15
			G-A	-2.36	0.20
			A-G	-2.13	0.25
Non-Word Repetition	0.78	G-A	G-G	-0.95	0.74
			A-G	-1.13	0.60
			A-A	-4.57	0.35
Pseudohomophone Judgement Task	0.73	A-A	G-G	0.72	0.64
			G-A	-0.28	0.86
			A-G	-0.08	0.96
Rapid Digit Naming	0.33	A-A	G-G	-0.52	0.67
			G-A	-1.56	0.24
			A-G	-1.75	0.20
Rapid Picture Naming	0.28	A-A	G-G	0.68	0.66
			G-A	-0.02	0.99
			A-G	-0.88	0.60
IQ	0.47	G-A	G-G	-5.07	0.32
			A-G	-4.99	0.20
			A-A	3.58	0.69

Table 5.23 A quantitative analysis of each component phenotype including only individuals classified as controls. *Includes children 12 years and under only.

Component Phenotype	p	Reference Haplotype	Haplotype	B	p
Single Word Reading	0.53	G-A	G-G	-0.85	0.16
			A-G	-0.07	0.90
			A-A	-0.47	0.65
Reading Rate*	0.36	G-G	G-A	-0.54	0.53
			A-G	-1.63	0.09
			A-A	-0.05	0.96
Reading Comprehension*	0.91	G-A	G-G	-0.36	0.66
			A-G	-0.64	0.47
			A-A	0.20	0.85
Reading Accuracy*	0.59	G-G	G-A	-0.55	0.23
			A-G	-0.50	0.31
			A-A	0.34	0.58
Spelling	0.40	G-A	G-G	-0.99	0.10
			A-G	-0.28	0.61
			A-A	-1.31	0.20
Rhyme Oddity	0.61	G-A	G-G	-0.67	0.65
			A-G	1.38	0.31
			A-A	-0.49	0.85
Phoneme Deletion	0.52	A-A	G-G	-1.03	0.24
			G-A	-1.17	0.26
			A-G	-0.76	0.50
Task of Auditory Analysis	0.56	G-G	G-A	-0.62	0.70
			A-G	0.72	0.68
			A-A	-3.02	0.23
Non-Word Reading	0.80	A-A	G-G	-0.35	0.68
			G-A	-0.27	0.79
			A-G	-0.78	0.48
Non-Word Repetition	0.85	G-A	G-G	-0.89	0.64
			A-G	-1.01	0.57
			A-A	-2.40	0.46
Pseudohomophone Judgement Task	0.51	A-A	G-G	-0.54	0.57
			G-A	-0.81	0.47
			A-G	-0.05	0.97
Rapid Digit Naming	0.31	A-A	G-G	-1.16	0.20
			G-A	-0.68	0.52
			A-G	-1.50	0.19
Rapid Picture Naming	0.16	A-A	G-G	0.00	1.00
			G-A	0.56	0.59
			A-G	-0.67	0.55
IQ	0.80	G-A	G-G	1.64	0.63
			A-G	0.91	0.77
			A-A	-2.93	0.62
ADHD - Connors' Questionnaire	0.96	A-A	G-G	-0.01	1.00
			G-A	-0.47	0.74
			A-G	-0.39	0.79
ADHD - Du Paul Questionnaire	0.65	A-A	G-G	0.96	0.39
			G-A	0.27	0.84
			A-G	0.18	0.90

Table 5.24 A quantitative analysis of each component phenotype including only individuals classified as DD cases. *Includes children 12 years and under only.

5.5.5 DISCUSSION

Association between continuous measures of component phenotypes of DD and rs4504469/rs6939076 was tested in order to answer the question of whether association observed with categorical definitions of component phenotypes extends to continuous measures of the same phenotypes and to try and independently replicate the associations observed between groups in analysis one using a quantitative within groups design.

When a sample across the reading ability spectrum was tested for association with component measures of DD, global evidence of association was observed between rs4504469/rs6939076 and single word reading, reading rate, reading accuracy, phoneme deletion, non-word reading, pseudohomophone judgement task and rapid digit naming. These observations suggest that continuous measures may be influenced by variation on the rs4504469/rs6939076 haplotype. However, doubt is cast on this observation when controls are considered separately.

No global evidence of association was observed between continuous componential phenotypes and rs4504469/rs6939076 in controls. This would suggest that the association observed in analysis of the whole reading ability spectrum may reflect the underlying categorical association observed in analysis one and would further suggest that DD forms a distinct pathological group and does not represent the lower tail of a normal reading ability continuum.

If variation on the rs4504469/rs6939076 haplotype did influence reading ability in general, an association would be expected in the controls sample, as seen in the whole sample.

Given the evidence of quantitative association between single word reading and rs4504469/rs6939076, it is of interest that an association is also observed

between single word reading and the rs4504469A/rs6939076G haplotype in the controls. This could imply that ability in single word reading (a measure often used in the diagnosis of DD) is influenced by variation on the rs4504469A/rs6939076G haplotype. This result and those of other component phenotype analyses could suggest that DD (as defined by single word reading), does lie on a continuum with normal reading ability but that other component processes of reading do not fall on such a continuum and problems with any of the components results in a pathological group. However, it is noteworthy that the direction of effect is different in the whole sample compared to that in just the controls for single word reading. In the whole sample the A-G haplotype is associated with poor single word reading ability whereas in the controls it is associated with better performance in single word reading. The association could be the result of multiple testing since p-values would not stand correction for multiple testing based on Bonferroni correction.

Previous studies have shown association between variation within *KIAA0319* and continuous measures of component phenotypes of DD, particularly orthographic coding, phonological awareness, phonological decoding, word recognition and a discrepancy score measuring overall reading ability to the region (Kaplan et al. 2002; Deffenbacher et al. 2004; Francks et al. 2004). The overlap between the associations observed by Deffenbacher and Francks and colleagues (Deffenbacher et al. 2004; Francks et al. 2004) and our continuous analysis suggest that all component phenotypes may be influenced by genetic variation in the region. It is also possible that a variation may influence a component of reading which is important in all aspects of reading component phenotype.

In order to independently replicate the between groups association observed in analysis one, DD cases and controls were considered separately and a quantitative

analysis undertaken within each group. Only one association was replicated in analysis two, that of single word reading in the controls. The absence of replication in this analysis could be due to:

1) The results in analysis one represent type I errors
2) These results represent type II errors. The sample size of DD cases and controls may lack power to detect subtle quantitative differences that may have been observed in the case-control design. Additionally, the non-normal distribution of some of the quantitative measures resulted in the use of non-parametric tests which can also lower the power of an association study.

5.5.5.1 LIMITATIONS OF THE STUDY

Quantitative analysis assumes the componential phenotypes under study are normally distributed. Often, these measures do not follow such a distribution and so violate the assumptions of such an analysis. Although steps to overcome the problem can be taken often this does not result in the normalisation of the data. Consequently, as in this study, the continuous nature of some of the component phenotypes is lost in the categorisation approach to allow analysis by binary logistic regression. However, the regression technique did allow the whole reading ability spectrum to be included in the analysis.

Another limitation in the quantitative study is that of power. There may be a lack of power to detect subtle associations in the sample used. Indeed the sample of controls used in this analysis is particularly small. In addition, unlike reading ability, which is normally distributed across the sample used in the quantitative study, a normal distribution of component phenotype scores is not represented in the sample (see distributions in Appendix 2). This could affect the power to detect

associations, with over-representation of one end of the distribution and under representation of scores at the other end of the distribution.

5.6 ANALYSIS THREE – ASSOCIATION BETWEEN rs4504469 AND rs6939076 AND ATTENTION DEFICIT HYPERACTIVITY DISORDER

In analysis one an association was observed between ADHD symptoms and rs4504469/rs6935076. Global haplotypic association was observed with ADHD symptoms reported using the Du Paul questionnaire and the specific haplotypes G-A and A-G. The haplotype A-G was associated with ADHD symptoms reported by the Connors' questionnaire of parent reported symptoms.

This observation is interesting since twin studies have suggested the possibility of common genetic influences predisposing children to both ADHD and DD (Gilger et al. 1992; Willcutt et al. 2000). Further, Willcutt and colleagues (Willcutt et al. 2002) have shown, through linkage analyses, that the DYX2 locus may in part explain the comorbidity of DD and ADHD. The DYX2 locus lies next to the HLA region on chromosome 6p, which has shown some evidence for association with ADHD (Warren et al. 1995; Odell et al. 1997).

Given this information, the haplotype rs4504469/rs6035976 was tested for association with DD without comorbid ADHD and in a sample of parent ADHD-proband trios, with aims of answering the questions below (see Section 5.6.1).

5.6.1 AIMS AND HYPOTHESIS

The aim of this analysis was to answer the questions:

1) Given the association between DD and comorbid ADHD and rs4504469/rs6935076 in Section 5.4.3, is the association observed in Chapter four the result of the presence of ADHD within the DD sample or is the association the result of DD (or related components) *per se*?

2) Does variation in *KIAA0319* influence susceptibility to ADHD given the tentative association observed in analysis one between rs4504469/rs6939076 and ADHD symptoms and the weak evidence for linkage between chromosome 6p and ADHD (Warren et al. 1995; Odell et al. 1997)?

In order to determine whether *KIAA0319* influences DD *per se* and/or susceptibility to ADHD, two analyses were undertaken. If variation on the haplotype rs4504469/rs6935076 influences DD, an association should be observed even if DD-probands who also have comorbid ADHD are removed from either a case-control or family-based association analysis. If the presence of ADHD in DD-probands is involved in the association, no association should be observed in the DD-probands without comorbid ADHD. This forms the basis of analysis 3A.

Analysis 3B tests directly for association between association between ADHD and rs4504469/rs6935076.

5.6.2 ANALYSIS THREE A – A CASE-CONTROL AND FAMILY BASED TEST OF ASSOCIATION BETWEEN DEVELOPMENTAL DYSLEXIA WITHOUT COMORBID ADHD AND rs4504649/rs6935076

5.6.2.1 PARTICIPANTS AND TEST BATTERY

A sample of 142 DD cases without ADHD symptoms reported on the Du Paul questionnaire (see Chapter three) and a sample of 172 DD cases without ADHD symptoms reported on the Connors' questionnaire (see Chapter three) were used to test for association with rs4504469/rs6935076 in conjunction with a control sample of 273. It should be noted that the two samples of DD cases without

comorbid ADHD are not independent samples.

Further analyses were undertaken using 72 parent DD-proband trios where the proband did not have comorbid ADHD based on ADHD symptoms reported using the Du Paul questionnaire and 60 parent DD-proband trios where the proband did not have ADHD based on ADHD symptoms reported on the Connors' questionnaire.

5.6.2.2 VARIATIONS AND GENOTYPING

The SNPs rs4504469 and rs6935076 were genotyped using the methods described in Chapter four.

5.6.2.3 STATISTICAL PROCEDURE

Association was tested in the case-control sample using the program COCAphase in the software package UNPHASED (Dudbridge 2003). Association in parent DD-proband trios was tested using TDTphase, also part of the UNPHASED software package (Dudbridge 2003).

5.6.3 RESULTS

Table 5.25 shows an association between DD and rs4504469/rs6935076 in the DD case-control sample where cases were excluded if they scored above 35 on the Du Paul questionnaire. Table 5.26 shows the same result but where cases were excluded if they scored 15 or more on the Connors' ADHD questionnaire. Both analyses revealed global evidence of association, with the A-G haplotype showing the most significant result. This association was in the same direction (i.e. was more common in controls) as that reported in Chapter four (see Tables 4.20 and 4.21).

Haplotype	Frequency in Cases	Frequency in Controls	χ^2	p
G-G	0.34	0.26	1.77	0.18
G-A	0.35	0.29	2.74	0.098
A-G	0.23	0.38	10.23	**0.001**
A-A	0.08	0.05	0.29	0.59

Table 5.25 Two-marker (rs4504469/rs6939076) haplotype association with DD in the case-control sample. ADHD symptoms were measured using the Du Paul questionnaire. The analysis contains 142 cases with DD and without ADHD and 273 controls. Globally the two marker haplotype yielded evidence for association with DD (p = 0.015).

Haplotype	Frequency in Cases	Frequency in Controls	χ^2	p
G-G	0.32	0.26	0.74	0.39
G-A	0.35	0.30	3.30	0.069
A-G	0.24	0.39	10.14	**0.001**
A-A	0.09	0.05	1.28	0.26

Table 5.26 Two-marker (rs4504469/rs6939076) haplotype association with DD in the case-control sample. ADHD symptoms were measured using the Connors' questionnaire. The analysis contains 172 cases with DD and without ADHD and 273 controls. Globally the two marker haplotype yielded evidence for association with DD (p = 0.01).

Tables 5.27 and 5.28 show association between rs4504469/rs6935076 and DD in a sample of parent DD-proband trios. Table 5.27 shows the association when probands with DD are excluded if they have a diagnosis of ADHD based on the Du Paul questionnaire and Table 5.28 shows the same analysis with probands excluded if they have ADHD based on Connors' questionnaire reported ADHD symptoms.

Haplotype	Frequency of Transmitted Haplotype	Frequency of Non-Transmitted Haplotype	p
G-G	0.36	0.29	0.25
G-A	0.36	0.24	**0.04**
A-G	0.22	0.39	**0.004**
A-A	0.06	0.08	0.48

Table 5.27 Two-marker (rs4504469/rs6939076) haplotype association with DD in the parent DD-proband trio sample. A sample of 72 parent DD-proband trios is included in the analysis. ADHD diagnosis was determined using the Du Paul questionnaire. Globally the two-marker haplotype yielded evidence of association with DD (p = 0.02).

Haplotype	Frequency of Transmitted Haplotype	Frequency of Non-Transmitted Haplotype	p
G-G	0.33	0.31	0.70
G-A	0.37	0.21	**0.004**
A-G	0.23	0.39	**0.0039**
A-A	0.07	0.09	0.57

Table 5.28 Two-marker (rs4504469/rs6939076) haplotype association with DD in the parent DD-proband trio sample. A sample of 60 parent DD-proband trios was used for this analysis. ADHD diagnosis was determined using the Connors' questionnaire. Globally the two-marker haplotype yielded evidence of association with DD ($p = 0.0098$).

In both association analyses using parent DD-proband trios, the direction of the effect was in the same direction for the G-A and A-G haplotypes as observed in the case-control sample (see Tables 5.25 and 5.26) and in Chapter four. All studies suggest the G-A haplotype is more common in DD cases whereas the A-G haplotype is more common in controls.

5.6.4 DISCUSSION

In analysis one an association was observed between DD cases with ADHD and rs4504469/rs6935076. This suggested the possibility that the association observed in Chapter four could be the result of the co-occurrence of ADHD in the DD cases. If the observed association is due to ADHD presence, when association is tested between rs4504469/rs6935076 and DD cases without ADHD, no significant association should be observed. The results of analysis 3A suggest that the association detected in Chapter four is due to DD or components of DD rather than ADHD.

Although the association with DD remains when DD cases with comorbid ADHD are removed and suggests the observed association is with DD, it does not imply that component processes common to both DD and ADHD are not influenced

by variation on the rs4504469/rs6935076 haplotype or indeed that ADHD itself is not influenced by the same (or different) variation on the same haplotype.

In order to determine whether ADHD *per se* is associated with the same haplotype a second association analysis was undertaken (see Section 5.6.5).

5.6.5 ANALYSIS THREE B – AN ASSOCIATION STUDY BETWEEN ADHD AND rs4504469/rs6935076 IN A SAMPLE OF PARENT ADHD-PROBAND TRIOS

5.6.5.1 PARTICIPANTS AND TEST BATTERY

A sample of 144 parent ADHD-proband trios and 115 single parent ADHD-proband duos were ascertained as described in Chapter three.

5.6.5.2 VARIATIONS AND GENOTYPING

The SNPs rs4504469 and rs6935076 were genotyped using the methods described in Chapter four.

5.6.5.3 STATISTICAL PROCEDURE

Association between ADHD and rs4504469/rs6935076 was tested using TDTphase, part of the UNPHASED software (Dudbridge 2003).

5.6.6 RESULTS

No evidence for global haplotype association was observed between rs4504469/rs6935076 and ADHD (see Table 5.29). Neither was any association observed between specific haplotypes (under 1 df) and ADHD.

Haplotype rs4504469/rs6939076	Frequency of Transmitted Haplotype	Frequency of Non-Transmitted Haplotype	p
G-G	0.27	0.29	0.39
G-A	0.37	0.36	0.73
A-G	0.29	0.29	0.96
A-A	0.07	0.06	0.45

Table 5.29 Haplotype analysis of rs4504469/rs6939076 in ADHD proband families. No evidence of global association was observed (p = 0.78).

5.6.7 DISCUSSION

Evidence of association between probands with DD and ADHD was observed with the haplotype rs4504469/rs6935076 in the categorical analysis undertaken in analysis one (see Section 5.4). Global evidence of association was observed when ADHD was defined using a cut-off on the Du Paul questionnaire. The A-G haplotype revealed evidence for association when ADHD was defined using cut-off scores on both the Du Paul questionnaire and the Connors' questionnaire.

In a sample of parent ADHD-proband trios, no evidence of association was observed between ADHD and rs4504469/rs6935076. These results propose that the locus does not influence susceptibility to ADHD. Whilst these results suggest that the association observed in the categorical analysis represents a type I error, it is also possible that the associations observed in this family-based analysis represent type II errors. It is however plausible that component phenotypes common to both ADHD and DD are influenced by rs4504469/rs6935076 and that the definition of ADHD in analysis one includes these components and that the definition of ADHD in analysis 3B does not. It is of note that the diagnosis of ADHD in the parent ADHD-proband trios sample is much more reliable as a full battery of tests for ADHD diagnosis was used rather than a screen for ADHD. Both the Du Paul and Connors' questionnaires screen for ADHD but rely on parental reports of ADHD symptoms. Problems with

recall bias and reporter error can occur in such questionnaires. This could indicate that it is more likely that the results of analysis one represent type I errors. Further, the definition of ADHD from both the screening questionnaires is subject to the arbitrariness of cut-off scores in the categorisation of ADHD presence or absence.

5.7 GENERAL CONCLUSION OF THE COMPONENTIAL ANALYSIS OF DEVELOPMENTAL DYSLEXIA

The sample utilised in this thesis was collected with the aim of detecting genetic variation increasing susceptibility to DD. As a result an extremes approach was used, whereby the extreme ends of what is assumed to be a normal ability continuum were sampled. This method was chosen in order to maximise the differences between populations (cases and controls) and increase the efficiency of the study to detect associations. It is important to note that the approach does not rule out the possibility that continuous measures of reading ability are associated with the same genetic variation. Whilst the results of analysis one suggest that the sampling technique used in this thesis was appropriate to sample differences between the extreme ends of the ability spectrum, the results for single word reading in analysis two suggest that variation on the rs4504469/rs6935076 haplotype increasing susceptibility to DD may act across the normal reading ability spectrum rather than act within the DD cases only.

The description and measurement of skills, which make up reading, are based on theoretical models of the reading process. It would be useful however for measurements to be developed which reflect the neurological processes involved in reading that are likely to be influenced by genes (Pennington 1997).

Although fractionation of the DD phenotype into components and processes

may allow a more precise definition of the deficits associated with the global DD disorder and ultimately allow the determination of the underlying aetiological genetic variants, the task is far from easy. The highly correlated nature of the componential measures of reading, make it difficult to determine the exact components related to different variants, assuming that there is specificity of the genetic effects. However, given the evidence from twin studies, it is unlikely that the loci involved in DD influence one single measure of reading ability given the interdependence of reading related measures. Furthermore, Francks (Francks et al. 2004) has suggested that protein function in CNS development is multiregional and interaction dependant and as a result it would be unlikely that a single genetic loci could encode a single cognitive phenotype.

In a recent paper, Plomin and Kovas (Plomin and Kovas 2005) predict that genes showing association to any learning disability, including reading disability, will be associated with disability and normal variation in ability. It was also suggested that genes associated with any component of a learning disability will also be associated with other components of the ability/disability and with other learning abilities and disabilities (Plomin and Kovas 2005). Although both the categorical and quantitative associations observed in this thesis go someway to support that view with respect to all components of DD and *KIAA0319* variation, with the exception of single word reading, the associations do not suggest that variation within *KIAA0319* influences normal ability in these component phenotypes. It may be that the study lacks the power to detect such associations and without further analysis on larger samples it would not be appropriate to suggest that normal ability in the component phenotypes is not influenced by *KIAA0319*. Since DD is often defined using single word reading ability, there is therefore evidence that DD may

lie on a continuum with normal reading ability given the quantitative association between rs4504469/rs6939076 in both the whole reading ability distribution sample and in the controls.

Given the associations observed with a number of component phenotypes and the correlation between many of these phenotypes, it is difficult to decipher whether the chromosome 6p locus influences many components of reading or whether a core deficit such as phonological awareness skills (believed to be the core deficit of DD – see Chapter one) causes a general inhibitory effect on other component phenotype measures/abilities rather than the locus affecting additional components too (Francks et al. 2004). If a number of reading components are influenced by the QTL, it would suggest that reading components are at least partly interdependent as argued by twin studies.

CHAPTER SIX

ASSOCIATION STUDY BETWEEN DEVELOPMENTAL DYSLEXIA AND THE DYSLEXIA CANDIDATE SUSCEPTIBILITY GENE, *EKN1*, ON CHROMOSOME 15q

CHAPTER SIX

6. ASSOCIATION STUDY BETWEEN DEVELOPMENTAL DYSLEXIA AND THE DYSLEXIA CANDIDATE SUSCEPTIBILITY GENE, *EKN1*, ON CHROMOSOME 15Q

Although many candidate genes for complex disorders are being identified by linkage and association studies, chromosomal translocations present in individuals with a specific disorder such as DD provide another method of localising disease susceptibility genes. Recently, identification of a translocation breakpoint near a dyslexia linkage region on chromosome 15q has identified a possible candidate gene for developmental dyslexia.

6.1 IDENTIFICATION OF A TRANSLOCATION BETWEEN CHROMOSOMES 15 AND 2

In 2000, a research group in Finland (Nopola-Hemmi et al. 2000) identified two families in which translocations were identified that co-segregated with reading problems (see Figure 6.1). The first was a t(2;15)(q11;q21) translocation co-segregating with reading problems in four individuals (Family 1); the second was a t(2;15)(p13;q22) translocation present in one family member with reading difficulties (Family 2).

Fluoresence in situ hybridisation (FISH) analysis revealed that the breakpoints of both the t(2;15)(q11;q21) and t(2;15)(p13;q22) translocations on chromosome 15 were located between markers D15S143 and D15S1029, a region spanning approximately 6-8Mb (Nopola-Hemmi et al. 2000). This region overlaps

the *DYX1* locus around 15q21 (see Chapter 2, Figure 2.1). The breakpoints on chromosome 2 each involve different arms of the chromosome and localisation has not been reported.

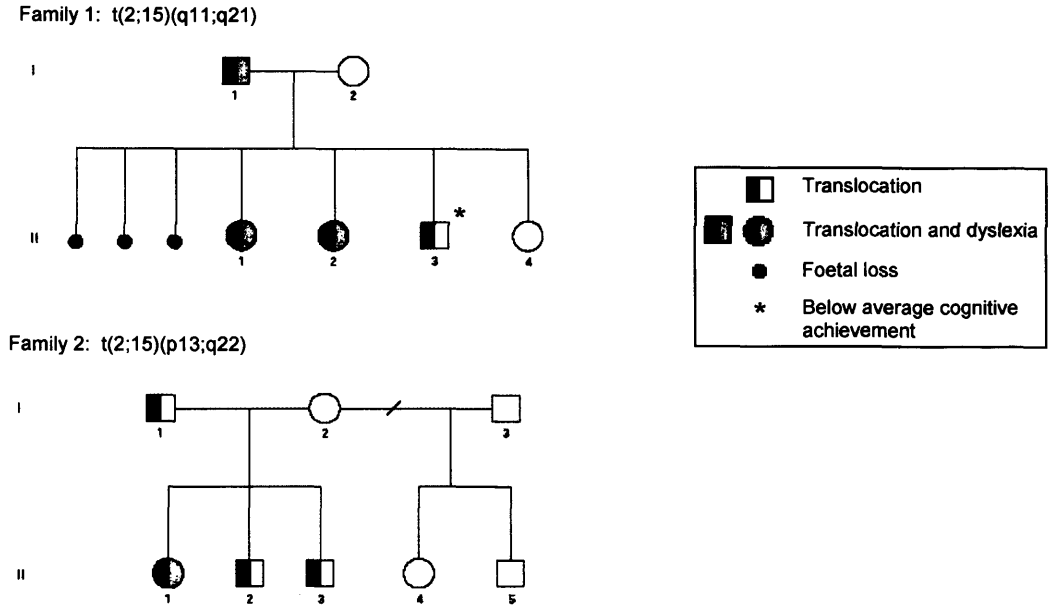


Figure 6.1 The Presence of a Translocation and DD in Two Families. Family one co-segregates a t(2; 15)(q11;q21) translocation and DD, whilst a second translocation, t(2;15)(p13;q22), is observed in an individual with DD in family 2. Adapted from Napola-Hemmi and colleagues 2000 (Nopola-Hemmi et al. 2000).

6.2 EVIDENCE SUPPORTING A ROLE OF *EKN1* IN THE SUSCEPTIBILITY OF DEVELOPMENTAL DYSLEXIA

Following identification of the two translocations on chromosome 15, the t(2;15)(q11;q21) breakpoint, was further refined using fluorescent *in situ* hybridisation and Southern hybridisation (Taipale et al. 2003), to a region of 3,229bp including exons 8 and 9 of the gene Dyslexia Susceptibility 1 Candidate 1, *DYX1C1* (*EKN1*; LocusID: 161582). Specifically, the translocation breakpoint localised to a tetratricopeptide repeat (TPR)-domain coding region of the gene, which may influence protein function. TPR-domains are suggested to be protein

interaction modules with no homology to other known proteins. Proteins containing TPR-domains are associated with multiprotein complexes and have been shown to be involved in axonal cargo transport and clustering of postsynaptic nicotine acetylcholine receptors. Within the 3,229bp containing the breakpoint, is a 301bp AT-rich region, with an almost complete 72bp inverted repeat. This suggests a repeat induced mechanism for the translocation (Taipale et al. 2003). AT-rich regions have previously been shown to occur at many chromosomal rearrangement sites (Edelmann et al. 2001).

EKN1 (or *DYX1C1*) encodes a nuclear tetratricopeptide repeat domain protein, which is expressed most abundantly in brain, lung, kidney and testis (based on RT-PCR on multiple-tissue cDNA panels). Linkage Disequilbrium across the gene based on data taken from the 33 SNPs in HapMap, suggests that there are five blocks of LD across the gene based on D' and r^2 values between SNPs (see Figure 6.2).

EKN1, 420 amino acids in length, encodes a 48kDa protein and contains three TPR motifs, suggesting the protein is likely to be involved in protein-protein interactions. The studies of Taipale and colleagues suggest EKN1 is a nuclear protein (Taipale et al. 2003).

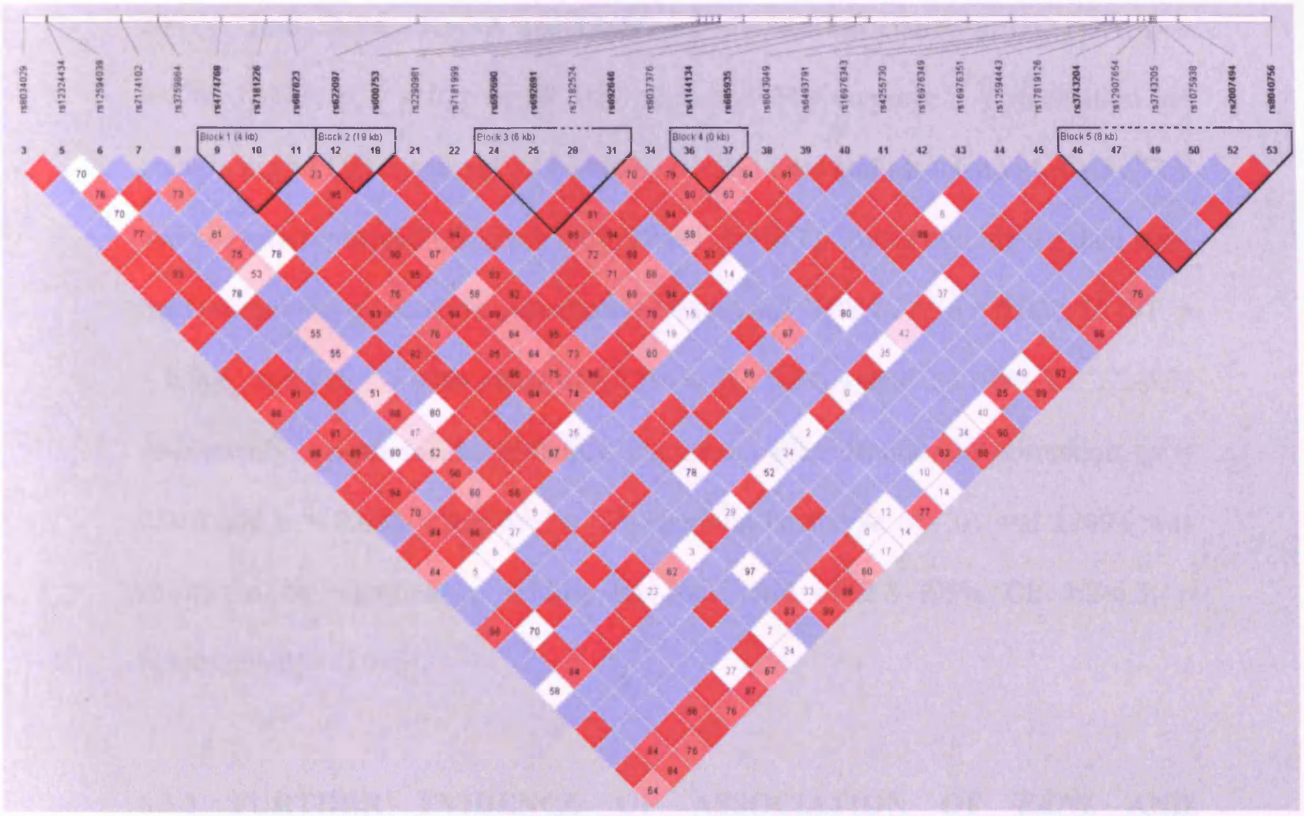


Figure 6.2 Linkage Disequilibrium Across *EKN1* Based on Data Taken From HapMap.

6.3 ASSOCIATION ANALYSIS OF *EKN1* AND DEVELOPMENTAL DYSLEXIA

Taipale and colleagues carried out an association study between *EKN1* and DD in a sample containing a mixture of unrelated cases and multiple cases from single families, as well as mixtures of unrelated and related controls (Taipale et al. 2003). The classification of DD in participants required a normal performance intelligence quotient (PIQ > 85) and a deviation in reading skills, depending on age, of at least 2 years.

Sequencing the 10 exons of *EKN1* in 20 dyslexic individuals revealed 8 single nucleotide polymorphisms, including 4G→T (P2S), 271G→A (V90I), 572G→A (G.191E), 1249G→T (E417X) and 1259C→G (S420C.) and -164C→T, -3G→A and -2G→A located in the 5' untranslated region (UTR) (Taipale et al.

2003). Two SNPs, -3G→A and 1249G→T showed association to DD (-3A, p = 0.006; 1249T, p = 0.02) in 55 DD cases and 113 controls. A replication set comprising 54 DD cases and 82 controls was also used and association between -3A and DD was replicated (-3A, p = 0.02; 1249T, p = 0.1). By combining the data from the two samples association was observed with both SNPs (-3A, p = 0.002; 1249T, p = 0.006) and yielded odds ratios of 3.2 (95% CI: 1.5-6.9) and 2.3 (95% CI: 1.2-4.2) respectively. Both results remained significant after Bonferroni correction (p = 0.016 and p = 0.048 respectively). A common haplotype of -3A and 1249T was shown to be significant, yielding an odds ratio of 2.8 (95% CI: 1.2-6.5; p (uncorrected) = 0.015).

6.3.1 FURTHER EVIDENCE OF ASSOCIATION OF *EKN1* AND DEVELOPMENTAL DYSLEXIA

Following the finding of Taipale and colleagues (Taipale et al. 2003), Wigg and colleagues provided evidence to support *EKN1* as a susceptibility gene for dyslexia (Wigg et al. 2004). They tested for association between variants in *EKN1* and dyslexia defined both categorically and quantitatively using measures of reading and reading-related processes. Classification of dyslexia required an IQ of above 80 and a score below 1.5 standard deviations from the mean on two of three core-reading tests, or 1 standard deviation on the average of three tests. In a categorical analysis, 83 families were studied including 83 probands and 18 siblings; for quantitative analysis, 148 families were included containing 202 children. Six polymorphic markers were tested for association with DD.

Rs11629841, based on a categorical definition of DD, showed the most significant association with DD (p = 0.018, corrected p = 0.036) (Wigg et al. 2004).

Neither the -3G→A or the 1249G→T SNP shown to be associated to DD in the Finnish study (Taipale et al. 2003) showed significant association (p = 0.145 and p = 0.317 respectively). Analysis of association between markers and DD defined quantitatively revealed associations (before correction for multiple testing) between the marker -3G→A and word identification (*WRAT-III*, p = 0.027; *WRMT-R Word ID*, p = 0.032), spelling (p = 0.027), phonological decoding (p = 0.02), receptive language skills (p = 0.03), expressive language skills (p = 0.03), rapid automatised naming (p = 0.04), verbal short term memory (p = 0.02) and phonological awareness (p = 0.02). Each association showed the G allele was specifically involved with poorer performance.

Using a sliding window approach to haplotype analysis, a number of associations were observed with DD (see Table 6.1).

Marker 1	Marker 2	Marker 1 Allele	Marker 2 Allele	Biased (non)-Transmission	Global p	p
-3G→A	1249G→T	G	G	Biased Transmission	0.036	0.026
rs3743204	rs11629841	G	C	Biased Transmission	0.055	0.0089
*rs11629841	rs692691	T	T	Biased Non-Transmission	0.013	0.0058
*rs11629841	rs692691	G	T	Biased Transmission	0.013	0.0389

Table 6.1 Significant haplotypic associations with DD. * Haplotype involving the same SNPs but different allele combinations. No correction for multiple testing was undertaken by Wigg and colleagues (Wigg et al. 2004).

Although the findings of Wigg and colleagues do not replicate the results of Taipale et al (Taipale et al. 2003) and in the analysis of the -3G→A/1249G→T haplotype, show association in the opposite direction to that of the first study (before correction for multiple testing), they do provide some evidence in support of the hypothesis that the gene *EKN1* may be a DD susceptibility gene. In this chapter, I

attempt to replicate the specific associations observed in the two previous studies reporting association between DD and variants in *EKN1* in an independent sample of UK origin.

6.4 MATERIALS AND METHODS

6.4.1 THE SAMPLE

In order to test whether SNPs in *EKN1* are associated to DD, a sample of 254 parent DD-proband trios were ascertained as discussed previously (see Chapter three). In this analysis, eleven componential measures of reading were used for quantitative analysis of the DD phenotype (see Chapter three, Section 3.1.2). The eleven measures test the same reading related processes as Wigg and colleagues (2004) using similar or equivalent tests and thus allows the specific associations observed previously (Wigg et al. 2004) to be examined in our sample. Componential measures included scores from the rhyme oddity test, phoneme deletion test, the non-word reading test, spelling test, single word reading test, pseudohomophone judgement task, accuracy scores from the NARA reading test, the non-word repetition test, rapid digit naming and rapid picture naming tasks and inattention symptoms measured using an abbreviated Connors ADHD screen (parent response).

6.4.2 SUBPHENOTYPIC ANALYSES

Component phenotypes were regressed for age to account for variance in scores that may have been attributable to the age of the child. Each of the component phenotype measures was tested for normality by calculating skewness and kurtosis scores (SPSS 11, 2001). Values between -1 and 1 suggested that the

data were normally distributed. Variables that did not follow a normal distribution were transformed with an appropriate transformation. Since measurement of spelling, single word reading and reading accuracy gave individuals a functional age score (an age based on reading accuracy, an age based on single word reading and an age based on spelling accuracy), a discrepancy measure was used to analyse the data quantitatively (chronological age – spelling age, chronological age – single word reading age or chronological age – accuracy age) and age was not regressed from these scores. Due to the nature of the NARA reading test, only children aged 12 years and under were included in analyses with this componential phenotype.

Table 6.2 shows the mean scores, standard deviations and the range for each of the quantitative traits analysed in this analysis, before transformation and regression for age.

Test	Mean Score	SD	Minimum Score	Maximum Score
Rhyme Oddity	15.85	5.09	0.00	29.00
Phoneme Deletion	19.58	5.56	0.00	28.00
Non-Word Reading	21.02	8.72	0.00	39.00
Non-Word Repetition	27.18	6.25	11.00	40.00
Rapid Digit Naming	29.02	9.10	13.00	77.00
Rapid Picture Naming	44.96	12.09	23.00	110.00
Inattention Symptoms	15.49	6.64	1.00	27.00
Pseudohomophone Judgement Task	52.36	13.05	4.00	68.00
Spelling	4.38	1.95	0.16	9.92
Reading Accuracy	5.03	1.70	2.50	9.92
Single Word Reading	4.41	2.00	0.19	0.31

Table 6.2 Descriptive statistics for the componential measures in DD-probands from a sample of 254 parent DD-proband trios used in a quantitative analysis of DD and variants in *EKN1*.

Table 6.3 shows the skewness and kurtosis scores for each of the components tested and the transformation used if necessary. Means and standard deviations are given for the age regressed and normalised trait values.

The correlation coefficients of the eleven componential measure results are shown in Table 6.4.

Quantitative Trait	Transformation	Mean	SD	Skewness	Kurtosis
Rhyme oddity	-	-0.04	4.84	-0.38	-0.21
Phoneme deletion	lngamma	0.0008	13.01	-0.25	0.16
Non word reading	-	-0.002	8.07	-0.08	-0.80
Non word repetition	-	0.0004	6.13	-0.30	-0.26
Rapid digit naming	artan	0.00001	0.008	-0.19	0.53
Rapid picture naming	artan	0.000007	0.005	-0.04	0.38
Inattention symptoms	sqrt	3.83	0.92	-0.57	-0.35
Pseudohomophone judgement task	lngamma	0.03	41.02	-1.05	1.48
BAS Spelling*	-	4.38	1.95	0.27	-0.05
NARA accuracy score*	-	5.03	1.70	0.70	-0.18
BAS Reading*	-	4.41	2.00	0.19	0.31

Table 6.3 Normality statistics for the subphenotypic measures in a sample of 254 parent DD-proband trios used in a quantitative analysis of DD and variants in *EKN1* after age regression.

* Based on a discrepancy score – age not regressed out. Where a transformation is given, skewness and kurtosis scores are after transformation.

	RO	PD	NWR	PHJT	NWRT	RDN	RPN	IS	BASS	NARA	BASR
RO	1	0.368	0.398	0.296	0.291	-0.352	-0.374	-0.025	-0.220	-0.331	-0.341
PD		1	0.386	0.202	0.207	-0.271	-0.219	-0.015	-0.274	-0.268	-0.329
NWR			1	0.142	0.322	-0.219	-0.231	0.060	-0.254	-0.374	-0.380
PHJT				1	0.071	-0.172	-0.175	-0.221	-0.404	-0.417	-0.366
NWRT					1	-0.031	-0.095	-0.021	-0.083	-0.230	-0.193
RDN						1	0.692	0.156	0.243	0.301	0.241
RPN							1	0.156	0.269	0.255	0.197
IS								1	0.027	0.021	0.014
BASS									1	0.784	0.718
NARA										1	0.799
BASR											1

Table 6.4 Pearson's Correlation Coefficient Between Componential Measures of Reading Related Processes Based on DD-Proband Scores from 254 Parent DD-Proband Trios. Values shown are the Pearson Product Moment Correlation Coefficients.
RO = Rhyme Oddity Test, PD = Phoneme Deletion Test, NWR = Non-Word Reading Test, BAS = BASS Spelling Test, BASR = BAS Single Word Reading Test, PHJT = Pseudohomophone Judgement Task, NARA = Accuracy Scores from the NARA Reading Test, NWRT = Non-Word Repetition Test, RDN = Rapid Digit Naming Task, RPN = Rapid Picture Naming Task, IS = Inattention Symptoms.

6.4.3 VARIANTS AND GENOTYPING

The DNA markers chosen for analysis were based upon previous reported associations to -3G→A, 1249G→T (Taipale et al. 2003) and rs11629841 (Wigg et al. 2004). Due to the information given to us by Dr Cathy Barr and Miss Karen Wigg (Toronto, Canada), it was calculated that the key haplotype information could be captured with these three markers and markers rs3743204 and rs692691 did not need to be genotyped.

In the study by Wigg and colleagues (Wigg et al. 2004), two SNPs (rs3743204 and rs692691) yielded evidence for association in combination with rs11629841. Rs3743204 and rs692691 were not genotyped in this study, however, we can be confident that the same haplotypes are not over-transmitted in our sample. The associated rs3743204C/rs11629841G haplotype observed by Wigg and colleagues (Wigg et al. 2004) is almost perfectly defined by the rs11629841G, with the alternative haplotype, rs3743204A/rs11629841G having a frequency estimated at 0.006. The rs3743204/rs11629841 haplotype analysed by Wigg and colleagues results in an over-transmitted haplotype carrying the G allele and two under-transmitted haplotypes carrying the T allele of rs11629841 (see Table 6.5). This pattern is not observed in our own sample, where the G allele was under-transmitted, though not at a significant level (see Table 6.6).

rs3743204	rs11629841	Frequency	Observed	Expected	χ^2	P
C	T	0.410	69.074	76.501	2.037	0.1535
A	T	0.251	42.926	47.354	1.311	0.2522
C	**G**	**0.333**	**82.926**	**70.549**	**6.856**	**0.0089**
A	G	0.006	1.074	1.597	0.216	-

Table 6.5 Results of haplotype analysis by Wigg and colleagues (Wigg et al. 2004). Due to the infrequent nature of the A/G haplotype, when a G is present at rs11629841, 99% of the time there is a C allele present at SNP rs3743204. In this table the C/G haplotype is overtransmitted to DD-probands. Global p = 0.055

SNP	Allele	Allele Frequency	Observed	Expected	P value	χ^2
rs11629841	T	0.68	237	233	0.50	0.46
	G	0.32	105	109		

Table 6.6 Results of rs11629841 in our sample of 254 parent DD-proband trios. The G allele is undertransmitted to DD-probands, thus showing a trend opposite of that of Wigg and colleagues (Wigg et al. 2004). It should be noted that undertransmission was not significant.

Although forming a significant haplotype with rs11629841, rs692691 was not genotyped since the 'risk' haplotype for DD is characterised by the G allele of rs11629841. The associated rs11629841G/rs692691T haplotype observed by Wigg and colleagues (Wigg et al. 2004) is almost perfectly defined by the rs11629841G allele, with the alternative haplotype, rs11629841G/rs692691C haplotype having a frequency estimated at 0.04 (see Table 6.7). As shown previously (see Table 6.6) the G allele of rs11629841 is not significantly overtransmitted to DD probands.

The rs11629841T/rs692691T allele showing association with DD by Wigg and colleagues (Wigg et al. 2004) is not overtransmitted to DD probands, indeed this haplotype shows undertransmission to probands by Wigg and colleagues (Wigg et al. 2004) and therefore does not provide evidence that the variations increase susceptibility to DD and was not tested in our sample.

rs11629841	rs692691	Frequency	Observed	Expected	χ^2	p
T	C	0.547	91.164	95.829	0.860	0.3537
G	C	0.041	8.336	7.574	0.192	
T	T	0.119	16.836	23.450	7.622	0.0058
G	T	0.293	71.664	61.147	4.268	0.0389

Table 6.7 Results of rs11629841/rs692691 haplotype analysis by Wigg and colleagues (Wigg et al. 2004). Global p = 0.013

SNPs were genotyped using Amplifluor™ (Serologicals' Corporation). Primers were designed for Amplifluor™ using the sequence NT-010194 (Table 6.8). For all three genotyping reactions, Amplifluor™ reactions were carried out

in 11µl reactions containing 48ng of DNA according to manufacturer's instructions.

SNP	Bases	Primer	Primer Sequences
-3G→A	A	Allele 1 Primer	5'-gaaggtcggagtcaacggatttcgct aacctgaagaggcattct-3'
	G	Allele 2 Primer	5'-gaaggtgaccaagttcatgctgctaa cctgagaggcattcc-3'
		Reverse Primer	5'-caagcaggcgcaagaagcaa-3'
rs11629841	T	Allele 1 Primer	5'-gaaggtgaccaagttcatgctggtatgag ttgctagttgtctatttaatgccata-3'
	G	Allele 2 Primer	5'-gaaggtcggagtcaacggattgag ttgctagttgtctatttaatgccatc-3'
		Reverse Primer	5'-cagggcatgtgtattcactgat-3'
1249G→T	G	Allele 1 Primer	5'-gaaggtgaccaagttcatgcttcggaat gtaattcaaggaacag-3'
	T	Allele 2 Primer	5'-gaaggtcggagtcaacggattcgga atgtaattcaaggaacat-3'
		Reverse Primer	5'-gtacaaagatgcctccagttgttt-3'

Table 6.8 Assay details of Amplifluor™ genotyping reactions used to type SNPs in the association study.

PCR reactions were undertaken as described in Chapter three. SNPs -3G→A and 1249G→T were genotyped using method D and rs11629841 using method A (see Chapter three, Table 3.18).

6.4.4 STATISTICS

All Genotypes were tested for Hardy-Weinberg Equilibrium using a chi square goodness of fit test (in-house HW program written by P. McGuffin and J. Williams (modified by Marian Hamshere)). Association between SNPs and DD was calculated using the graphical interface (GLUE) of the statistical package TRANSMIT (Clayton 1999) available at the MRC Human Genome Mapping Project website (http://www.hgmp.mrc.ac.uk/). Haplotype analysis was also undertaken using TRANSMIT. HAPLOVIEW was used to analyse LD between markers. LD

between SNPs and marker D15S994 was calculated using UNPHASED (http://www.hgmp.mrc.ac.uk/). Quantitative measures were analysed in affected DD-probands using ANOVA and linear regression in SPSS 11 (2001) and in trios using FBAT (Family Based Association Test; Horvath et al. 2001).

6.5 RESULTS

All genotypes were in Hardy-Weinberg Equilibrium in both DD-probands and parents, as shown in Table 6.9.

No evidence of association was shown between DD, defined categorically, and any of the SNPs analysed (See Table 6.10).

	Parents			Probands		
Marker	**χ^2**	**p**	**Hardy Weinberg**	**χ^2**	**p**	**Hardy Weinberg**
-3G→A	1.53	0.22	Yes	0.36	0.55	Yes
rs11629841	3.67	0.06	Yes	1.12	0.29	Yes
1249G→T	3.73	0.06	Yes	1.55	0.21	Yes

Table 6.9 Hardy-Weinberg calculations for each of the SNPs genotyped for this analysis.

SNP	Allele	Allele Frequency	Observed	Expected	P value	χ^2
-3G→A	G A	0.95 0.05	382 16	380 18	0.42	0.65
rs11629841	T G	0.68 0.32	237 105	233 109	0.50	0.46
1249G→T	G T	0.92 0.08	399 33	397 35	0.65	0.20

Table 6.10 Categorical analysis of the three SNPs, in 254 parent DD-proband trios, previously shown to have association with DD (Taipale et al. 2003; Wigg et al. 2004), using TRANSMIT.

Analysis of all two- and three-markers haplotypes also yielded no evidence for association, both globally and under 1 degree of freedom analysis of specific haplotypes (See Table 6.11).

Markers	Haplotype	Frequency	Observed	Expected	p	Global p
-3/1249	AT	0.05	19.54	22.39	0.37	0.69
	GT	0.03	14.74	15.15	0.87	
	GG	0.92	419.17	415.82	0.41	
-3/ rs11629841	GG	0.33	129.32	133.26	0.52	0.84
	AT	0.05	17.10	19.33	0.44	
	GT	0.63	263.58	257.41	0.33	
rs11629841/ 1249	GG	0.32	136.51	140.84	0.48	0.76
	TT	0.08	33.62	35.24	0.68	
	TG	0.60	269.87	263.92	0.35	
-3/ rs11629841/ 1249	GGG	0.32	143.08	147.36	0.49	0.78
	ATT	0.05	19.55	22.32	0.38	
	GTT	0.03	14.75	15.07	0.90	
	GTG	0.59	278.08	270.61	0.24	

Table 6.11 Categorical analysis of the 2- and 3-marker haplotypes using TRANSMIT. Haplotypes with a frequency less than 1% are not shown.

Table 6.12 below shows very limited evidence for ancestral recombination between the three markers (high D') and for 2 of the 3 pairings, genotypes at each locus were only very weakly correlated (low r^2), reflecting the large differences in allele frequencies between the markers. Previously our sample had shown evidence of association to chromosome 15q, in particular, marker D15S994 (Morris et al. 2000). No significant evidence of LD was shown between marker D15S994 and the three markers in *EKN1* (D' $\leq$ 0.52) (see Table 6.13).

Marker 1	Marker 2	D'	95% CI (D')	r^2
-3G→A	rs11629841	1.0	0.59-1.0	0.02
-3G→A	1249G→T	0.96	0.84-1.0	0.49
rs11629841	1249G→T	1.0	0.73-1.0	0.04

Table 6.12 LD between markers measured by both D' and r^2.

Marker 1	Marker 2	D' (Transmitted Allele)	D' (Nontransmitted Allele)	p
-3G→A	D15S994	0.41	0.52	0.48
rs11629841	D15S994	0.26	0.25	0.78
1249G→T	D15S994	0.38	0.34	0.47

Table 6.13 LD between markers measured by D' between SNPs and microsatellite marker D15S994, previously shown to be associated to DD in this sample.

Componential phenotype analysis of DD, had in a previous study yielded significant associations, even in the absence of association between markers and DD defined categorically (Wigg et al. 2004). Due to the observed associations, a componential analysis of DD was undertaken using eleven componential measures of DD and reading related processes in our sample. This analysis was untaken using two methods; FBAT and ANOVA. FBAT was used as it is based on parent DD-proband trios and overcomes the problem of population stratification and allows continuous measures of DD and reading to be included in an analysis. However, since only informative parents are used in FBAT analysis, the reduction in sample size used results in a lower power for the detection of quantitative traits. ANOVA allows the inclusion of all DD-probands, increasing power.

Componential phenotype analysis revealed only one nominally significant association ($p = 0.02$) between rs11629841 and inattention symptoms using ANOVA (Table 6.16). No prior hypothesis exists for this specific association based on existing studies, and after correction for multiple phenotypic tests ($n = 11$), association is not observed.

Component Phenotype	Marker	Genotype	Mean	F	p
Rhyme Oddity	-3G→A	A.G G.G	1.04 -0.45	1.39	0.24
Phoneme Deletion	-3G→A	A.G G.G	-1.24 -0.50	0.04	0.84
Non-Word Reading	-3G→A	A.G G.G	-0.82 -0.85	0.00	0.99
Non-Word Repetition	-3G→A	A.G G.G	2.83 -0.39	3.63	0.06
Rapid Digit Naming	-3G→A	A.G G.G	-0.001 0.0003	0.48	0.49
Rapid Picture Naming	-3G→A	A.G G.G	-0.001 0.0001	0.88	0.35
Inattention Symptoms	-3G→A	A.G G.G	0.18 -0.06	0.73	0.39
Pseudohomophone Judgement Task	-3G→A	A.G G.G	-3.82 0.71	0.16	0.69
Spelling	-3G→A	A.G G.G	4.51 4.45	0.01	0.92
Single Word Reading	-3G→A	A.G G.G	4.64 4.53	0.05	0.83
Reading Accuracy	-3G→A	A.G G.G	3.65 3.75	0.075	0.79

Table 6.14 ANOVA analysis of componential measures of DD/reading in SNP -3G→A.

Component Phenotype	Marker	Genotype	Mean	F	p
Rhyme Oddity	1249G→T	T.G G.G	0.03 -0.28	0.12	0.73
Phoneme Deletion	1249G→T	T.G G.G	-0.10 0.39	0.04	0.85
Non-Word Reading	1249G→T	T.G G.G	-0.54 -0.34	0.02	0.90
Non-Word Repetition	1249G→T	T.G G.G	1.93 -0.33	3.56	0.06
Rapid Digit Naming	1249G→T	T.G G.G	0.0007 0.0003	0.05	0.83
Rapid Picture Naming	1249G→T	T.G G.G	0.0001 0.0002	0.002	0.96
Inattention Symptoms	1249G→T	T.G G.G	-0.69 0.007	0.74	0.39
Pseudohomophone Judgement Task	1249G→T	T.G G.G	-2.51 1.61	0.25	0.62
Spelling	1249G→T	T.G G.G	4.42 4.35	0.04	0.84
Single Word Reading	1249G→T	T.G G.G	4.64 4.18	0.58	0.44
Reading Accuracy	1249G→T	T.G G.G	3.42 3.76	1.66	0.20

Table 6.15 ANOVA analysis of componential measures of DD/reading in SNP 1249G→T.

Component Phenotype	Marker	Genotype	Mean	F	p
Rhyme Oddity	rs11629841	G.G G.T T.T	1.28 -0.78 -0.74	1.04	0.36
Phoneme Deletion	rs11629841	G.G G.T T.T	3.79 -0.40 -1.28	0.75	0.48
Non-Word Reading	rs11629841	G.G G.T T.T	-0.39 -0.63 -0.70	0.008	0.99
Non-Word Repetition	rs11629841	G.G G.T T.T	-0.39 -0.92 -0.17	0.27	0.76
Rapid Digit Naming	rs11629841	G.G G.T T.T	-0.002 0.0005 0.0006	0.92	0.40
Rapid Picture Naming	rs11629841	G.G G.T T.T	-0.0001 0.0002 0.0004	0.08	0.92
Inattention Symptoms	rs11629841	G.G G.T T.T	-0.07 0.25 -0.20	4.01	**0.02**
Pseudohomophone Judgement Task	rs11629841	G.G G.T T.T	7.12 3.47 1.24	0.12	0.88
Spelling	rs11629841	G.G G.T T.T	4.07 4.14 4.83	2.77	0.07
Single Word Reading	rs11629841	G.G G.T T.T	4.11 4.37 4.91	2.11	0.12
Reading Accuracy	rs11629841	G.G G.T T.T	3.27 3.98 3.71	1.73	0.19

Table 6.16 ANOVA analysis of componential measures of DD/reading in SNP rs11629841. For SNPS -3G→A and 1249G→T 1df was used, for rs11629841, 2df.

Quantitative component phenotype analysis using ANOVA was used as a screening test to determine whether there were any differences in the component phenotype based on genotype groups. In order to determine the underlying genetic model, linear regression was undertaken. Since the only significant association observed was between inattention symptoms and rs11629841, linear regression was

undertaken for this component phenotype. Inattention symptom scores were used as the dependant variable and additive and dominance codings for rs11629841 were utilised as the independent variable (see Chapter 4, Section 4.4.5 for a description). Linear regression suggested that there was a genotypic effect rather than an allelic effect at rs11629841 influencing inattention symptoms. Specifically individuals who were homozygous for any allele at rs11629841 were more likely to show more symptoms of inattention (see Table 6.17).

Component Phenotype	SNP	p	Model	B	p
Inattention Symptoms	rs11629841	0.02	Additive Dominance	-0.06 0.78	0.74 0.078

Table 6.17 Linear regression analysis of inattention symptoms and additive and dominance coding for rs11629841. A continuous measure of inattention symptoms in the DD-probands was used as the dependant variable and additive and dominance codings to determine the genetic model underlying the association observed in the ANOVA screen, as the independent variable. Results suggest that individuals who are homozygous for either allele at rs11629841 show more symptoms of inattention.

Quantitative analysis was also undertaken using FBAT. The results of which are shown in Tables 6.18 – 6.20.

Quantitative Trait	Allele	Frequency	Informative Families	Z	P
Rhyme Oddity	G	0.92	42	0.73	0.46
Phoneme Deletion	G	0.92	42	0.20	0.84
Non-Word Reading	G	0.92	42	1.77	0.08
Non-Word Repetition	G	0.92	41	-0.65	0.51
Rapid Digit Naming	G	0.92	37	0.52	0.60
Rapid Picture Naming	G	0.92	25	-0.16	0.87
Inattention Symptoms	G	0.92	33	1.81	0.07
Pseudohomophone Judgement Task	G	0.92	38	0.18	0.86
Spelling	G	0.92	43	0.50	0.62
Single Word Reading	G	0.92	43	0.31	0.75
Reading Accuracy	G	0.92	13	2.10	0.35

Table 6.18 Quantitative analysis of DD and DD related phenotypes using FBAT. The table shows results for the SNP 1249G→T.

Quantitative Trait	Allele	Frequency	Informative Families	Z	P
Rhyme Oddity	G	0.96	22	-1.66	0.10
Phoneme Deletion	G	0.96	20	-0.41	0.68
Non-Word Reading	G	0.96	21	-0.36	0.72
Non-Word Repetition	G	0.96	20	-0.92	0.36
Rapid Digit Naming	G	0.96	18	1.21	0.23
Rapid Picture Naming	G	0.96	17	0.04	0.97
Inattention Symptoms	G	0.96	14	-1.61	0.11
Pseudohomophone Judgement Task	G	0.96	18	-0.42	0.67
Spelling	G	0.96	22	1.21	0.23
Single Word Reading	G	0.96	22	0.09	0.35
Reading Accuracy	G	-	-	-	-

Table 6.19 Quantitative analysis DD related phenotypes using FBAT. The table shows results for the SNP -3G→A.

Quantitative Trait	Allele	Frequency	Informative Families	Z	P
Rhyme Oddity	T	0.67	80	-0.44	0.66
Phoneme Deletion	T	0.67	77	-0.66	0.51
Non-Word Reading	T	0.67	79	0.93	0.35
Non-Word Repetition	T	0.67	74	0.92	0.36
Rapid Digit Naming	T	0.67	63	1.38	0.17
Rapid Picture Naming	T	0.67	59	1.57	0.12
Inattention Symptoms	T	0.67	61	-1.43	0.15
Pseudohomophone Judgement Task	T	0.67	71	-0.16	0.88
Spelling	T	0.67	81	1.24	0.22
Single Word Reading	T	0.67	81	1.26	0.21
Reading Accuracy	T	0.67	26	-1.58	0.12

Table 6.20 Quantitative analysis of DD and DD related phenotypes using FBAT. The table shows results for the SNP rs11629841.

Haplotype analysis (Tables 6.21 – 6.24) between componential phenotypes and both two- and three-marker haplotypes revealed nominal evidence for association between rhyme oddity and the three marker haplotype -3G>A/1249G>T/rs11629841 (haplotype A-T-T, p = 0.05; see Table 6.21) and the two marker haplotype rs11629841/-3G>A (haplotype T.A, p = 0.05; see Table 6.24) and evidence of association between reading accuracy in children aged 12 years and under and the two marker haplotype 1249G>T/-3G>A (haplotype G.G, p = 0.03; see Table 6.23). None of the observed associations stand correction for multiple testing

using the Bonferroni correction.

Component Phenotype	Haplotype	Frequency	Informative Families	Z	P
Rhyme Oddity	G-G-T	0.60	78	-0.73	0.46
	G-G-G	0.32	78	0.39	0.70
	A-T-T	0.04	15	1.92	**0.05**
	A-G-T	0.04	16	-1.39	0.16
	G-T-T	0.006	3	-	-
Phoneme Deletion	G-G-T	0.60	76	-1.05	0.29
	G-G-G	0.32	76	0.71	0.48
	A-T-T	0.04	14	0.70	0.48
	A-G-T	0.04	16	-0.09	0.93
	G-T-T	0.006	2	-	-
Non-Word Reading	G-G-T	0.60	78	1.05	0.29
	G-G-G	0.32	77	-0.93	0.35
	A-T-T	0.04	14	0.37	0.71
	A-G-T	0.04	16	-0.66	0.51
	G-T-T	0.006	3	-	-
Single Word Reading	G-G-T	0.60	78	1.37	0.17
	G-G-G	0.32	78	-1.26	0.21
	A-T-T	0.04	15	0.18	0.86
	A-G-T	0.04	16	0.14	0.89
	G-T-T	0.006	3	-	-
Pseudohomophone Judgement Test	G-G-T	0.60	69	-0.16	0.87
	G-G-G	0.32	70	0.46	0.65
	A-T-T	0.04	13	0.05	0.96
	A-G-T	0.04	14	-0.58	0.57
	G-T-T	0.006	1	-	-
Reading Accuracy	G-G-T	0.60	24	-0.65	0.51
	G-G-G	0.32	23	1.67	0.09
	A-T-T	0.04	2	-	-
	A-G-T	0.04	6	-	-
	G-T-T	0.006	2	-	-
Non-Word Repetition Test	G-G-T	0.60	72	0.13	0.89
	G-G-G	0.32	72	-0.71	0.48
	A-T-T	0.04	13	1.42	0.16
	A-G-T	0.04	16	0.52	0.61
	G-T-T	0.006	3	-	-
Rapid Digit Naming	G-G-T	0.60	61	1.64	0.10
	G-G-G	0.32	62	-1.23	0.22
	A-T-T	0.04	12	-1.19	0.23
	A-G-T	0.04	14	0.35	0.72
	G-T-T	0.006	3	-	-

Table 6.21 Continued ...

Component Phenotype	Haplotype	Frequency	Informative Families	Z	P
Rapid Picture Naming	G-G-T	0.60	58	1.09	0.28
	G-G-G	0.32	57	-1.41	0.16
	A-T-T	0.04	11	-0.47	0.64
	A-G-T	0.04	8	-	-
	G-T-T	0.006	3	-	-
Spelling	G-G-T	0.60	78	0.88	0.38
	G-G-G	0.32	78	-0.65	0.52
	A-T-T	0.04	15	0.59	0.55
	A-G-T	0.04	16	-0.44	0.66
	G-T-T	0.006	3	-	-
Hyperactive/ Impulse Dimensions: Connors' ADHD	G-G-T	0.60	55	1.24	0.22
	G-G-G	0.32	60	-0.66	0.51
	A-T-T	0.04	9	-	-
	A-G-T	0.04	13	-0.99	0.32
	G-T-T	0.006	3	-	-

Table 6.21 Componential phenotype analysis of the haplotype -3G>A/1249G>T/rs11629841. Haplotypes with a frequency less than 1% are not shown.

Component Phenotype	Haplotype	Frequency	Informative Families	Z	P
Rhyme Oddity	G-T	0.61	76	-0.63	0.53
	G-G	0.32	79	0.39	0.69
	T-T	0.07	32	0.36	0.72
Phoneme Deletion	G-T	0.61	75	-1.02	0.31
	G-G	0.32	77	0.67	0.50
	T-T	0.07	31	0.42	0.68
Non-Word Reading	G-T	0.61	76	1.31	0.19
	G-G	0.32	78	-0.90	0.36
	T-T	0.07	31	-0.71	0.48
Single Word Reading	G-T	0.61	76	0.98	0.33
	G-G	0.32	79	-1.20	0.23
	T-T	0.07	32	0.44	0.66
Pseudohomophone Judgement Test	G-T	0.61	69	-0.41	0.69
	G-G	0.32	71	0.61	0.55
	T-T	0.07	28	-0.42	0.68
Reading Accuracy	G-T	0.61	22	-1.14	0.25
	G-G	0.32	24	1.82	0.07
	T-T	0.07	10	-1.42	0.16
Non-Word Repetition Test	G-T	0.61	70	-0.009	0.99
	G-G	0.32	73	-0.68	0.49
	T-T	0.07	30	1.49	0.14
Rapid Digit Naming	G-T	0.61	59	1.61	0.11
	G-G	0.32	63	-1.34	0.18
	T-T	0.07	27	-0.30	0.77
Rapid Picture Naming	G-T	0.61	56	1.39	0.17
	G-G	0.32	58	-1.48	0.14
	T-T	0.07	20	0.25	0.80
Spelling	G-T	0.61	76	0.43	0.67
	G-G	0.32	79	-0.55	0.58
	T-T	0.07	32	0.24	0.81
Hyperactive /Impulse Dimensions: Connors' ADHD	G-T	0.61	53	0.77	0.44
	G-G	0.32	61	-0.60	0.55
	T-T	0.07	25	-0.21	0.83

Table 6.22 Componential phenotype analysis of the haplotype 1249G>T/rs11629841.

.Component Phenotype	Haplotype	Frequency	Informative Families	Z	P
Rhyme Oddity	G-G	0.92	38	-0.16	0.87
	T-A	0.04	19	1.50	0.13
	T-G	0.03	20	-1.73	0.08
	G-A	0.006	3	-	-
Phoneme Deletion	G-G	0.92	36	-0.17	0.86
	T-A	0.04	18	0.54	0.59
	T-G	0.03	20	-0.26	0.79
	G-A	0.006	2	-	-
Non-Word Reading	G-G	0.92	37	0.66	0.51
	T-A	0.04	18	0.12	0.90
	T-G	0.03	20	-1.09	0.27
	G-A	0.006	3	-	-
Single Word Reading	G-G	0.92	38	0.62	0.54
	T-A	0.04	19	-0.55	0.58
	T-G	0.03	20	0.18	0.86
	G-A	0.006	3	-	-
Pseudohomophone Judgement Test	G-G	0.92	32	0.38	0.71
	T-A	0.04	17	0.72	0.47
	T-G	0.03	18	-0.77	0.44
	G-A	0.006	1	-	-
Reading Accuracy	G-G	0.92	11	2.22	**0.03**
	T-A	0.04	3	-	-
	T-G	0.03	6	-	-
	G-A	0.006	2	-	-
Non-Word Repetition Test	G-G	0.92	36	-0.44	0.66
	T-A	0.04	17	1.03	0.30
	T-G	0.03	20	-0.23	0.82
	G-A	0.006	3	-	-
Rapid Digit Naming	G-G	0.92	32	0.46	0.65
	T-A	0.04	15	-1.34	0.18
	T-G	0.03	18	0.70	0.48
	G-A	0.006	3	-	-
Rapid Picture Naming	G-G	0.92	23	-1.32	0.19
	T-A	0.04	14	-0.41	0.69
	T-G	0.03	10	1.91	0.06
	G-A	0.006	3	-	-
Spelling	G-G	0.92	38	1.01	0.32
	T-A	0.04	19	-0.34	0.73
	T-G	0.03	20	-0.37	0.71
	G-A	0.006	3	-	-
Hyperactive /Impulse Dimensions: Connors' ADHD	G-G	0.92	30	1.45	0.14
	T-A	0.04	11	-0.09	0.93
	T-G	0.03	16	-1.24	0.22
	G-A	0.006	3	-	-

Table 6.23 Componential phenotype analysis of the haplotype 1249G>T/-3G>A Haplotypes with a frequency less than 1% are not shown.

Component Phenotype	Haplotype	Frequency	Informative Families	Z	P
Rhyme Oddity	T-G	0.64	82	-1.30	0.19
	G-G	0.32	80	0.44	0.66
	T-A	0.05	18	1.96	**0.05**
Phoneme Deletion	T-G	0.64	79	-0.96	0.34
	G-G	0.32	77	0.66	0.51
	T-A	0.05	16	0.56	0.58
Non-Word Reading	T-G	0.64	82	0.72	0.47
	G-G	0.32	79	-0.93	0.35
	T-A	0.05	17	0.64	0.52
BAS Single Word Reading	T-G	0.64	82	1.44	0.15
	G-G	0.32	80	-1.33	0.19
	T-A	0.05	18	-0.26	0.79
Pseudohomophone Judgement Test	T-G	0.64	73	-0.04	0.97
	G-G	0.32	71	0.16	0.88
	T-A	0.05	14	-0.39	0.70
Reading Accuracy	T-G	0.64	25	-1.12	0.26
	G-G	0.32	25	1.42	0.16
	T-A	0.05	4	-	-
Non-Word Repetition Test	T-G	0.64	76	0.49	0.63
	G-G	0.32	74	-0.92	0.36
	T-A	0.05	16	1.28	0.20
Rapid Digit Naming	T-G	0.64	64	1.83	0.07
	G-G	0.32	63	-1.38	0.17
	T-A	0.05	15	-1.05	0.29
Rapid Picture Naming	T-G	0.64	59	1.64	0.10
	G-G	0.32	59	-1.57	0.12
	T-A	0.05	14	-0.05	0.96
Spelling	T-G	0.64	82	0.78	0.44
	G-G	0.32	80	-0.74	0.47
	T-A	0.05	18	-0.21	0.83
Hyperactive/ Impulse Dimensions: Connors' ADHD	T-G	0.64	59	0.68	0.50
	G-G	0.32	61	-0.54	0.59
	T-A	0.05	12	-0.33	0.74

Table 6.24 Componential phenotype analysis of the haplotype rs11629841/-3G>A.

6.6 DISCUSSION

Previous studies have reported association between the gene *EKN1* and DD (Taipale et al. 2003; Wigg et al. 2004). Originally identified from a translocation breakpoint on chromosome 15q, *EKN1* maps towards the distal end of DYX1, which has previously shown linkage to DD (Grigorenko et al. 1997; Schulte-Körne et al.

1998; Nöthen et al. 1999; Chapman et al. 2004), and in our own sample, association (Morris et al. 2000). Since *EKN1* maps approximately 15Mb from our LD signal, it is not *a priori* likely that *EKN1* can explain our previous association findings. However, given the association evidence from two studies (Taipale et al. 2003; Wigg et al. 2004), the specific hypotheses concerning *EKN1* and DD was tested in our sample of 254 parent-DD proband trios, one of the largest samples yet to be examined in regard to this hypothesis.

In total 15 SNPs have been reported in the association analyses of *EKN1* and DD. The predicted amino acids encoded by the SNPs are shown in Table 6.25 and the location of each of these SNPs is shown in Figure 6.3. A general summary of all the studies undertaken between *EKN1* and DD can be seen in Tables 6.26 and 6.27. It should be noted that there are many other variants that have been reported in this gene however these have not been identified in a sample of DD probands or tested for association with DD.

6.6.1 COMPARISON OF RESULTS TO ASSOCIATION STUDIES BETWEEN *EKN1* AND DEVELOPMENTAL DYSLEXIA PRIOR TO OUR RESEARCH

In the analysis of Taipale and colleagues (Taipale et al. 2003), association was observed between the -3A (combined p = 0.002) and 1249T (combined p = 0.006) alleles and between the haplotype -3A/1249T (p = 0.015) and DD. The two SNPs could influence the function of the EKN1 protein. Unlike 1249G→T which alters the amino acid sequence (truncating the protein by four amino acids), -3G→A does not alter protein amino acid sequence but is located in the binding sequence of the transcription factors Elk-1, HSTF and TFII-I.

SNP	Position from ATG	Alleles	Amino Acid	Amino Acid Change
rs2007494	-1718	A→T	INTRONIC	-
rs1075938	-164	C→T	5'-UTR	-
-159A→G	-159	A→G	5'-UTR	-
-129A→C	-129	A→C	5'-UTR	-
-3G→A	-3	G→A	5'-UTR	-
-2G→A	-2	G→A	5'-UTR	-
4G→T (C→T)	4	G→T (C→T)	PROLINE	SERINE
13C→T→G	13	C→T→G	INTRONIC	-
rs3743204	218	T→G	INTRONIC	-
271G→A	271	G→A	VALINE	ISOLEUCINE
572G→A	572	G→A	GLYCINE	GLUTAMATE
1249G→T	1249	G→T	GLUTAMATE	STOP
1259C→G	1259	C→G	SERINE	CYSTEINE
rs11629841	12889	T→G	INTRONIC	-
rs692691	29952	T→C	INTRONIC	-

Table 6.25 Possible functional effects of variants found in *EKN1* during association analyses with DD.

At both the single SNP level and at the haplotype level, we were unable to replicate any of the findings of Taipale and colleagues (Taipale et al. 2003) and trends seen in our data, are in the direction of undertransmission of the alleles involved in the reported association. The associations observed by Wigg and colleagues (Wigg et al. 2004) between SNPs and DD defined categorically, also failed to replicate the results of the original study by Taipale and colleagues (Taipale et al. 2003). However, Wigg and colleagues did observe association between categorical DD and the G allele of SNP rs11629841. We were also unable to replicate this association in our study.

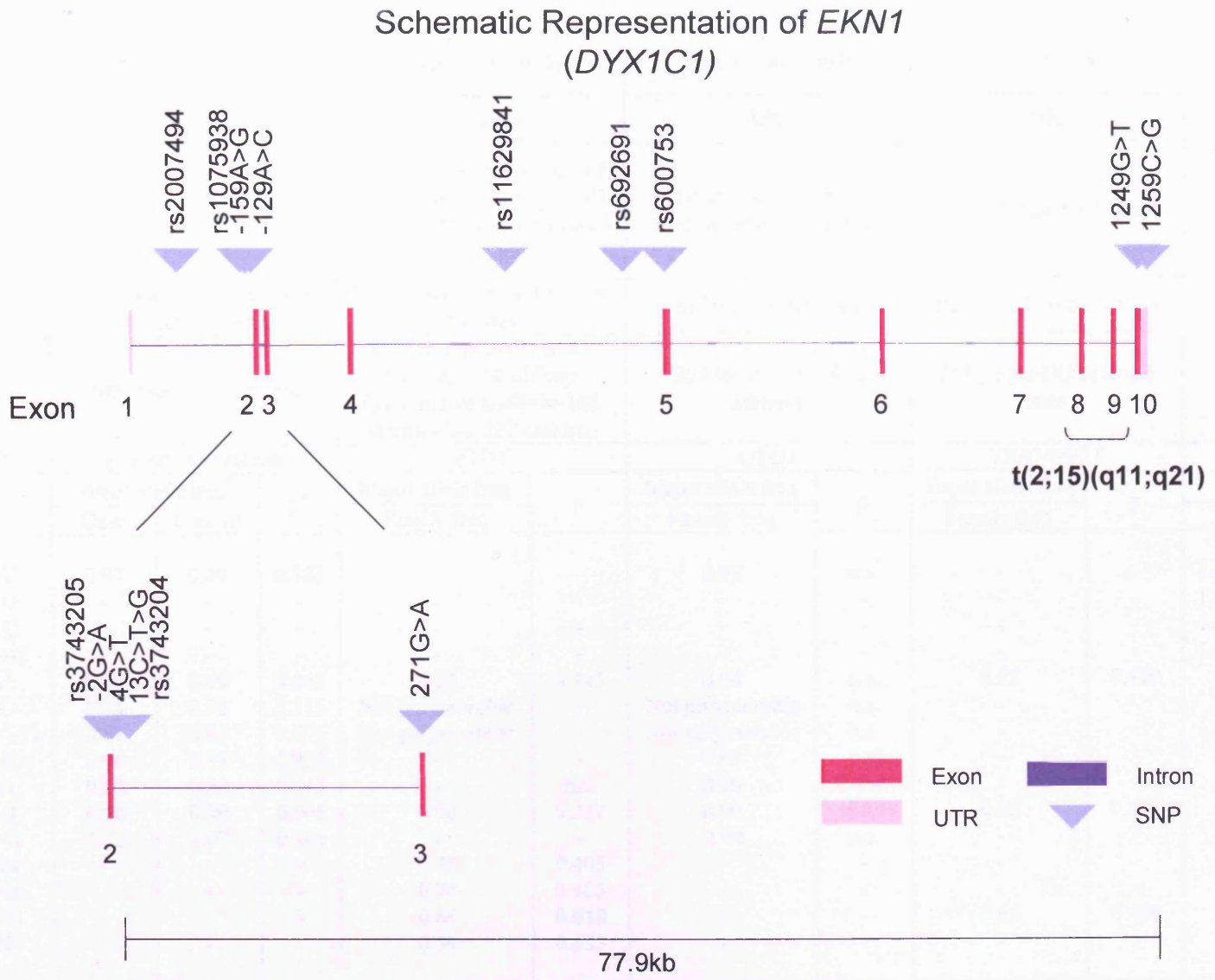


Figure 6.3 A schematic representation of EKN1. Indicated in the diagram are all the polymorphisms tested for association with DD in previous studies. The translocation breakpoint lies between exons 8 and 9.

		Study									
		Taipale *et al* (2003)			**Wigg *et al* (2004)**		**Scerri *et al* (2004)**		**This Study**		**Marino *et al* (2005)**	
Population		Finnish			Canadian		UK		UK		Italian	
DD Definition Criteria		PIQ > 85; RD > 2 years			1.5 SD below mean on 2 of 3 core reading tests or 1 SD below on the average of all 3 reading tests		Quantitative measures of reading related processes		IQ ≥ 85; RD ≥ 2.5 years		DSM IV (DD); 2 SD below mean on speed or accuracy (text reading) or 1.5 SD below mean on 1+ parameters and 2 SD below on accuracy or speed, IQ ≥ 85	
Sample		Family and Individual cases: Case-control design			Parent DD proband trios and siblings		Sibling pair families		Parent DD proband trios		Parent DD proband trios + any affected siblings	
Sample Size		109 cases; 195 controls			Categorical analysis: 83 families + 18 siblings Quantitative analysis: 148 families inc. 202 children		264 families inc. 630 siblings		254 parent-DD proband trios		158 nuclear families inc. 171 children	
Analysis Method		χ^2, Fishers exact test			eTDT		QTDT		TRANSMIT		FBAT (TDT), TRANSMIT	
		Major allele freq[#]		p[#]	Major allele freq	p	Major allele freq	p	Major allele freq	p	Major allele freq	p
		Case	Control		Family freq		Family freq		Family freq		Family freq	
SNP	-164C→T	0.97	0.99	0.383	-	-	0.98	n.s.	-	-	Low Informativeness	-
	-159A→G	-	-	-	-	-	-	-	-	-	Low Informativeness	-
	-129A→C	-	-	-	-	-	-	-	-	-	Low Informativeness	-
	13C→T→G	-	-	-	-	-	-	-	-	-	Low Informativeness	-
	-3G→A[¶]	0.97	0.99	**0.002**	0.92	0.145	0.94	n.s.	0.95	0.419	0.92	0.65
	-2G→A	0.94	0.98	0.715	Not polymorphic	-	Not polymorphic	n.s.	-	-	-	-
	4G→T (C→T)	0.81	0.86	0.875	Not polymorphic	-	Not polymorphic	n.s.	-	-	Low Informativeness	-
	271G→A	0.94	0.99	0.925	-	-	0.93	n.s.	-	-	-	-
	572G→A	0.57	0.59	0.645	-	n.s.	0.50	**0.064**^	-	-	-	-
	1249G→T	0.98	0.99	**0.006**	0.90	0.317	0.90	n.s.^	0.92	0.655	0.87	0.29
	1259C→G	0.72	0.67	0.161	-	-	0.90	n.s.	-	-	0.88	0.26
	rs2007494	-	-	-	0.75	0.405	-	-	-	-	-	-
	rs3743204	-	-	-	0.74	0.405	-	-	-	-	-	-
	rs11629841	-	-	-	0.66	**0.018**	-	-	0.68	0.496	-	-
	rs692691	-	-	-	0.59	0.623	-	-	-	-	-	-

Table 6.26 Summary of the association reports between DD and variation in *EKN1*. [#] Combined sample; ^ READ reading measure; [¶] Showed significant association with a number of quantitative measures by Wigg *et al* (2004); [#] Marino *et al* study (2005)

		Haplotype: -3G→A/1249G→T								
		AT Frequency	**GT Frequency**	**GG Frequency**	**AG Frequency**	**Odds Ratio**	**95% CI – lower limit**	**95% CI – upper limit**	**P (haplotype)**	**Global p**
Study	Taipale *et al* (2003)	0.13* (0.05^)				2.8	1.2	6.5	**0.015 (AT)**	
	Wigg *et al* (2004)	0.07	**0.03**	0.89	**0.02**				**0.026 (GG)**	**0.036**
	Scerri *et al* (2004)¶								**0.016 (GG)**	**0.035**
	This Study	**0.05**	**0.03**	0.92	0				n.s.	0.69
	Marino *et al* (2005)								n.s.	0.41

Table 6.27 Summary of the associations observed between -3G→A/1249G→T and DD.
* Refers to frequency in cases
^ Refers to frequency in controls
¶ Refers to OC-choice quantitative reading measure

It is interesting to see in our sample, that the low level of evidence for recombination (D' = 0.96) between markers means that the -3G/1249G haplotype is almost perfectly defined by the G allele of the SNP 1249G→T (see frequencies in Tables 6.10 and 6.11). It is therefore, of no surprise, that in the absence of association with SNP 1249G, we also find no association with the -3G/1249G haplotype associated to DD in the Wigg and colleagues (Wigg et al. 2004) study. Although Wigg and colleagues (Wigg et al. 2004) also failed to observe association with the 1249G allele (p = 0.32), association with the haplotype -3G/1249G was observed (global p = 0.036). It is most likely that the discrepancy between the two studies reflects the random fluctuations in the estimate of D' between the two markers. This is consistent with the fact that their estimate of D' = 0.84, is included within our own estimate of the 95% confidence interval for D' (see Table 6.12).

In the study of Wigg and colleagues (Wigg et al. 2004), several quantitative aspects of the DD phenotype (phonological awareness, phonological decoding, spelling, word identification, phonological memory, verbal short-term memory and rapid automated naming) were associated with the -3G allele. None of these findings replicated in our sample. With individual SNP analysis, only one analysis achieved nominal levels of significance between rs11629841 and inattention symptoms, based upon ANOVA including all DD-probands. This finding was not supported by analysis using FBAT. Although FBAT may lose power, by being based only on heterozygous parents, it is most likely that our finding is simply due to chance as result of multiple testing. Indeed the result does not stand correction for multiple testing using the Bonferroni correction. To further clarify any associations between *EKN1* and ADHD or aspects of ADHD further analyses will be required on samples of children with ADHD. Interestingly *EKN1* has shown some

evidence for association to ADHD in a Canadian sample (Barr 2004; see Section 6.6.6).

Analysis of two- and three marker haplotypes revealed nominal evidence for association between rhyme oddity and the haplotypes -3A/1249T/rs11629841T and rs11629841T/-3A and reading accuracy in children 12 years and under and the haplotype 1249G/-3G. For both haplotypes showing association with rhyme oddity score, the associated haplotype was associated with better performance on the test and therefore would seem to suggest that variations on the haplotypes have a 'protective' effect against problems with phonological awareness. Given the few informative families included in these results and the observation that none of the associations remain significant after correction for multiple testing, it is most likely that these results represent type I errors.

6.6.2 WHY DO STUDIES FAIL TO REPLICATE THE INITIAL ASSOCIATIONS BETWEEN EKN1 AND DEVELOPMENTAL DYSLEXIA?

The main question raised by the results of association studies between DD and *EKN1* is why do they fail to replicate the associations observed in the original report of association by Taipale and colleagues (Taipale et al. 2003)? Chapter two, Section 2.8.4 discussed reasons why studies fail to replicate earlier observations.

This study has over 80% power to detect association between DD and SNPs rs11629841 and 1249G→T and over 70% power to detect an association with -3G→A based on a relative risk of 2 which approximates to the effect sizes of the SNPs in the two previous studies of association between DD and *EKN1* (Taipale et al. 2003; Wigg et al. 2004). Since previous studies have utilised smaller sample sizes (Taipale et al. 2003; Wigg et al. 2004) this is one of the most powerful studies

undertaken so far. However it is important to note that since the effect size is based on an initial study, which tend to inflate true effect sizes (Ioannidis et al. 2001), the power calculation should be regarded as illustrative only. More importantly, our analysis benefits from a sample that has previously shown strong evidence for association between markers on chromosome 15q and DD (Morris et al. 2000).

Population history has been shown to influence LD between markers and SNPs (see Chapter two, Section 2.8.1). Any differences in LD could alter the power to detect associations between markers and DD, particularly if the variant is in strong LD with the causal variant in one sample population and not in a second sample population. It is known that the Finnish population underwent a population bottleneck in its history. As a result, extensive LD may be observed due to the loss of some haplotypes within the population, thus increasing the chances of observing an association between DD and variants further away from the causal variant.

Since the original association between DD and *EKN1* was observed in a case-control sample the inability to replicate association in our study, given that a family-based sample was used, could indicate that the original result was obtained due to population stratification. By using a family-based sample, population stratification is avoided since the nature of the analyses means that a separate control sample is not required. If population stratification is present in the original study, the results would likely represent a type I error. Although some attempt was made to determine if the results were the result of population stratification in the original study by including a family-based analysis, this was based on only 9 parent DD-proband trios and would not be a representative result of the general population and would lack reasonable power to detect replicable associations. It is noteworthy, that the most significant associations observed by Taipale and colleagues (Taipale et al.

2003) were observed in a case-control sample obtained from only 20 families. Non-independence of the alleles in the sample could result in type I error due to over representation of the disease allele in cases. Indeed the replication set did not replicate the association with the SNP 1249G>T ($p = 0.1$).

Phenotypic difference across all studies testing for association between *EKN1* and DD could be a major factor in the non-replication of results. Although Taipale and colleagues (Taipale et al. 2003) use a similar threshold for diagnosing DD categorically as in this thesis, a different battery of tests was utilised to determine IQ and reading age and indeed the test batteries were undertaken in a different language. In the study by Wigg and colleagues (Wigg et al. 2004) a different criteria was used in diagnosing DD. By defining DD differently, one study may include part of a phenotype not covered by another one, resulting in the observation of association in one study and not in another. Even down to componential measures of reading ability, the different test batteries used may correlate differently with the genotypes even if they appear to capture the 'same' componential phenotype. Replication of associations observed on chromosome 6p have suggested that the loci influences susceptibility to the most severe cases of DD (Francks et al. 2004; Cope et al. 2005; Meng et al. 2005; Schumacher et al. 2005) recommending that phenotype definition can be important in the replication of associations. It has also been shown that correlated phenotypes can result in few shared genetic effects (Deng et al. 2002; Recker and Deng 2002).

The quality of other studies genotyping is unknown given the data from their papers. Missing genotypes (especially in family-based samples), miss-classification of genotypes in genotype calling and non-replication of genotypes when they are tested twice can result in skewed results. In order to achieve high quality results,

genotypes in this thesis were called by two individuals, all SNPs were checked for Mendelian errors within the parent DD-proband trios and 10% of the genotypes were retyped to ensure the genotyping assay was reproducible. This study tries to eliminate the problems associated with quality control issues as much as possible. It is unknown whether previous results from studies were the result of quality control issues.

6.6.3 *EKN1* AND THE INTERNATIONAL HAPMAP PROJECT

It is now feasible, given data available from the HapMap project, to select a minimum number of SNPs across a gene to extract most information about variation in the gene. Data that has become available suggests that to extract maximum information from *EKN1*, it would be necessary to individually genotype 17 SNPs (based on an $r^2 > 0.8$). One of these SNPs, rs3743205 corresponds to the SNP -3G>A, whilst the SNP rs11629841 and 1249G>T were not included in the tag SNP selection. From data available it is unknown whether these SNPs are in LD with the 17 tag SNPs and indeed if they would be captured by the tag SNPs. Given that variation in *EKN1* was not *a priori* likely to explain our previous association to chromosome 15q and given that the results obtained in this study did not support the role of *EKN1* variations in DD, it was not deemed appropriate to spend time and money genotyping the tag SNPs in our sample of parent DD-proband trios. It is more appropriate to undertake positional candidate gene studies around our previous report of association (between markers D15S146 and D15S994), particularly given the success in narrowing the candidate region of chromosome 6p reported in this thesis (see Chapter four) using a similar method and given the overlap between this region and that of linkage studies on chromosome 15 (see Chapter 2, Figure 2.1). A

tag SNP study to extract all the LD information across the association region (D15S146-D15S994) is also feasible now.

Whilst previous studies have attempted to extract some information on the LD across *EKN1* by sequencing individuals and identifying sequence variations, this study does not attempt to extract all information across the gene, just to replicate the initial associations. Consequently, the study lacks the power to detect associations between DD and variations that are not in LD with the variations genotyped. Previous studies by Taipale and colleagues and Wigg and colleagues (Taipale et al. 2003; Wigg et al. 2004) have most likely extracted more information with regards to LD (as they have typed more SNPs), however neither study has genotyped enough variation to have picked up all the information which is now available from HapMap data. The power to detect associations is also questionable in the Finnish sample given the individuals in the case-control sample are not independent.

6.6.4 COMPARISON OF RESULTS TO ASSOCIATION STUDIES BETWEEN *EKN1* AND DEVELOPMENTAL DYSLEXIA PUBLISHED AFTER OUR RESEARCH

Scerri and colleagues (Scerri et al. 2004) examined the *EKN1* gene for evidence of association to DD. Only six of the eight polymorphisms found in the Taipale and colleagues study were polymorphic in this sample of UK families. The two SNPs not shown to be polymorphic, -2G→A and 4C→T, were also the SNPs with the lowest frequency in the Taipale *et al* (2003) study. These variants were also not polymorphic in the Wigg *et al* study (Wigg et al. 2004). Scerri and colleagues tested for association between variants and six reading related component measures including orthographic coding (OC) (using two complementary tests, a

forced choice task (OC-choice), and a test involving reading of irregular words (OC-irreg)), phonological decoding, single word reading, spelling and phoneme awareness. Association was observed between single markers and one reading related measure; the 1249G allele showed association with poor performance on the OC-choice quantitative measure ($p = 0.02$). This is in contrast with the findings of Taipale and colleagues who found association between the rarer, T allele and categorical dyslexia, and thus poorer performance. Analysis of two-marker haplotypes and the OC-choice component phenotype measure suggested an association with the two-marker haplotype -3G→A/1249G→T ($p = 0.0351$) and more specifically the -3G/1249G haplotype and OC-choice ($p = 0.0158$). This was the most common haplotype, found in 90.7% of this UK sample. This result follows Wigg and colleagues data, who also found an association between the -3G/1249G haplotype and DD (88% of the Canadian sample). The original association observed by Taipale and colleagues involved the opposite alleles in this haplotype, -3A/1249T. Trends in our data suggest that there is under-transmission of the -3A (22 transmissions, 29 non-transmissions) and 1249T alleles (9 transmissions and 15 non-transmissions). The haplotype -3G/1249G is the most frequent haplotype (92% of the population sampled), however there was no evidence for association with DD.

Scerri and colleagues also carried out analyses on a subgroup of their sample, comprising families that contained at least one proband whose average psychometric score was less than 1 standard deviation below the mean of the normative population. The resulting 264 families, containing a proband on average performing poorly across all the component measures, revealed one significant association between OC-choice and 1249G→T ($p = 0.0076$), where the G allele was associated with poorer performance, against the findings of Taipale and colleagues

(Taipale et al. 2003). Globally the haplotype -3G→A/1249G→T yielded significant association (p = 0.014) with OC-choice. The haplotypes -3G /1249G and -3A/1249T yielded associations (p = 0.014 and p = 0.0182 respectively). It is noteworthy that none of the associations observed remain significant after correction for multiple testing.

The most recent study was undertaken using 158 families (including 171 DD probands) in a sample of Italian children (Italian speaking) confirmed as having DD based on DSM IV criteria (Marino et al. 2005). Marino and colleagues (Marino et al. 2005) suggest that it is important to note that Italian words can be read both via the indirect phonological route or the direct lexical route. The latter is probably used more in the case of highly familiar, high frequency words. Spelling of Italian words requires a more orthographic knowledge compared to languages where a phonetic knowledge is required for successful spelling. This was accounted for in the study, which used only word spelling and sentence writing under dictation to represent the use of lexical orthographic knowledge. However, it may not be wise to directly compare this study with those using subjects whose first language is English, particularly on phenotypes of reading-related measures. Marino and colleagues (Marino et al. 2005) sequenced exons 2 and 10, since most of the previously reported SNPs lay within these two exons. Eight SNPs were identified, five of which had previously been reported (Marino et al. 2005). Dropping uninformative markers, three SNPs were tested for association with DD. TDT analysis (using FBAT) did not show evidence of association between any of the three SNPs -3G→A, 1249G→T or 1259C→G and DD defined categorically or with component phenotypes. Haplotype analysis revealed no significant associations either with subphenotypic measures of DD or as a categorical analysis (Marino et al. 2005).

The haplotype -3G→A/1249G→T reported by Taipale and colleagues and Wigg and colleagues (Taipale et al. 2003; Wigg et al. 2004) also showed no evidence of association ($p = 0.41$). This study was calculated to have 80% power to detect an association based upon a genetic additive model, a disease allele frequency of 0.06 and an attributable fraction of 0.13 at the 5% level (Marino et al. 2005). The componential phenotype analysis, based on a heritability for each component of 0.4 and 120 individuals, yielded 87% to detect an association.

6.6.5 ASSOCIATION BETWEEN *EKN1* AND DEVELOPMENTAL DYSLEXIA IN LIGHT OF ALL STUDIES: EXPLANATION OF THE RESULTS

DD is a heterogeneous neurological syndrome (Francks et al. 2002). Therefore it is not necessary to find an association between every componential phenotype and marker analysed. That said, few of the associations reported by Taipale and colleagues (Taipale et al. 2003) show replication in independent samples. Any replication reported tend to show trends in the opposite directions to the original study (Taipale et al. 2003) and fail to withstand correction for multiple testing.

Variations within *EKN1*, may be responsible for the problems of the family carrying the translocation, however the problems may not extrapolate to more common forms of DD. Indeed the phenotypic analysis of the family pedigree carrying the t(2;15)(q11;q22) translocation used to identify *EKN1*, do not have a phenotype akin to that used in later studies. Whilst most of the individuals in the family pedigree showed symptoms of DD, child II.3 (see Figure 6.1) differed from other children in that his overall cognitive achievement was below the normal range

in both verbal and non-verbal performance. It could be that the general phenotype in the family is not DD as defined in other studies, but a slightly different disorder or manifestation of dyslexia, and associations observed are for the alternative disorder. It is also possible that the translocation reflects linkage between DD and chromosome 15, rather than affecting DD directly. It is noteworthy that the linkage is not significant in this study (Taipale et al. 2003).

Subphenotypic analysis was not undertaken by Taipale and colleagues (Taipale et al. 2003), so subtle difference in the DD phenotype cannot be distinguished in relation to *EKN1*. The componential phenotype analyses undertaken by Wigg and colleagues (Wigg et al. 2004) enabled the specific components involved in their association between *EKN1* and DD to be identified. Such an analysis could identify the common components of reading that drive the associations observed by Wigg and colleagues and Taipale and colleagues (Taipale et al. 2003; Wigg et al. 2004).

It is important to point out that the functional significance of the SNPs implicated and suggested by Taipale and colleagues (Taipale et al. 2003), is based on association analyses and bioinformatic predictions only. As yet, no studies have been performed to assess the true functional significance of the two variants, -3G→A and 1249G→T. It is unknown whether Elk-1 binds to the -3G→A region of *EKN1 in vivo*, or whether the -3A variant alters binding efficiency. It is possible that the 1249G→T SNP introducing truncation of the EKN1 protein by four amino acids at the C terminus, may have no functional effect. Further analyses are needed in order to test whether the hypotheses that these SNPs are directly influencing DD susceptibility are true. Indeed the evidence suggests that these variants, if *EKN1* is truly involved in DD, are not themselves the susceptibility alleles conferring

susceptibility to DD given that both alleles at the loci -3G→A and 1249G→T have shown association to DD (Taipale et al. 2003; Wigg et al. 2004). If the association between *EKN1* and DD is not the result of a type I error, the susceptibility loci for DD (including the region DYX1) are in fact in LD with these *EKN1* variants. The conflicting results may in part be attributable to the distinct samples used in this and the three other studies to date (Taipale et al. 2003; Scerri et al. 2004; Wigg et al. 2004; Marino et al. 2005). Consequently, each of the samples analysing the *EKN1* association hypothesis may contain the same 15q susceptibility gene for DD, but the variants involved lay on different haplotype backgrounds.

Association previously observed between DD and chromosome 15 variations in our sample, lies centromeric to *EKN1* (Morris et al. 2000). In particular association is observed between DD and D15S994. Being approximately 15Mb away from *EKN1*. Given the lack of evidence of LD between any of the three SNPs analysed in *EKN1* and D15S994, it is unlikely that the *EKN1* gene is responsible for the associations we have previously shown (Morris et al. 2000).

The genomic region 15q21.3 (including *EKN1*) shows preferential sharing amongst siblings (Scerri et al. 2004). Consequently linkage and association analyses could be affected. Since association tested in our sample did not include siblings, it is less likely that the increased sharing of this chromosomal region between sibling pairs influenced our study. The increased sharing of this region and the use of related individuals in the study by Taipale and colleagues (Taipale et al. 2003) could explain the association observed in their sample and the inability to replicate the association in samples excluding related individuals.

6.6.6 *EKN1* AND COMORBID DISORDER ASSOCIATION ANALYSES

There are many disorders that are often shown to be comorbid with DD (see Chapter two, Section 1.6). It has also been pointed out that learning disorders and neuropsychiatric disorders of childhood onset (such as language disorders, autism and attention disorders) may have a similar origin given that component phenotypes may be common across disorders (Grigorenko 2003; Plomin and Kovas 2005). Consequently *EKN1* has also been examined for association with ADHD and autism.

In a poster presented at the World Congress of Neuropsychiatric Genetics, Dublin (2004), Dr C Barr and colleagues reported association between *EKN1* and a sample of 186 nuclear families containing an ADHD diagnosed proband (Barr 2004). A trend was noted for linkage to the ADHD phenotype as a categorical trait, as well as a biased transmission of a haplotype (uncorrected $p = 0.009$). Association was shown also between the same haplotype and inattention and hyperactive/impulsive symptom dimensions reported by parents and inattention symptoms reported by teachers. Our data showed one significant association between rs11629841 and inattention symptoms (one of the measures used in the diagnosis of ADHD), however, this is most likely due to multiple testing.

Recently, another study into autism and *EKN1* was undertaken. Autism is a severe neurodevelopmental disorder characterised by severe social and communication impairments, repetitive and ritualistic behaviours and restricted pattern of interests (2000). It usually presents before the age of 3 years with delayed speech development being one of the characteristic components. Ylisaukko-Oja and colleagues, tested for association between *EKN1* and autism, under the hypothesis that the influences of variations in *EKN1* extend to language-related disorders other

than DD (Ylisaukko-Oja et al. 2004).

No evidence for association was shown between any of the SNPs analysed and infantile autism (-3G→A, uninformative; rs3743204, p = 1; 572G→A, p = 0.404; 1249G→T, p = 0.394). Haplotype analysis using a sliding window method for two-, three- and four marker haplotypes also failed to show association between infantile autism and *EKN1*.

6.7 GENERAL CONCLUSION

It is unlikely that any study can exclude the hypothesis that a gene plays a role in a complex cognitive disorder. However, given the lack of support for association between DD and variations in *EKN1*, it is unlikely that variations within *EKN1* explain the linkage and association evidence suggesting a gene on chromosome 15q increasing susceptibility to DD. A lack of power, sample-related heterogeneity, with different susceptibility alleles or different patterns of LD in different samples could explain differences in the results obtained across studies.

If *EKN1* does influence susceptibility to DD, the data so far however, seem to suggest that the variants from the original study (Taipale et al. 2003) that seem plausibly functional are not themselves susceptibility variants for DD, given that in the follow up studies, any associations or trends for association, were in the opposite direction from the original study. Whether there are different patterns of as yet undiscovered variants marked by different patterns of LD is a much more difficult question to answer. In the original paper, despite sequencing all the exons of *EKN1*, no other associated variants were detected. This suggests that the susceptibility variants themselves if they exist in this gene, are located in non-exonic sequence. This hypothesis can ultimately only be tested by detailed population-specific

sequencing and genotyping across the whole gene (including introns, and distal regulatory elements) or by utilising data available from the HapMap project in a haplotype tagging study. Given that even excluding distal regulatory elements, the gene represents 80kb of sequence, both methods are far from trivial endeavours, which are probably not justified by either the current association support for *EKN1*, or by the *a priori* case for *EKN1*. Although the breakpoint of the translocation that originally pointed to *EKN1* as a candidate gene disrupts the gene, the evidence for co-segregation between the translocation and DD is not statistically significant, with a maximum possible LOD score of 1.2. Moreover, even if the evidence for co-segregation were statistically significant, this may simply reflect genetic linkage, not direct gene disruption. Finally, if we assume that the translocation is indeed directly relevant to DD in the family, the mechanism might involve position effects affecting the expression of other genes on 15, or even genes on 2q.

Acknowledging that the exclusion of any gene for a complex disorder is not possible, based upon the weak *a priori* case and the patterns of non-replication, *EKN1* is unlikely to be a susceptibility gene for DD. It can also be concluded with considerably greater confidence, that it is not responsible for our previous reports of association evidence between DD and chromosome 15q.

CHAPTER SEVEN

GENERAL DISCUSSION

CHAPTER SEVEN

7. GENERAL DISCUSSION

DD has a highly heritable component (see Chapter two). Linkage and association studies have been undertaken to determine regions of the genome which are more likely to harbour susceptibility genes for DD (see Chapter two). As of April 2005, nine putative linkage regions for DD have been identified and designated DYX1 – DYX9 (OMIM: http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id= 300509, HUGO: http://www.gene.ucl.ac.uk/cgi-bin/ nomenclature/searchgenes.pl). DYX2 on chromosome 6p shows most evidence for linkage to DD, with all but one sample tested, showing linkage to the region.

Given the strength of evidence for linkage between DD and chromosome 6p21.3-22 (Cardon et al. 1994; Cardon et al. 1995; Grigorenko et al. 1997; Fisher et al. 1999; Gayán et al. 1999; Grigorenko et al. 2000; Smith et al. 2001; Kaplan et al. 2002; Willcutt et al. 2002; Grigorenko et al. 2003; Deffenbacher et al. 2004; Francks et al. 2004) the main aim of my work was to identify a gene in this region, that increases vulnerability to DD. Further, given the published evidence of a susceptibility gene for DD on chromosome 15 (*EKN1*; see Chapter six), I also aimed to determine whether this gene was likely to account for the association previously observed on chromosome 15 in the sample used in this thesis (Morris et al. 2000).

7.1 CHROMOSOME 6p

In order to identify susceptibility genes in the chromosome 6p linkage and association region (DYX2), I tested 11 positional candidate genes on chromosome 6p

for association with DD using either a direct or indirect association analysis.

Direct association analysis of *SOX4* revealed two SNPs, one synonymous and one non-synonymous. Despite repeated efforts to test these SNPs for association to DD in DNA pools, I was unable to optimise an assay which would allow PCR amplification of the regions of DNA containing the SNPs.

Indirect association analysis based on a dense map of SNPs across ten genes revealed evidence for association between SNPs in *MRS2L* ($p = 0.003$), *KIAA0319* ($p \leq 0.0001$), and *THEM2* ($p \leq 0.0.002$). Logistic regression techniques were used to determine which minimal set of SNPs could account for the association signal observed with DD. Following this, the resulting SNPs were used to construct haplotypes. All haplotypes comprising the 7 SNPs identified in the logistic regression analyses were tested for association with DD. This analysis identified a highly significant association between DD and a haplotype defined by SNPs rs4504469 and rs6935076 spanning *KIAA0319* ($p = 0.0001$). Specifically the A-G haplotype (rs4504469 and rs6935076 respectively) was under-transmitted to the cases suggesting a protective effect against DD ($p = 0.00003$). A second haplotype defined by alleles G-A also yielded evidence of association ($p = 0.02$), with over transmission to DD cases, suggesting this haplotype is associated with increased risk.

In order to determine whether there were any potential *cis*-acting influences on *KIAA0319* gene expression, allele specific expression analyses were undertaken. Assays were performed which allow expression differences of *KIAA0319*, based on the alleles present at the SNP rs4504469 to be determined. No differences in *KIAA0319* expression were observed, however this does not exclude the possibility that expression differences exist with other transcripts of *KIAA0319*, not containing the SNP rs4504469.

A detailed examination of the DD phenotype (Chapter five) revealed a number of component phenotypes which could be influenced by variation tagged by the rs4504469/rs6935076 haplotype. It is therefore likely that the association observed is not specific to one component phenotype. Associations observed with quantitative measures of component phenotypes, suggested that variation on the rs4504469/rs6935076 haplotype may influence single word reading across the ability spectrum. However analyses of other component phenotypes suggested that normal ability was not influenced by variation on the rs4504469/rs6935076 haplotype and problems in these component processes form distinct pathological groups from normal reading ability.

Categorical analysis (Chapter five, analysis one) suggested that in people with DD, the symptoms of ADHD are also associated with rs4504469/rs6935076. However, ADHD symptoms do not drive the association between the haplotype and DD since association with DD was still observed when probands with comorbid ADHD were removed from the analysis (Chapter five, analysis 3a). Further, the absence of association between rs4504469/rs6935076 and ADHD (Chapter five, analysis 3b) suggested that susceptibility to ADHD *per se* is not influenced by rs4504469/rs6935076. The analyses do not exclude the possibility that phenotypic components common to both DD and ADHD are influenced by *KIAA0319*.

7.2 DISCUSSION OF THE TECHNIQUES EMPLOYED IN IDENTIFYING SUSCEPTIBILITY GENES FOR DEVELOPMENTAL DYSLEXIA ON CHROMOSOME 6p

This study was mainly based upon an indirect association approach to positional candidate genes. Although indirect approaches do not aim to directly test

the actual causal variants, association can be detected through LD between SNPs tested for association with DD and the causal variation.

The use of DNA pooling as a screening step allowed considerable reductions in the cost, labour and the amount of DNA required to identify associations. The error rates resulting from the use of DNA pooling have been shown to be very low, even when applied as in this study, to samples mainly from mouthwashes (Norton et al. 2002). It is a cost effective technique, essentially being the cost of primers and therefore provides a cheap alternative to individual genotyping. The DNA pooling strategy employed in this study proved to be highly accurate at determining association to DD. For the 12 SNPs showing association with DD in the DNA pools and chosen for individual genotyping, 9 were confirmed by individual genotyping. Although association was not confirmed for two SNPs, they did show trends towards significance ($p = 0.06$ and $p = 0.07$).

Accuracy of pooling is usually improved by correction of unequal representation of alleles in the SNaPshot™ reaction. However, in order to identify heterozygotes from which to derive this factor, given the low minor allele frequency of some of the SNPs, a relatively large number of individuals would be required to be genotyped. To overcome this problem, I applied a range of values for the correction factor *k*. These were based on empirically derived estimates of *k* from a number of assays for each combination of alleles at polymorphic loci (Moskvina et al. 2005). This strategy then required heterozygotes to be identified only for those markers showing association to DD based upon one or more values for *k*, thus cutting down on the amount of individual genotyping required.

One problem resulting from selecting SNPs from public databases is that in some instances, the allele frequencies are unknown. The result is that some markers

turn out not to be not polymorphic. However, it was a fairly simple matter to fill in the gaps with replacement SNPs. It is now also possible to select SNPs from the HapMap project for which individual genotyping data are available, cutting down the number of non-informative markers.

The approach to DNA pooling adopted here does not allow the formation of haplotypes. Although suggestions have been made to overcome this problem by (Barratt et al. 2002), haplotype association analysis in DNA pools remains a problem as it relies on deducing haplotype frequencies from allele frequencies in DNA pools. The aim of DNA pooling is to accurately detect differences in allele frequencies between DNA pools rather than accurately estimate allele frequencies within DNA pools. This could result in problems in estimating haplotype frequencies. However, this problem is offset by the higher density of markers that would be required to acquire the same information across the region without the use of haplotypes.

The approach I adopted meant that DNA pooling could not be applied to componential phenotype analysis. DNA pooling was employed to keep the amount of pooled genotyping to a minimum with two replicates each of three DD case pools and four control pools. In principle, DNA pools could have been made to allow analysis of component phenotypes. Thus, I could have constructed multiple pools each one containing individuals performing poorly on a specific component task. However this would require genotyping at least 15 pools, and even more if a number of different cut-off scores to form the categories for component phenotypes needed to be tested. Given the cost of SNaPshot™ compared to other methods discussed in Chapter five, the use of so many pools would have resulted in the loss of any economic advantage. Also, the examination of multiple component phenotypes either by pooling or by individual genotyping would have reduced the power of the

study as a result of an increased requirement to consider the effects of multiple testing. For these reasons, DNA pooling was used to detect associations between variations on chromosome 6p with DD and individual genotyping used to follow up these associations and allow haplotype and componential phenotype analyses.

7.2.1 Indirect versus Direct Association Design

Direct association studies are usually targeted to the coding sequence and their immediate flanking regions because of the high cost both financially and in time of *de novo* polymorphism discovery. As a result, when association is tested for, unless LD exists between the variants in these areas and those variants of functional importance in the unscreened sequence, association that results from the influence of the latter will not be observed. In my case, if I had adopted a direct approach to *KIAA0319*, it would be expected that rs4504469 located in exon 4 would have been detected and included. However, this is the only exonic variant associated in this sample and so only a single association signal would have been observed. Under these circumstances it is unlikely that intron one (and other introns) would have been included in *de novo* polymorphism discovery and consequently the associations in intron one observed in Chapter 4 would not have been detected. The association with rs4504469 was observed in the DNA pooling analysis. The number of other association signals observed using the strategy strengthened my enthusiasm for further genetic analyses of *KIAA0319*. Given the single association that would have been observed in a direct study, further genetic analysis of KIAA0319 would not have been undertaken.

It has been suggested that the subtle, general effects of complex phenotypes such as DD are unlikely to be the result of mutations in exons, which are often the

cause of rare and more severe disorders (Plomin and Kovas 2005). Given this and that under direct association analysis the signal in *KIAA0319* would not have been followed up, it is important that in the future, non-coding sequence within genes is included in direct association analyses.

The results of this study suggest the combination of DNA pooling and a high density SNP grid can successfully detect the influence of susceptibility variants but that does not mean it will also be successful when applied to DD or other complex disorders. A successful outcome here was aided by the low recombination rate across the region. In regions of high recombination, a direct association approach may be a more suitable alternative to an indirect approach given that LD would not be expected to extend far between SNPs and consequently the genotyping of a large number of SNPs would be required to extract the LD information across the region. Nevertheless, a design similar to that employed in this thesis is likely to be effective in many other circumstances since randomly selected SNPs spaced 2-3kb apart as in this thesis capture >80% of the common SNPs within a region in a Caucasian population with an $r^2 \geq 0.8$ (Consortium 2005),

Given genotyping data made available by the International HapMap project, it is now possible to identify tag SNPs across genes which detect greater than 90% of the LD information within a gene. This data my allow a reduction in the number of SNPs required to be genotyped across regions (both of low and high recombination rate), making indirect association studies more feasible option for association design even in regions with a high recombination rate. It is also possible to identify tag SNPs that do not require the formation of haplotypes and which therefore allow DNA pooling methods to be undertaken using tag SNPs.

7.3 REPLICATION OF AN ASSOCIATION BETWEEN DEVELOPMENTAL DYSLEXIA AND *EKN1*

In order to replicate the association between *EKN1* and DD reported in other studies (Taipale et al. 2003; Wigg et al. 2004) and determine whether this gene could explain the association observed between DD and chromosome 15 previously reported in the sample used for this thesis, I tested variations within *EKN1* for association with DD in a large sample of parent DD-proband trios (Chapter six). I also tested for associations between variations and component phenotypes of DD (both quantitative and categorical definitions). Since this was a replication study, only associations reported in previous studies were tested. No evidence of association was observed between DD and SNPs within *EKN1* ($p \geq 0.42$) and further, no association was observed between the haplotype -3/1249 and DD. Subsequent studies have also failed to yield evidence of association between *EKN1* and DD (Scerri et al. 2004; Marino et al. 2005).

Quantitative component phenotype measures of DD were previously reported to show some evidence of association with variations in *EKN1* (Wigg et al. 2004). I tested for these associations using the same methods as Wigg and colleagues (Wigg et al. 2004). Although some associations were observed, they are likely to be attributable to issues regarding multiple testing. The small number of informative families used in the analysis chosen by Wigg and colleagues (Wigg et al. 2004) could result in a reduction of power to detect associations, as a result I used all the available component phenotype data to screen for differences in component phenotypes based on genotypic groups. Although an association was observed with inattention symptoms based ($p = 0.02$), given the large number of tests performed, it is likely that the association represents a type I error.

The associations detected in the previous studies (Taipale et al. 2003; Wigg et al. 2004) are unlikely to have identified the actual causal variations increasing susceptibility to DD if this gene is involved in increasing susceptibility to DD, since the 'risk' allele in the study by Taipale and colleagues (Taipale et al. 2003) has the opposite effect in later studies (Wigg et al. 2004). Given that the exons of *EKN1* have been selected for *de novo* polymorphism discovery in other studies, it is more likely that variation within the non-coding sequence of *EKN1* influences DD susceptibility, if this is a true susceptibility gene for DD. Although a direct association approach could be undertaken to assess association between non-coding sequence in *EKN1* and DD, it is possible to reduce the cost and time required to undertake this approach by performing an indirect association approach using tag SNPs selected using data from the HapMap project.

Although the study presented in this thesis does not allow the exclusion of *EKN1* as a susceptibility gene for DD since not all variation in *EKN1* was tested in our sample, given the results and that previously, association to chromosome 15 in this sample was observed over 15Mb proximal to the *EKN1* gene, it is unlikely that *EKN1* can explain the association previously observed (Morris et al. 2000). Systematic approaches will help identify genes that may be responsible for the linkage and association observed on chromosome 15.

7.4 FURTHER WORK

7.4.1 *DCDC2*

Two studies have recently suggested that the association between DYX2 and DD may be the result of variation within *DCDC2* (Meng et al. 2005; Schumacher et al. 2005). Meng and colleagues (Meng et al. 2005) observed association between

SNPs in *DCDC2* and three components of DD. In addition, a deletion was identified within a region that encodes a putative brain-related transcription factor binding site (Meng et al. 2005). Schumacher and colleagues (Schumacher et al. 2005) observed evidence of association between DD and markers within *DCDC2* in two samples of independent families broadly supporting the association observed by Meng and colleagues (Meng et al. 2005). In neither study was association with *KIAA0319* observed.

I observed no evidence for association between variation within *DCDC2* and DD using DNA pooling and individual genotyping. However, due to the time and cost to analyse the 60-90 SNPs required to fully cover introns 2, 7 and 8 of *DCDC2*, these regions were not comprehensively assessed for association with DD. Given the associations now observed between *DCDC2* and DD (Meng et al. 2005; Schumacher et al. 2005), it will be important that markers within these regions and the deletion in intron 2, are tested for association with DD to exclude the possibility that the association observed in this thesis is not the result of a causal variation in *DCDC2*. It is unlikely that the association with *KIAA0319* is the result of variation within *DCDC2* given the lack of LD between SNPs within the genes based on HapMap data. It is possible that there are two susceptibility genes for DD on chromosome 6p or that population specific genetic heterogeneity could account for differing results across the studies given the samples are from different populations.

7.4.2 *KIAA0319*

One of the main aims of future work will be to see if replication of the results presented in this thesis can be achieved. Evidence consistent with the results of this thesis has been achieved by other groups (Francks et al. 2004; Couto 2005) and to

some extent our sample achieves some evidence since the parental control genotypes in the parent DD-proband trios are independent of those used in the case-control sample (Plomin et al. 2004). Currently, our laboratory has no independent replication set available for DD research and so collaborations will be required with other research groups to achieve this aim.

Although rs4504469 is a non-synonymous SNP, the absence of association in two samples showing association with the gene (Francks et al. 2004; Couto 2005; personal communication) suggest it is not responsible for the observed association. It will be important to try and elucidate the causal variation(s) responsible for the association and determine the effects on the gene and/or protein.

Little is known about *KIAA0319* and the protein which it encodes. There is only one known transcript of *KIAA0319*. Alternative splicing of exons can affect the functional properties of a protein and so it will be interesting to identify novel transcripts of *KIAA0319* (if they exist). Protein isoforms resulting from alternative splicing can be tissue specific, result in membrane bound or soluable forms, alter intracellular localisation and can alter protein function. SNPs altering splicing may result in differences in these protein isoforms between DD cases and controls. Identification of alternative transcripts of *KIAA0319* may also allow the detection of allele expression differences with different transcripts, especially if the tag SNP rs4504469 is not present in some transcripts of *KIAA0319*. Further expression analysis could allow the localisation of KIAA0319 and any alternative transcripts within the brain and more specifically within cell(s). It would also be interesting to investigate the expression patterns of *KIAA0319* in development. The gene *FOXP2* has been shown to be associated with severe speech and language disorder (SSLD) (Lai et al. 2001) and association between early expression of *FOXP2* and later

pathology in the brain has been shown to result in SSLD (Lai et al. 2003).

By determining the interactions of KIAA0319 with other proteins and how causal variations influence these interactions, it may be possible to determine the pathways in the brain that enable reading and why and how the causal variations influence DD pathology. In the future this might allow intervention pharmacologically and educationally to alleviate the problems of DD.

It has been suggested that KIAA0319 interacts with KIAA0369 and KIAA1299 (Nakayama et al. 2002). Both map close, although not within, DD linkage regions; *KIAA0396* to the DYX1 locus and *KIAA1299* to a chromosome 16 locus weakly implicated in a genome scan by Raskind and colleagues (Raskind et al. 2005). In future studies it would be interesting to determine whether polymorphisms present in *KIAA0319* alter the interactions with KIAA1299 and KIAA0369.

7.4.3 Chromosome 15

Given the association observed in the sample utilised in this thesis between chromosome 15 and DD (Morris et al. 2000) it will be important identify the gene(s) responsible for this association. It is unlikely that variation within *EKN1* is responsible for our previous association. Association analyses will be required in order to refine this region of association. Further assessment of the approaches most appropriate to the region (direct association analysis vs indirect association analysis, tag SNPs vs SNP grids, LD across the region, HapMap data available and the (putative) function of genes in the region) will be required in order that the most appropriate methods are used to maximum efficiency.

APPENDIX 1

SOX4 Fragment	Region of Gene	Annealing Temperature (°C)	DHPLC Temperatures (°C)	Fragment Size (bp)
A	Promotor	58	57, 59	345
B	Promotor	60	55, 60, 62	381
C	Promotor	58	56, 61, 63	366
D	5' UTR	58	61, 63	391
E	5' UTR	58	61, 63	341
F	Exon 1	61	61, 66, 68	399
G	Exon 1	58	59, 64, 66	399
H	Exon 1	61	65, 67	307
I	Exon 1	61	66, 68	399
J	Exon 1	61	57, 62, 67, 69	367
K	3' UTR	58	58, 63, 65	364
L	3' UTR	58	58, 63, 65	384
M	3' UTR	61	61, 63	357
N	3' UTR	59	53, 58, 63, 65	389
O	3' UTR	59	53, 58, 60	321

Table 1 The sequence of *SOX4* was split into 15 PCR fragments. DHPLC was undertaken at a range of temperatures to allow maximum efficiency in detecting variations.

SOX4 Fragment	Forward PCR Primer	Reverse PCR Primer
A	cggagcactacctaatgtgttt	gctttaagaaagagctgggga
B	tgatgacaggtgtctaggtaccac	cacacacacagcaaaaggaaa
C	catttccgtctctcccctta	gtctgctctaagctgcagca
D	ctgcagccaagactgtgaaa	tttcctgagctaccgagagc
E	tcctggagccacagtctttt	gagaggcggagaactccttc
F	gcctgggaactataactcctct	gagctggtgcaagacccc
G	aaaccaacaatgccgagaac	caggaagaaggtgaagtccg
H	tggctgactaccccgactac	caagctcatcctggcagg
I	agagctgcggctccaaagt	tgtacgaggaggagggcg
J	gtgaagcgcgtctacctgtt	gaaccccagctcaaactttg
K	gagttcgaagacgacctgct	aaaagtaagcagggctggct
L	gaaacgaaaaggacagacgaa	gacttgaaggagtctccccc
M	ccagcaagaaggcgagttag	gctaggaaatgacccgagaa
N	cggaggaggagatgttgagg	aatcggaatcgtgatggtgt
O	gtggtacaggggcagtcagt	ggattcaaacgcaactcaaa

Table 2 The PCR primers utilised in the PCR of the 15 fragments of *SOX4*. All primers are described 5' → 3'.

Gene	rs Number	Alleles		Forward PCR Primer (5' → 3')	Reverse PCR Primer (5' → 3')	Fragment Size	Annealing Temperature	Location
ID4	1980461	G	C	ttccatgacaaaaggatttgg	ttggtccatgaaaccagaga	291	Touchdown	19942988
ID4	3193769	G	C	ggtcatcacggctttctagc	acaatcctgcacgttttggt	387	54	19947317
PRL	1205961	T	C	tttgttctacccccctcacctt	ggtggtattgtcaggctggt	133	60	22393991
PRL	6239	T	C	tgtgggcttagcagttgttg	aaatgagatctaccctgtctggtc	181	60	22395724
PRL	1205957	G	A	tgtagacgccgatctgctaa	tgtcttagggctcaagtaccaa	251	60	22397119
PRL	849886	T	C	gcctgcctaggttttgctta	gcaatgctcccagaaaaact	244	60	22399346
PRL	1205954	G	A	ctgtggaggcccttgattta	cccttgttccctccttcct	287	60	22401833
PRL	2244502	A	T	aagccttattgctccccact	agaagcacttctgggaaacg	181	60	22402966
PRL	6237	T	C	ccaaaacatctcccgacatt	gatcgccatggaaaggtatg	245	60	22405134
PRL	849876	T	C	gcaactgaggggtgatgatt	ccgccctgtgaaggattt	288	56	22407617
PRL	849875	G	A	gaggcaggaggatcaagagat	ttgcagtagacacctcctaggtt	285	64	22407979
VMP	12208318	G	C	cttcgagagacatgggaagc	tccacactcttctcctggatg	265	58	24243823
VMP	10946676	C	T	taggatgtgggctctggttg	caggcaccatctcacagaac	328	58	24252846
DCDC2	1277155	T	C	gatcgagacgatcctggcta	ggagtcttgctttgtcacca	202	56	24276123
DCDC2	1832709	A	C	gggttggtaggcaaatagacc	gaaatcattaactccagacccagt	400	60	24280941
DCDC2	3789219	T	C	tacttctggccttggtggtc	aaatgagtcagggcctttca	287	56	24283103
DCDC2	1419228	T	C	gtttgtgtctgcccttctgc	cccaagtgcctaatgtgttg	330	64	24286285
DCDC2	2996452	T	C	gggcttgcagtgagcaga	ccagctttcagagccaagac	372	64	24288345
DCDC2	1277192	A	C	catttgtcaatttcggcttt	accaactgagcagcacagag	345	60	24308388
DCDC2	1277194	A	C	aactggaactggacccctttt	catttgtcaatttcggcttt	200	Touchdown	24311703
DCDC2	793861	A	T	tccagtccaatgaggagaca	tggacacaaaatgtcatccaa	300	56	24314595
DCDC2	793862	T	C	gatggctgtctctgggtcat	cagcaaagctcaacgctgt	400	60	24315179
DCDC2	870601	T	C	ttgcctcacagttgaacagc	tgtgggctgtcattgttctg	335	58	24369039
DCDC2	807698	T	C	gggtagtcggattgcttgag	ttttatcttcatgcccacca	330	54	24384232

Table 3 Continued ...

Gene	rs Number	Alleles		Forward PCR Primer (5' → 3')	Reverse PCR Primer (5' → 3')	Fragment Size	Annealing Temperature	Location
DCDC2	807726	T	C	gctgctgtttttatgcttgg	ccatctctcacctgctagtgc	376	64	24385979
DCDC2	3789224	T	C	accccagatacatggctgag	cactcaatacggtgtgagtttga	284	56	24388080
DCDC2	3789227	T	C	tttaacctcaccccacaagg	ccacaccaaaaaggaccatc	359	60	24395508
DCDC2	2296539	G	A	tgaatgcaaatgccagtgtt	cccacctggaagtcaactgt	366	60	24397431
DCDC2	2274305	T	C	ggtgaaaccctgtacagcaag	acccaaaaggccttctcatc	398	64	24399182
DCDC2	6907864	T	C	gccaatggcctatatccaag	ccttgggaggatcacttgag	254	60	24401303
DCDC2	807709	T	C	taaaggatgatggcctccag	ctaccgtttgacccagcaat	268	Touchdown	24405282
DCDC2	807704	G	A	cgtcagatgggagggttaca	atatttccgggatgatgctg	379	60	24408825
DCDC2	807703	T	C	aggcgagaagctgggtttat	caacgtgtcagctcgcttta	216	60	24410066
DCDC2	3857541	T	C	acatctacccagtggcaagg	taagccctaggcacactgct	372	60	24410512
DCDC2	707862	G	A	gcataggagtgggaaggaca	ttaacagggctgggtgtgtt	400	60	24412730
DCDC2	807685	A	T	ttgcctcacagttgaacagc	tgtgggctgtcattgttctg	558	58	24418623
DCDC2	793704	G	C	ttaatggcttggtcctggtg	ggctgctttatcaggctgtg	537	58	24444623
DCDC2	793722	T	C	gcctgggtgacagaagga	aaacctggacttccgcaata	201	60	24460342
DCDC2	793720	T	C	gggaaggttccctgtcaagt	ccagggtttcagaagctacg	315	64	24461259
DCDC2	1277350	G	A	atgcacagggaaagttttgg	gacaattcgtgggctttcat	252	60	24463129
DCDC2	1277349	G	C	tgctacccttccaaacagc	acctgggcaactacatccag	119	60	24466462
DCDC2	2792666	G	C	tggttatggagcttgtggaa	cagctcatctctgccagtgt	364	54	24469168
DCDC2	793663	C	T	aatggggaaaagagtaatcaaaatg	tctatcagtttggccgttacactc	477	58	24477494
MRS2L	9393553	G	A	tggaacttgtgtctgccttg	ggctgaggcataagaattgc	537	58	24509633
KIAA0319	2817241	T	C	ggctcctggatcatcttcaa	ggtcatttgaggccaggagt	371	60	24650242
KIAA0319	807526	C	T	ggccaaatcctgtctcaaaa	gggattacaaagagccacga	351	60	24651699
KIAA0319	2817243	G	A	ggtgaccagggttatggttg	catccatcaaggcaggagtt	346	51	24653575
KIAA0319	2817245	G	A	gtcaaagctgggacagaagc	ccatcttttctcggctgaag	307	52	24655585

Table 3 Continued …

Gene	rs Number	Alleles		Forward PCR Primer (5' → 3')	Reverse PCR Primer (5' → 3')	Fragment Size	Annealing Temperature	Location
KIAA0319	807532	G	A	ctatgttgcactgcctctgc	ggtggacaggtgcctgtaat	369	60	24657647
KIAA0319	2076313	G	A	agatgcataatggccctctg	gctgactccagcaaatcgtt	375	54	24659603
KIAA0319	807536	C	A	ttcacttttgccgctgtcta	gcttctggacagccaaacat	357	54	24662253
KIAA0319	4083411	G	A	cttcagcccaggaagttgag	gccagcaatagtgagccaat	353	54	24664281
KIAA0319	2760167	A	T	gccaacatggctaaggctct	ttcagtagactccagggaaatga	273	53	24666498
KIAA0319	807542	G	A	aggaaaatgccaaccaaacc	attagcacctcctccccaca	589	57	24668312
KIAA0319	807544	C	A	ggatgcagtggctcatacc	ctgcctcctaggttcaagca	345	53	24670281
KIAA0319	2744549	A	G	ctaggaagagtggcctggtg	ggccatactactgcccttg	320	54	24671531
KIAA0319	2760161	G	A	tagacaggcgtggtggtg	gagtggtcacagggaaggaa	196	53	24673389
KIAA0319	2817195	T	C	tccagggtagcactttccac	tcactccaggattccaaagg	219	53	24678323
KIAA0319	807521	G	T	cctttggaatcctggagtga	ttgatctcctgacctcgtga	228	53	24678410
KIAA0319	3846835	A	C	ctcattgccaggaactggat	tgaaggggctggttcttatg	240	56	24680310
KIAA0319	2744556	C	G	aacctccacccccctgagtt	ggacacttggatgacagtgg	354	60	24682096
KIAA0319	807525	C	T	tccactggtcgctttttagg	taattgcaaggcatgttcca	382	53	24684016
KIAA0319	2817199	G	A	cctaaaaagcgaccagtgga	caattggctctgtgctgaaa	203	56	24684227
KIAA0319	2760158	A	T	agtgggtccctgcctttatt	ccaggctgagtcaggacact	341	56	24686016
KIAA0319	2760157	T	C	cccatgaaagccaaatctct	agtgggtccctgcctttatt	118	56	24686251
KIAA0319	807507	G	C	cagccgtgtgtatcctgaga	cacaaagtgattttccaactgc	315	50	24687846
KIAA0319	807509	G	A	tcaagtgatctgccctcctt	ggcaaattgtgacctggaat	366	59	24690011
KIAA0319	2817201	A	C	actcccaggccagtttcttt	gtgggccatgtattttggag	131	56	24693193
KIAA0319	4504469	G	A	gtgcctggagggaatgagta	aaaagcccagtgctcacagt	230	56	24696863
KIAA0319	5026394	G	T	caactggtgaaacctgcaac	acaaacccagatgcctgaaa	197	56	24698526
KIAA0319	4576240	G	T	gctcccggtagtcatctgag	ctgctggactatggggacat	153	60	24704457
KIAA0319	4352670	A	C	cccgacagatgtctccagtc	ggtgttaggctaggcagcag	259	60	24712047

Table 3 Continued ...

Gene	rs Number	Alleles		Forward PCR Primer (5' → 3')	Reverse PCR Primer (5' → 3')	Fragment Size	Annealing Temperature	Location
KIAA0319	4712831	C	T	caggaaaagatgctgctgag	agaactgggtcattcggaga	170	Touchdown	24713548
KIAA0319	4236032	G	T	cttctttgggtgacggtgat	gagcctcattctctgcatcc	223	56	24714706
KIAA0319	6911855	T	C	tggacagccaacctttgaat	cccacctcagcctctcaag	285	Touchdown	24715290
KIAA0319	4712833	T	C	caagcctccctcaaaacaaa	tgactggcttctggcttctt	267	56	24719122
KIAA0319	6939068	A	T	ctcgatctcctgacctggtg	aagctctcagcgaccaaaag	155	56	24721274
KIAA0319	7763790	A	G	tccaagcagttaggtgtagca	aatgggatgtatgggaacca	140	56	24723042
KIAA0319	2745335	T	C	aaatggaacctcaaatgtcca	ttttccccagcaatctcaag	244	56	24725393
KIAA0319	2745334	T	C	aaatggaacctcaaatgtcca	ttttccccagcaatctcaag	244	Touchdown	24725509
KIAA0319	2817206	A	T	ggtcgggagtagtggagtca	caagttctggagtgccacag	128	56	24728381
KIAA0319	7751357	G	C	cagaatgagctcccctctct	gcatttcacatttaggccttt	233	56	24733377
KIAA0319	6917660	G	A	gtgtgcctctgggtcaaaat	aggacaggcaaatttcagga	235	56	24737085
KIAA0319	6456622	G	A	cctctttccttgccgtatca	catgccacctattgctcctt	225	56	24739537
KIAA0319	7766230	G	A	ttccaaaccctccttttca	ggaatgggatctccctgatt	234	56	24741408
KIAA0319	2206525	T	C	tgaggacaatgacccaaaga	cccttgcagattgatgtgg	122	56	24744882
KIAA0319	7755563	G	A	acccaaaatgcaattcaagg	tcgatctcctgaccttgtga	213	56	24746690
KIAA0319	7755579	A	C	ctcctgatggcttcaaaagg	cccaggttcacaccattctc	258	56	24746719
KIAA0319	6935076	G	A	aaacccacaccctctgacct	catcacagggtgcactttca	284	56	24752301
KIAA0319	4363021	T	C	taagttcttgcctcggcatt	tgctcactttccaaatgctt	297	56	24753130
KIAA0319	2038137	G	T	ctttcctgaatgcccagaag	acgacgaggaggaacaagtg	240	58	24753922
KIAA0319	3756821	T	C	ccacccaagttttctgtggt	cagtctgggggatgaacact	352	60	24754800
KIAA0319	1555089	G	A	caggcttagcccagaatgtt	aaatgtcccaaccaaaagca	296	56	24756385
TTRAP	3212236	T	C	aaatgtcccaaccaaaagca	ctgattttgtccccaacgat	369	51	24756434
TTRAP	3087943	T	C	atcactggggtcttctgtgc	tatttcggtgcctgaatgga	375	56	24758740
TTRAP	2294691	G	A	tgctgtatgcctaggggaag	gtggcaggaaaagtgggtaa	308	51	24760822

Table 3 Continued …

Gene	rs Number	Alleles		Forward PCR Primer (5' → 3')	Reverse PCR Primer (5' → 3')	Fragment Size	Annealing Temperature	Location
TTRAP	3181238	T	C	aaccattgcctttagagaaaaa	cctgtggcacttatgccttt	330	51	24762422
TTRAP	3212234	G	C	ccttggcttttggttcacat	tgttagcccacggctgatag	365	56	24763930
TTRAP	3212232	T	C	tgttgtgggatttccttcatt	tcctgaaaaatcatccatttcc	390	50	24765876
TTRAP	2143340	G	A	ggtcaacactggcaagatca	gcagcgtgattaatgctcaa	201	56	24767050
TTRAP	2056999	A	G	gcaacaggatgcttggtagg	tttcccattcctggtgatgt	375	52	24769490
TTRAP	3756819	G	T	ggccagtcagatcctctcct	gggcctattacaggaaatcca	387	57	24773319
TTRAP	1061925	A	C	ttgtaacaggggtcatgcag	cctctccagttcctgcactt	331	57	24773986
TTRAP	3756815	G	C	cagcttggggctttaagaca	cctctccagttcctgcactt	230	56	24774173
THEM2	3181227	T	C	ctctgcgggaggtgataaag	agcaatgcagacgtttgttg	358	51	24775778
THEM2	2223589	G	C	caaccatttgtctcccacct	ctccagatccgattctcctg	343	56	24777451
THEM2	9393576	T	C	tcccagatttcatctgatcg	agccctgcaaactgaagcta	298	56	24777924
THEM2	7765052	G	A	tgccttttaaaggggagag	ggccaacgtagtgacacctc	298	56	24781599
THEM2	7451561	T	C	cctgaaaagccacttctgct	gagggttccttggcataaca	148	60	24785967
THEM2	2143338	G	A	gggtttgtatcccgatcctt	cccctgacctttgacaaatc	294	56	24791260
THEM2	1555086	T	C	atcaggcagacatggtctca	tagctcactgcagccttgaa	232	56	24793000
THEM2	1205120	G	T	ggcagagttgctagggataaa	cgtgtggatcatgaggtcag	348	50	24793537
THEM2	926529	G	A	ccatgtggactccaacacac	ctctccagtgtgttcctcctct	139	56	24795744
THEM2	1885209	G	A	tgtccaaccatgggagaatta	caaagtgcatagtgcctgaca	203	56	24796937
THEM2	1885211	G	A	aggccctcagactttccatt	ccaccatccccatgtttagt	304	60	24800158
THEM2	3777664	G	A	cagcagaatgcaacccttttt	agactcaggttgccatagcc	270	56	24801825
THEM2	2092404	A	C	gcccagccaaaaatcataaa	ccacccagctgagatctgtt	326	59	24805367
THEM2	3777663	T	C	gctgcctacgcttcaaaaac	attaagtccggggaatttgg	309	51	24808214
THEM2	1056319	A	C	ggtccaatagaggggaaga	ttgaggctatgggaagctgt	213	56	24811025
THEM2	1053598	T	C	aggaagcacaaagtgtgcaa	tggctcttaaggcctatgct	265	56	24812517

Table 3 Continued …

Gene	rs Number	Alleles		Forward PCR Primer (5' → 3')	Reverse PCR Primer (5' → 3')	Fragment Size	Annealing Temperature	Location
C6orf62	3756814	T	C	gtcttcccatttgtgctggt	tggggaaagagatctggaaa	309	57	24813814
C6orf62	2294686	G	A	tcatgtgaccgttgagcatt	tgcttacctgcaagtgctgt	385	50	24817181
C6orf62	1065364	G	A	ctgggaaaggatttgaagca	tgccgttacgtcaaaaacag	330	53	24818379
C6orf62	6913673	A	C	agtgggaaaggcagggtatt	catcacaagtcgtggagcat	264	60	24821123
C6orf62	3813687	G	A	tttttcccccactgaacctt	agggagacttccctgtggat	304	56	24827957
C6orf62	1923187	A	T	gggacagaggcaagtagcac	cccgatgtcgtcataaaagc	370	59	24829646

Table 3 PCR primers and PCR conditions for DNA pooling assays. All primers are described 5' → 3'.

Gene	rs Number	Direction of Extension Primer	Extension Primer
ID4	3193769	Reverse	tagacacttacatgacatcaaagctg
ID4	1980461	Reverse	acacttacatgacatcaaagctg
PRL	1205961	Reverse	gctttagcagggatttctagaaaa
PRL	6239	Forward	gttataataagcagaaaggcgaga
PRL	1205957	Reverse	cacacacgcgcaccacacac
PRL	849886	Forward	gatttaaaatgcttactttgtagatgtc
PRL	1205954	Forward	ggaaggaagggaggagggaagga
PRL	2244502	Forward	cgccttccactgtgtaaatgtagat
PRL	6237	Forward	atccccccacaggagtgttgatacaa
PRL	849876	Reverse	ccaacctgatgttacttaatgttaaaaa
PRL	849875	Forward	aggagaatcgcttgaacccaggaggc
VMP	12208318	Reverse	accatctccactaacgcca
VMP	10946676	Forward	gccatcatgagaaagaagtagg
DCDC2	1277155	Reverse	agctgggactacaggcgcct
DCDC2	1832709	Reverse	ctgggaccgggtgctgtggctcacac
DCDC2	3789219	Forward	gtttggttcacaaaggtaatgacat
DCDC2	1419228	Forward	agccattgacccatccatcta
DCDC2	2996452	Forward	cagaaactatgacactttatataacttttct
DCDC2	1277192	Forward	tggttttaggtcttacatttaagtcttt
DCDC2	1277194	Reverse	tggttttaggtcttacatttaagtcttt
DCDC2	793861	Forward	gaaggaatgaaggaaagaaggaagga
DCDC2	793862	Forward	gccttttttgatcactgtttattctta
DCDC2	870601	Forward	gtcaatgatggtgctcaa
DCDC2	807698	Forward	caaaacaaataaaaaaaattagtggggca
DCDC2	807726	Forward	tcatcctctgctcccactgata

Table 4 Continued ...

Gene	rs Number	Direction of Extension Primer	Extension Primer
DCDC2	3789224	Reverse	ctgtactagtcactccataagc
DCDC2	3789227	Reverse	gatgcaattcctcccctcct
DCDC2	2296539	Forward	ctggaaaaatcattattcacatttggaa
DCDC2	2274305	Reverse	cgttgacttgtcaaaaagtaactcac
DCDC2	6907864	Forward	cttgtttcctctgcaggtaagtaaaa
DCDC2	807709	Forward	ccacttttttctttatccagtctacca
DCDC2	807704	Reverse	atgagactgtggatattaatatgtgtaa
DCDC2	807703	Reverse	taactgtttcagaaggtttgcc
DCDC2	3857541	Reverse	actgctcatgatctgtggaaaaaaaaaa
DCDC2	707862	Reverse	ataaaataaaatgtttagcaggtaatgcac
DCDC2	807685	Reverse	actccatcacagttcagc
DCDC2	793704	Reverse	gaattgttcttcacatgtttag
DCDC2	793722	Reverse	agacaacaggtgtgcaatagatttat
DCDC2	793720	Forward	cgagttaagcatagttcgtgaaag
DCDC2	1277350	Forward	gcagcgatttgatcagttccatttg
DCDC2	1277349	Forward	tacccctcaatctcctgctt
DCDC2	2792666	Forward	ggtttgaactctctctctctctttc
DCDC2	793663	Reverse	ttaaaacaatggttaccctatac
MRS2L	9393553	Forward	gctatcttacattttattaaatatggaata
KIAA0319	2817241	Reverse	tgattcttgtgcctcagcct
KIAA0319	807526	Reverse	gaggttgcagtgagcagaga
KIAA0319	2817243	Reverse	agtgaacatttaccgcttta
KIAA0319	2817245	Reverse	cagctggaagtgaatcacgg
KIAA0319	807532	Reverse	gttctggaggctgggaaggc
KIAA0319	2076313	Reverse	atgactgcctggcaccttta

Table 4 Continued ...

Gene	rs Number	Direction of Extension Primer	Extension Primer
KIAA0319	807536	Reverse	tcagatgccagttgcatacc
KIAA0319	4083411	Forward	agctggttcacttggcttca
KIAA0319	2760167	Forward	ctctactaaaaataccaaaaaaaaagaaaagaa
KIAA0319	807542	Reverse	agcaactcaagtgtccacca
KIAA0319	807544	Reverse	ggaggccaaggcaggcagat
KIAA0319	2744549	Forward	gtaagagctggaattttgta
KIAA0319	2760161	Forward	aggagaatggcttgaacctggga
KIAA0319	2817195	Reverse	gctgcctggtctgaagatgagtaa
KIAA0319	807521	Reverse	gaacttggactgttcaactcct
KIAA0319	3846835	Reverse	taatggtctttgccttagaacttttct
KIAA0319	2744556	Reverse	ctcctgacttcatgatccac
KIAA0319	807525	Forward	aaacaccaagctggttagca
KIAA0319	2817199	Reverse	ctaatctaaacctgattatcagaatttttttt
KIAA0319	2760158	Reverse	ttactgctgctaacccatga
KIAA0319	2760157	Reverse	ctacttacattttgactgtttctttaaattaa
KIAA0319	807507	Forward	tccctatcaaatggccaggc
KIAA0319	807509	Forward	tctgcctgcctacttgcatg
KIAA0319	2817201	Forward	tctttctctctatggaagctca
KIAA0319	4504469	Reverse	tcccaacacctcccactagc
KIAA0319	5026394	Reverse	ttgatcaatacagcttaatttcattatt
KIAA0319	4576240	Forward	ccaatctttgcctagaaagg
KIAA0319	4352670	Forward	cttgtttttgagtttaatcgagaa
KIAA0319	4712831	Reverse	tggatgaacttcacagaactatcaaa
KIAA0319	4236032	Reverse	gatccatcagtacttgttgaatacc
KIAA0319	6911855	Forward	ttctaagaagccaggccaggtg

Table 4 Continued ...

Gene	rs Number	Direction of Extension Primer	Extension Primer
KIAA0319	4712833	Forward	ggcatagtaataggcaatataaaggaa
KIAA0319	6939068	Forward	cgcctcggctgagaatttttt
KIAA0319	7763790	Forward	ttgttttaccagccagcccctc
KIAA0319	2745335	Forward	agcacttaatggtaggatacaatgt
KIAA0319	2745334	Forward	agcacttaatggtaggatacaatgt
KIAA0319	2817206	Forward	caaaactgcttcacaaattttt
KIAA0319	7751357	Forward	tctgccaaatagaacaagaataaaa
KIAA0319	6917660	Reverse	aggttatgagagggccctttatcc
KIAA0319	6456622	Reverse	tcccctccttcctttctagc
KIAA0319	7766230	Forward	gtgaagggttagcaaagttctcagaa
KIAA0319	2206525	Reverse	tcaatatcatctcatccatttttaaaa
KIAA0319	7755563	Forward	ataaaaagaattagaaagaggcagcc
KIAA0319	7755579	Reverse	tgcccacctcagcctcccaaagtg
KIAA0319	6935076	Reverse	cgcagacatgaggagaatga
KIAA0319	4363021	Forward	ctgacacctccagttttcctcat
KIAA0319	2038137	Forward	cctctttcctatttctcggccaggcgc
KIAA0319	3756821	Forward	tgtccccagcttgtgtcaccc
KIAA0319	1555089	Forward	cctttcacttcttttcagcattca
TTRAP	3212236	Forward	gccagacactgcaagaggtag
TTRAP	3087943	Reverse	cctgtcagagttttcaacggtgctta
TTRAP	2294691	Reverse	caatggaaagattggaccgtg
TTRAP	3181238	Forward	gtcagcaaactcaaatattcttaagaaaa
TTRAP	3212234	Forward	ctactcagctcccttcccct
TTRAP	3212232	Reverse	tgagaaaaataagcaagacacagaatac
TTRAP	2143340	Forward	ttacagacaaattttaaaagagcccta

Table 4 Continued ...

Gene	rs Number	Direction of Extension Primer	Extension Primer
TTRAP	2056999	Reverse	aaaaaaagaacccatactgatataaatacc
TTRAP	3756819	Reverse	gagagaaagcccttaactgccc
TTRAP	1061925	Reverse	ctgatgtcctaagtctgggtaag
TTRAP	3756815	Forward	cccgttggaaaaaagtaaaagtttgatag
TTRAP	2223589	Forward	gcttggttttatgcgttttaggga
THEM2	3181227	Reverse	ccagccttggaactgtctcca
THEM2	9393576	Forward	agaaggctgtgaactcccatgtc
THEM2	7765052	Forward	ctcttataattttttttttgaatagagtctc
THEM2	7451561	Forward	tagtgtgtataatggcctgactattt
THEM2	2143338	Reverse	tcctaggccaactaagaatcccta
THEM2	1555086	Reverse	agtctcaaactcctagcctcaag
THEM2	1205120	Forward	gacatggtctctttccttttactg
THEM2	926529	Forward	tccaaaggaaggcaaagtgtaaagtc
THEM2	1885209	Forward	tgtgcaccttaaaaacttagaaatttca
THEM2	1885211	Reverse	gctacaagaatgctcctttgaaaatata
THEM2	3777664	Forward	gcccttctccctatctatcttt
THEM2	2092404	Forward	ccttatacttactataaatgggccaa
THEM2	3777663	Forward	gagtgaatcatgaagtgttaactgac
THEM2	1056319	Forward	actagccccttttcttgtaccgag
THEM2	1053598	Forward	acacaacagctttcatattactctgg
C6orf62	3756814	Forward	atgagatgcgtttttcatttaagatttc
C6orf62	2294686	Forward	acaagttatactacctatctgcatct
C6orf62	1065364	Reverse	tgtacttgtcctcttcagcagtat
C6orf62	6913673	Forward	tccttaactaaaaaacaacaaaatccta
C6orf62	3813687	Forward	ttatcttcttgtgagcattgccgag

Table 4 Continued ...

Gene	rs Number	Direction of Extension Primer	Extension Primer
C6orf62	1923187	Reverse	atctccaacacacactttccttttcc

Table 4 Extension primers for DNA pooling assays using SNaPshot™ technology. Assay protocols are shown Chapter three. All primers are described 5' → 3'.

Gene	rs Number	Allele 1	Allele 2	Allele 1 Primer	Allele 2 Primer
PRL	2244502	A	T	gaaggtcggagtcaacggattcgccttccactgtgtaaatgtagata	gaaggtgaccaagttcatgctgccttccactgtgtaaatgtagatt
DCDC2	793722	T	C	ccttgcatgaaattggtgaa (RFLP Forward PCR Primer)	cctgggtgacagaaggagac (RFLP Reverse PCR Primer)
MRS2L	2793422	A	G	gaaggtcggagtcaacggattctgggattgaccatgcagaagaa	gaaggtgaccaagttcatgctttgggattgaccatgcagaagag
KIAA0319	4504469	G	A	gaaggtcggagtcaacggattcaacacctcccactagcg	gaaggtgaccaagttcatgcttccaacacctcccactagca
KIAA0319	6911855	T	C	tggacagccaacctttgaat (RFLP Forward PCR Primer)	cttgagaggctgaggtggg (RFLP Reverse PCR Primer)
KIAA0319	6939068	A	T	See Pooling Data for SNaPshot Conditions	
KIAA0319	2179515	C	T	gaaggtcggagtcaacggattagcctccttccaaatactgcc	gaaggtgaccaagttcatgctgagcctccttccaaatactgct
KIAA0319	6935076	G	A	gaaggtcggagtcaacggattcactgtctatgtggagtctgc	gaaggtgaccaagttcatgcttccactgtctatgtggagtctgt
KIAA0319	2038137	G	T	gaaggtgaccaagttcatgctatttctcggccaggcgct	gaaggtcggagtcaacggatttatttctcggccaggcgcg
TTRAP	2143340	A	G	gaaggtcggagtcaacggattccctgtaaggacagtgtcacttt	gaaggtgaccaagttcatgcttcctgtaaggacagtgtcacttct
THEM2	926529	G	A	gaaggtcggagtcaacggattggaaggcaaagtgtaaagtcg	gaaggtgaccaagttcatgctaaggaaggcaaagtgtaaagtca
THEM2	3777664	A	G	gaaggtgaccaagttcatgcttgcagcccttctccctatctatcttta	gaaggtcggagtcaacggattcagcccttctccctatctatctttg
THEM2	1053598	C	T	gaaggtgaccaagttcatgctacacaacagctttcatattactctggc	gaaggtcggagtcaacggattgattacacaacagctttcatattactctggt

Table 5 Amplifluor® primers for individual genotyping reactions. Assay protocols are shown Chapter three. All primers are described 5' → 3'. For SNPs genotyped using SNaPshot™ technology, the same assay protocols and primers were used as for DNA pooling. PCR primers for rs793722 and rs6911855 are shown. These SNPs were genotyped using RFLPs. Characteristics of each RFLP are shown in Figures 5 and 6.

Gene	rs Number	Reverse Primer	Amplifluor® Method
PRL	2244502	ggaggtaacccacttctcaaa	B
DCDC2	793722	RFLP – Restriction Enzyme = TaqI	-
MRS2L	2793422	cggtagtagttttccaacagcaa	B
KIAA0319	4504469	agcagtggtaggagatatgggta	A
KIAA0319	6911855	RFLP – Restriction Enzyme = XcmI	-
KIAA0319	6939068	See Pooling Data for SNaPshot Conditions	-
KIAA0319	2179515	caattactcagttcattttgccctagaa	C
KIAA0319	6935076	gaaaccgaagcccagagaaaa	C
KIAA0319	2038137	agtgtcgccagcagtga	A
TTRAP	2143340	tgtagccctcattttacagacaaatt	A
THEM2	926529	tccagtgtgttcctcctctt	A
THEM2	3777664	gagtctctcctttcctcactttt	A
THEM2	1053598	ggcattttggcagccattgttt	B

Table 6 Reverse primers for Amplifluor® assays used for individual genotyping. All primers are described 5' → 3'. The Amplifluor method refers to cycling parameters noted in Table 3.18 in Chapter 3.

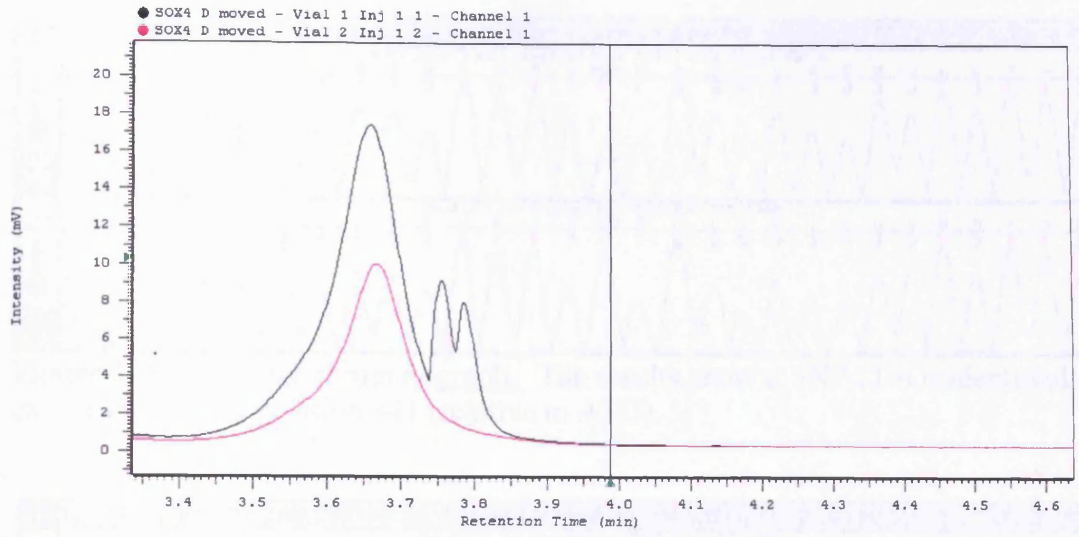


Figure 1 A chromatograph showing the two different traces obtained for two individuals using the PCR fragment which resulted in the identification of two SNPs in *SOX4*. The pink line represents the chromatogram obtained for all samples, the black line shows the chromatogram from an individual containing the two SNPs.

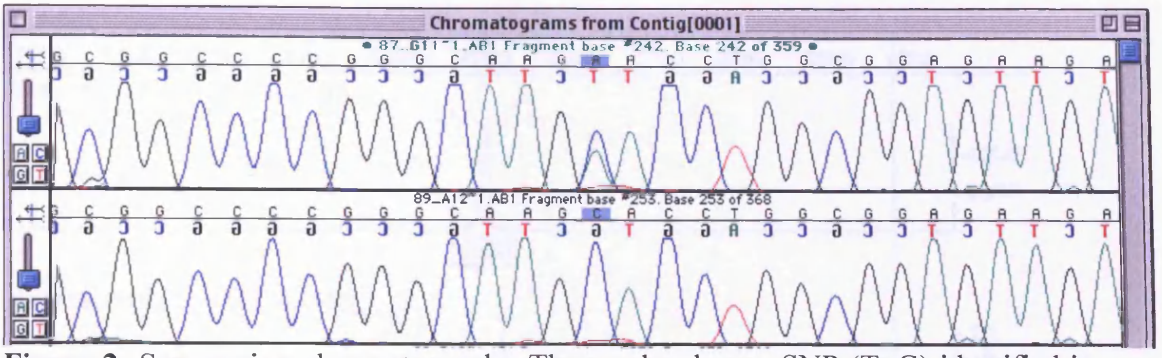


Figure 2 Sequencing chromatograph. The results show a SNP (T>G) identified in exon 1 of *SOX4* at position 841 (relative to ATG).

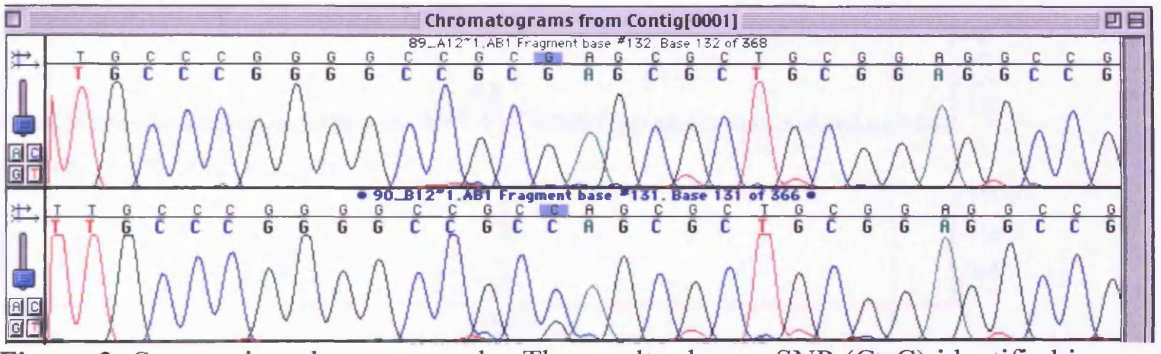


Figure 3 Sequencing chromatograph. The results show a SNP (G>C) identified in the 3' UTR of *SOX4* at position 858 (relative to ATG).

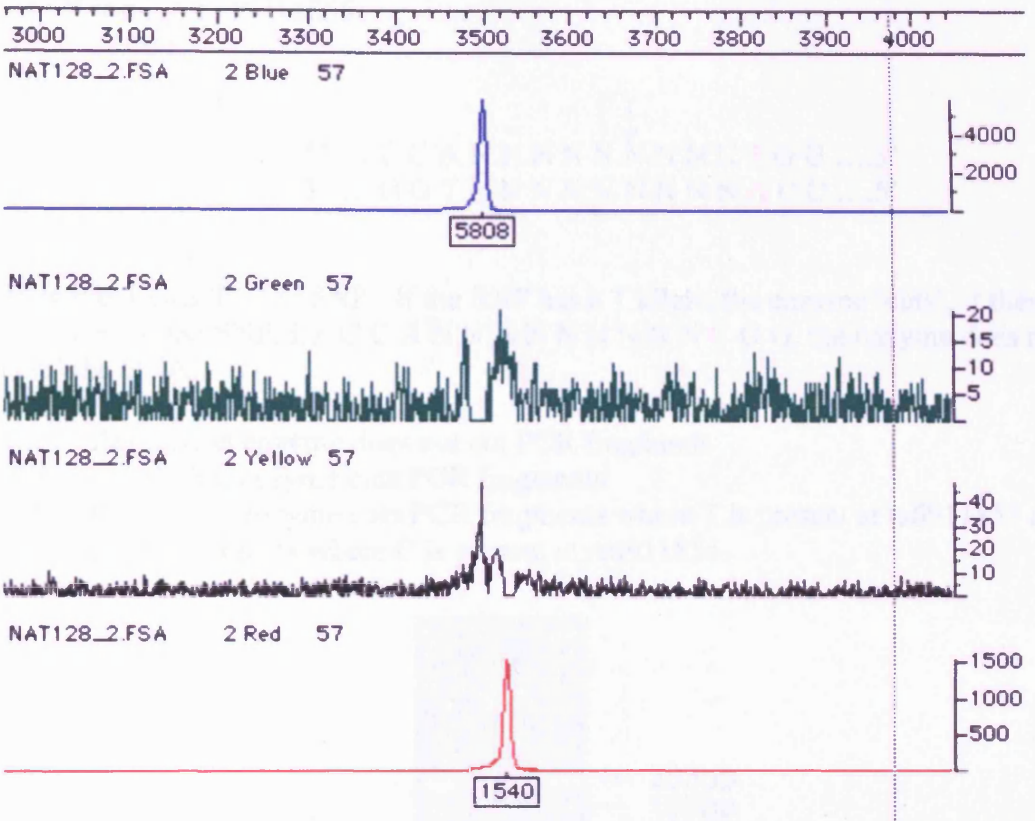


Figure 4 Chromatograph of a G (blue) → T (red) SNP genotypes using SNaPshot chemistry. Peak heights are indicated in the boxes beneath the peaks representing the base extended at the end of the extension primer in a SNaPshot reaction. Both pooled and individual genotypes will result in chromatographs like this.

RFLP result for rs6911855

XcmI:

5'.... C C A N N N N N N N N N N T G G ....3'
3'.... G G T N N N N N N N N N N A C C ....5'

Pink highlights T → C SNP. If the SNP has a T allele, the enzyme 'cuts', if there is a C allele at the SNP, i.e. C C A N N N N N N N N N N C G G, the enzyme does not 'cut' the DNA.

C/C = Restriction enzyme does not cut PCR fragments
T/T = Restriction enzyme cuts PCR fragments
T/C = Restriction enzyme cuts PCR fragments where T is present at rs6911855 and does not cut fragments where C is present at rs6911855

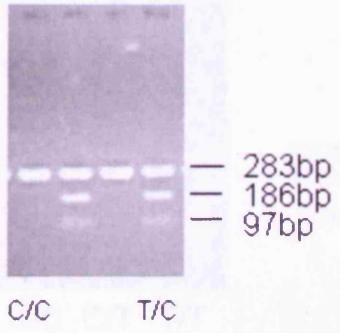


Figure 5 Restriction fragment length polymorphism analysis of rs6911855.

RFLP result for rs793722

TaqI :

5'....T C G A....3'

3'....A G C T ....5'

Pink highlights C→T SNP. If the SNP has a C allele, the enzyme 'cuts', if there is a T allele at the SNP, i.e. T T G A, the enzyme does not 'cut' the DNA.

T/T = Restriction enzyme does not cut PCR fragments

C/C = Restriction enzyme cuts PCR fragments

C/T = Restriction enzyme cuts PCR fragments where C is present at rs793722 and does not cut fragments where T is present at rs793722

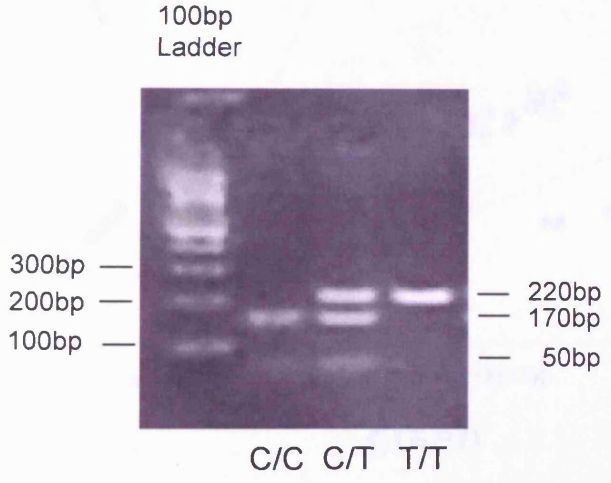


Figure 6 Restriction fragment length polymorphism analysis of rs793722.

Amplifluor results for rs2179515
Allele 1 Primer (C) = Red
Allele 2 Primer (T) = Green

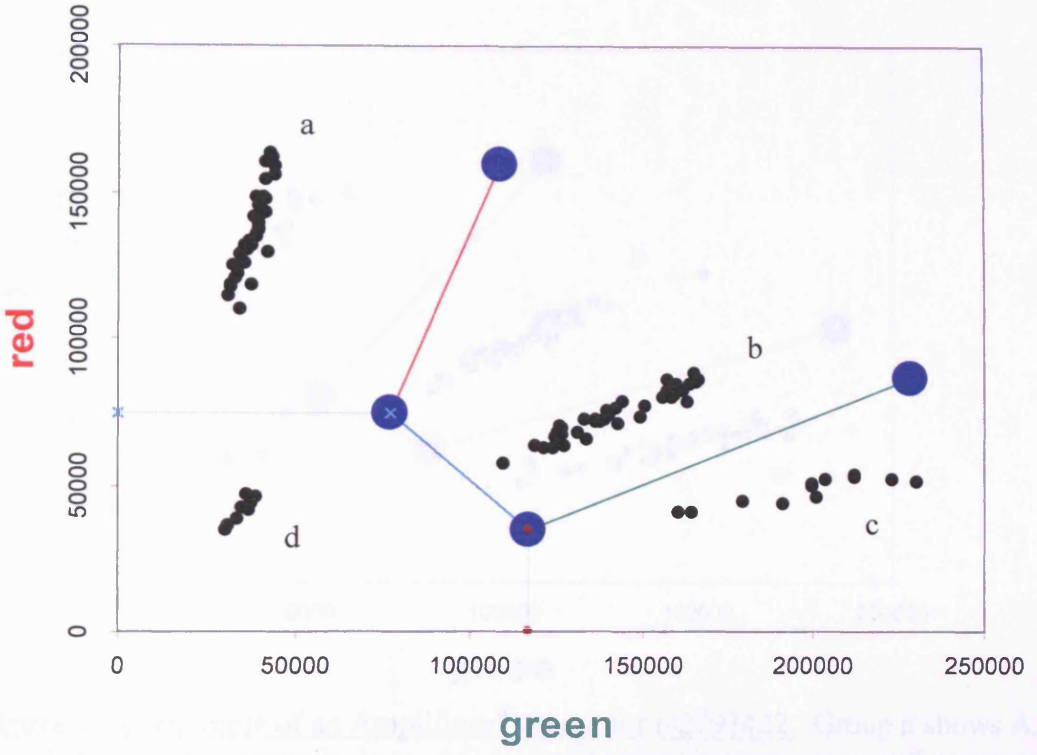


Figure 7 An example of an Amplifluor® output for rs2179515. Group a shows CC (red) homozygotes, group b shows CT (red and green) heterozygotes and group c shows TT (green) homozygotes. Group D represents negative (water) controls.

Amplifluor results for rs2793422
Allele 1 Primer (A) = Red
Allele 2 Primer (G) = Green

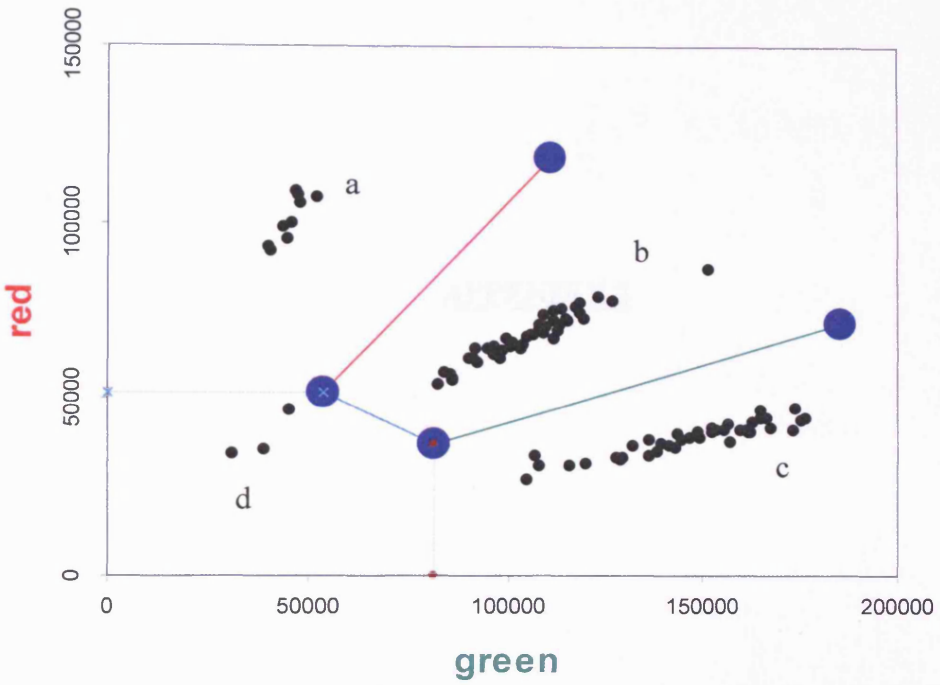
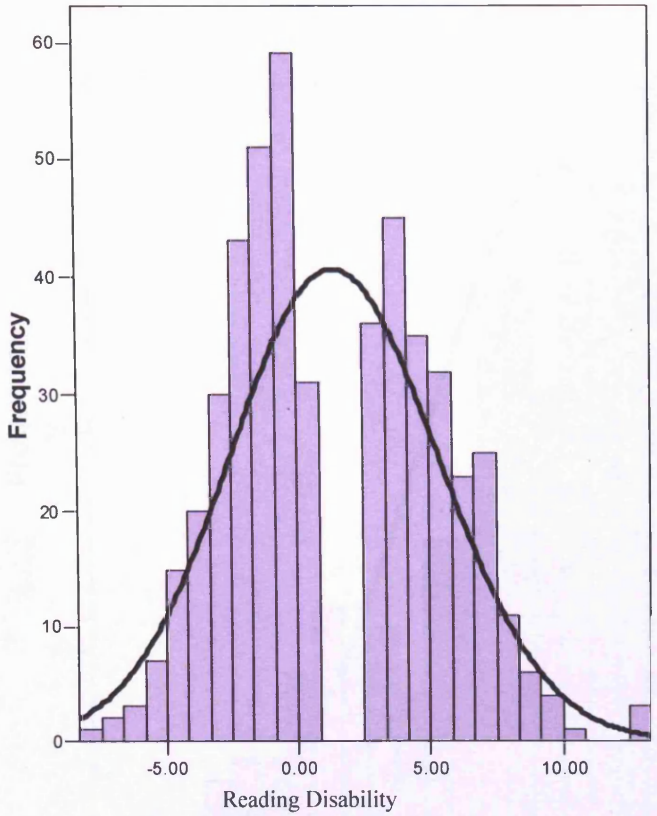


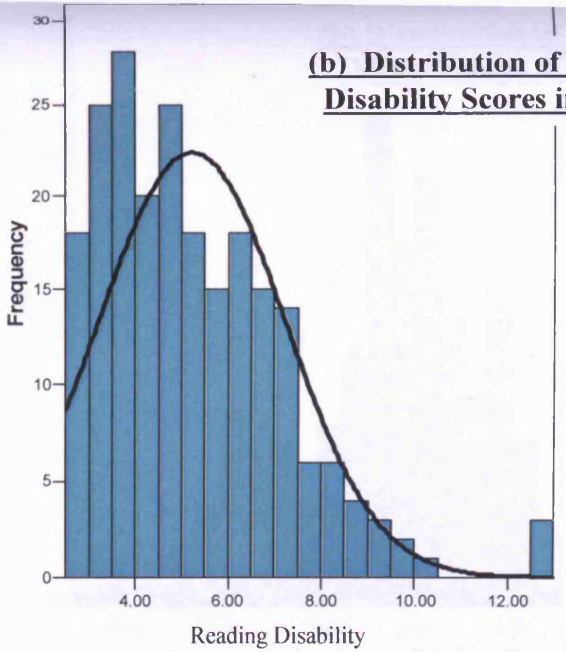
Figure 8 An example of an Amplifluor® output for rs2793422. Group a shows AA (red) homozygotes, group b shows AG (red and green) heterozygotes and group c shows GG (green) homozygotes. Group D represents negative (water) controls.

APPENDIX 2

(a) Distribution of Reading Disability Scores in Cases and Controls



(b) Distribution of Reading Disability Scores in Cases



(c) Distribution of Reading Disability Scores in Controls

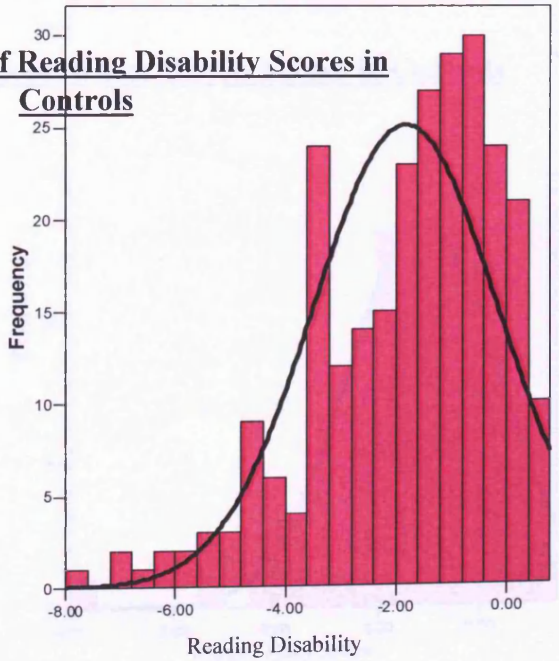


Figure 1 Historgrams showing the distribution of reading diasability scores in (a) cases and controls, (b) cases, (c) controls. All scores based on either accuracy score from NARA or single word reading score. Negative numbers represent reading ahead of their chronological age; positive numbers represent reading lag i.e. reading age below chronological age

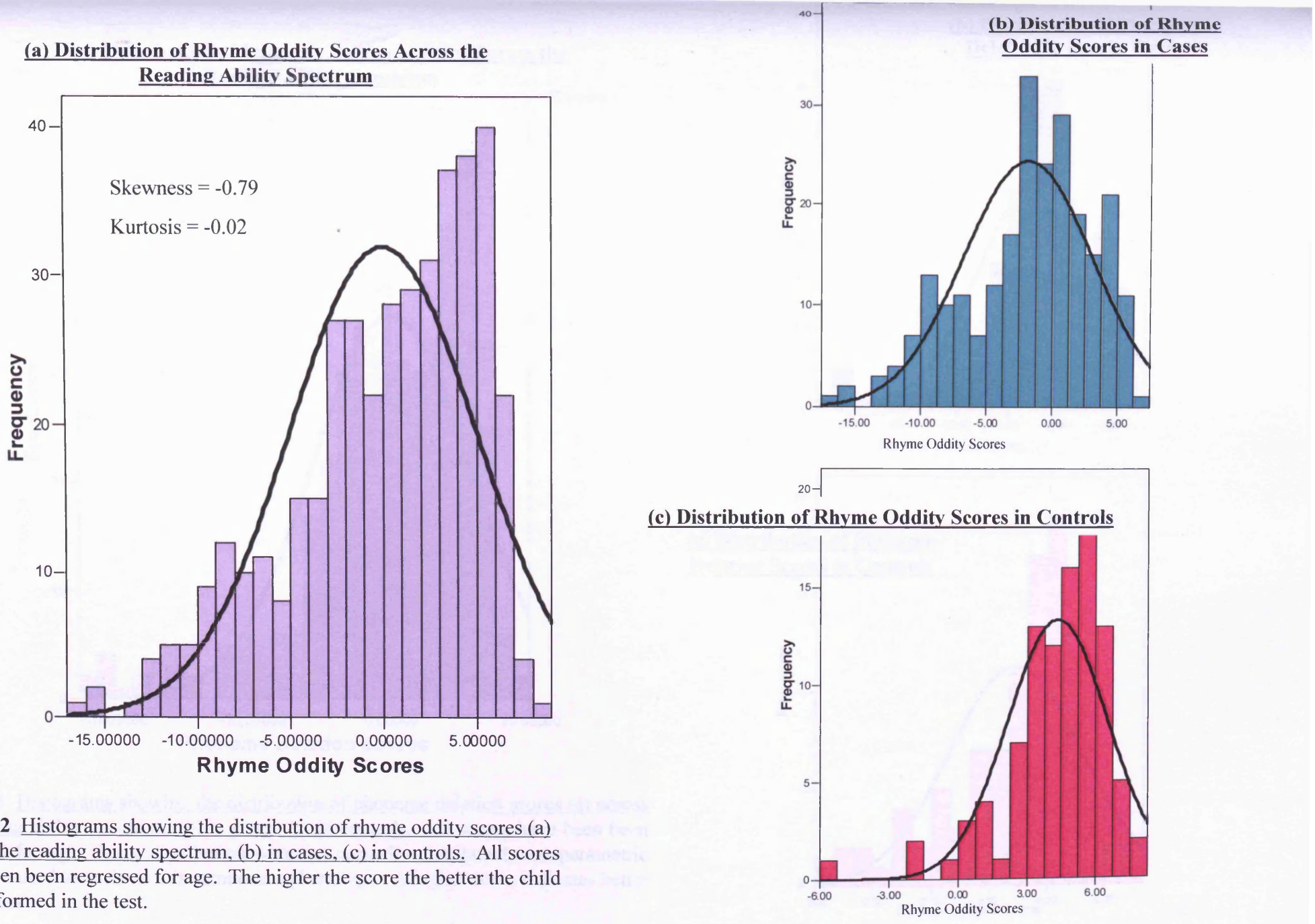


Figure 2 Histograms showing the distribution of rhyme oddity scores (a) across the reading ability spectrum, (b) in cases, (c) in controls. All scores have been been regressed for age. The higher the score the better the child has performed in the test.

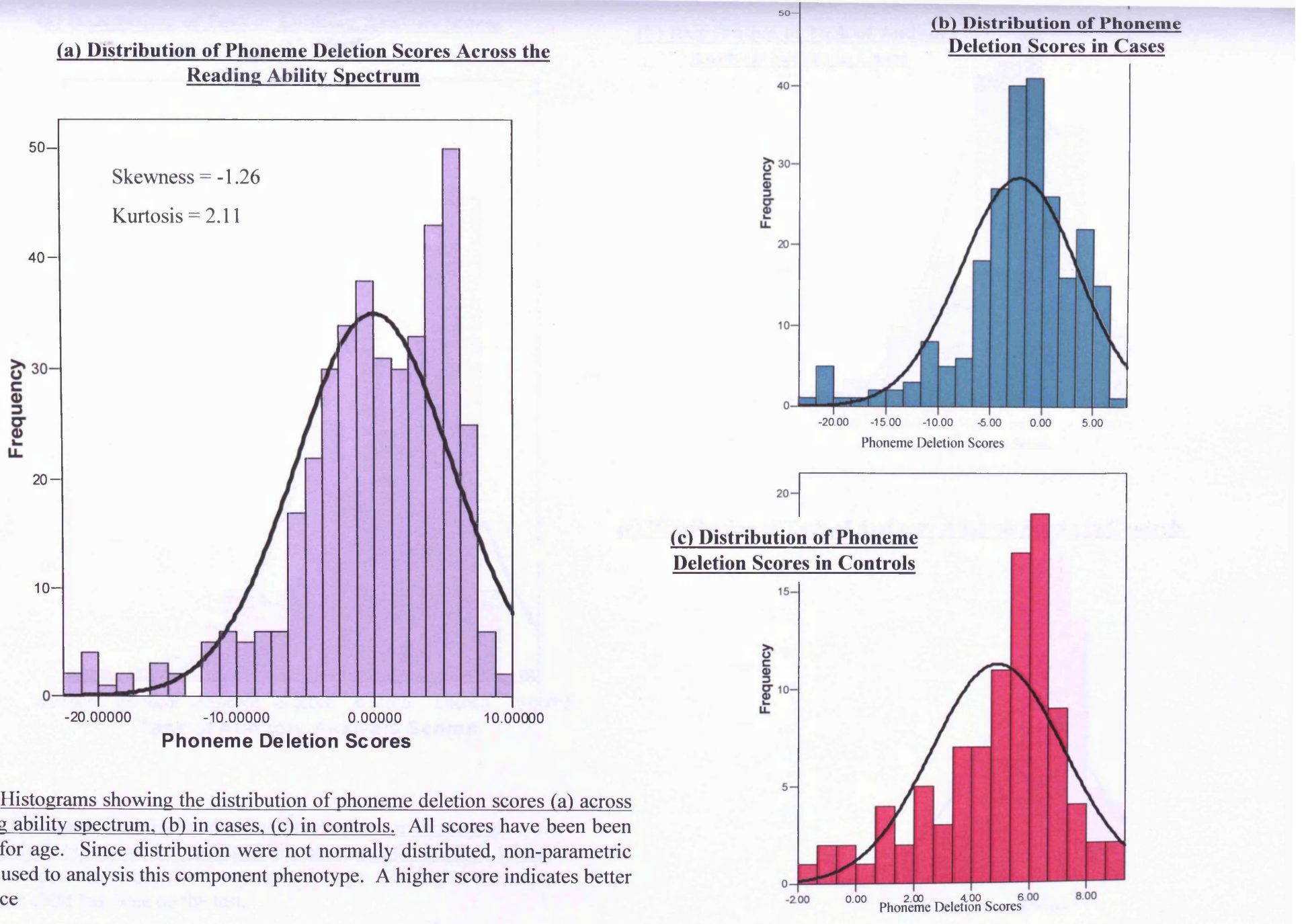
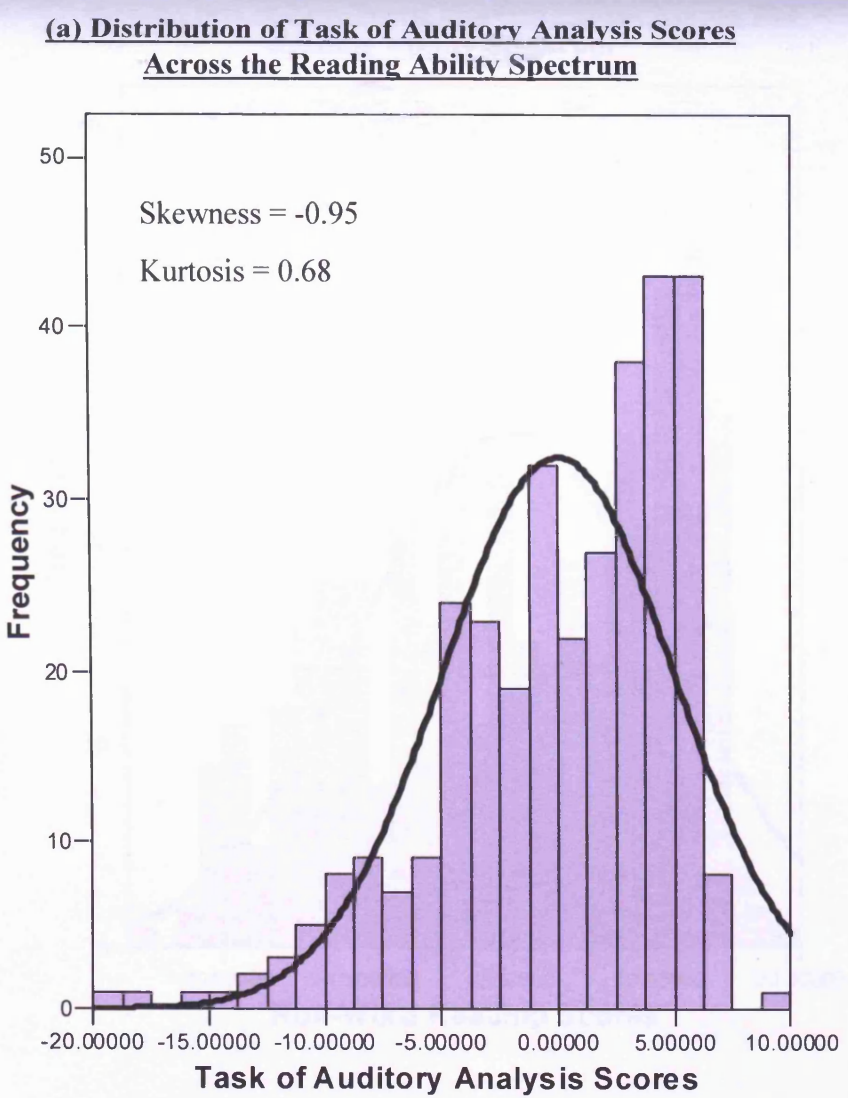
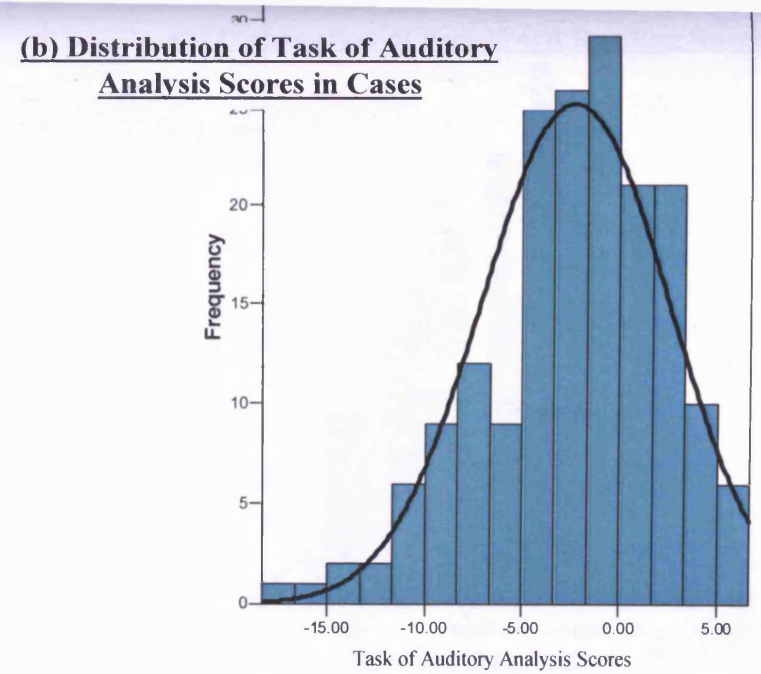


Figure 3 Histograms showing the distribution of phoneme deletion scores (a) across the reading ability spectrum, (b) in cases, (c) in controls. All scores have been been regressed for age. Since distribution were not normally distributed, non-parametric tests were used to analysis this component phenotype. A higher score indicates better performance





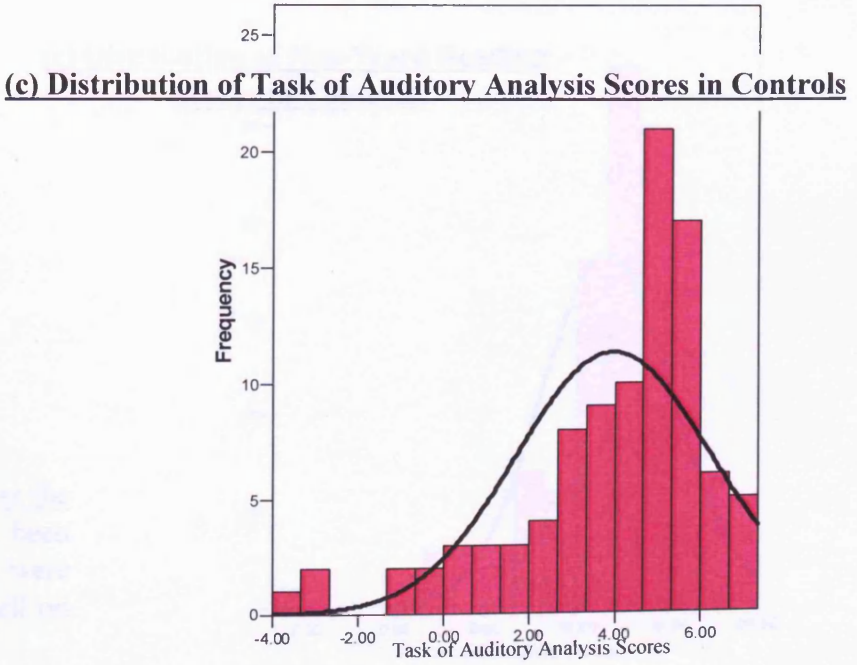
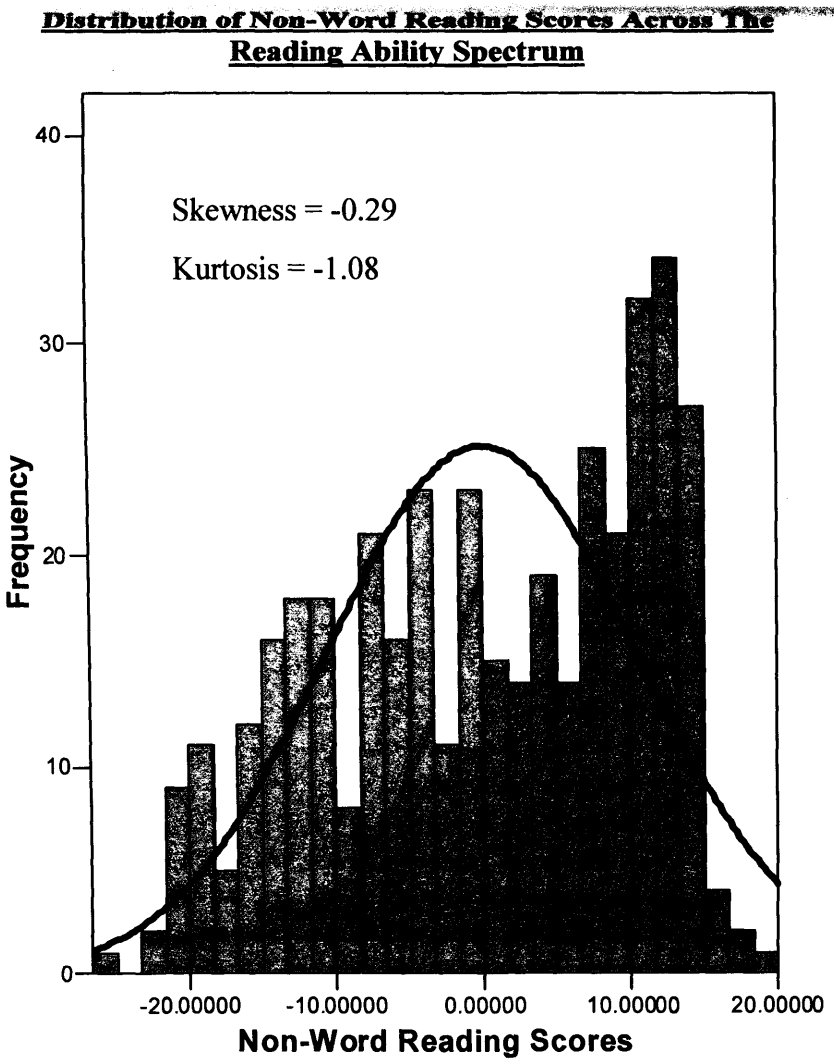
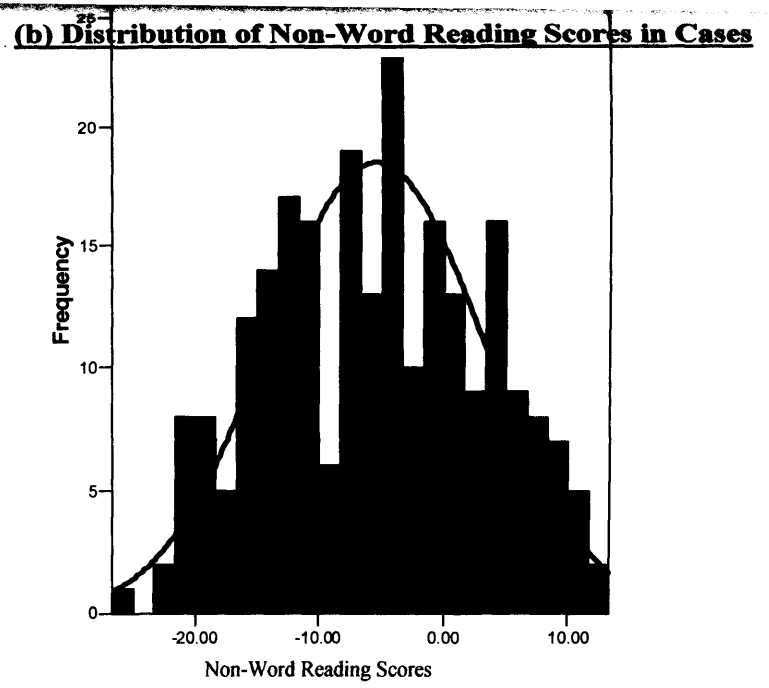


Figure 4 Histograms showing the distribution of scores on the task of auditory analysis (a) across the reading ability spectrum, (b) in cases, (c) in controls. All scores have been been regressed for age. The higher the score the better the child has done on the test.





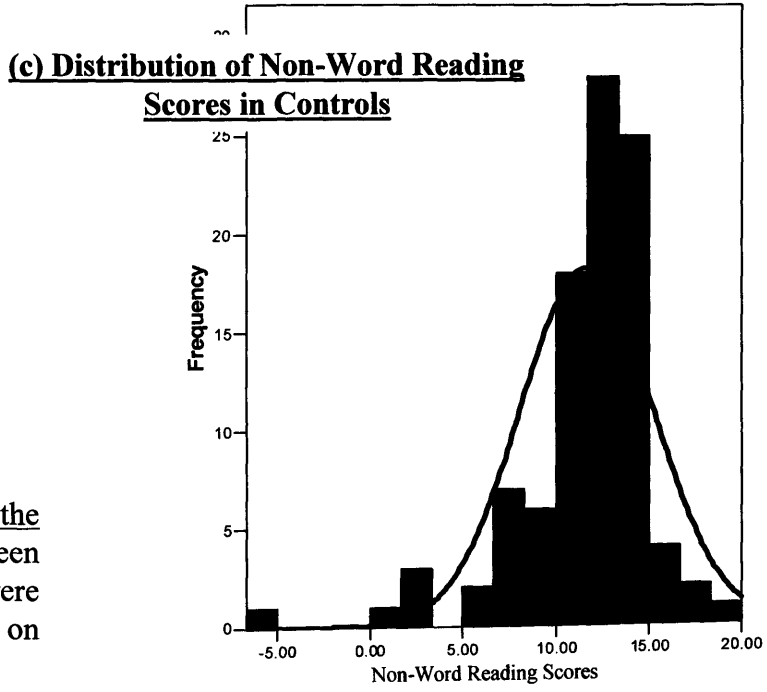
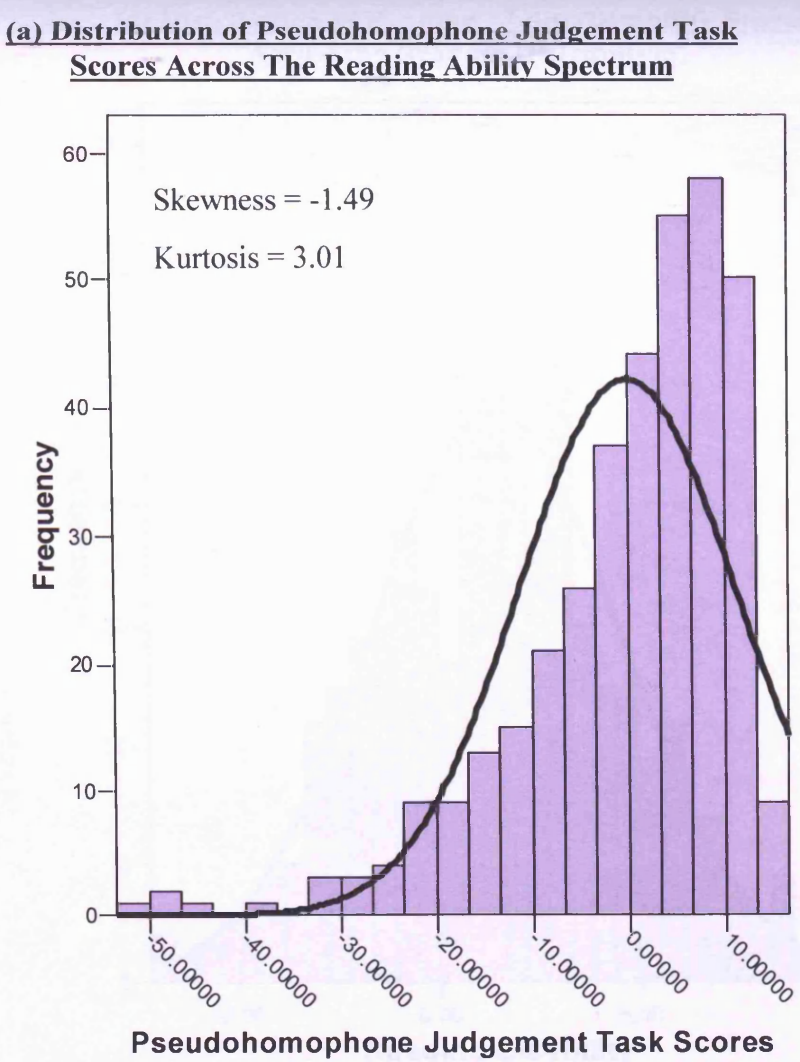
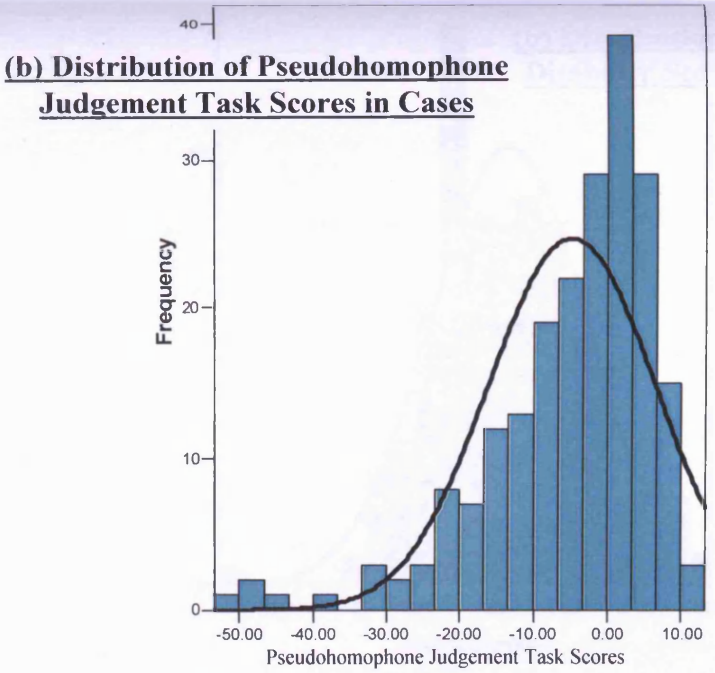


Figure 5 Histograms showing the distribution of non-word reading scores (a) across the reading ability spectrum, (b) in cases, (c) in controls. All scores have been been regressed for age. Since data were not normally distributed non-parametric tests were used in analysis of this component phenotype. A higher score indicates doing well on the test.





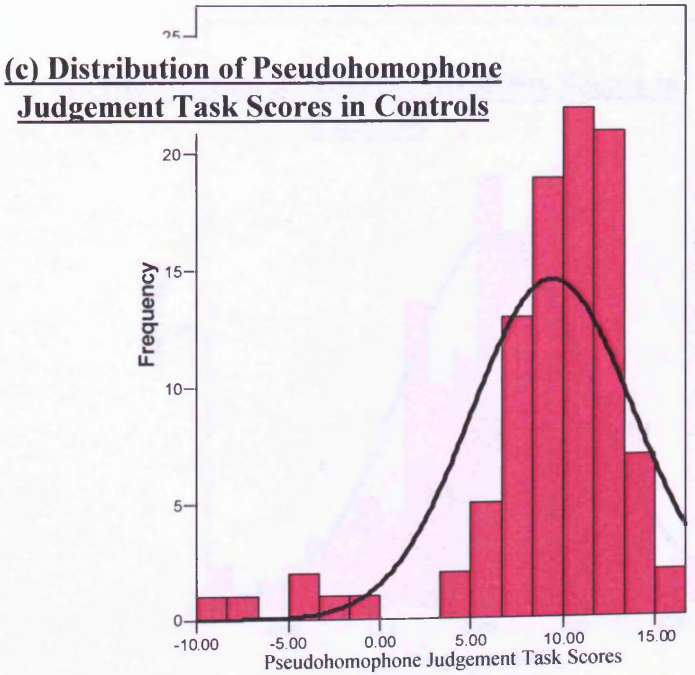
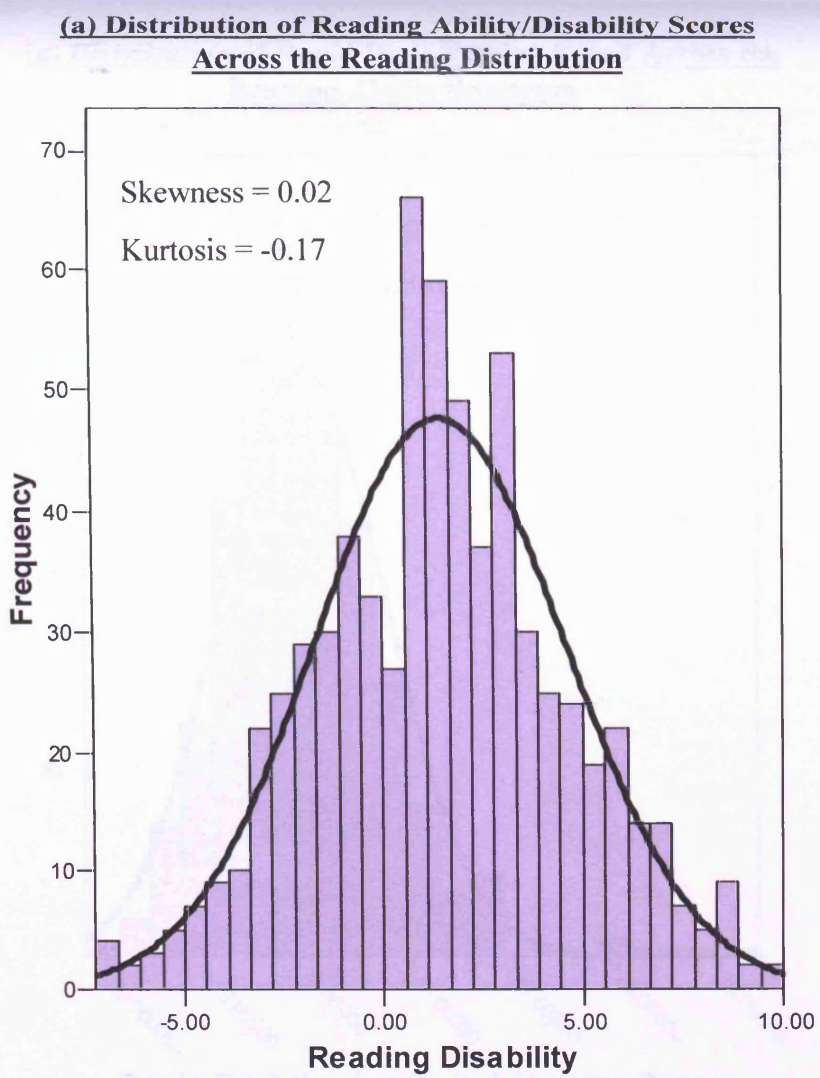
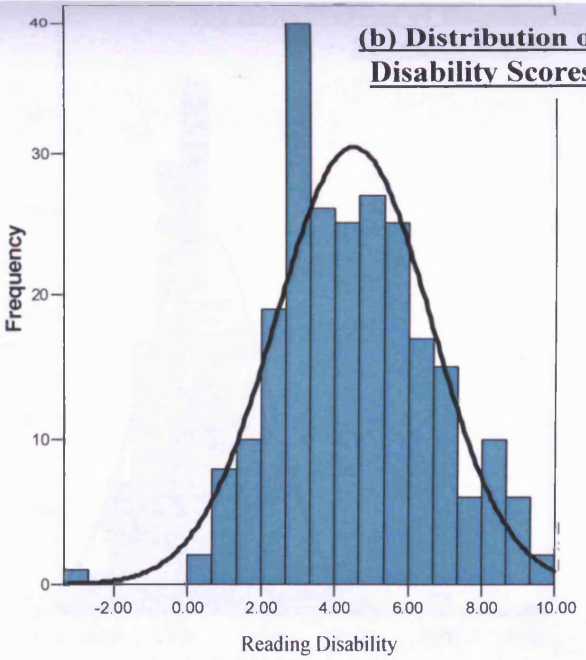


Figure 6 Histograms showing the distribution of pseudohomophone judgement task scores (a) across the reading ability spectrum, (b) in cases, (c) in controls. All scores have been been regressed for age. Since data were not normally distributed, non-parametric tests were used in analysis of this component phenotype. Good performance on this test requires a high score.





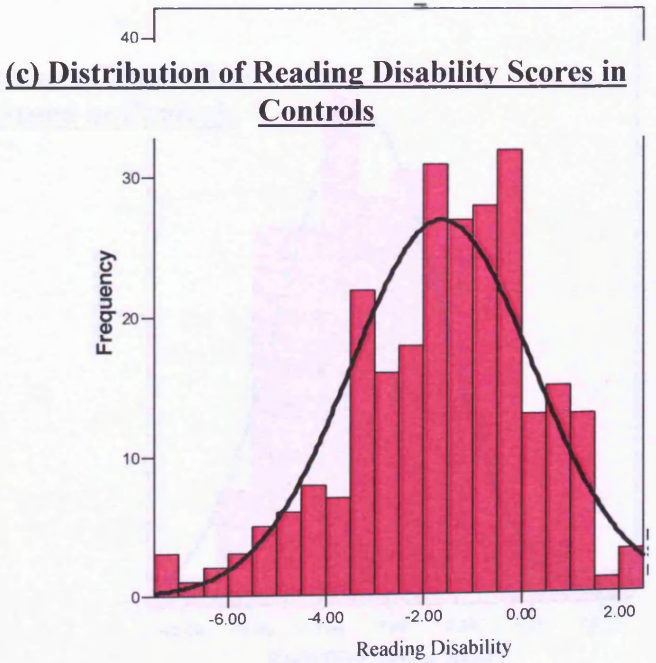


Figure 7 Histograms showing the distribution of single word reading/word recognition scores (a) across the reading distribution, (b) in cases, (c) in controls. All scores based on discrepancy between chonological age and reading age based on the single word reading score. Negative numbers represent reading ahead of their chronological age; positive numbers represent reading lag i.e. reading age below chronological age

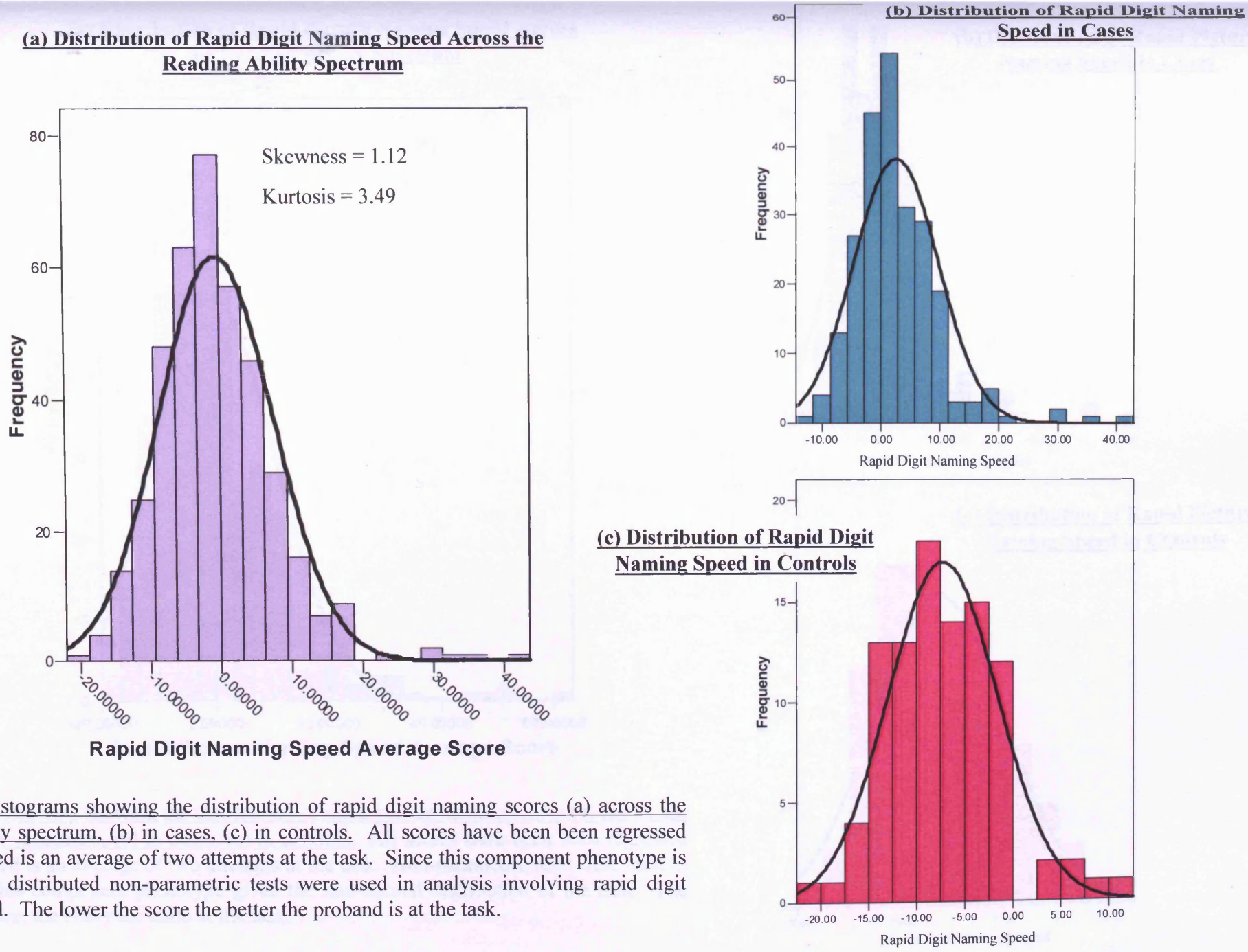


Figure 8 Histograms showing the distribution of rapid digit naming scores (a) across the reading ability spectrum, (b) in cases, (c) in controls. All scores have been been regressed for age. Speed is an average of two attempts at the task. Since this component phenotype is not normally distributed non-parametric tests were used in analysis involving rapid digit naming speed. The lower the score the better the proband is at the task.

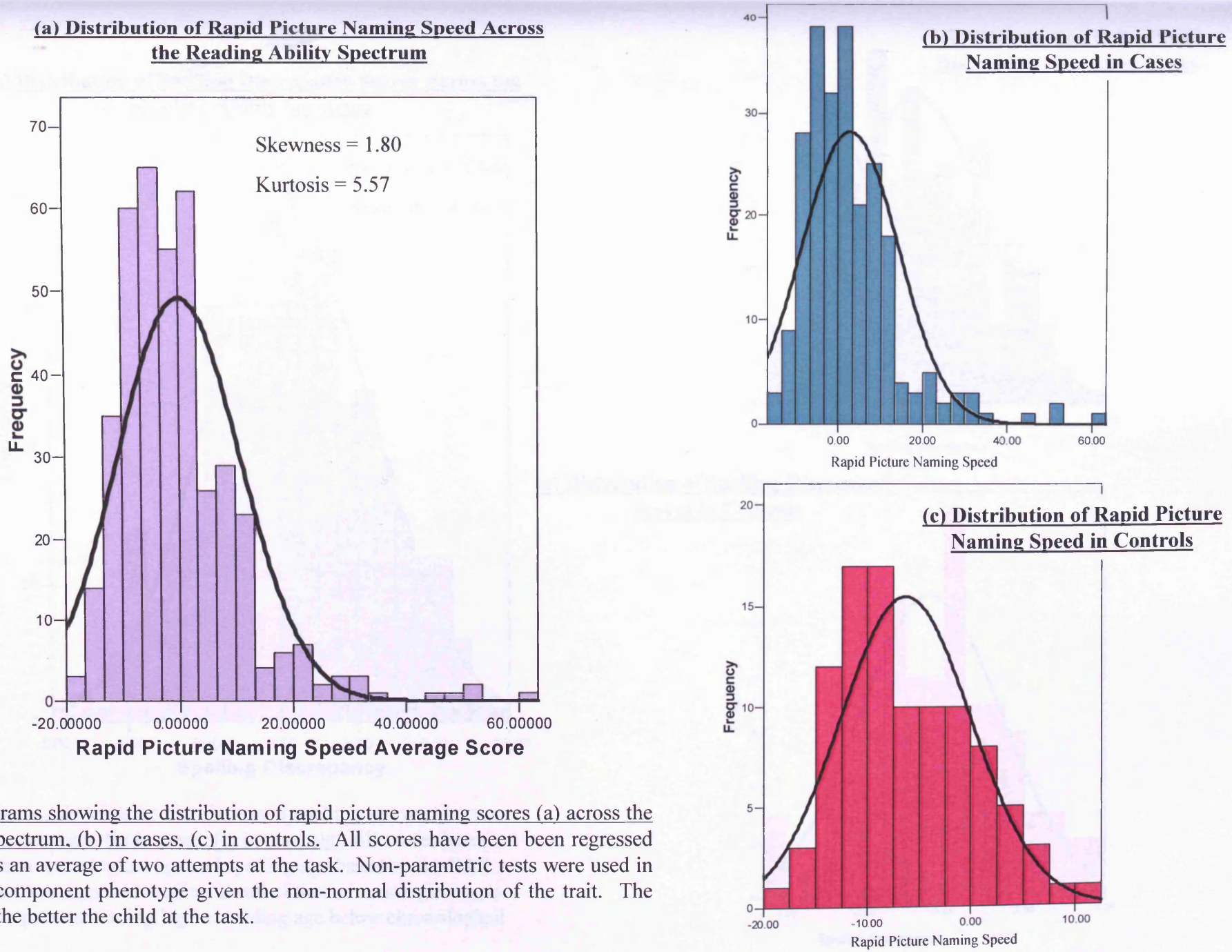
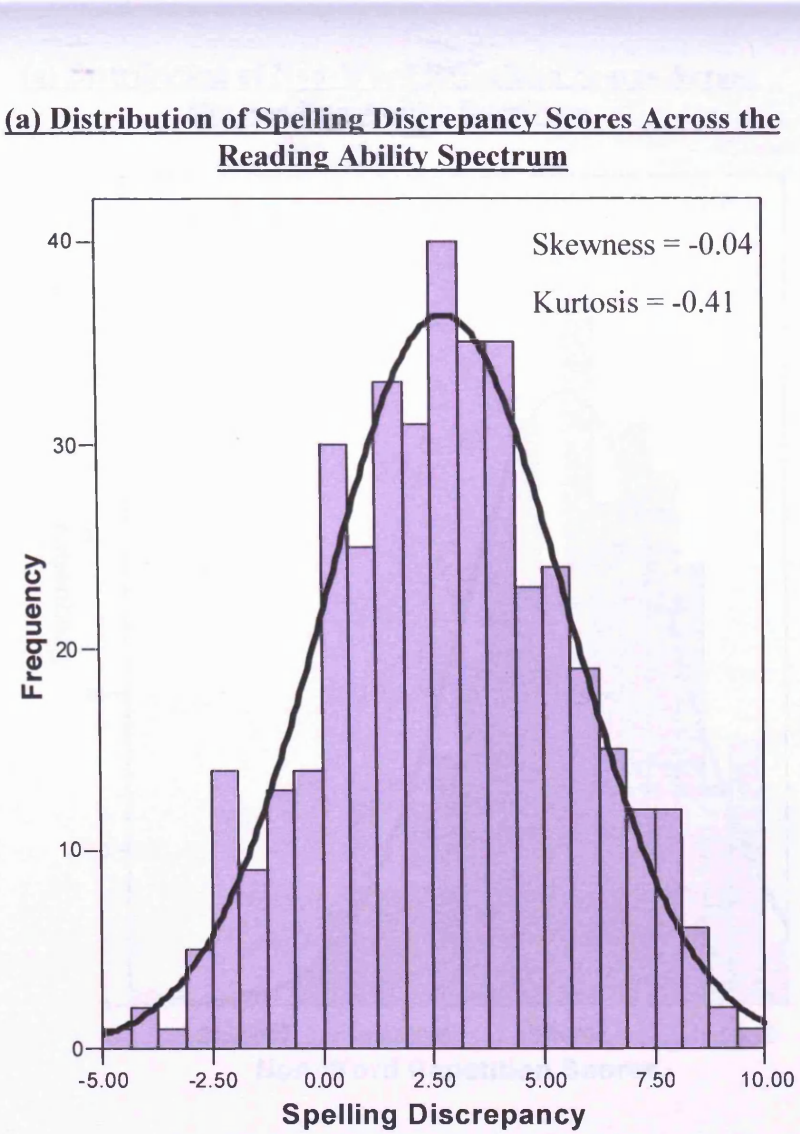
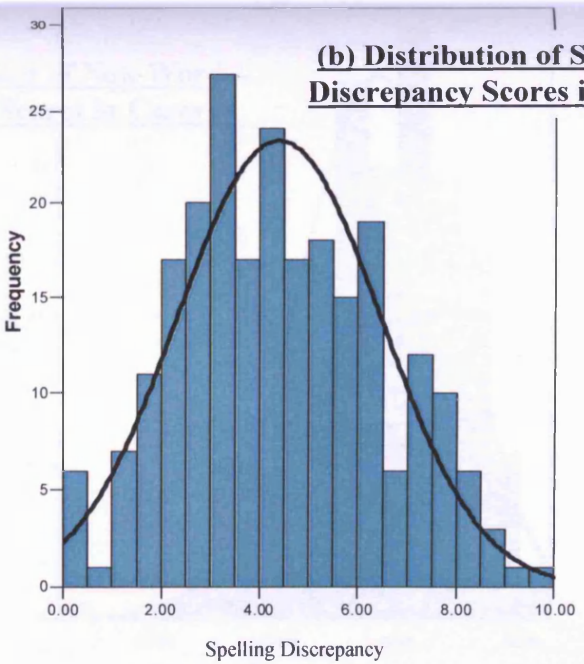


Figure 9 Histograms showing the distribution of rapid picture naming scores (a) across the reading ability spectrum, (b) in cases, (c) in controls. All scores have been been regressed for age. Speed is an average of two attempts at the task. Non-parametric tests were used in analysis of this component phenotype given the non-normal distribution of the trait. The lower the score, the better the child at the task.





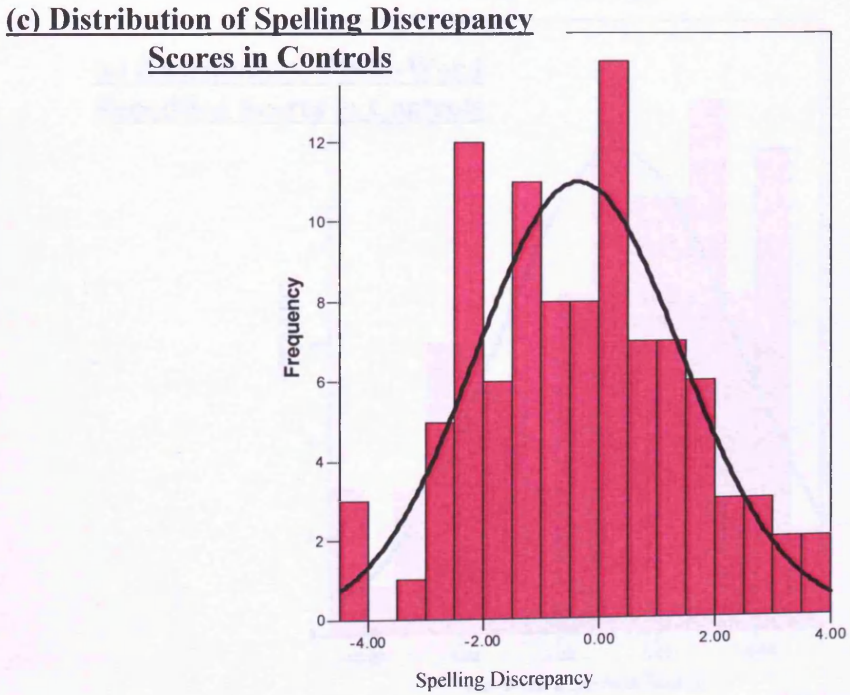
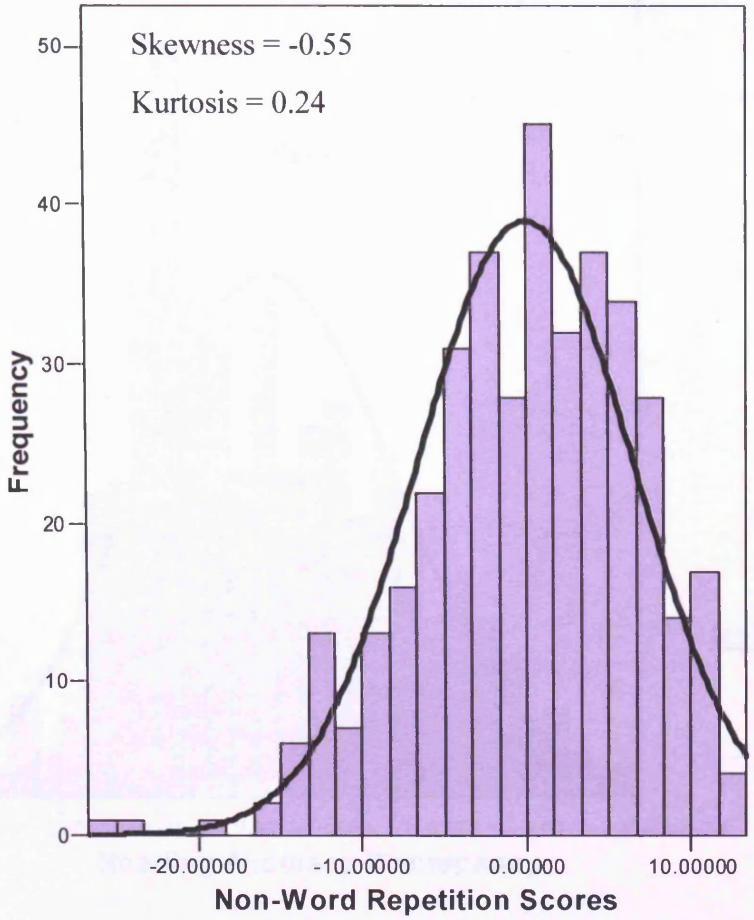
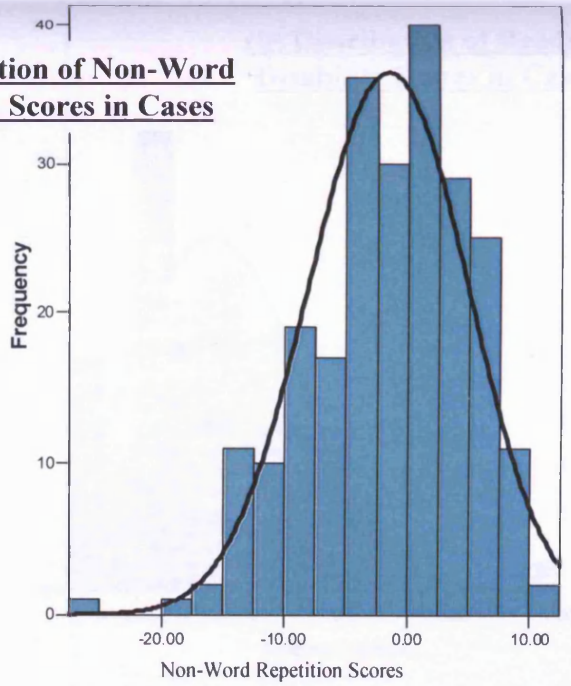


Figure 10 Histograms showing the distribution of spelling scores (a) across the reading ability spectrum, (b) in cases, (c) in controls. All scores based on discrepancy between chonological age and spelling age based on the BAS test. Negative numbers represent reading ahead of their chronological age; positive numbers represent reading lag i.e. reading age below chronological age

(a) Distribution of Non-Word Repetition Scores Across the Reading Ability Spectrum



(b) Distribution of Non-Word Repetition Scores in Cases



(c) Distribution of Non-Word Repetition Scores in Controls

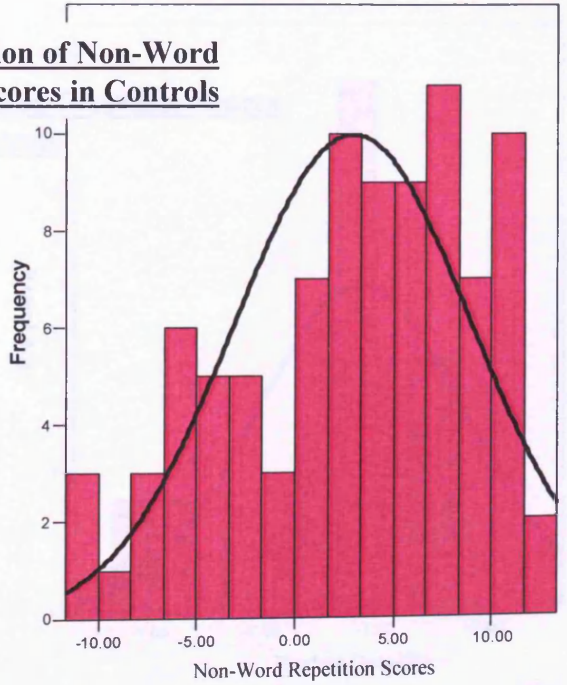
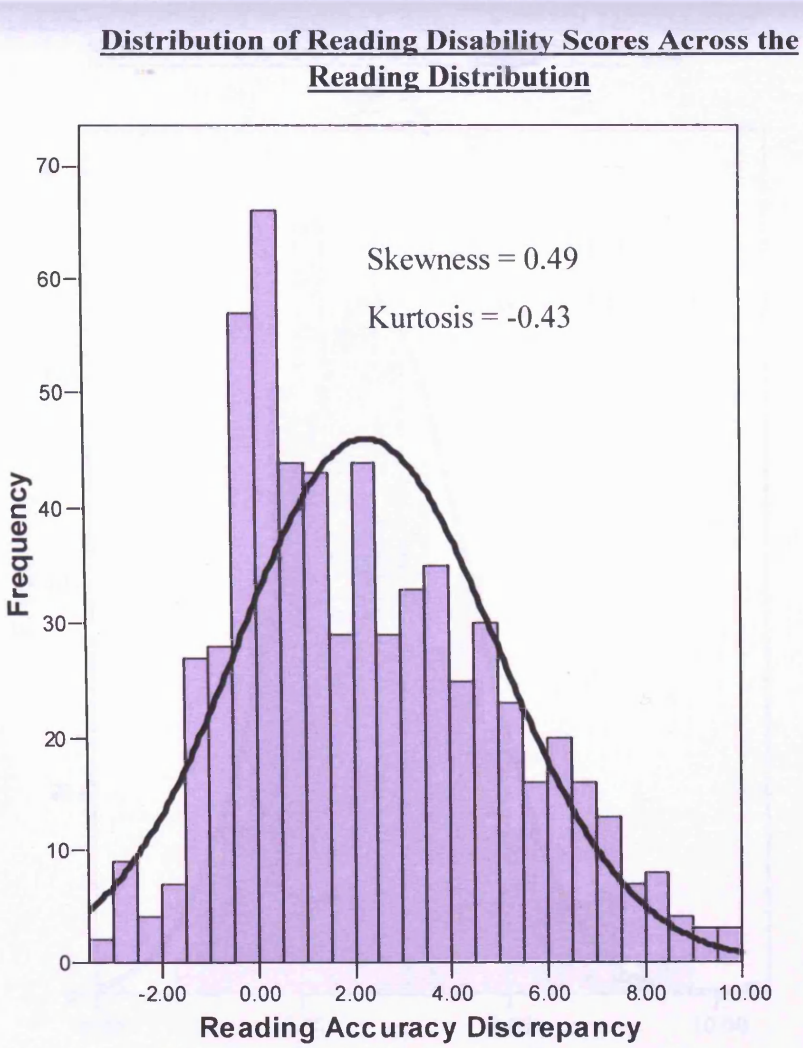
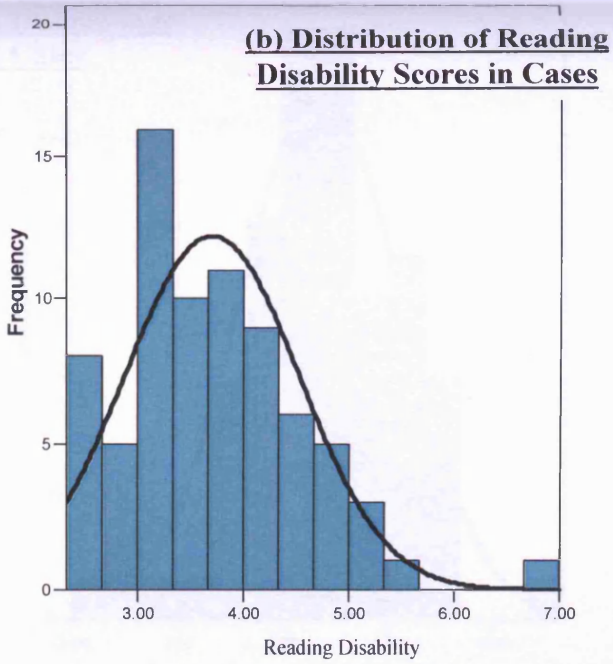


Figure 11 Histograms showing the distribution of non-word repetition scores (a) across the reading ability spectrum, (b) in cases, (c) in controls. All scores have been been regressed for age. The higher the score the better the child has performed on the test.





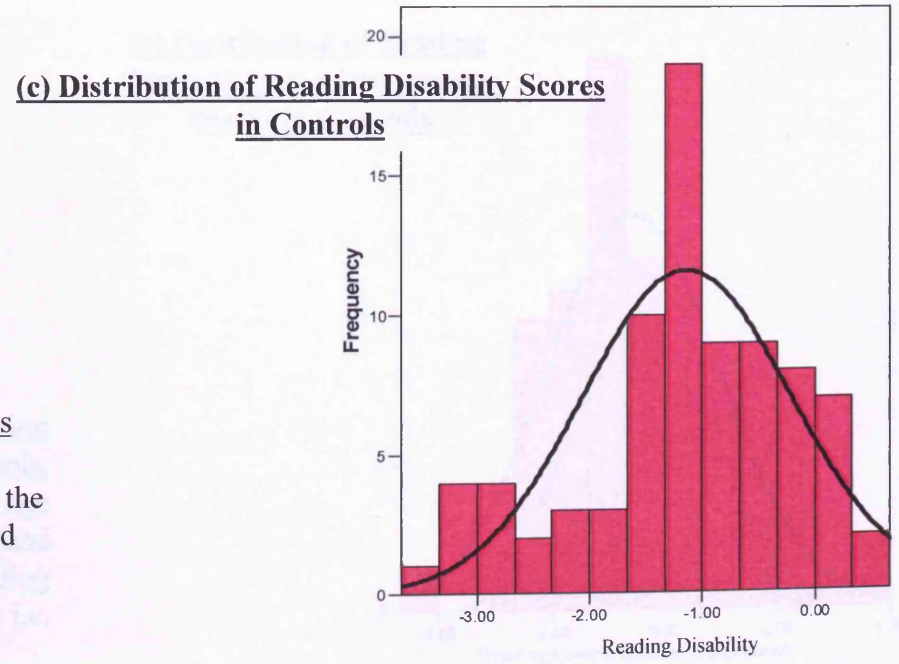
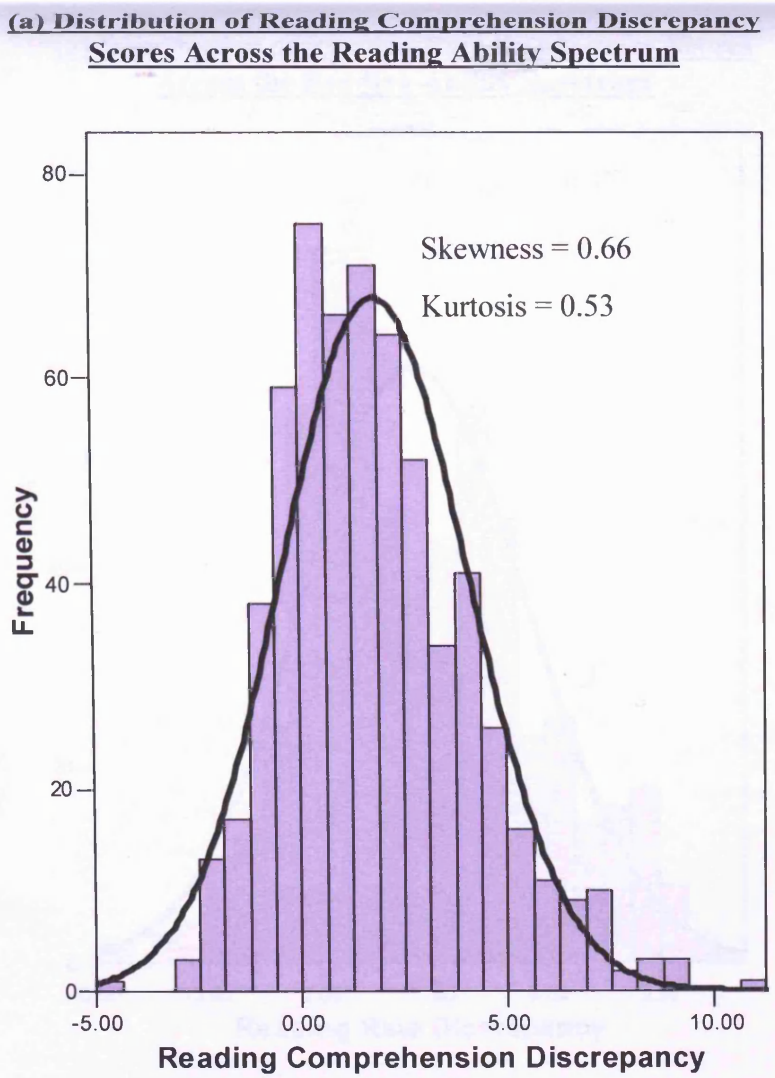
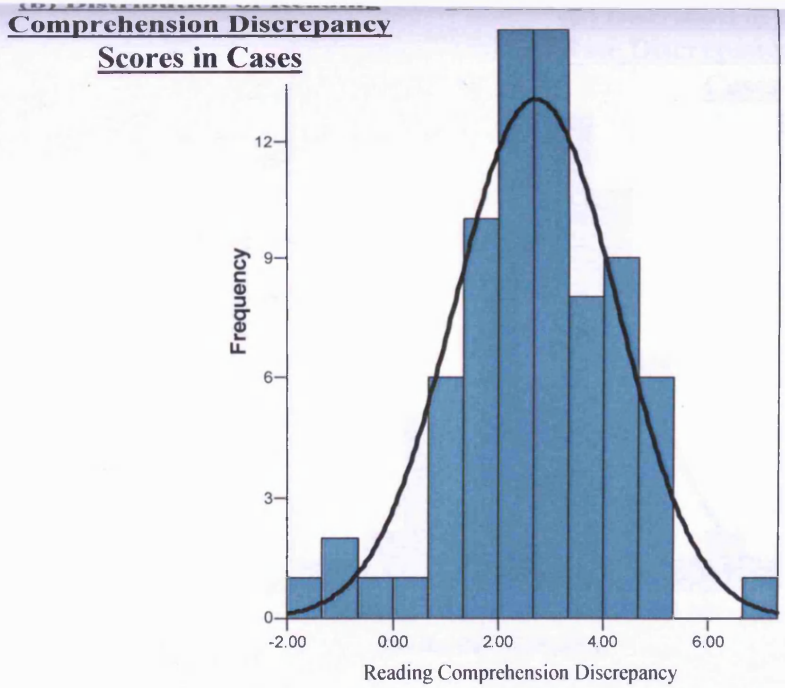


Figure 12 Histograms showing the distribution of reading accuracy scores (a) across the reading distribution, (b) in cases, (c) in controls. All scores based on discrepancy between chonological age and reading age based on the accuracy score from NARA. Only individuals 12 years and under included in histograms. Negative numbers represent reading ahead of their chronological age; positive numbers represent reading lag i.e. reading age below chronological age





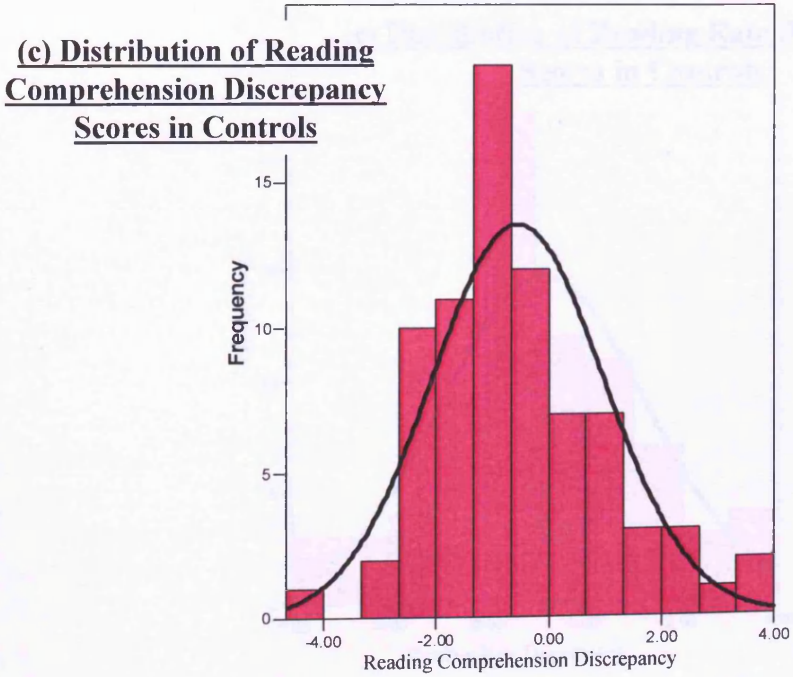
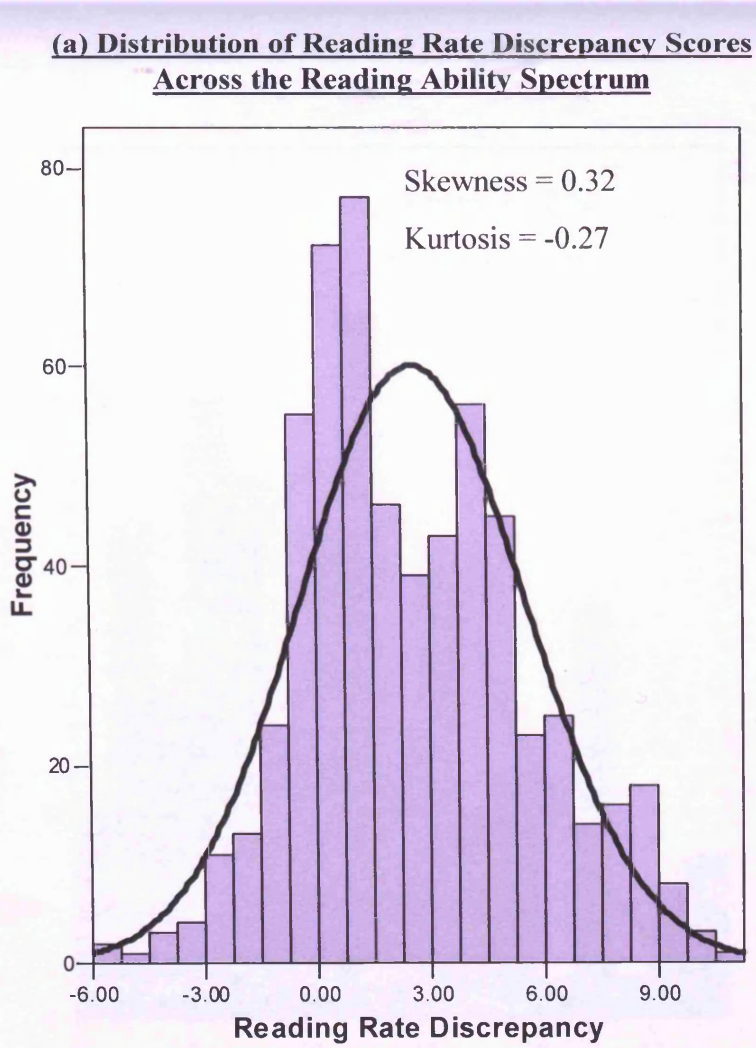
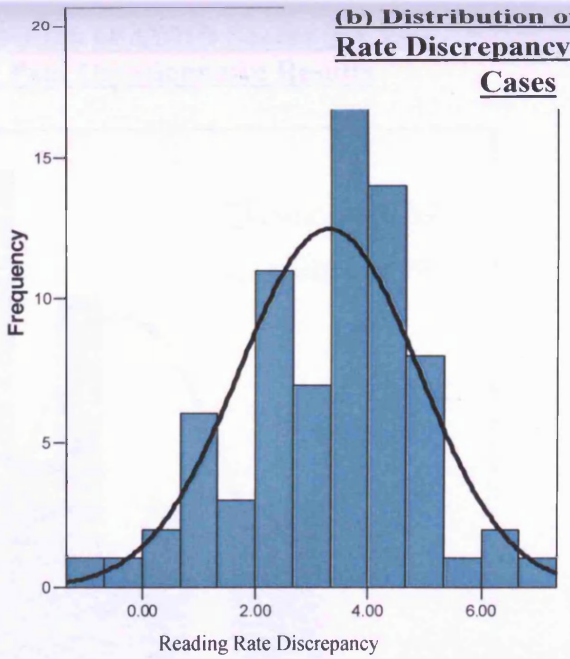


Figure 13 Histograms showing the distributions of reading comprehension scores (a) across the reading ability spectrum, (b) in cases, (c) in controls. All scores based on discrepancy between chonological age and reading age based on the comprehension score of NARA. Only individual 12 years and under are included in the histograms. Negative numbers represent reading ahead of their chronological age; positive numbers represent reading lag i.e. reading age below chronological age





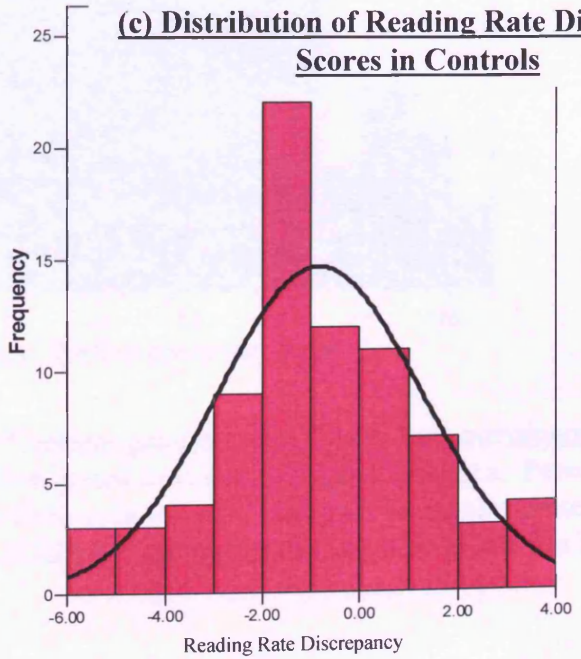


Figure 14 Histograms showing the distributions of reading rate scores (a) across the reading ability spectrum, (b) in cases, (c) in controls. All scores based on discrepancy between chonological age and reading age based on the rate score of NARA. Only individual 12 years and under are included in the histograms. Negative numbers represent reading ahead of their chronological age; positive numbers represent reading lag i.e. reading age below chronological age

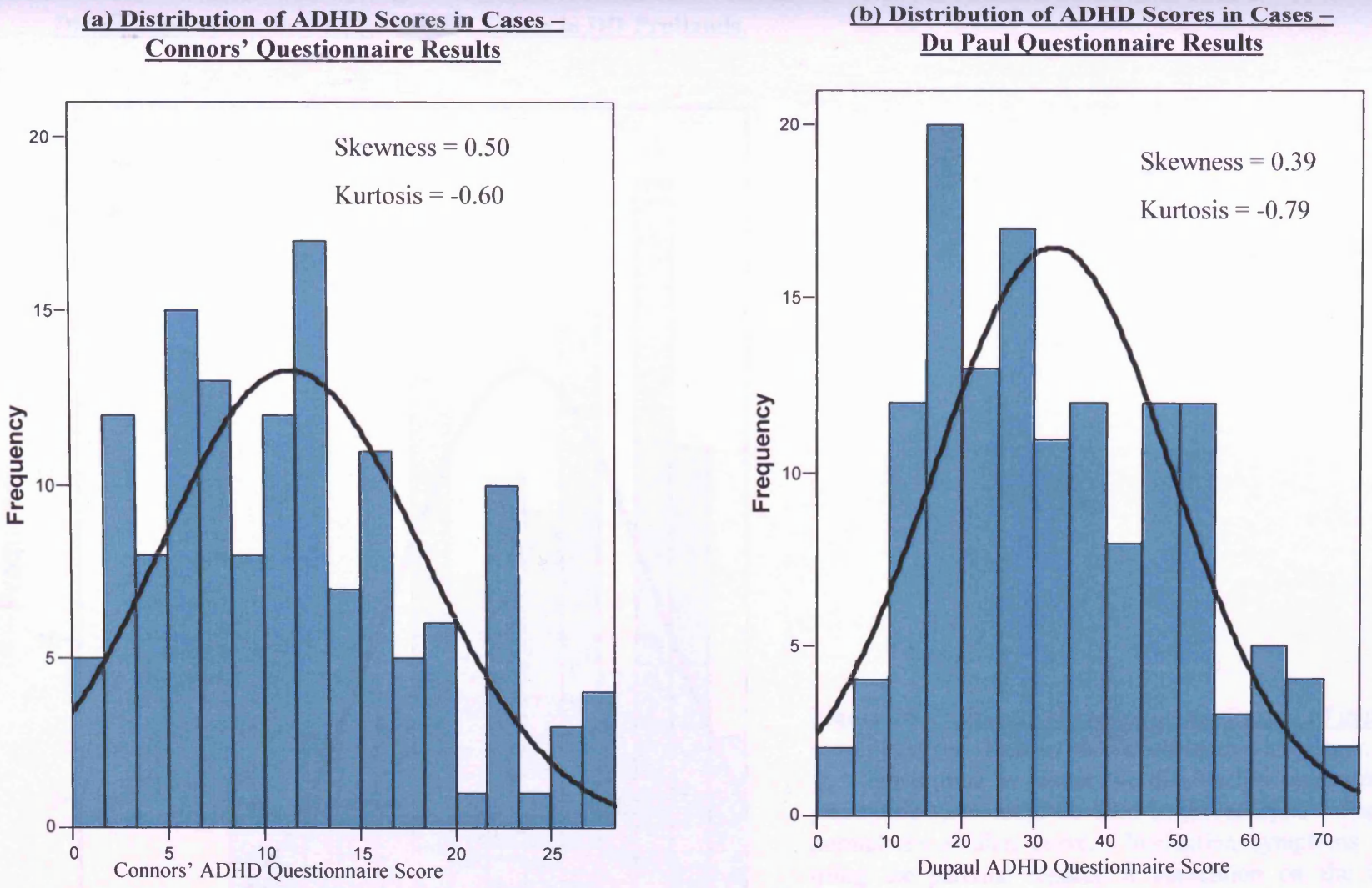


Figure 15 Histograms showing the distribution of ADHD questionnaire scores in cases using (a) the Connors questionnaire, (b) Du Paul questionnaire. Distributions show raw scores reported by parents on ADHD symptoms using the Connors' and Du Paul parent report ADHD questionnaires. Parents of controls children did not report on ADHD symptoms in their child. For categorical analyses (chapter five) a cut off of 15 was used to suggest presence of ADHD in probands using the Connors' questionnaire. A cut off of 35 was used to indicate presence of ADHD from symptoms reported on the Du Paul Questionnaire. The higher the score the more ADHD symptoms are present in the proband.

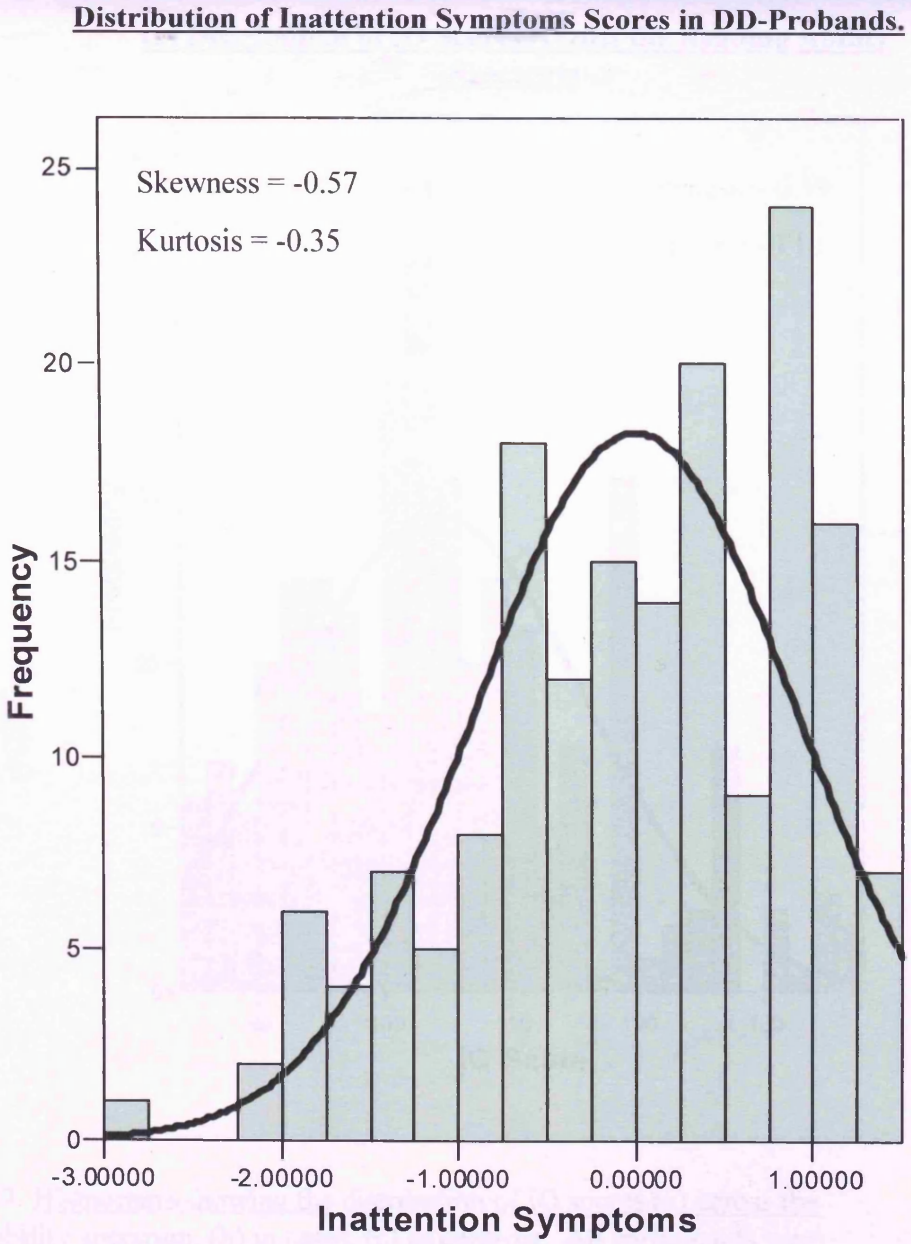


Figure 16 Histogram showing the distribution of inattention symptom scores in cases. Raw scores for inattention have been regressed for age and transformed to ensure the distribution was normal in order that parametric tests could be used in the analysis. The line indicates a normal distribution curve. Inattention symptoms were determined using the parental reports of inattention on the Connors' ADHD questionnaire. The higher the score the more inattention symptoms reported by the parent.

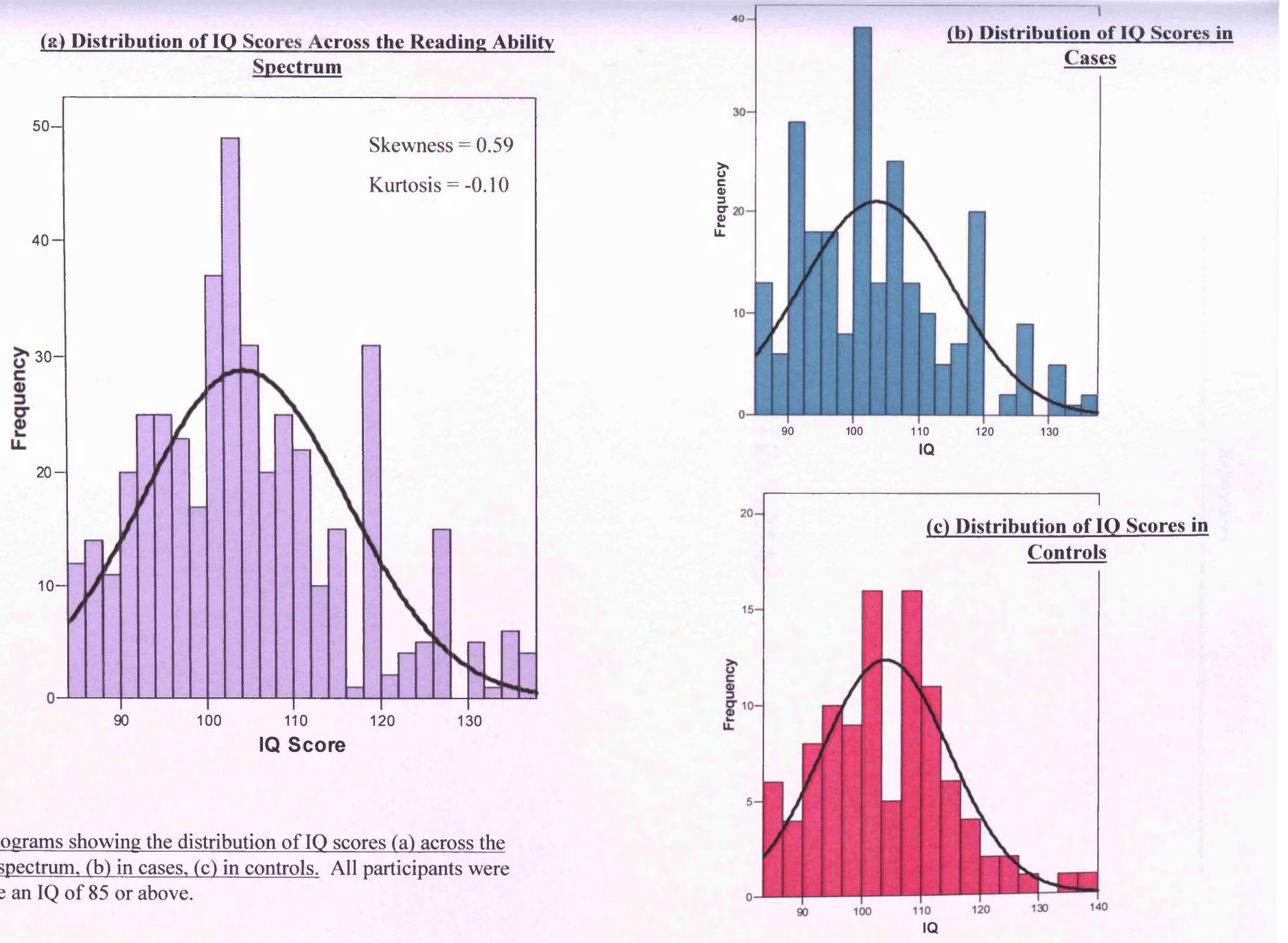


Figure 17 Histograms showing the distribution of IQ scores (a) across the reading ability spectrum, (b) in cases, (c) in controls. All participants were required to have an IQ of 85 or above.

REFERENCES

REFERENCES

Abdullaev YG, Melnichuk KV (1997) Cognitive Operations in the Human Caudate Nucleus. *Neuroscience Letters* 234:151-155

Adlard A, Hazan V (1998) Speech Perception in Children with Specific Reading Difficulties (Dyslexia). *Quarterly Journal of Experimental Psychology* 51A:153-177

Ahissar M, Protopapas A, Reid M, Merzenich MM (2000) Auditory Processing Parallels Reading Abilities in Adults. *Proceedings of the National Academy of Sciences USA* 97:6832-6837

Ahn J-I, Lee K-H, Shin D-M, Shim J-W, Lee J-S, Chang SY, Lee Y-S, Brownstein MJ, Lee S-H, Lee Y-S (2004) Comprehensive Transcriptome Analysis of Differentiation of Embryonic Stem Cells into Midbrain and Hindbrain Neurons. *Developmental Biology* 265:491-501

Alarcón M, DeFries JC (1997) Reading Performance and General Cognitive Ability in Twins with Reading Difficulties and Control Pairs. *Personality and Individual Differences* 22:793-803

Alarcon M, DeFries JC, Fulker DW (1995) Etiology of individual differences in reading performance: a test of sex limitation. *Behavior Genetics* 25:17-23

Amitay S, Ben-Yehudah G, Banai K, Ahissar M (2002) Disabled Readers Suffer from Visual and Auditory Impairments but not from a Specific Magnocellular Deficit. *Brain* 125:2272-2285

Andolfatto P, Depaulis F, Navarro A (2001) Inversion Polymorphisms and Nucleotide Variability in *Drosophila*. *Genetic Research* 77:1-8

Andres-Barquin PJ, Hernandez M-C, Israel MA (1998) Injury Selectivity Down-Regulates the Gene Encoding for the Id4 Transcription Factor in Primary Cultures of Forebrain Astrocytes. *NeuroReport* 9:4075-4080

Andres-Barquin PJ, Hernandez MC, Israel MA (2000) ID Genes in Nervous System Development. *Histology and Histopathology* 15:603-618

August GJ, Garfinkel BD (1990) Comorbidity of ADHD and Reading Disability Among Clinic-Referred Children. *Journal of Abnormal Child Psychology* 18:29-45

Backes W, Vuurman E, Wennekes R, Spronk P, Wuisman M, Engelshoven JV, Jolles J (2002) Atypical Brain Activation of Reading Processes in Children with Developmental Dyslexia. *Journal of Child Neurology* 17:867-871

Baddeley AD (1986) Working Memory. Clarendon, Oxford, England

Baddeley AD (1992) Is Working Memory Working? *Quarterly Journal of Experimental Psychology* 44A:1-31

Badian NA (1996) Dyslexia: A Validation of the Concept at Two Age Levels. *Journal of Learning Disabilities* 29:102-112

Bakwin H (1973) Reading disability in twins. *Developmental Medicine and Child Neurology* 15:184-7

Balogh SA, Sherman GF, Hyde LA, Denenberg VH (1998) Effects of Neocortical Ectopias Upon the Acquisition and Retention of a Non-Spatial Reference Memory Task in BXSB Mice. *Brain Research Developmental Brain Research* 111:291-293

Barnard R, Futo V, Pecheniuk N, Slattery M, Walsh T (1998) PCR Bias Toward the Wild-Type k-*ras* and *p53* Sequences: Implications for PCR Detection of Mutations and Cancer Diagnosis. *Biotechniques* 24:684-691

Baron J (1979) Orthographic and Word Specific Mechanism in Children's Reading of Words. *Child Development* 50:60-72

Barr C (2004) Support for *EKN1* as the Susceptibility Locus for Reading Disabilities and Attention Deficit Hyperactivity Disorder on 15Q21. Paper presented at XIIth World Congress of Psychiatric Genetics. Dublin, Ireland

Barrat BJ, Payne F, Rance HE, Nutland S, Todd JA, Clayton DG (2002) Identification of the Sources of Error in Allele Frequency Estimations from Pooled DNA Indicates an Optimal Experimental Design. *Annals of Human Genetics* 66: 393-405

Barron RW (1986) Word-Recognition in Early Reading: A Review of the Direct and Indirect Access Hypothesis. *Cognition* 24:93-119

Beitchman JH, Young AR (1997) Learning Disorders with a Special Emphasis on Reading Disorders: A Review of the Past Ten Years. *Journal of the American Academy of Child and Adolescent Psychiatry* 36:1020-1032

Ben-Jonathan N, Mershon JL, Allen DL, Steinmetz RW (1996) Extrapituitary Prolactin: Distribution, Regulation, Functions and Clinical Aspects. *Endocrine Reviews* 17:639-669

Ben-Yehudah G, Ahissar M (2004) Sequential Spatial Frequency Discrimination is Consistently Impaired Among Adult Dyslexics. *Vision Research* 44:1047-1063

Ben-Yehudah G, Banai K, Ahissar M (2004) Patterns of Deficit in Auditory Temporal Processing Among Dyslexic Adults. *NeuroReport* 15:627-631

Berninger V (1994) Reading and Writing Acquisition. A Developmental Neuropsychological Disorder. W C Brown and Benchmark, Madison, WI, USA

Berninger VW, Abbot RD (1994) Multiple Orthographic and Phonological Codes in Literacy Acquisition: An Evolving Research Program. In: Berninger VW (ed) The Varieties of Orthographic Knowledge II: Theoretical and Developmental Issues. Kluwer Academic, Dordrecht, The Netherlands, pp 1-32

Berninger VW, Abbott RD, Thomson JB, Raskind WH (2001) Language Phenotype for Reading and Writing Disability: A Lifespan Approach. *Scientific Studies of Reading* 5:59-105

Bisgaard ML, Eiberg H, Møller N, Niebuhr E, Mohr J (1987) Dyslexia and Chromosome 15 Heteromorphism: Negative LOD Score in a Danish Material. *Clinical Genetics* 32:118-119

Bishop DV, Bishop SJ, Bright P, James C, Delaney T, Tallal P (1999) Different Origin of Auditory and Phonological Processing Problems in Children with Language Impairment: Evidence from a Twin Study. *Journal of Speech, Language and Hearing Research* 42:155-168

Blachman BA (1984) Relationship of Rapid Naming Ability and Language Analysis Skills to Kindergarten and First-Grade Reading Achievement. *Journal of Educational Psychology* 76:610-622

Boetsch EA, Green PA, Pennington BF (1996) Psychosocial Correlates of Dyslexia Across the Lifespan. *Development and Psychopathology* 8:539-562

Borecki IB, Suarez BK (2001) Linkage and Association: Basic Concepts. *Advances in Genetics* 42:45-66

Bounpheng MA, Dimas JJ, Dodds SG, Christy BA (1999) Degradation of ID Proteins by the Ubiquitin-Proteasome Pathway. *The FASEB Journal* 13:2257-2264

Bowers PG, Swanson LB (1991) Naming Speed Deficits in Reading Disability: Multiple Measures of a Singular Process. *Journal of Experimental Child Psychology* 51:195-219

Bowey JA, Patel RK (1988) Metalinguistic Ability and Early Reading Achievement. *Applied Psycholinguistics* 9:367-383

Bradley L, Bryant PE (1983) Categorising Sounds and Learning to Read: A Causal Connection. *Nature* 301:419-421

Bradlow AR, Kraus N, Nicol TG, McGee TJ, Cunningham J, Zecker SG (1999) Effects of Lengthened Formant Transition Duration on Discrimination and Neural Representation of Synthetic CV Syllables by Normal and Learning-Disabled Children. *Journal of the Acoustic Society of America* 106:2086-2096

Bray NJ, Buckland PR, Owen MJ, O'Donovan MC (2003) *Cis*-Acting Variation in the Expression of a High Proportion of Genes in Human Brain. *Human Genetics* 113:149-153

Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive Human Genetic Maps: Individual and Sex-Specific Variation in Recombination. *American Journal of Human Genetics* 63:861-869

Brown G (1987) Resolving Inconsistency: A Computational Model of Word Naming. *Journal of Memory and Language* 26:1-23

Brown GM, Leversha M, Hulten M, Ferguson-Smith MA, Affara NA, Furlong RA (1998) Genetic Analysis of Meiotic Recombination in Humans by Use of Sperm Typing: Reduced Recombination Within a Heterozygous Paracentric Inversion of Chromosome 9q32-q34.3. *American Journal of Human Genetics* 62:1484-1492

Brown WE, Eliez S, Menon V, Rumsey JM, White CD, Reiss AL (2001) Preliminary Evidence of Widespread Morphological Variations of the Brain in Dyslexia. *Neurology* 56:781-783

Brunswick N, McCrory E, Price CJ, Frith CD, Frith U (1999) Explicit and Implicit Processing of Words and Pseudowords by Adult Developmental Dyslexics: A Search for Wernicke's Wortschatz? *Brain* 122:1901-1917

Bryant PE, Impey L (1986) The Similarities Between Normal Readers and Developmental and Acquired Dyslexic Children. *Cognition* 24:121-137

Cardon LR, Bell JI (2001) Association Study Designs for Complex Diseases. *Nature Reviews Genetics* 2:91-99

Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC (1994) Quantitative Trait Locus for Reading Disability on Chromosome 6. *Science* 266:276-279

Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC (1995) Quantitative Trait Locus for Reading Disability: Correction. *Science* 268:1553

Carroll JM, Snowling MJ (2004) Language and Phonological Skills in Children at High Risk of Reading Difficulties. *Journal of Child Psychology and Psychiatry* 45:631-640

Castles A, Coltheart M (1993) Varieties of Developmental Dyslexia. *Cognition* 47:149-180

Castles A, Coltheart M (2004) Is There a Causal Link from Phonological Awareness to Success in Learning to Read? *Cognition* 91:77-111

Castles A, Datta H, Gayan J, Olson RK (1999) Varieties of Developmental Reading Disorder: Genetic and Environmental Influences. *Journal of Experimental Child Psychology* 72

Chapman NH, Igo RP, Thomson JB, Matsushita M, Brkanac Z, Holzman T, Berninger VW, Wijsman EM, Raskind W (2004) Linkage Analyses of Four Regions Previously Implicated in Dyslexia: Confirmation of a Locus on Chromosome 15q. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 131:67-75

Charlesworth B, Morgan MT, Charlesworth D (1983) The Effect of Deleterious Mutations on Neutral Molecular Variation. *Genetics* 134:1289-1303

Cheng C, Xu J, Ye X, Dai J, Wu Q, Zeng L, Wang L, Zhao W, Ji C, Gu S, Xie Y, Mao Y (2002) Cloning, Expression and Characterization of a Novel Human VMP Gene. *Molecular Biology Reports* 29:281-286

Childs B, Finucci JM (1983) Genetics, Epidemiology and Specific Reading Disability. In: Rutter M (ed) Developmental Psychiatry. Guilford, New York, pp 507-519

Claridge G, Clark K, Davis C (1997) Nightmares, Dreams and Schizotypy. *British Journal of Clinical Psychology* 36:377-386

Clayton D (1999) A Generalization of the Transmission/Disequilibrium Test for Uncertain Haplotype Transmission. *American Journal of Human Genetics* 65:1170-1177

Cohen M, Campbell R, Yaghmai F (1989) Neuropathological Abnormalities in Developmental Dyslexia. *Annals of Neurology* 25:567-570

Collins A, Lonjou C, Morton NE (1999) Genetic Epidemiology of Single-Nucleotide Polymorphisms. *Proceedings of the National Academy of Sciences USA* 96:15173-15177

Coltheart M (1978) Lexical Access in Simple Reading Tasks. In: Underwood G (ed) Strategies of Information Processing. Academic Press, London, pp 151-216

Coltheart M (1987) Varieties of Developmental Dyslexia: A Comment on Bryant and Impey. *Cognition* 27:97-101

Connors CK (1973) Rating Scale for Use in Drug Studies with Children. *Psychopharmacology Bulletin* 9:24-29

Consortium IHGS (2001) Initial Sequencing and Analysis of the Human Genome. *Nature* 409:860-921

Consortium TIH (2005) A Haplotype Map of the Human Genome. *Nature* 437:1299-1320

Cope N, Harold D, Hill G, Moskvina V, Stevenson J, Holmans P, Owen MJ, O'Donovan MC, Williams J (2005) Strong Evidence That *KIAA0319* on Chromosome 6p is a Susceptibility Gene for Developmental Dyslexia. *American Journal of Human Genetics* 76:581-591

Cordell HJ (2002) Epistasis: What it Means, What it Doesn't Mean and Statistical Methods to Detect it in Humans. *Human Molecular Genetics* 11(20):2463-8

Cordell HJ, Clayton DG (2002) A Unified Stepwise Regression Procedure for Evaluating the Relative Effects of Polymorphisms within a Gene using Case/Control or Family Data: Application to HLA in Type I Diabetes. *American Journal of Human Genetics* 70:124-141

Cordell HJ, Clayton DG (2005) Genetic Association Studies. *The Lancet* 366:1121-1131

Cornelissen P, Richardson A, Mason A, Fowler S, Stein J (1995) Contrast Sensitivity and Coherent Motion Detection Measured at Photopic Luminance Levels in Dyslexics and Controls. *Vision Research* 35:1483-1494

Couto J (2005). In: Cope N (ed), Toronto, Canada

Critchley M (1981) Dyslexia: An Overview. In: Pavlidis GT, Miles TR (eds) Dyslexia Research and its Applications to Education. Wiley, Chichester, UK

Crosson B (1999) Subcortical Mechanisms in Language: Lexical-Semantic Mechanisms and the Thalamus. *Brain and Cognition* 40:414-438

Crowe RR (1993) Candidate Genes in Psychiatry: An Epidemiological Perspective. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 48:74-77

Cruchten Iv, Cinato E, Fox M, King ER, Newton JS, Riechmann V, Sablitzky F (1998) Structure, Chromosomal Localisation and Expression of the Murine Dominant Negative Helix-Loop-Helix ID4 Gene. *Biochimica et Biophysica Acta* 1443:55-64

Daal VV, Leij AVD (1999) Developmental Dyslexia: Related to Specific or General Deficits. *Annals of Dyslexia* 49:71-104

Decker SN, Vandenberg SG (1985) Colorado Twin Study of Reading Disability. In: Gray DB, Kavanagh JF (eds) Biobehavioural Measures of Dyslexia. York Press, Parkton, MD, pp 123-135

Deffenbacher KE, Kenyon JB, Hoover DM, Olson RK, Pennington BF, DeFries JC, Smith SD (2004) Refinement of the 6p21.3 Quantitative Trait Locus Influencing Dyslexia: Linkage and Association Analyses. *Human Genetics* 115:128-138

DeFries JC, Alarcón M, Olson RK (1997a) Genetic Aetiologies of Reading and Spelling Deficits: Developmental Differences. In: Hulme, Snowling (eds) Dyslexia: Biology, Cognition and Intervention. Whurr, London, pp 20-37

DeFries JC, Decker SN (1982) Genetic aspects of Reading Disability: A Family Study. In: Malatesha RN, Aaron PG (eds) Reading Disorders: Varieties and Treatments. Academic Press, New York, pp 255

DeFries JC, Filipek PA, Fulker DW, Olson RK, Pennington BF, Smith SD, Wise BW (1997b) Colorado Learning Disabilities Research Center. *Learning Disabilities* 8:7-19

DeFries JC, Fulker DW (1985) Multiple Regression Analysis of Twin Data. *Behavior Genetics* 15:467-473

DeFries JC, Fulker DW, LaBuda MC (1987) Evidence for a Genetic Aetiology in Reading Disability of Twins. *Nature* 329:537-539

DeFries JC, Gillis JJ (1991) Etiology of Reading Deficits in Learning Disabilities: Quantitative Genetic Analysis. In: Obrzut JE, Hynd GW (eds) Neuropsychological Foundations of Learning Disabilities: A Handbook of Issues, Methods and Practice. Academic Press, Inc, London, pp 29-47

DeFries JC, Gillis JJ (1993) Genetics of Reading Disability. In: Plomin R, McClearn GE (eds) Nature, Nurture and Psychology. American Psychological Association, Washington DC, USA, pp 121-145

DeFries JC, Olson RK, Pennington BF, Smith SD (1991) Colorado Reading Project: Past, Present and Future. *Learning Disabilities* 2:37-46

DeFries JC, Stevenson J, Gillis JJ, Wadsworth SJ (1991) Genetic Etiology of Spelling Deficits in the Colorado and London Twin Studies of Reading Disability. *Reading and Writing: An Interdisciplinary Journal* 3:271-283

Denckla M, Rudel R (1976) Rapid "Automatized" Naming (R.A.N.): Dyslexia Differentiated from other Learning Disabilities. *Neuropsychologia* 14:471-479

Denenberg VH, Sherman GF, Schrott LM, Rosen GD, Galaburda AM (1991) Spatial Learning, Discrimination Learning, Paw Preference and Neocortical Ectopias in Two Autoimmune Strains of Mice. *Brain Research* 562:98-104

Deng HW, Mahaney MC, Williams JT, Li J, Conway T, Davies KM, Li JL, Deng H, Recker RR (2002) Relevance of the Genes for Bone Mass Variation to Susceptibility to Osteoporotic Fractures and its Implications to Gene Search for Complex Human Diseases. *Genetic Epidemiology* 22:12-25

Doyon J, Penhume V, Ungerleider LG (2003) Distinct Contribution of the Cortico-Striatal and Cortico-Cerebellar Systems to Motor Skill Learning. *Neuropsychologia* 41:252-262

DSM-IV, 2000: Diagnostic and Statistical Manual of Mental Disorders. American Psychiatric Association, Washington DC

Duan J, Wainwright MS, Comeron JM, Saitou N, Sanders AR, Gelernter J, Gejman PV (2003) Synonymous Mutations in the Human Dopamine Receptor D2 (DRD2) Affect mRNA Stability and the Synthesis of the Receptor. *Human Molecular Genetics* 12:205

Dudbridge F (2003) Pedigree Disequilibrium Tests for Multilocus Haplotypes. *Genetic Epidemiology* 25:115-221

Edelmann L, Spiteri E, Koren K, Pulijaal V, Bialer MG, Shanske A, Goldberg R, Morrow BE (2001) AT-Rich Palindromes Mediate the Constitutional t(11;22) Translocation. *American Journal of Human Genetics* 68:1-13

Edwards JH (1965) The Meaning of Associations Between Blood Groups and Disease. *Annals of Human Genetics* 29:77-83

Ehri LC (1989) The Development of Spelling Knowledge and its Role in Reading Acquisition and Reading Disability. *Journal of Learning Disabilities* 22:356-365

Ehri LC (1992) Reconceptualizing the Development of Sight Word Reading and its Relationship to Recoding. In: Gough P, Ehri L, Treiman R (eds) Reading Acquisition. Erlbaum, Hillsdale, NJ, USA, pp 107-144

Elliot CD (1983) British Ability Scales. NFER-Nelson, Windsor

Ellis AW (1985) The Production of Spoken Words: A Cognitive Neuropsychological Perspective. In: Ellis AW (ed) Progress in the Psychology of Language. Vol 2. Wrlbaum, Hillsdale, NJ, USA

Eustis RS (1947) Specific Reading Disability. *New England Journal of Medicine* 237:243-249

Eynde BJVD, Gaoler B, Probst-Kepper M, Michaux L, Devuyst O, Lorge F, Weynants P, Boon T (1999) A New Antigen Recognized by Cytolytic T Lymphocytes on a Human Kidney Tumour Results from Reverse Strand Transcription. *Journal of Experimental Medicine* 190:1793-1799

Facoetti A, Paganoni P, Lorusso ML (2000) The Spatial Distribution of Visual Attention in Developmental Dyslexia. *Experimental Brain Research* 132:531-538

Fagerheim T, Raeymaekers P, Tønnessen FE, Pedersen M, Tranebjærg L, Lubs HA (1999) A New Gene (DYX3) for Dyslexia is Located on Chromosome 2. *Journal of Medical Genetics* 36:664-669

Fagerheim T, Toennessen F-E, Lubs HA (2002) Refinement of the Dyslexia 3 Locus (*DYX3*) on Chromosome 2p15-p16 and Mutation Screening of Candidate Genes. *American Journal of Human Genetics* 71:461

Faraone SV, Doyle AE, Mick E, Biederman J (2001) Meta-Analysis of the Association Between the 7-Repeat Allele of the Dopamine D4 Receptor Gene and Attention Deficit Hyperactivity Disorder. *American Journal of Psychiatry* 158:1052-1057

Farrag AF, Khedr EM, Abel-Naser W (2002) Impaired Parvocellular Pathway in Dyslexic Children. *European Journal of Neurology* 9:359-363

Fawcett AJ, Nicholson RI (2001) Dyslexia: The Role of the Cerebellum. In: Fawcett AJ (ed) Dyslexia: Theory and Good Practice. Whurr, London, pp 89-105

Fawcett AJ, Nicholson RI, Dean P (1996) Impaired Performance of Children with Dyslexia on a Range of Cerebellar Tasks. *Annals of Dyslexia* 46:259-283

Fedorova L, Fedorov A (2003) Introns in Gene Evolution. *Genetica* 118:123

Field LL, Kaplan BJ (1998) Absence of Linkage of Phonological Coding Dyslexia to Chromosome 6p23-p21.3 in a Large Family Data Set. *American Journal of Human Genetics* 63:1448-1456

Finucci JM, Childs B (1981) Are There Really More Dyslexic Boys Than Girls? In: Ansara A, Geschwind N, Galaburda A, Albert M, Gatrell N (eds) Sex Differences in Dyslexia. Orton Dyslexia Society, Towson, Maryland, USA

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

Fisher JH (1905) Case of Congenital Word Blindness (Inability to Learn to Read). *Opthalmological Review* 24:315

Fisher SE, Francks C, Marlow AJ, MacPhie IL, Newbury DF, Cardon LR, Ishikawa-Brush Y, Richardson AJ, Talcott JB, Gayán J, Olson RK, Pennington BF, Smith SD, DeFries JC, Stein JF, Monaco AP (2002) Independent Genome-Wide Scans Identify a Chromosome 18 Quantitative-Trait Locus Influencing Dyslexia. *Nature Genetics* 30:86-91

Fisher SE, Marlow AJ, Lamb J, Maestrini E, Williams DF, Richardson AJ, Weeks DE, Stein JF, Monaco AP (1999) A Quantitative-Trait Locus on Chromosome 6p Influences Different Aspects of Developmental Dyslexia. *American Journal of Human Genetics* 64:146-156

Fisher SE, Smith SD (2001) Progress Towards the Identification of Genes Influencing Developmental Dyslexia. In: Dyslexia: Theory and Good Practice. Whurr

Flowers DL (1993) Brain Basis of Dyslexia: A Summary of Work in Progress. *Journal of Learning Disabilities* 26:575-582

Flowers DL, Wood FB, Naylor CE (1991) Regional Cerebral Blood Flow Correlates of Language Processes in Reading Disability. *Archives of Neurology* 48:637-643

Flynn JM, Rahbar MH (1994) Prevalence of Reading Failure in Boys Compared with Girls. *Psychology in the Schools* 31:66-71

Francks C, Fisher SE, Marlow AJ, MacPhie IL, Taylor KE, Richardson AJ, Stein JF, Monaco AP (2003) Familial and Genetic Effects on Motor Coordination, Laterality and Reading-Related Cognition. *American Journal of Psychiatry* 160:1970-1977

Francks C, Fisher SE, Olson RK, Pennington BF, Smith SD, DeFries JC, Monaco AP (2002) Fine Mapping of the Chromosome 2p12-16 Dyslexia Susceptibility Locus: Quantitative Association Analysis and Positional Candidate Genes *SEMA4F* and *OTX1*. *Psychiatric Genetics* 12:35-41

Francks C, MacPhie IL, Monaco AP (2002) The Genetic Basis of Dyslexia. *The Lancet Neurology* 1:483-490

Francks C, Paracchini S, Smith SD, Richardson AJ, Scerri TS, Cardon LR, Marlow AJ, MacPhie IL, Walter J, Pennington BF, Fisher SE, Olson RK, DeFries JC, Stein JF, Monaco AP (2004) A 77-Kilobase Region of Chromosome 6p22.2 is Associated with Dyslexia in Families from the United Kingdom and from the United States. *American Journal of Human Genetics* 75:1046-1058

Frank Y, Pavlakis SG (2001) Brain Imaging in Neurobehavioural Disorders. *Pediatric Neurology* 25:278-287

Fried I, Wilson CL, Morrow JW, Cameron KA, Behnke ED, Ackerson LC, Maidment NT (2001) Increased Dopamine Release in the Human Amygdala During Performance of Cognitive Tasks. *Nature Neuroscience* 4:201-206

Friedman MC, Chhabildas N, Budhiraja N, Willcutt EG, Pennington BF (2003) Etiology of the Comorbidity Between RD and ADHD: Exploration of the Non-Random Mating Hypothesis. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 120B:109-115

Frisse L, Hudson RR, Bartoszewicz A, Wall JD, Donfack J, Rienzo AD (2001) Gene Conversion and Different Population Histories May Explain the Contrast Between Polymorphism and Linkage Disequilibrium Levels. *American Journal of Human Genetics* 69:831-843

Frith U (1985) Beneath the Surface of Developmental Dyslexia. In: Patterson K, Marshall JC, Coltheart M (eds) Surface Dyslexia. Erlbaum, London, pp 301-330

Froster U, Schulte-Körne G, Hebebrand J, Remschmidt H (1993) Cosegregation of Balanced Translocation (1;2) With Retarded Speech Development and Dyslexia. *Journal of Medical Genetics* 342:179

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The Structure of Haplotype Blocks in the Human Genome. *Science* 296:2225-2229

Galaburda A, Sherman G, Rosen G, Aboitiz F, Geschwind N (1985) Developmental Dyslexia: Four Consecutive Patients with Cortical Anomalies. *Annals of Neurology 18*:222-233

Galaburda AM, Kemper TL (1979) Cytoarchitectonic Abnormalities in Developmental Dyslexia: A Case Study. *Annals of Neurology* 6:94-100

Galaburda AM, Menard MT, Rosen GD, Aboitiz F, Geschwind N (1994) Evidence for Aberrant Auditory Anatomy in Developmental Dyslexia. *Proceedings of the National Academy of Sciences USA* 91:8010-1013

Gathercole SE, Baddeley AD (1993) Working Memory and Language. Erlbaum, Hove, UK

Gathercole SE, Willis CS, Baddeley AD, Emslie H (1994) The Children's Test of Nonword Repetition: A Test of Phonological Memory. *Memory* 2:103-127

Gayán J, Olson RK (1999) Reading Disability: Evidence for a Genetic Etiology. *European Child and Adolescent Psychiatry* 8:S52-S55

Gayán J, Olson RK (2001) Genetic and Environmental Influences on Orthographic and Phonological Skills in Children with Reading Disability. *Developmental Neuropsychology* 20:483-507

Gayán J, Olson RK, Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC (1995) Quantitative Trait Locus for Different Measures of Reading Disability. *Behavior Genetics* 25:266

Gayán J, Smith SD, Cherny SS, Cardon LR, Fulker DW, Brower AM, Olson RK, Pennington BF, DeFries JC (1999) Quantitative-Trait Locus for Specific Language and Reading Deficits on Chromosome 6p. *American Journal of Human Genetics* 64:157-164

Geiger G, Lettvin JY, Fahle M (1994) Dyslexic Children Learn a New Visual Strategy for Reading: A Controlled Experiment. *Vision Research* 34:1223-1233

Geschwind N, Behan P (1982) Left-Handedness: Association with Immune Disease, Migraine and Developmental Learning Disorder. *Proceedings of the National Academy of Sciences USA* 79:5097-5100

Geschwind N, Galaburda AM (1985) Cerebral Lateralization: Biological Mechanisms, Associations and Pathology: I. A Hypothesis and a Program for Research. *Archives of Neurology* 42:428-459

Gilger JW, Pennington BF, DeFries JC (1991) Risk for Reading Disability as a Function of Parental History in Three Family Studies. *Reading and Writing: An Interdisciplinary Journal* 3:205-217

Gilger JW, Pennington BF, DeFries JC (1992) A Twin Study of the Etiology of Comorbidity: Attention-Deficit Hyperactivity Disorder and Dyslexia. *Journal of the American Academy of Child and Adolescent Psychiatry* 31:343-348

Gilger JW, Pennington BF, Green P, Smith SM, Smith SD (1992b) Reading Disability, Immune Disorders and Non-Right-Handedness: Twin and Family Studies of their Relations. *Neuropsychologia* 30:209-227

Gleeson JG, Allen KM, Fox JW, Lamperti ED, Berkovic S, Scheffer I, Cooper EC, Dobyns WB, Minnerath SR, Ross ME, Walsh CA (1998) Doublecortin, A Brain-Specific Gene Mutated in Human X-Linked Lissencephally and Double Cortex Syndrome, Encodes a Putative Signalling Protein. *Cell* 92:63-72

Glushko RJ (1979) The Organization and Activation of Orthographic Knowledge in Reading Aloud. *Journal of Experimental Psychology* 5:674-691

Glushko RJ (1981) Principles for Pronouncing Print: The Psychology of Phonography. In: Lesgold AM, Perfetti CA (eds) Interactive Processes in Reading. Erlbaum, Hillsdale, NJ, USA, pp 61-84

Golding GB (1984) The Sampling Distribution of Linkage Disequilibrium. *Genetics* 108:257-274

Goswami U (1986) Children's Use of Analogy in Learning to Read: A Developmental Study. *Journal of Experimental Child Psychology* 42:73-83

Goswami U (1988) Orthographic Analogies and Reading Development. *Quarterly Journal of Experimental Psychology* 40A:239-268

Gough PB, Tunmer W (1986) Decoding, Reading and Reading Disability. *Remedial and Special Education* 7:6-10

Gough PB, Walsh MA (1991) Chinese, Phoenicians and the Orthographic Cipher of English. In: Brady SA, Shankweiler DP (eds) Phonological Processes in Literacy. Erlbaum, Hillsdale, NJ, USA, pp 199-209

Grigorenko EL (2003) The First Candidate Gene for Dyslexia: Turning the Page of a New Chapter of Research. *Proceedings of the National Academy of Sciences* 100:11190-11192

Grigorenko EL, Wood FB, Golovyan L, Meyer M, Romano C, Pauls D (2003) Continuing the Search for Dyslexia Genes on 6p. *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)* 118B:89-98

Grigorenko EL, Wood FB, Meyer MS, Hart LA, Speed WC, Shuster A, Pauls DL (1997) Susceptibility Loci for Distinct Components of Developmental Dyslexia on Chromosome 6 and 15. *American Journal of Human Genetics* 60:27-39

Grigorenko EL, Wood FB, Meyer MS, Pauls DL (2000) Chromosome 6p Influences of Different Dyslexia-Related Cognitive Processes: Further Confirmation. *American Journal of Human Genetics* 66:715-723

Grigorenko EL, Wood FB, Meyer MS, Pauls JED, Hart LA, Pauls DL (2001) Linkage Studies Suggest a Possible Locus for Developmental Dyslexia on Chromosome 1p. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 105:120-129

Gross-Glenn K, Duara R, Barker WW, Lowenstein D, Chang JY, Yoshii F, Apicella AM, Pascal S, Boothe T, Sevush S, Jallard BJ, Novoa L, Lubs HA (1991) Positron Emission Tomographic Studies During Serial Word Reading by Normal and Dyslexic Adults. *Journal of Clinical and Experimental Neuropsychology* 13:531-544

Guze BH (1991) Magnetic Resonance Spectroscopy: A Technique for Functional Brain Imaging. *Archives of General Psychiatry* 48:572-574

Habib M (2000) The Neurobiological Basis of Developmental Dyslexia. An Overview and Working Hypothesis. *Brain* 123:2373-2399

Haff LA, Smirnov IP (1997) Single-Nucleotide Polymorphism Identification Assays Using a Thermostable DNA Polymerase and Delayed Extraction MALDI-TOF mass Spectrometry. *Genome Research* 7:378-388

Hairston WD, Burdette JH, Flowers DL, Wood FB, Wallace MT (2005) Altered Temporal Profile of Visual-Auditory Multisensory Interaction in Dyslexia. *Experimental Brain Research* 166:474-480

Hallgren B (1950) Specific Dyslexia (Congenital Word Blindness). A Clinical and Genetic Study. *Acta Psychiatrica et Neurologica Supplementum* 65:1-287

Hari R, Renvall H, Tanskanen T (2001) Left Minineglect in Dyslexic Adults. *Brain* 124:1373-1380

Harris EL (1986) The Contribution of Twin Research to the Study of the Etiology of Reading Disability. In: Smith SD (ed) Genetics and Learning Disabilities. Taylor and Francis Ltd, London, pp 3-19

Hartl DL, Clark AG (1997) Principles of Population Genetics. Sinauer Associates, Sunderland Massachusetts

Hatcher PJ, Hulme C, Snowling MJ (2004) Explicit Phoneme Training Combined with Phonic Reading Instruction Helps Young Children at Risk of Reading Failure. *Journal of Child Psychology and Psychiatry* 45:338-358

Hermann K (1956) Congenital Word Blindness. *Acta Psychiatrica et Neurologica Scandinavica* 108:177-184

Hermann K, Norrie E (1958) Is Congenital Word Blindness a Hereditary Type of Gerstmann's Syndrome? *Psychiatria et Neurologia* 136:59-73

Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR (2005) Whole-Genome Patterns of Common DNA Variation in Three Human Populations. *Science* 307:1072-1079

Hinshelwood J (1900) Letter-, Word- and Mind-Blindness. Lewis and Son, London

Hinshelwood J (1907) Four Cases of Congenital Word-Blindness Occurring in the Same Family. *British Medical Journal* 2:1229-1232

Hinshelwood J (1911) Two Cases of Hereditary Word-Blindness. *British Medical Journal* 1:608-609

Hinshelwood J (1917) Congenital Word-Blindness. H K Lewis, London

Hodge SE (1993) Linkage Analysis Versus Association Analysis: Distinguishing Between Two Models that Explain Disease-Marker Associations. *American Journal of Human Genetics* 53:367-384

Hohnen B, Stevenson J (1999) The structure of genetic influences on general cognitive, language, phonological, and reading abilities. *Developmental Psychology* 35:590-603

Hoogendoorn B, Norton N, Kirov G, Williams N, Hamshere ML, Spurlock G, Austin J, Stephens MK, Buckland PR, Owen MJ, O'Donovan MC (2000) Cheap, Accurate and Rapid Allele Frequency Estimation of Single Nucleotide Polymorphisms by Primer Extension and DHPLC in DNA Pools. *Human Genetics* 107:488-493

Horvath S, Xu X, Laird NM (2001) The Family Based Association Test Method: Strategies for Studying Genera Genotype-Phenotype Associations. *European Journal of Human Genetics* 9:301-306

Horwitz B, Rumsey JM, Donohue BC (1998) Functional Connectivity of the Angular Gyrus in Normal Reading and Dyslexia. *Proceedings of the National Academy of Sciences USA* 95:8939-8944

Hsiung G-YR, Kaplan BJ, Petryshen TL, Lu S, Field LL (2004) A Dyslexia Susceptibility Locus (*DYX7*) Linked to Dopamine D4 Receptor (DRD4) Region on Chromosome 11p15.5. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 125:112-119

Hudson RR, Kaplan NL (1985) Statistical Properties of the Number of Recombination Events in the History of a Sample of DNA Sequences. *Genetics* 111:147-164

Hulme C, Snowling M (1992) Phonological Deficits in Dyslexia: A "Sound" Reappraisal of the Verbal Deficit Hypothesis? In: Singh NN, Beale IL (eds) Learning Disabilities: Nature, Theory and Treatment. Springer-Verlag, New York, USA, pp 270-301

Humphreys GW, Evett LJ (1985) Are there Independent Lexical and Nonlexical Routes in Word Processing? An Evaluation of the Dual-Route Theory of Reading. *The Behavioral and Brain Sciences* 8:689-739

Humphreys P, Kaufmann WE, Galaburda AM (1990) Developmental Dyslexia in Women: Neuropathological Findings in Three Patients. *Annals of Neurology* 28:727-738

Hur E-H, Hur W, Choi J-Y, Kim I-K, Kim H-Y, Yoon SK, Rhim H (2004) Functional Identification of the Pro-Apoptotic Effector Domain in Human SOX4. *Biochemical and Biophysical Research Communications* 325:59-67

Hurst LD (1999) The Evolution of Genomic Anatomy. TREE 14:108-112

Hyde LA, Hoplight BJ, Harding S, Sherman GF, Mobraaten LE, Denenberg VH (2001) Effects of Ectopias and their Cortical Location on Several Measures of Learning in BXSB Mice. *Developmental Psychobiology* 39:286-300

Hyde LA, Stavnezer AJ, Bimonte HA, Sherman GF, Denenberg VH (2002) Spatial and Nonspatial Morris Maze Learning: Impaired Behavioral Flexibility in Mice with Ectopias Located in the Prefrontal cortex. *Behavioral Brain Research* 133:247-259

Hynd GW, Hall J, Novey ES, Eliopulos D, Black K, Gonzalez JJ, Edmonds JE, Riccio C, Cohen M (1995) Dyslexia and Corpus Callosum Morphology. *Archives of Neurology* 52:32-38

Hynd GW, Lorys AR, Semrud-Clikeman M, Nieves N, Huettner MIS, Lahey BB (1991) Attention-Deficit Disorder Without Hyperactivity: A Distinct Behavioural and Neurocognitive Syndrome. *Journal of Child Neurology* 6:35-41

Hynd GW, Semrud-Clikeman M, Lorys AR, Novey ES (1990) Brain Morphology in Developmental Dyslexia and Attention Deficit Disorder/Hyperactivity. *Archives of Neurology* 47:919-926

Ioannidis JPA, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG (2001) Replication Validity of Genetic Association Studies. *Nature Genetics* 29:306-309

Ivanov DK, Moskvina V, Owen MJ, Kirov G Design of Optimal Primers for SNP Genotyping Assays Based on Primer Extension.

Jaskowski P, Rusiak P (2005) Posterior Parietal Cortex and Developmental Dyslexia. *Acta Neurobiologiae Experimentalis (Wars)* 65:79-94

Jen Y, Manova K, Benezra R (1996) Expression Patterns of ID1, ID2 and ID3 are Highly Related but Distinct From that of ID4 During Mouse Embryogenesis. *Developmental Dynamics* 207:235-52

Jobard G, Crivello F, Tzourio-Mazoyer N (2003) Evaluation of the Dual Route Theory of Reading: A Meta Analysis of 35 Neuroimaging Studies. *NeuroImage* 20:693-712

Joseph J, Noble K, Eden G (2001) The Neurobiological Basis of Reading. *Journal of Learning Disabilities* 34:566-579

Juel C, Griffith P, Gough P (1986) Acquisition of Literacy: A Longitudinal Study of Children in First and Second Grade. *Journal of Educational Psychology* 78

Julier C, Hyer RN, Davies J (1991) Insulin-IGF2 Region on Chromosome 11p Encodes a Gene Implicated in HLA-DR4-dependant Diabetes Susceptibility. *Nature* 354:155-159

Kadesjo B, Gillberg C (1999) Developmental Coordination Disorder in Swedish 7-Year-Old Children. *Journal of the American Academy of Child and Adolescent Psychiatry* 38:820-828

Kaminen N, Hannula-Jouppi K, Kestilä M, Lahermo P, Muller K, Kaaranen M, Myllyluoma B, Voutilainen A, Lyytinen H, Nopola-Hemmi J, Kere J (2003) A Genome Scan for Developmental Dyslexia Confirms Linkage to Chromosome 2p11 and Suggests a New Locus on 7q32. *Journal of Medical Genetics* 40(5):340-5

Kaplan B, Dewey D, Crawford S, Wilson B (2001) The Term Comorbidity is of Questionable Value in Reference to Developmental Disorders: Data and Theory. *Journal of Learning Disabilities* 34:555-565

Kaplan DE, Gayán J, Ahn J, Won T-W, Pauls D, Olson RK, DeFries JC, Wood F, Pennington BF, Page GP, Smith SD, Gruen JR (2002) Evidence for Linkage and Association with Reading Disability, on 6p21.3-22. *American Journal of Human Genetics* 70:1287-1298

Keogh BK, Margolis JS (1976) A Component Analysis of Attentional Problems of Educationally Handicapped Boys. *Journal of Abnormal Child Psychology* 4:349-359

Kerr J (1897) School Hygiene, in its Mental, Moral and Physical Aspects. *Journal of The Royal Statistical Society* 60:613-680

Kidd KK (1993) Associations of Disease with Genetic Markers: Déjà Vu all over Again. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 48:71-73

Klasen (1968) Legasthenia. Huber, Bern

Klingberg T, Hedehus M, Temple E, Salz T, Gabrieli JD, Moseley ME, Poldrack RA (2000) Microstructure of Temporoparietal White Matter as a Basis for Reading Ability: Evidence from Diffusion Tensor Magnetic Resonance Imaging. *Neuron* 25:493-500

Knopik VS, Alarcon M, DeFries JC (1997) Comorbidity of Mathematics and Reading Deficits: Evidence for a Genetic Etiology. *Behavior Genetics* 27:447-453

Knopik VS, Smith SD, Cardon L, Pennington BF, Gayan J, Olson RK, DeFries JC (2002) Differential Genetic Etiology of Reading Component Processes as a Function of IQ. *Behavior Genetics* 32:181-198

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K (2002) A High-Resolution Recombination Map of the Human Genome. *Nature Genetics* 31:241-247

Korhonen TT (1995) The Persistence of Rapid Naming Problems in Children with Learning Disabilities: A Nine-Year Follow-Up. *Journal of Learning Disabilities* 28:232-239

Kovel CGFD, Hol FA, Heister JGAM, Willemen JJHT, Sandkuijl LA, Franke B, Padberg GW (2004) Genomewide Scan Identifies Susceptibility Locus for Dyslexia on Xq27 in an Extended Dutch Family. *Journal of Medical Genetics* 41:652-657

Kronbichler M, Hutzler F, Wimmer H (2002) Dyslexia: Verbal Impairments in the Absence of Magnocellular Impairments. *NeuroReport* 13:617-620

Kruglyak L (1999) Prospects for Whole-Genome Linkage Disequilibrium Mapping of Common Disease Genes. *Nature Genetics* 22:139-144

Kujala T, Myllyviita K, Tervaniemi M, Alho K, Kallio J, Näätänen R (2000) Basic Auditory Dysfunction in Dyslexia as Demonstrated by Brain Activity Measurements. *Psychophysiology* 37:262-266

Kushch A, Gross-Glenn K, Jallad B, Lubs H, Rabin M, Feldman E, Duara R (1993) Temporal Lobe Surface Area Measurements on MRI in Normal and Dyslexic Readers. *Neuropsychologia* 31:811-821

Laan M, Pääbo S (1997) Demographic History and Linkage Disequilibrium in Human Populations. *Nature Genetics* 17:435-438

LaBuda MC, DeFries JC (1988) Genetic and Environmental Etiologies of Reading Disability: A Twin Study. *Annals of Dyslexia* 38:131-138

LaBuda MC, DeFries JC (1990) Genetic Etiology of Reading Disability: Evidence from a Twin Study. In: Pavlidis G (ed) Perspectives on Dyslexia. John Wiley and Sons, Chichester, UK, pp 47-76

LaBuda MC, DeFries JC, Pennington BF (1990) Reading Disability: A Model for the Genetic Analysis of Complex Behavioural Disorders. *Journal of Counselling and Development* 68:645-651

Lachmann T, Berti S, Kujala T, Schröger E (2005) Diagnostic Subgroups of Developmental Dyslexia Have Different Deficits in Neural Processing of Tones and Phonemes. *International Journal of Psychophysiology* 56:105-120

Lai CS, Gerrelli D, Monaco AP, Fisher SE, Copp AJ (2003) FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain* 126:2455-62

Lai CSL, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001) A Forkhead-Domain Gene is Mutated in Severe Speech and Language Disorder. *Nature* 413:519-523

Lander E, Kruglyak L (1995) Genetic Dissection of Complex Traits: Guidelines for Interpreting and Reporting Linkage Results. *Nature Genetics* 11:241-247

Larson JP, Hoien T, Odegaard H (1992) Magnetic Resonance Imaging of the Corpus Callosum in Developmental Dyslexia. *Cognitive Neuropsychology* 9:123-134

Lee C-J, Appleby VJ, Orme AT, Chan W-I, Scotting PJ (2002) Differential Expression of *SOX4* and *SOX11* in Medullablastoma. *Journal of Neuro-Oncology* 57:201-214

Leonard CM, Eckert MA, Lombardino LJ, Oakland T, Kranzler J, Mohr CM, King WM, Freeman A (2001) Anatomical Risk Factors for Phonological Dyslexia. *Cerebral Cortex* 11:148-157

Lerner JW (1989) Educational Interventions in Learning Disabilities. Journal of American Academy of Child and Adolescent Psychiatry 28:326-331

Lewis C, Hitch G, Walker P (1994) The Prevalence of Specific Arithmetic Difficulties and Specific Reading Difficulties in 9- to 10-Year-Old Boys and Girls. *Journal of Child Psychology and Psychiatry* 35:283-292

Lewitter FI, DeFries JC, Elston RC (1980) Genetic Models of Reading Disability. *Behavior Genetics* 10:9-30

Liberman IY, Shankweiler D, Fischer FW, Carter B (1974) Explicit Syllable and Phoneme Segmentation in the Young Child. *Journal of Experimental Child Psychology* 18:201-212

Light JG, DeFries JC (1995) Comorbidity of Reading and Mathematics Disabilities: Genetic and Environmental Etiologies. *Journal of Learning Disabilities* 28:96-106

Liu Q, Thorland EC, Sommer SS (1997) Inhibition of PCR Amplification by a Point Mutation Downstream of a Primer. *Biotechniques* 22:292-296

Livingstone MS, Rosen GD, Drislane FW, Galaburda AM (1991) Physiological and Anatomical Evidence for a Magnocellular Defect in Developmental Dyslexia. *Proceedings of the National Academy of Sciences USA* 88:7943-7947

Londin ER, Meng H, Gruen JR (2003) A Transcription Map of the 6p22.3 Reading Disability Locus Identifying Candidate Genes. *BMC Genomics* 4(1):25

Loo SK, Fisher SE, Francks C, Ogdie MN, MacPhie IL, Yang M, McCracken JT, McGough JJ, Nelson SF, Monaco AP, Smalley SL (2004) Genome-Wide Scan of Reading Ability in Affected Sibling Pairs with Attention-Deficit/Hyperactivity Disorder: Unique and Shared Genetic Effects. *Molecular Psychiatry* 9:485-493

Lorenzi C, Dumont A, Füllgrabe C (2000) Use of Temporal Envelope Cues by Children with Developmental Dyslexia. *Journal of Speech, Language and Hearing Research* 43:1367-1379

Lovegrove WJ, Bowling A, Badcock D, Blackwood M (1980) Specific Reading Disability: Differences in Contrast Sensitivity as a Function of Spatial Frequency. *Science* 210:439-440

Lubs HA, Duara R, Levin B, Jallad B, Lubs M-L, Rabin M, Kushch A, Gross-Glenn K (1991a) Dyslexia Subtypes: Genetics, Behavior and Brain Imaging. In: Duane DD, Gray DB (eds) The Reading Brain. The Biological Basis of Dyslexia. York Press Inc, Parkton, Maryland, USA, pp 89-117

Lubs HA, Rabin M, Carland-Saucier K, Wen XL, Gross-Glenn K, Duara R, Levin B, Lubs ML (1991b) Genetic Bases of Developmental Dyslexia: Molecular Studies. In: Neuropsychological Foundations of Learning Disabilities. Academic Press Inc, London, pp 49-77

Lubs HA, Rabin M, Feldman E, Jallad BJ, Kushch A, Gross-Glenn K (1993) Familial Dyslexia: Genetic and Medical Findings in Eleven Three-Generation Families. *Annals of Dyslexia* 43:44-60

Luria AR (1973) The Working Brain. Basic Books, New York

Lykken DT, Tellegen A, DeRubeis R (1978) Volunteer Bias in Twin Research: The Rule of Two-Thirds. *Social Biology* 25:1-9

Lyon GR, Rumsey JM (1996) Neuroimaging: A Window to the Neurological Foundations of Learning and Behaviour in Children. Paul H Brooks Publishing, Baltimore, USA

Maher BS, Marazita ML, Ferrell RE, Vanyukov MM (2002) Dopamine System Genes and Attention Deficit Hyperactivity Disorder: A Meta-Analysis. *Psychiatric Genetics* 12:207-215

Manis FR, McBride-Chang C, Seidenberg MS, Keating P, Doi LM, Munson B, Petersen A (1997) Are Speech Perception Deficits Associated with Developmental Dyslexia? *Journal of Experimental Child Psychology* 66:211-235

Manis FR, Seidenberg MS, Doi LM, McBride-Chang C, Petersen A (1996) On The Bases of Two Subtypes of Developmental Dyslexia. *Cognition* 58:157-195

Manis FR, Szeszulski PA, Holt LK, Graves K (1990) Variation in Component Word Recognition and Spelling Skills Among Dyslexic Children and Normal Readers. In: Carr T, Levy BA (eds) Reading and its Development: Component Skills Approaches. Academic Press, New York, pp 207-259

Mann VA (1984) Review: Reading Skill and Language Skill. *Developmental Review* 4:1-15

Marino C, Giorda R, Lorusso ML, Vanzin L, Salandi N, Nobile M, Citterio A, Beri S, Crespi V, Battaglia M, Molteni M (2005) A Family-Based Association Study Does Not Support DYX1C1 on 15q21.3 as a Candidate Gene in Developmental Dyslexia. *European Journal of Human Genetics*: 1-9

Marlow AJ, Fisher SE, Richardson AJ, Francks C, Talcott JB, Monaco AP, Stein JF, Cardon LR (2001) Investigation of Quantitative Measures Related to Reading Disability in a Large Sample of Sib-Pairs from the UK. *Behavior Genetics* 31:219-230

Marsh G, Friedman MP, Welch V, Desberg P (1981) A Cognitive-Developmental Theory of Reading Acquisition. In: Waler TG, MacKinnon GE (eds) Reading Research: Advances in Theory and Practice. Academic Press, London, pp 87-139

Marshall CM, Snowling MJ, Bailey PJ (2001) Rapid Auditory Processing and Phonological Ability in Normal Readers and Readers with Dyslexia. *Journal of Speech, Language and Hearing Research* 44:925-940

Martin RH (1999) Sperm Chromosome Analysis in a Man Heterozygous for a Paracentric Inversion of Chromosome 14 (q24.1q32.1. *American Journal of Human Genetics* 64:1480-1484

McAnally KI, Stein JF (1996) Auditory Temporal Coding in Dyslexia. *Proceedings of The Royal Society, Biological Sciences* 263:961-965

McCrory E, Frith U, Brunswick N, Price C (2000) Abnormal Functional Activation During a Simple Word Repetition Task: A PET Study of Adult Dyslexics. *Journal of Cognitive Neuroscience* 12:753-762

Meffert MK, Chang JM, Wiltgen BJ, Fanselow MS, Baltimore D (2003) NF-kB Functions in Synaptic Signaling and Behavior. *Nature Neuroscience* 6:1072-1078

Meng H, Smith SD, Hager K, Held M, Liu J, Olson RK, Pennington BF, DeFries JC, Gelernter J, O'Reilly-Pol T, Somlo S, Skudlarski P, Shaywitz SE, Shaywitz BA, Marchione K, Wang Y, Paramasivam M, LoTurco JJ, Page GP, Gruen JR (2005) DCDC2 is Associated with Reading Disability and Modulates Neuronal Development in the Brain. *Proceedings of the National Academy of Sciences* 102:17053-17058

Meyer MS, Wood FB, Hart LA, Felton RH (1998) The Selective Predictive Value of Rapid Automatized Naming in Poor Readers. *Journal of Reading Disabilities* 31:106-117

Miles TR, Miles E (1999) Dyslexia A Hundred Years On. Open University Press, Buckingham, UK

Mody M, Studdert-Kennedy M, Brady S (1997) Speech Perception Deficits in Poor Readers: Auditory Processing or Phonological Coding? *Journal of Experimental Child Psychology* 64:199-231

Molenaar PCM, Boomsma DI, Dolan CV (1993) A Third Source of Developmental Differences. *Behaviour Genetics* 23:519-531

Morgan WP (1896) A Case of Congenital Word-Blindness (Inability to Learn to Read). *British Medical Journal* 2:1543-1544

Morris DW, Ivanov D, Robinson L, Williams N, Stevenson J, Owen MJ, Williams J, O'Donovan MC (2004) Association Analysis of Two Candidate Phospholipase Genes that Map to the Chromosome 15q15.1-15.3 Region Associated with Reading Disability. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 129:97-103

Morris DW, Robinson L, Turic D, Duke M, Webb V, Milham C, Hopkin E, Pound K, Fernando S, Easton M, Hamshere M, Williams N, McGuffin P, Stevenson J, Krawczak M, Owen MJ, O'Donovan MC, Williams J (2000) Family-Based Association Mapping Provides Evidence for a Gene for Reading Disability on Chromosome 15q. *Human Molecular Genetics* 9:843-848

Morton J, Patterson KE (1980) A New Attempt at an Interpretation or an Attempt at a New Interpretation. In: Coltheart M, Patterson KE, Marshall JC (eds) Deep Dyslexia. Routledge and Kegan Paul, London, pp 91-118

Morton NE (1955) Sequential Test for the Detection of Linkage. *American Journal of Human Genetics* 7:277-318

Moskvina V, Norton N, Williams N, Holmans P, Owen M, O'Donovan M (2005) Streamlined Analysis of Pooled Genotype Data in SNP-Based Association Studies. *Genetic Epidemiology* 28:273-282

Muter V, Hulme C, Snowling M, Stevenson J (2004) Phonemes, Rimes, Vocabulary and Grammatical Skills as Foundations of Early Reading Development: Evidence From a Longitudinal Study. *Developmental Psychology* 40:665-681

Nagarajan S, Mahncke H, Salz T, Tallal P, Roberts T, Merzenich MM (1999) Cortical Auditory Signal Processing in Poor Readers. *Proceedings of the National Academy of Sciences USA* 96:6483-6488

Naiden N (1976) Ratio of Boys to Girls among Disabled Readers. *The Reading Teacher* 29(5)439-442

Naidoo S (1972) Specific Dyslexia. Pitman, London

Nakayama M, Kikuno R, Ohara O (2002) Protein-Protein Interactions Between Large Proteins: Two-Hybrid Screening Using a Functionally Classified Library Composed of Long cDNAs. *Genome Research* 12:1773-1784

Nass R (1993) Sex Differences in Learning Abilities and Disabilities. *Annals of Dyslexia* 43:61-77

Neale MD (1989) Analysis of Reading Ability. In: Revised British Edition. NFER Nelson Publishing Co. Ltd., Windsor

Niall HD, Hogan ML, Sauer R, Rosenblum IY, Greenwood FC (1971) Sequences of Pituitary and Placental Lactogenic and Growth Hormones: Evolution from a Primordial Peptide by Gene Duplication. *Proceedings of the National Academy of Sciences USA* 68:866-870

Nicholson RI, Fawcett AJ (1990) Automaticity: A New Framework for Dyslexia Research? *Cognition* 35:159-182

Nicholson RI, Fawcett AJ, Berry EL, Jenkins IH, Dean P, Brooks DJ (1999) Association of Abnormal Cerebellar Activation with Motor Learning Difficulties in Dyslexic Adults. *Lancet* 353:1662-1667

Nicholson RI, Fawcett AJ, Dean P (1995) Time Estimation Deficits in Developmental Dyslexia: Evidence of Cerebellar Involvement. *Proceedings of The Royal Society London B Biological Sciences* 259:43-47

Nittrouer S (1999) Do Temporal Processing Deficits Cause Phonological Processing Problems? Journal of Speech, *Language and Hearing Research* 42:925-942

Njiokiktjien C, Sonneville LD, Vaal J (1994) Callosal Size in Children with Learning Disabilities. *Behavioural Brain Research* 64:213-218

Nopola-Hemmi J, Myllyluoma B, Haltia T, Taipale M, Ollikainen V, Ahonen T, Voutilainen A, Kere J, Widen E (2001) A Dominant Gene for Developmental Dyslexia on Chromosome 3. *Journal of Medical Genetics* 38:658-664

Nopola-Hemmi J, Taipale M, Haltia T, Lehesjoki A-E, Voutilainen A, Kere J (2000) Two Translocations of Chromosome 15q Associated with Dyslexia. *Journal of Medical Genetics* 37:771-775

Nordberg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D (2002) The Extent of Linkage Disequilibrium in *Arabidopsis Thaliana*. *Nature Genetics* 30:190-193

Norton N, Williams NM, Williams HJ, Spurlock G, Kirov G, Morris DW, Hoogendoorn B, Owen MJ, O'Donovan MC (2002) Universal, Robust, Highly Quantitative SNP Allele Frequency Measurement in DNA Pools. *Human Genetics* 110:471-478

Nöthen MM, Schulte-Körne G, Grimm T, Cichon S, Vogt IR, Müller-Myhsok B, Propping P, Remschmidt H (1999) Genetic Linkage Analysis with Dyslexia: Evidence for Linkage of Spelling Disability to Chromosome 15. *European Child and Adolescent Psychiatry* 8:III/56-III/59

Odell JD, Warren RP, Warren WL, Burger RA, Maciulis A (1997) Association of Genes Within the Major Histocompatibility Complex with Attention-Deficit Hyperactivity Disorder. *Neuropsychobiology* 35:181-186

Olson R, Wise B, Conners F, Rack J, Fulker D (1989) Specific Deficits in Component Reading and Language Skills: Genetic and Environmental Influences. *Journal of Learning Disabilities* 22:339-348

Olson RK, Datta H, Gayan J, DeFries JC (1999) A Behavioural Genetic Analysis of Reading Disabilities and Component Processes. In: Klein RM, McMullen P (eds) Converging Methods for Understanding Reading and Dyslexia. MIT Press, Boston, Massachusetts

Olson RK, Forsberg H, Wise B (1994) Genes, Environment and the Development of Orthographic Skills. In: Berninger VW (ed) The Varieties of Orthographic Knowledge. 1: Theoretical and Developmental Issues. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 1-31

Olson RK, Rack JP, Conners FA, DeFries JC, Fulker DW (1991) Genetic Etiology of Individual Differences in Reading Disability. In: Feagans L, Short E, Meltzer L (eds) Subtypes of Reading Disabilities. Lawrence Erlbaum, Hillsdale, New Jersey, USA, pp 113-135

Olson RK, Wise B, Conners F, Rack J (1990) Organization, Heritability and Remediation of Component Word Recognition and Language Skills in Disabled Readers. In: Carr TH, Levy BA (eds) Reading and its Development: Component Skills Approaches. Academic Press, New York, pp 261-322

Omenn GS, Weber BA (1978) Dyslexia: Search for Phenotypic and Genetic Heterogeneity. *American Journal of Medical Genetics* 1:333

Orton ST (1930) Familial Occurrence of Disorders in the Acquisition of Language. *Eugenics* 3:140-147

Orton ST (1937 / 1989) Reading, Writing and Speech Problems in Children and Selected Papers. Pro-Ed, Austin, Texas, USA

Orton ST (1937) Reading, Writing and Speech Problems in Children. Norton, New York

Owen F, Adams P, Forrest T, Stolz L, Fisher S (1971) Learning Disorders in Children: Sibling Studies. *Monographs of the Society for Research in Child Development* 36:1-77

Paracchini S (2005). In: Cope N (ed), Oxford, UK

Parsch J, Meiklejohn CD, Hartl DL (2001) Patterns of DNA Sequence Variation Suggest the Recent Action of Positive Selection in the Janus-Ocnus Region of *Drosophila Simulans*. *Genetics* 159:647-657

Paulesu E, Démonet J-F, Fazio F, McCrory E, Chanoine V, Brunswick N, Cappa SF, Cossu G, Habib M, Frith CD, Frith U (2001) Dyslexia: Cultural Diversity and Biological Unity. *Science* 291:2165-2167

Paulesu E, Frith U, Snowling M, Gallagher A, Morton J, Frackowiak RSJ, Frith CD (1996) Is Developmental Dyslexia a Disconnection Syndrome? Evidence from PET Scanning. *Brain* 119:143-157

Pennington B, Orden GCv, Smith SD, Green PA, Haith MM (1990) Phonological Processing Skills and Deficits in Adult Dyslexics. *Child Development* 61:1753-1778

Pennington BF (1989) Using Genetics to Understand Dyslexia. *Annals of Dyslexia* 39:81-93

Pennington BF (1990) Annotation: The Genetics of Dyslexia. *Journal of Child Psychology and Psychiatry* 31:193-201

Pennington BF (1997) Using Genetics to Dissect Cognition. *American Journal of Human Genetics* 60:13-16

Pennington BF (1999) Toward an Integrated Understanding of Dyslexia: Genetic, Neurological and Cognitive Mechanisms. *Development and Psychopathology* 11:629-654

Pennington BF, Filipek PA, Lefly D, Churchwell J, Kennedy DN, Simon JH, Filley CM, Galaburda A, Alarcon M, DeFries JC (1999) Brain Morphometry in Reading-Disabled Twins. *Neurology* 53:723-729

Pennington BF, Gilger JW (1996) How is Dyslexia Transmitted? In: Chase CH, Rosen GD, Sherman GF (eds) Developmental Dyslexia. Neural, Cognitive and Genetic Mechanisms. York Press, Baltimore, Maryland, pp 41-61

Pennington BF, Gilger JW, Pauls D, Smith SA, Smith SD, DeFries JC (1991) Evidence for Major Gene Transmission of Developmental Dyslexia. *Journal of the American Medical Association* 266:1527-1534

Perfetti CA (1985) Reading Ability. Oxford Press, New York

Perfetti CA (1992) The Representation Problem in Reading Acquisition. In: Gough PB, Ehri LC, Treiman R (eds) Reading Acquisition. Erlbaum, Hillsdale, NJ, USA, pp 145-174

Perfetti CA (1994) Psycholinguistics and Reading Ability. In: Gernsbacher M (ed) Handbook of Psycholinguistics. Academic Press, New York, pp 849-894

Perfetti CA, Bell L (1991) Phonemic Activation During the First 40ms of Word Identification: Evidence from Backward Masking and Masked Priming. *Journal of Memory and Language* 30:473-485

Petryshen TL, Kaplan BJ, Hughes ML, Tzenova J, Field LL (2002) Supportive Evidence for the DYX3 Dyslexia Susceptibility Gene in Canadian Families. *Journal of Medical Genetics* 39:125-126

Petryshen TL, Kaplan BJ, Liu MF, Field LL (2000) Absence of Significant Linkage Between Phonological Coding Dyslexia and Chromosome 6p23-21.3, as Determined by Use of Quantitative-Trait Methods: Confirmation of Qualitative Analyses. *American Journal of Human Genetics* 66:708-714

Petryshen TL, Kaplan BJ, Liu MF, French NSD, Tobias R, Hughes ML, Field LL (2001) Evidence for a Susceptibility Locus on Chromosome 6q Influencing Phonological Coding Dyslexia. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 105:507-517

Petryshen TL, Kaplan BL, Hughes ML, Field LL (2000b) Evidence for the Chromosome 2p15-p16 Dyslexia Susceptibility Locus (DYX3) in a Large Canadian Data Set. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 96:472

Peyrard-Janvid M, Anthoni H, Onkamo P, Lahermo P, Zucchelli M, Kaminen N, Hannula-Jouppi K, Nopola-Hemmi J, Voutilainen A, Lyytinen H, Kere J (2004) Fine Mapping of the 2p11 Dyslexia Locus and Exclusion of TACR1 as a Candidate Gene. *Human Genetics* 114:510-516

Plaut D, McClelland JL, Seidenberg MS, Patterson KE (1996) Understanding Normal and Impaired Reading: Computational Principles in Quasi-Irregular Domains. *Psychological Review* 103:56-115

Plomin R, Kovas Y (2005) Generalist Genes and Learning Disabilities. *Psychological Bulletin* 131:592-617

Plomin R, Owen MJ, McGuffin P (1994) The Genetic Basis of Complex Human Behaviours. *Science* 264:1733-1739

Plomin R, Turic DM, Hill L, Turic DE, Stephens M, Williams J, Owen MJ, O'Donovan MC (2004) A Functional Polymorphism in the Succinate-Semialdehyde Dehydrogenase (Aldehyde Dehydrogenase 5 Family, Member A1) Gene is Associated with Cognitive Ability. *Molecular Psychiatry* 9:582-586

Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JD (1999) Functional Specialization for Semantic and Phonological Processing in the Left Inferior Prefrontal Cortex. *Neuroimage* 10:15-35

Posthuma D, Luciano M, Geus EJCD, Wright MJ, Slagboom PE, Montgomery GW, Boomsa DI, Martin NG (2005) A Genomewide Scan for Intelligence Identifies Quantitative Trait Loci on 2q and 6p. *American Journal of Human Genetics* 77:318-326

Pritchard JK, Przeworski M (2001) Linkage Disequilibrium in Humans: Models and Data. *American Journal of Human Genetics* 69:1-14

Pugh KR, Mencl WE, Shaywitz BA, Shaywitz SE, Fulbright RK, Constable RT, Skudlarski P, Marchione KE, Jenner AR, Fletcher JM, Liberman AM, Shankweiler DP, Katz L, Lacadie C, Gore JC (2000) The Angular Gyrus in Developmental Dyslexia: Task-Specific Differences in Functional Connectivity within Posterior Cortex. *Psychological Science* 11:51-56

Purvis KL, Tannock R (1997) Language Abilities in Children with Attention Deficit Hyperactivity Disorder, Reading Disabilities and Normal Controls. *Journal of Abnormal Child Psychology* 25:133-144

Pype S, Declercq W, Ibrahimi A, Michiels C, Rietschoten JGIV, Dewulf N, Boer MD, Van den Abeele P, Huylebroeck D, Remacle JE (2000) TTRAP, a Novel Protein that Associates with CD40, Tumor Necrosis Factor (TNF) Receptor-75 and TNF Receptor-Associated Factors (TRAFs) and the Inhibits Nuclear Factor-kB Activation. *Journal of Biological Chemistry* 275:18586-18593

Quintana PJE, Neuwirth EAH, Grosovsky AJ (2001) Interchromosomal Gene Conversion at an Endogenous Human Cell Locus. *Genetics* 158:757-767

Rabin M, Wen XL, Hepburn M, Lubs HA (1993) Suggestive Linkage of Developmental Dyslexia to Chromosome 1p34-p36. *The Lancet* 342:178

Rae C, Lee M, Dixon RM, Blamire AM, Thompson CH, Styles P, Talcott J, Richardson AJ, Stein JF (1998a) Metabolic Abnormalities in Developmental Dyslexia Detected by 1H Magnetic Resonance Spectroscopy. *The Lancet* 351:1849-1852

Ramus F (2003) Developmental Dyslexia: Specific Phonological Deficit or General Sensorimotor Dysfunction? *Current Opinion in Neurobiology* 13:212-218

Ramus F, Pidgeon E, Frith U (2003a) The Relationship Between Motor Control and Phonology in Dyslexic Children. *Journal of Child Psychology and Psychiatry* 44(5):712-22

Ramus F, Rosen S, Dakin SC, Day BL, Castellote JM, White S, Frith U (2003b) Theories of Developmental Dyslexia: Insights from a Multiple Case Study of Dyslexic Adults. *Brain* 126:841-865

Raskind WH, Hsu L, Berninger VW, Thomson JB, Wijsman EM (2000) Familial Aggregation of Dyslexia Phenotypes. *Behavior Genetics* 30:385-396

Raskind WH, Igo RP, Chapman NH, Berninger VW, Thomson JB, Matsushita M, Brkanac Z, Holzman T, Brown M, Wijsman EM (2005) A Genome Scan in Multigenerational Families with Dyslexia: Identification of a Novel Locus on Chromosome 2q That Contributes to Phonological Decoding Efficiency. *Molecular Psychiatry* 10:699-711

Raymer AM, Moberg P, Crosson B, Nadeau S, Rothi LJ (1997) Lexical-Semantic Deficits in Two Patients with Dominant Thalamic Infarction. *Neuropsychologia* 35:211-219

Recker RR, Deng HW (2002) Role of Genetics in Osteoporosis. *Endocrine* 17:55-66

Reed MA (1989) Speech Perception and the Discrimination of Brief Auditory Cues in Reading Disabled Children. *Journal of Experimental Child Psychology* 48:270-292

Reppe S, Rian E, Jemtland R, Olstad OK, Gautvik VT, Gautvik KM (2000) SOX-4 Messenger RNA is Expressed in the Embryonic Growth Plate and Regulated via the Parathyroid Hormone/Parathyroid Hormone-Related Protein Receptor in Osteoblast-Like Cells. *Journal of Bone and Mineral Research* 15:2402-2412

Richards TL, Berninger VW, Aylward EH, Richards AL, Thomson JB, Nagy WE, Carlisle JF, Dager SR, Abbott RD (2002) Reproducibility of Proton MR Spectroscopic Imaging (PEPSI): Comparison of Dyslexic and Normal-Reading Children and Effects of Treatment on Brain Lactate Levels During Language Tasks. *American Journal of Neuroradiology* 23:1678-1685

Richards TL, Corina D, Serafini S, Steury K, Echelard DR, Dager SR, Marro K, Abbott RD, Maravilla KR, Berninger VW (2000) Effects of a Phonologically Driven Treatment for Dyslexia on Lactate Levels Measured by Proton MR Spectroscopic Imaging. *American Journal of Neuroradiology* 21:916-922

Richards TL, Dager SR, Corina D, Serafini S, Heide AC, Steury K, Strauss W, Hayes CE, Abbott RD, Craft S, Shaw D, Posse S, Berninger VW (1999) Dyslexic Children Have Abnormal Brain Lactate Response to Reading-Related Language Tasks. *American Journal of Neuroradiology* 20:1393-398

Richardson AJ (1994) Dyslexia, Handedness and Syndromes of Psychosis Proneness. *International Journal of Psychophysiology* 18:251-263

Richardson AJ, Ross MA (2000) Fatty Acid Metabolism in Neurodevelopment Disorder: A New Perspective on Associations Between Attention-Deficit/Hyperactivity Disorder, Dyslexia, Dyspraxia and the Autistic Spectrum. *Prostaglandins, Leukotrienes and Essential Fatty Acids* 63:1-9

Riechmann V, Sablitzky F (1995) Mutually Exclusive Expression of Two Dominant-Negative Helix-Loop-Helix (dnHLH) Genes, Id4 and Id3, in the Developing Brain of the Mouse Suggests Distinct Regulatory Roles of these dnHLH Proteins During Cellular Proliferation and Differentiation of the Nervous System. *Cell Growth Differentiation* 6:837-843

Rigolet M, Rich T, Gross-Morand MS, Molina-Gomes D, Wiegas-Pequignot E, Junien C (1998) cDNA Cloning, Tissue Distribution and Chromosomal Localisation of the Human ID4 Gene. *DNA Research* 5:309-313

Risch N, Merikangas K (1996) The Future of Genetic Studies of Complex Human Diseases. *Science* 273:1516-1517

Roberts PA (1976) The Genetics of Chromosome Aberration. In: Ashburner M, Novitski E (eds) The Genetics and Biology of *Drosophila*. Academic Press, London, pp 67-184

Robinson L (2001) A Phenotypic Analysis of Developmental Dyslexia for Molecular Genetic Research. PhD, University of Wales College of Medicine, Cardiff

Rodgers B (1983) The Identification and Prevalence of Specific Reading Retardation. *British Journal of Educational Psychology* 53:369-373

Rodrigues-Lima F, Josephs M, Katan M, Cassinat B (2001) Sequence Analysis Identifies TTRAP, a Protein that Associates with CD40 and TNF Receptor-Associated Factors, as a Member of a Superfamily of Divalent Cation-Dependent Phosphodiesterases. *Biochemical and Biophysical Research Communications* 285:1274-1279

Rosen GD, Waters NS, Galaburda AM, Denenberg VH (1995) Behavioral Consequences of Neonatal Injury of the Neocortex. *Brain Research* 681:177-189

Rosen S, Manganari E (2001) Is There a Relationship Between Speech and Nonspeech Auditory Processing in Children with Dyslexia? *Journal of Speech, Language and Hearing Research* 44:720-736

Rosenberg MS, Wilson RJ, Legenhausen E (1989) The Assessment of Hyperactivity in Preschool Populations: A Multidisciplinary Approach. *Topics in Early Childhood Special Education* 9:90-105

Ruff S, Cardebat D, Marie N, Demonet J (2002) Enhanced Response of the Left Frontal Cortex to Slowed Down Speech in Dyslexia: An fMRI Study. *NeuroReport* 13:1285-1289

Rumsey JM, Dorwart R, Vermess M, Denckla MB, Kruesi MJP, Rapoport JL (1986) Magnetic Resonance Imaging of Brain Anatomy in Severe Developmental Dyslexia. *Archives of Neurology* 43:1045-1046

Rumsey JM, Nace K, Donohue B, Wise D, Maisong JM, Andreason PA (1997) Positron Emission Tomographic Study of Impaired Word Recognition and Phonological Processing in Dyslexic Men. *Archives of Neurology* 54:562-573

Rumsey LM, Andreason P, Zametkin J, Aquino T, King AC, Hamburger SD, Pikus A, Rapoport JL, Cohen RM (1992) Failure to Activate Left Temporoparietal Cortex in Dyslexia: A 15-O PET Study. *Archives of Neurology* 49:527-534

Rumsey LM, Zametkin J, Andreason P, Hanahan AP, Hamburger SD, Aquino T, King AC, Pikus A, Cohen RM (1994) Normal Activation of Frontotemporal Language Cortex in Dyslexia, as Measured with Oxygen 15 Positron Emission Tomography. *Archives of Neurology* 51:27-38

Rutter M, Tizard J, Whitmore K (1970) Education, Health and Behaviour: Psychological and Medical Study of Childhood Development. Wiley, New York, USA

Sablitzky F, Moore A, Bromley M, Deed RW, Newton JS, Norton JD (1998) Stage- and Subcellular-Specific Expression of ID Proteins in Male Germ and Sertoli Cells Implicates Distinctive Regulatory Roles for ID Proteins During Meiosis, Spermatogenesis and Sertoli Cells Function. *Cell Growth and Differentiation* 9:1015-1024

Sanger F, Nicklen S, Coulson AR (1977) DNA Sequencing with Chain-Terminating Inhibitors. *Proceedings of the National Academy of Sciences USA* 74:5463

Sauver JLS, Katusic SK, Barbaresi WJ, Colligan RC, Jacobsen SJ (2001) Boy/Girl Differences in Risk for Reading Disability: Potential Clues? *American Journal of Epidemiology* 154:787-794

Savin HB (1972) What the Child Knows About Speech When He Starts to Learn to Read. In: Kavanagh JF, Mattingly IG (eds) Language by Ear and by Eye. MIT Press, Cambridge, MA, USA, pp 319-326

Scarborough HS (1989) Prediction of Reading Disability from Familial and Individual Differences. *Journal of Educational Psychology* 81:101-108

Scerri TS, Fisher SE, Francks C, MacPhie IL, Paracchini S, Richardson AJ, Stein JF, Monaco AP (2004) Putative Functional Alleles of *DYX1C1* are not Associated with Dyslexia Susceptibility in a Large Sample of Sibling Pairs From the UK. *Journal of Medical Genetics* 41:11

Schlaug G, Jäncke L, Huang Y, Steinmetz H (1995) In Vivo Evidence of Structural Brain Asymmetry in Musicians. *Science* 267:699-701

Schmidt U, Maue I, Lehmann K, Belcher SM, Stahl U, Perlman PS (1998) Mutant Alleles of the MRS2 Gene of Yeast Nuclear DNA Suppress Mutations in the Catalytic Core of a Mitochondrial Group II Intron. *Journal of Molecular Biology* 282:525-541

Schneider W, Graham DJ (1992) Introduction to Connectionist Modelling in Education. *Educational Psychologist* 27:513-530

Schonhaut S, Satz P (1983) Prognosis for Children With Learning Disabilities: A Review of Follow-Up Studies. In: Rutter M (ed) Developmental Psychiatry. The Guildford Press, New York, pp 542-563

Schrott LM, Denenberg VH, Sherman GF, Waters NS, Rosen GD, Galaburda AM (1992) Environmental Enrichment, Neocortical Ectopias and Behavior in the Autoimmune NZB Mouse. *Brain Research* 67:85-93

Schulte-Körne G, Deimel W, Bartling J, Remschmidt H (1998a) Auditory Processing and Dyslexia: Evidence for a Specific Speech Processing Deficit. *NeuroReport* 9:337-340

Schulte-Körne G, Deimel W, Bartling J, Remschmidt H (1998b) Role of Auditory Temporal Processing for Reading and Spelling Disability. *Perception and Motor Skills* 86:1043-1047

Schulte-Körne G, Deimel W, Müller K, Gutenbrunner C, Remschmidt H (1996) Familial Aggregation of Spelling Disability. *Journal of Child Psychology and Psychiatry* 37:817-822

Schulte-Körne G, Grimm T, Nöthen MM, Müller-Myhsok B, Cichon S, Vogt IR, Propping P, Remschmidt H (1998) Evidence for Linkage of Spelling Disability to Chromosome 15. *American Journal of Human Genetics* 63:279-282

Schultz RT, Cho NK, Staib LH, Kier LE, Fletcher JM, Shaywitz SE, Shankweiler DP, Katz L, Gore JC, Duncan JS, Shaywitz BA (1994) Brain Morphology in Normal and Dyslexic Children: The Influence of Sex and Age. *Annals of Neurology* 35:732-742

Schumacher J, Anthoni H, Dahdouh F, König IR, Hillmer AM, Kluck N, Manthey M, Plume E, Warnke A, Remschmidt H, Hülsmann J, Cichon S, Lindgren CM, Propping P, Zucchelli M, Ziegler A, Peyrard-Janvid M, Schulte-Körne G, Nöthen MM, Kere J (2005) Strong Genetic Evidence of *DCDC2* as a Susceptibility Gene for Dyslexia. *American Journal of Human Genetics* 78:52-62

Seidenberg MS, McClelland JL (1989) A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review* 96:523-568

Semrud-Clikeman M (1997) Evidence from Imaging on the Relationship Between Brain Structure and Developmental Language Disorders. *Seminars in Pediatric Neurology* 4:117-124

Sernoclaes W, Sprenger-Charolles L, Carré R, Démonet J-F (2001) Perceptual Discrimination of Speech Sounds in Developmental Dyslexia. *Journal of Speech Language and Hearing Research* 44:384-399

Seymour PHK, MacGregor CJ (1984) Developmental Dyslexia: A Cognitive Experimental Analysis of Phonological, Morphemic and Visual Impairments. *Cognitive Neuropsychology* 1:43-83

Sham PC, Lin M-W, Zhao JH, Curtis D (2000) Power Comparison of Parametric and Nonparametric Linkage Tests in Small Pedigrees. *American Journal of Human Genetics* 66:1661-1668

Shapiro BK (2001) Specific Reading Disability: A Multiplanar View. *Mental Retardation and Developmental Disabilities* 7:13-20

Share DL, Jorm AF, MacLean R, Matthews R (2002) Temporal Processing and Reading Disability. *Reading and Writing: An Interdisciplinary Journal* 15:151-178

Share DL, Silva PA (2003) Gender Bias in IQ-Discrepancy and Post-Discrepancy Definitions of Reading Disability. *Journal of Learning Disabilities* 36:4-14

Shaywitz BA, Fletcher JM, Shaywitz SE (1995) Defining and Classifying Learning Disabilities and Attention Deficit/Hyperactivity Disorder. *Journal of Child Neurology* 10:S50-S57

Shaywitz BA, Shaywitz SE, Pugh KR, Fulbright RK, Mencl WE, Constable RT, Skudlarski P, Fletcher JM, Lyon GR, Gore JC (2001) The Neurobiology of Dyslexia. *Clinical Neuroscience Research* 1:291-299

Shaywitz BA, Shaywitz SE, Pugh KR, Mencl WE, Fulbright RK, Skudlarski P, Constable RT, Marchione KE, Fletcher JM, Lyon GR, Gore JC (2002) Disruption of Posterior Brain Systems for Reading in Children with Developmental Dyslexia. *Biological Psychiatry* 52:101-110

Shaywitz SE (1998) Dyslexia. *New England Journal of Medicine* 338:307-312

Shaywitz SE, Escobar MD, Shaywitz BA, Fletcher JM, Makuch R (1992) Evidence that Dyslexia May Represent the Lower Tail of a Normal Distribution of Reading. *New England Journal of Medicine* 326:145-150

Shaywitz SE, Fletcher JM, Shaywitz BA (1994) Issues in the Definition and Classification of Attention Deficit Disorder. *Topics in Language Disorders* 14:1-25

Shaywitz SE, Shaywitz BA, Fletcher JM, Escobar MD (1990) Prevalence of Reading Disability in Boys and Girls. Results of the Connecticut Longitudinal Study. *Journal of the American Medical Association* 264:998-1002

Shaywitz SE, Shaywitz BA, Pugh KR, Fulbright RK, Constable RT, Mencl WE, Shankweiler DP, Liberman AM, Skudlarski P, Fletcher JM, Katz L, Marchione KE, Lacadie C, Gatenby C, Gore JC (1998) Functional Disruption in the Organisation of the Brain for Reading in Dyslexia. *Proceedings of the National Academy of Sciences USA* 95:2636-2641

Simos PG, Breier JI, Fletcher JM, Bergman E, Papanicolaou AC (2000) Cerebral Mechanisms Involved in Word Reading in Dyslexic Children: A Magnetic Source Imaging Approach. *Cerebral Cortex* 10:809-810

Skottun BC (2000) The Magnocellular Deficit Theory of Dyslexia: The Evidence From Contrast Sensitivity. *Vision Research* 40:111-127

Skottun BC (2001) On the Use of Meta Contrast to Assess Magnocellular Function in Dyslexic Readers. *Perception and Psychophysics* 63:1272-1274

Skottun BC (2005) Magnocellular Reading and Dyslexia. *Vision Research* 45:133-134

Sladen BK (1970) Inheritance of Dyslexia. *Bulletin of the Orton Society* 20:30-40

Smith SD, Kelley PM, Askew JW, Hoover DM, Deffenbacher KE, Gayán J, Brower AM, Olson RK (2001) Reading Disability and Chromosome 6p21.3: Evaluation of MOG as a Candidate Gene. (Myelin Oligodendrocyte Glycoprotein). *Journal of Learning Disabilities* 34:512-518

Smith SD, Kelley PM, Brower AM (1998) Molecular Approaches to the Genetic Analysis of Specific Reading Disability. *Human Biology* 70:239-256

Smith SD, Kimberling WJ, Pennington BF (1991) Screening for Multiple Genes Influencing Dyslexia. *Reading and Writing: An Interdisciplinary Journal* 3:285-298

Smith SD, Kimberling WJ, Pennington BF, Lubs HA (1983) Specific Reading Disability: Identification of an Inherited Form Through Linkage Analysis. *Science* 219:1345-1347

Smith SD, Pennington BF, Kimberling WJ, Ing PS (1990) Familial Dyslexia: Use of Genetic Linkage Data to Define Subtypes. *Journal of the American Academy of Child and Adolescent Psychiatry* 29:204-213

Snowling MJ (2000) Dyslexia. Blackwell, Oxford

Snowling MJ, Gallagher A, Frith U (2003) Family Risk of Dyslexia is Continuous: Individual Differences in the Precursors of Reading Skill. *Child Development* 74:358-373

Snyder L, Downey D (1995) Serial Rapid Naming Skills in Children with Reading Disabilities. *Annals of Dyslexia* XLV:31-50

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission Test for Linkage Disequilibrium: The Insulin Gene Region and Insulin-Dependant Diabetes Mellitus. *American Journal of Human Genetics* 52:506-516

Spring C, Davis JM (1988) Relations of Digit Naming Speed with Three Components of Reading. *Applied Psycholinguistics* 9:315-334

Stanovich KE (1991) Discrepancy Definitions of Early Reading Disability: Has Intelligence Led us Astray? *Reading Research Quarterly* 26:7-29

Stanovich KE, Siegel LS, Gottardo A (1997) Converging Evidence for Phonological and Surface Subtypes of Reading Disability. *Journal of Educational Psychology* 89:114-127

Stanovich KE, West RF (1989) Exposure to Print and Orthographic Processing. *Reading Research Quarterly* 24:402-433

Stein CM, Schick JH, Taylor HG, Shriberg LD, Millard C, Kundtz-Kluge A, Russo K, Minich N, Hansen A, Freebairn LA, Elston RC, Lewis BA, Lyengar SK (2004) Pleiotropic Effects of a Chromosome 3 Locus on Speech-Sound Disorder and Reading. *American Journal of Human Genetics* 74:283-297

Stein J (2000) The Neurobiology of Reading Difficulties. *Prostaglandins, Leukotrienes and Essential Fatty Acids* 63:109-116

Stein J, Talcott J, Witton C (2001) The Sensorimotor Basis of Developmental Dyslexia. In: Fawcett AJ (ed) Dyslexia: Theory and Good Practice. Whurr, London, pp 65-88

Stein J, Walsh V (1997) To See But Not to Read; The Magnocellular Theory of Dyslexia. *Trends in Neuroscience* 20:147-152

Steinman SB, Steinman BA, Garzia R (1998) Vision and Attention: II: Is Visual Attention a Mechanism Through Which a Deficient Magnocellular Pathway Might Cause Reading Disability? *Optometry and Visual Science* 75:674-681

Stephenson S (1907) Six Cases of Congenital Word-Blindness Affecting Three Generations of One Family. *Opthalmoscope* 5:482-484

Stevenson J (1991a) Which Aspects of Processing Text Mediate Genetic Effects. *Reading and Writing: An Interdisciplinary Journal* 3:249-269

Stevenson J (1991b) Which Aspects of Reading-Ability Show a "Hump" in their distribution. *Applied Cognitive Psychology* 2:77-85

Stevenson J, Graham P, Fredman G, McLoughlin V (1984) The Genetics of Reading Disability. In: Turner CJ, Miles HB (eds) The Biology of Human Intelligence. Nafferton Books Limited, Nafferton, England, pp 85-97

Stevenson J, Graham P, Fredman G, McLoughlin V (1987) A Twin Study of Genetic Influences on Reading and Spelling Ability and Disability. *Journal of Child Psychology and Psychiatry* 28:229-247

Streets AJ, Newby LJ, O'Hare MJ, Bukanov NO, Ibraghimov-Beskrovnaya O, Ang AC (2003) Functional Analysis of PKD1 Transgenic Lines Reveals a Direct Role for Polycystin-1 in Mediating Cell-Cell Adhesion. *Journal of the American Society of Nephrology* 14:1804-1815

Stuart M, Coltheart M (1988) Does Reading Develop in a Sequence of Stages? *Cognition* 30:139-181

Swanson HL (1999) Reading Comprehension and Working Memory in Learning-Disabled Readers: Is the Phonological Loop More Important Than the Executive System? *Journal of Experimental Child Psychology* 72:1-31

Swanson HL, Alexander J (1997) Cognitive Processes that Predict Reading in Learning Disabled Readers: Revisiting the Specificity Hypothesis. *Journal of Educational Psychology* 89:128-158

Swanson HL, Berninger V (1995) The Role of Working Memory in Skilled and Less Skilled Readers' Comprehension. *Intelligence* 21:83-108

Symmes JS, Rapoport JL (1972) Unexpected reading failure. *American Journal of Orthopsychiatry* 42:82-91

Taipale M, Kaminen N, Nopola-Hemmi J, Haltia T, Myllyluoma B, Lyytinen H, Muller K, Kaaranen M, Lindsberg PJ, Hannula-Jouppi K, Kere J (2003) A Candidate Gene for Developmental Dyslexia Encodes a Nuclear Tetratricopeptide Repeat Domain Protein Dynamically Regulated in Brain. *Proceedings of the National Academy of Sciences* 100:11553-11558

Talcott JB, Gram A, Ingelghem MV, Witton C, Stein JF, Toennessen FE (2003) Impaired Sensitivity to Dynamic Stimuli in Poor Readers of a Regular Orthography. *Brain and Language* 87:259-266

Talcott JB, Hansen PC, Elikem LA, Stein JF (2000a) Visual Motion Sensitivity in Dyslexia: Evidence for Temporal and Motion Energy Integration Deficits. *Neuropsychologia* 38:935-943

Talcott JB, Hansen PC, Willis-Owen C, McKinnell IW, Richardson AJ, Stein JF (1998) Visual Magnocellular Impairment in Adult Developmental Dyslexics. *Neuro-Opthalmology* 20:187-201

Talcott JB, Witton C, McLean MF, Hansen PC, Rees A, Green GGR, Stein JF (2000b) Dynamic Sensory Sensitivity and Children's Word Decoding Skills. *Proceedings of the National Academy of Sciences* 97:2952-2957

Tallal P (1980) Auditory Temporal Perception, Phonics and Reading Disabilities in Children. *Brain and Language* 9:182-198

Tallal P, Miller S, Fitch RH (1993) Neurobiological Basis of Speech: A Case for the Pre-Eminence of Temporal Processing. *Annals of the New York Academy of Science* 682:27-47

Taylor J (1958) Selected Writings of John Hughling Jackson. Basic Books, New York

Taylor KE, Richardson AJ (2000) Visual Function, Fatty Acids and Dyslexia. *Prostaglandins, Leukotrienes and Essential Fatty Acids* 63:89-93

Temple E, Poldrack RA, Protopapas A, Nagarajan S, Salz T, Tallal P, Merzenich MM, Gabrieli JD (2000) Disruption of the Neural Response to Rapid Acoustic Stimuli in Dyslexia: Evidence from Functional MRI. *Proceedings of the National Academy of Sciences USA* 97:13907-13912

Temple E, Poldrack RA, Salidis J, Deutsch GK, Tallal P, Merzenich MM, Gabrieli JD (2001) Disrupted Neural Responses to Phonological and Orthographic Processing in Dyslexic Children: An fMRI Study. *NeuroReport* 12:299-307

Terwilliger JD, Weiss KM (1998) Linkage Disequilibrium Mapping of Complex Disease: Fantasy or Reality? *Current Opinion in Biotechnology* 9:578-594

Terwilliger JD, Zollner S, Laan M, Paabo S (1998) Mapping Genes Through the Use of Linkage Disequilibrium Generated by Genetic Drift: 'Drift Mapping' in Small Populations with no Demographic Expansion. *Human Heredity* 48:138-154

Thomas CJ (1905) Congenital "Word-Blindness" and its Treatment. *Opthalmoscope* 3:380-385

Thomson M (1990) Developmental Dyslexia: Studies in Disorders of Communication. Whurr Publishers Ltd, London

Tsunoda T, Lathrop GM, Sekine A, Yamada R, Takahashi A, Ohnishi Y, Tanaka T, Nakamura Y (2004) Variation of Gene-Based SNPs and Linkage Disequilibrium Patterns in the Human Genome. *Human Molecular Genetics* 13:1623-1632

Turic D, Robinson L, Duke M, Morris DW, Webb V, Hamshere M, Milham C, Hopkin E, Pound K, Fernando S, Grierson A, Easton M, Williams N, Bree Mvd, Chowdhury R, Gruen J, Stevenson J, Krawczak M, Owen MJ, O'Donovan MC, Williams J (2003) Linkage Disequilibrium Mapping Provides Further Evidence of a Gene for Reading Disability on Chromosome 6p21.3-22. *Molecular Psychiatry* 8:176-185

Tzenova J, Kaplan BJ, Petryshen TL, Field LL (2004) Confirmation of a Dyslexia Susceptibility Locus on Chromosome 1p34-p36 in a Set of 100 Canadian Families. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 127:117-124

Vellutino FR (1979) Dyslexia: Research and Theory. MIT Press, Cambridge, MA, USA

Vellutino FR, Fletcher JM, Snowling MJ, Scanlon DM (2004) Specific Reading Disability (Dyslexia): What Have we Learned in the Past Four Decades? *Journal of Child Psychology and Psychiatry* 45:2-40

Verrelli BC, Eanes WF (2001) Clinal Variation for Amino Acid Polymorphisms at the *Pgm* Locus in *Drosophila Melanogaster. Genetics* 157:1649-1663

Vogel S (1990) Gender Differences in Intelligence, Language Abilities and Academic Achievement in Students with Learning Disabilities: A Review of the Literature. *Journal of Learning Disabilities* 23:44-52

Vogler GP, DeFries JC, Decker SN (1985) Family History as an Indicator of Risk for Reading Disability. *Journal of Learning Disabilities* 18:419-421

Wadsworth SJ, DeFries JC, Stevenson J, Gilger JW, Pennington BF (1992) Gender ratios among reading-disabled children and their siblings as a function of parental impairment. *Journal of Child Psychology and Psychiatry* 33:1229-39

Wadsworth SJ, Gillis JJ, DeFries JC, Fulker DW (1989) Differential Genetic Aetiology of Reading Disability as a Function of Age. *The Irish Journal of Psychology* 10:509-520

Wagner R, Torgesen J (1987) The Nature of Phonological Processing and its Causal Role in the Acquisition of Reading Skills. *Psychological Bulletin* 101:192-212

Wagner RK, Torgesen JK, Laughon PL, Simmons K, Rashotte CAS (1993) Development of Young Readers' Phonological Processing Abilities. *Journal of Educational Psychology* 85:83-103

Wagner RK, Torgesen JK, Rashotte CA (1994) Development of Reading-Related Phonological Processing Abilities: New Evidence of Bi-Directional Causality from a Latent Variable Longitudinal Study. *Developmental Psychology* 30:73-87

Waites L (1968) Dyslexia International World Federation of Neurology. 'Report of Research Group on Developmental Dyslexia and World Literacy'. *Bulletin of the Orton Society* 18:21-22

Walker L, Cole E (1965) Familial Patterns of Expression of Specific Reading Disability in a Population Sample. *Bulletin of the Orton Society* 15:12-24

Wang W, Thornton K, Berry A, Long M (2002) Nucleotide Variation Along the *Drosophila Melanogaster* Fourth Chromosome. *Science* 295:134-137

Warren RP, Odell JD, Warren WL, Burger RA, Maciulis A, Daniels WW, Torres AR (1995) Reading Disability, Attention-Deficit Hyperactivity Disorder and the Immune System. *Science* 268:786-788

Wechsler D (1992) WISC III UK. In: The Psychological Corporation Ltd. Harcourt Brace and Co, London

Weiss KM, Clark AG (2002) Linkage Disequilibrium and the Mapping of Complex Human Traits. *Trends in Genetics* 18:19-24

Weiss KM, Terwilliger JD (2000) How Many Diseases Does it Take to Map a Gene with SNPs? *Nature Genetics* 26:151-157

WHO, 2003: World Health Organisation. The International Classification of Diseases, Vol. 10: Classification of Mental and Behavioural Disorders. World Health Organisation, Geneva, Switzerland

Wigg KG, Couto JM, Feng Y, Anderson B, Cate-Carter TD, Macciardi F, Tannock R, Lovett MW, Humphries TW, Barr CL (2004) Support for *EKN1* as the Susceptibility Locus for Dyslexia on 15q21. *Molecular Psychiatry* 9:1111-1121

Willcutt EG, Pennington BF (2000a) Comorbidity of Reading Disability and Attention-Deficit/Hyperactivity Disorder: Differences by Gender and Subtype. *Journal of Learning Disabilities* 33:179-191

Willcutt EG, Pennington BF (2000b) Psychiatric Comorbidity in Children and Adolescents with Reading Disability. *Journal of Child Psychology and Psychiatry* 41:1039-1048

Willcutt EG, Pennington BF, DeFries JC (2000) Twin Study of the Etiology of Comorbidity Between Reading Disability and Attention-Deficit/Hyperactivity Disorder. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 96:293-301

Willcutt EG, Pennington BF, Smith SD, Cardon LR, Gayán J, Knopik VS, Olson RK, DeFries JC (2002) Quantitative Trait Locus for Reading Disability on Chromosome 6p is Pleiotropic for Attention-Deficit/Hyperactivity Disorder. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 114:260-268

Wimmer H, Mayringer H, Landerl K (1998) Poor Reading: A Deficit in Skill-Automatization or a Phonological Deficit? *Scientific Studies of Reading* 2:321-340

Witton C, Talcott JB, Hansen PC, Richardson AJ, Griffiths TD, Rees A, Stein JF, Green GG (1998) Sensitivity to Dynamic Auditory and Visual Stimuli Predicts Nonword Reading Ability in Both Dyslexic and Normal Readers. *Current Biology* 8:791-797

Wolf M (1986) Rapid Alternating Stimulus Naming in the Developmental Dyslexias. *Brain and Language* 27:360-379

Wolf M, Bally H, Morris R (1986) Automaticity, Retrieval Processes and Reading: A Longitudinal Study in Average and Impaired Reading. *Child Development* 57:988-1000

Wolf M, Bowers PG (1999) The Double-Deficit Hypothesis for the Developmental Dyslexias. *Journal of Educational Psychology* 91:415-438

Wolf M, Pfeil C, Lotz R, Biddle K (1994) Towards a More Universal Understanding of the Developmental Dyslexias: The Contribution of Orthographic Factors. In: Berninger VW (ed) The Varieties of Orthographic Knowledge. 1: Theoretical and Developmental Issues. Kluwer Academic Press, Dordrecht, The Netherlands, pp 137-171

Wolff PH, Melngailis I (1994) Familial Patterns of Developmental Dyslexia: Clinical Findings. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 54:122-131

Yap RL, Leij AVD (1994) Testing the Automatization Deficit Hypothesis of Dyslexia via a Dual-Task Paradigm. *Journal of Learning Disabilities* 27:660-665

Ylisaukko-Oja T, Peyrard-Janvid M, Lindgren CM, Rehnström K, Vanhala R, Peltonen L, Järvela I, Kere J (2004) Family-Based Association Study of *DYX1C1* Variants in Autism. *European Journal of Human Genetics*: 13(1):127-30

Yokota Y (2001) Id and Development. *Oncogene* 20:8290-8298

Yu A, Zhao CF, Fan Y, Jang WH, Mungall AJ, Deloukas P, Olsen A, Doggett NA, Ghebranious N, Broman KW, Weber JL (2001) Comparison of Human Genetic and Sequence-Based Physical Maps. *Nature* 409:951-953

Yule W, Rutter M (1975) The Concept of Specific Reading Retardation. *Journal of Child Psychology and Psychiatry* 16:181-197

Yule W, Rutter M, Berger M, Thompson J (1974) Over and Under Achievement in Reading: Distribution in the General Population. *British Journal of Educational Psychology* 44:1-12

Yun K, Mantani A, Garel S, Rubenstein J, Israel MA (2004) *Id4* Regulates Neural Progenitor Proliferation and Differentiation *In Vivo*. *Development* 131:5441-5448

Zahalkova M, Vrzal V, Kloboukova E (1972) Genetical investigations in dyslexia. *Journal of Medical Genetics* 9:48-52

Zerbin-Rüdin E (1967) Congenital Word-Blindness. *Bulletin of the Orton Society* 17:47-54

Zhao JH, Curtis D, Sham PC (2000) Model-Free Analysis and Permutation Tests for Allelic Associations. *Human Heredity* 50:133-139

Zondervan KT, Cardon LR (2004) The Complex Interplay Among Factors That Influence Allelic Association. *Nature Reviews Genetics* 5:89-101

Zsurka G, Gregáñ J, Schweyen RJ (2001) The Human Mitochondrial Mrs2 Protein Functionally Substitutes for its Yeast Homologue, A Candidate Magnesium Transporter. *Genomics* 72:158-168