

From The Department of Biosciences and Nutrition
Karolinska Institutet, Stockholm, Sweden

IDENTIFICATION OF SUSCEPTIBILITY GENES FOR DYSLEXIA

Heidi Anthoni



**Karolinska
Institutet**

Stockholm 2007

Supervisor	Professor Juha Kere Department of Biosciences and Nutrition Karolinska Institutet Stockholm, Sweden and Department of Medical Genetics University of Helsinki Helsinki, Finland
Co-supervisor	Dr Myriam Peyrard-Janvid Department of Biosciences and Nutrition Karolinska Institutet Stockholm, Sweden
Faculty opponent	Professor Shelley D. Smith Munroe Meyer Institute University of Nebraska Medical Center Omaha, NE, USA
Examination board	Professor Niels Tommerup Wilhelm Johannsen Centre for Functional Genome Research Department of Cellular and Molecular Medicine University of Copenhagen Copenhagen, Denmark Professor Lennart von Wendt Department of Child Neurology Hospital for Children and Adolescents University of Helsinki Helsinki, Finland Docent Catharina Lavebratt Department of Molecular Medicine and Surgery Karolinska Institutet Stockholm, Sweden

All previously published papers were reproduced with permission from the publisher.
Cover illustration adapted from “one in 25” by Lawrence Cockrill, Dyslexia Institute
website (www.dyslexiaaction.org.uk)

Published and printed by Karolinska Institutet.

© Heidi Anthoni, 2007
ISBN 978-91-7357-345-0

*The process of scientific discovery is, in effect,
a continual flight from wonder*

Albert Einstein

To my family

ABSTRACT

Developmental dyslexia, also known as specific reading disability, is characterized by persistent difficulties in learning to read and spell in spite of adequate intelligence, education, social environment, and normal senses. It is the most common learning disability affecting 5-10% of school-aged children. The core deficit in dyslexia is believed to involve phonological processing. Dyslexia has a complex genetic basis, and family studies as well as extensive molecular genetic studies have proven the importance of genetic factors in the development of this disorder. To date, nine chromosomal regions have been identified as susceptibility loci for dyslexia; *DYX1–DYX9*. *DYX1C1* on chromosome 15q21 was the first candidate gene suggested based on the cloning of a translocation breakpoint co-segregating with dyslexia.

The aim of this thesis project was to identify susceptibility genes for dyslexia primarily by using a positional cloning approach. Specifically, three candidate loci for dyslexia were studied; *DYX1*, *DYX2*, and *DYX3*. Several rounds of genetic mapping within the *DYX3* region lead to the identification of overlapping dyslexia risk haplotypes in two independent sample sets. Carriers of the risk haplotype showed attenuated expression of two co-expressed genes within the region, *MRPL19* and *C2ORF3*, indicating a possible regulatory effect of the risk variants. Linkage disequilibrium mapping within the most replicated susceptibility for dyslexia, *DYX2*, revealed a strong genetic effect for *DCDC2* in dyslexic individuals, in particular in more severely affected cases. The effect of this gene as a susceptibility factor for dyslexia was confirmed by replication analysis in an independent sample set.

Replication efforts of *DYX1C1* have shown inconsistent results, and thus its role in the development of dyslexia has been considered unsettled. We refined the haplotype structure by analyzing additional variants within the *DYX1C1* locus. The haplotypes showed association with dyslexia in a large sample set, with possible sex-specific effects. Refined mapping of another translocation within the *DYX1* region co-segregating with dyslexia located the breakpoint to the complex promoter region of *CYP19A1* (aromatase). Genetic variation within *CYP19A1* associated with speech and language measures and dyslexia in three independent sample sets. Variation in the highly conserved brain promoter of *CYP19A1* altered transcription factor binding. An aromatase inhibitor reduced dendritic growth in cultured rat neurons. Brain morphology studies of aromatase-deficient mice showed increased cortical neuronal density and occasional cortical heterotopias, similar to those observed in human dyslexic brains.

To date, seven candidate susceptibility genes have been suggested for dyslexia. In addition to the ones studied in this thesis, *KIAA0319* within *DYX2* and *ROBO1* within *DYX5* have been indicated in dyslexia. Studies of the dyslexia candidate genes in rats and mice implicate neuronal migration and axon guidance as neurobiological mechanisms that likely mediate this disorder. Anatomical studies support this hypothesis as cortical abnormalities have been observed in dyslexic brains. Functional brain imaging studies show that these disrupted areas are involved in phonological processing and display abnormal activation in dyslexics. Taken together, our results and these studies implicate a biological basis for developmental dyslexia.

LIST OF PUBLICATIONS

This thesis is based on the following publications referred to in the text by their Roman numerals (I-V):

- I. 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.
Fine mapping of the 2p11 dyslexia locus and exclusion of *TACR1* as a candidate gene.
Hum Genet. 2004 Apr;114(5):510-6.
- II. **Anthoni H**, Zucchelli M, Matsson H, Müller-Myhsok B, Fransson I, Schumacher J, Massinen S, Onkamo P, Warnke A, Griesemann H, Hoffmann P, Nopola-Hemmi J, Lyytinen H, Schulte-Körne G, Kere J, Nöthen MM, Peyrard-Janvid M.
A locus on 2p12 containing the co-regulated *MRPL19* and *C2ORF3* genes is associated to dyslexia.
Hum Mol Genet. 2007 Mar;16(6):667-677.
- III. 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.
Strong genetic evidence of *DCDC2* as a susceptibility gene for dyslexia.
Am J Hum Genet. 2006 Jan;78(1):52-62.
- IV. Dahdouh F, **Anthoni H**, Tapia-Páez I, Peyrard-Janvid M, Schulte-Körne G, Warnke A, Remschmidt H, Ziegler A, Kere J, Müller-Myhsok B, Nöthen MM, Schumacher J, Zucchelli M.
Evidence of *DYX1C1* as a sex-specific susceptibility factor for dyslexia.
Submitted
- V. **Anthoni H**, Lewis BA, Fan X, Zucchelli M, Sucheston LEM, Tapia-Páez I, Taipale M, Stein CM, Hokkanen ME, Castrén E, Gilger JW, Hynd GW, Nopola-Hemmi J, Lyytinen H, Schoumans J, Nordenskjöld M, Simpson E, Mäkelä S, Gustafsson JÅ, Warner M, Peyrard-Janvid M, Iyengar S, Kere J.
Aromatase (*CYP19A1*) regulates development of cognitive functions.
Submitted

*Authors contributed equally to the study

CONTENTS

1	Introduction.....	1
1.1	Developmental dyslexia	1
1.2	Neurocognitive deficits in dyslexia.....	1
1.2.1	General reading ability	1
1.2.2	Reading disability	2
1.2.3	Comorbidity of dyslexia.....	4
1.2.4	Diagnosis of dyslexia	4
1.3	Neurobiology of dyslexia	6
1.3.1	Neurobiology of reading	7
1.3.2	Neurobiology of reading disability	8
1.4	Genetics of complex diseases.....	11
1.4.1	Genetic studies of complex diseases.....	12
1.4.2	Linkage analysis	13
1.4.3	Association analysis	15
1.4.4	Chromosomal rearrangements	17
1.4.5	Candidate gene studies	17
1.5	Genetics of dyslexia.....	18
1.5.1	<i>DYX1</i> on chromosome 15q21	20
1.5.2	<i>DYX2</i> on chromosome 6p21-p22.....	21
1.5.3	<i>DYX3</i> on chromosome 2p15-p12.....	23
1.5.4	<i>DYX4 – DYX9</i>	24
1.5.5	Other possible loci for dyslexia susceptibility.....	25
2	Aims of the study.....	27
3	Materials and methods.....	29
3.1	Samples	29
3.1.1	Finnish dyslexia cohort (I, II, V)	29
3.1.2	German dyslexia cohort (II, III, IV)	30
3.1.3	US dyslexia cohort (V).....	31
3.1.4	US speech sound disorder cohort (V).....	31
3.2	Genetic analysis	32
3.2.1	Genotyping (I – V)	32
3.2.2	Linkage analysis (I).....	32
3.2.3	Association analysis (I – V)	33
3.2.4	DNA re-sequencing (I – V).....	33
3.3	Gene identification.....	34
3.3.1	Fluorescent <i>in situ</i> hybridization (V)	34
3.3.2	Southern blot analysis (V).....	34
3.3.3	Gene predictions (II, V)	35
3.4	Gene expression studies	35
3.4.1	Transcript characterization (I – III, V).....	35
3.4.2	Quantitative mRNA expression (II, V).....	35
3.4.3	Northern blot analysis (I – III)	35
3.4.4	Allele-specific expression (II).....	36
3.4.5	Electrophoretic mobility shift assay (IV, V)	36
3.5	Functional studies (V)	36

3.5.1	Aromatase knock-out mice	36
3.5.2	Process outgrowth of rat hippocampal neurons	37
3.6	Evolutionary analysis (II, V).....	37
4	Results	39
4.1	Genetic analysis.....	39
4.1.1	<i>DYX3</i> (I, II).....	39
4.1.2	<i>DYX2</i> (III).....	42
4.1.3	<i>DYX1C1</i> (IV).....	43
4.1.4	t(2;15)(p12;q21) (V)	45
4.1.5	Causal variants in the dyslexia candidate genes (I – V)	47
4.2	Expression analysis	50
4.2.1	Transcript characterization (I – III, V)	50
4.2.2	Correlation of mRNA expression across various regions of human brain (II, V).....	50
4.3	Functional studies (V)	51
4.3.1	Aromatase knock-out mice	51
4.3.2	Hippocampal process outgrowth	51
4.4	Evolutionary analysis (II, V).....	51
5	Discussion.....	53
5.1	Genetic analysis of dyslexia.....	53
5.1.1	The <i>DYX3</i> locus and <i>MRPL19</i> and <i>C2ORF3</i>	53
5.1.2	Strong evidence for <i>DCDC2</i> as a susceptibility gene for dyslexia.....	54
5.1.3	Evidence for <i>DYX1C1</i> as a sex-specific risk factor for dyslexia.....	56
5.1.4	<i>CYP19A1</i> is associated with dyslexia and speech- and language disorders.....	58
5.1.5	General comments.....	60
5.2	Molecular mechanisms of dyslexia.....	61
5.3	Neurobiological mechanisms of dyslexia.....	63
5.3.1	Neuronal migration and dyslexia.....	63
5.3.2	A neuronal network model for dyslexia.....	65
5.4	Sex-specific effects of dyslexia.....	67
5.5	Evolution of speech, language and reading	68
6	Conclusions and future perspectives	71
7	Acknowledgements	73
8	References	75

LIST OF ABBREVIATIONS

ADHD	attention-deficit hyperactivity disorder
BAC	bacterial artificial chromosome
bp	base pair
cM	centimorgan
cDNA	complementary deoxyribonucleic acid
CI	confidence interval
CNS	central nervous system
cSNP	coding single nucleotide polymorphism
dN/dS	nonsynonymous/synonymous substitution ratio
DNA	deoxyribonucleic acid
DTI	diffusion tensor imaging
E17	embryonic day 17 (mouse or other model organism)
EGF	epidermal growth factor
EEG	electroencephalography
FISH	fluorescent <i>in situ</i> hybridization
EMSA	electrophoretic mobility shift assay
fMRI	functional magnetic resonance imaging
GRR	genotype relative risk
HLA	human leukocyte antigen
HPM	haplotype pattern mining
IBD	identical by descent
IQ	intelligence quotient
kb	kilobase
KO	knock-out (mouse or other model organism)
LD	linkage disequilibrium
LGN	lateral geniculate nucleus
LOD	logarithm of odds
LRT	likelihood ratio test
MALDI-TOF	matrix-assisted laser desorption/ionization time-of-flight
Mb	megabase
MEG	magnetoencephalography
MGN	medial geniculate nucleus
MRI	magnetic resonance imaging
mRNA	messenger ribonucleic acid
NeuN	neuron-specific nuclear protein
NPL	non-parametric linkage
OR	odds ratio
PCR	polymerase chain reaction
PDT	pedigree disequilibrium test
PET	positron emission tomography
RNAi	RNA interference
RT-PCR	reverse transcriptase PCR
SD	standard deviation
SLI	specific language impairment

SNP	single nucleotide polymorphism
SSD	speech sound disorder
TDT	transmission disequilibrium test
UTR	untranslated region

1 INTRODUCTION

1.1 DEVELOPMENTAL DYSLEXIA

Dyslexia was first described as “word blindness” in 1877, when Kussmaul reported a man who despite normal intelligence was unable to learn to read even though he received an adequate education (reviewed by Richardson 1992). In the end of the 19th century, Hinshelwood and Morgan both described word blindness as a congenital defect, occurring in children with otherwise normal brains. These reports were based on studies of acquired dyslexia, or alexia, where neurological damage to certain brain areas result in loss of reading ability. In 1925, Orton described the first theory of specific learning difficulty. He hypothesized that the children's reading problems stemmed from a failure of the left hemisphere to become dominant over the right, and that a deficit in visual processing was the cause for the reading difficulties. He termed the disorder “strephosymbolia”, i.e., “twisted symbols” from the Greek words [strepho]=twist, and [symbolon]=symbol. The disorder was later more appropriately called dyslexia, “difficulty with words” (from Greek [dys]=difficult, [lexis]=words), as it was recognized that the condition is language-related, rather than a visual problem. In 1968 The Research Group on Developmental Dyslexia from The World Federation of Neurology recommended the definitions of dyslexia used today, although they have since been somewhat specified. The definition of dyslexia set by The International Dyslexia Association (IDA) and adopted by the National Institute of Child Health and Human Development (NICHD) in 2002 is:

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

1.2 NEUROCOGNITIVE DEFICITS IN DYSLEXIA

1.2.1 General reading ability

Reading is a cognitively complex task, with various component processes contributing to the development of this skill. The two main components required for fluent reading are word recognition and language comprehension (Vellutino et al. 2004). Word recognition can be divided into two components of written language skills; phonological decoding and orthographic coding. Written words are the initial input to the reading system, which are then converted to explicit speech or sound-based mental

representations during silent reading (Snowling 2001). This conversion process for an alphabetic language involves storing words in phonologically coded temporal memory, language-specific correspondence between letters and sounds, and conscious awareness of the sounds in auditory words.

Reading performance depends on a number of phenotypically correlated and overlapping factors. *Phonological awareness* is an oral language skill characterized by the ability to dissect a spoken word into its individual speech sounds; the phonemes. In order to learn to read, children must be able to recognize and manipulate the individual phonemes that correspond to letters (graphemes) in written language. *Phonological decoding* is the conversion of written words into spoken words. This demonstrates an understanding of the letter-sound (grapheme-phoneme) correspondence, which is critical in learning to read and required for learning new words. *Orthographic coding* is the ability to encode the specific spelling pattern of a word and the recognition of an entire written word as a unit. It is required for development of fluent reading. *Rapid automatized naming* is the rapid access to a phonological name code during phonological decoding; it is also required for fluent reading. Deficits in one or more of these processes are manifested in impaired real-word-reading, or word recognition, an indicator of reading disability in children (Vellutino et al. 2004).

1.2.2 Reading disability

Reading ability is normally distributed in the general population, where disability (dyslexia) represents the lower tail of the continuous distribution (Shaywitz et al. 1992). The prevalence of dyslexia is 5-10% in school-aged children (Shaywitz et al. 1990; Shaywitz et al. 1992; Lyytinen et al. 1995), depending on the chosen diagnostic criteria as well as on the underlying language. The prevalence is correlated to the complexity of the orthographic rules for a given language. In languages with transparent or shallow orthography such as Finnish, Italian or even German, the letters or combinations of them correspond to the speech sounds occurring in the language. In these regular orthography languages, low reading speed is the major and persistent problem for dyslexics (Landerl et al. 1997; Zoccolotti et al. 1999; Leinonen et al. 2001). In English, which has a more complex orthography, the problems are prominent also in reading accuracy, although reading speed is also severely affected (Ziegler et al. 2003). Nevertheless, dyslexics in both regular and complex orthography languages show deficits across the same neurocognitive reading-related processes.

Dyslexia is characterized by deficits in accurate and fluent word recognition. Dyslexic children show early difficulties in learning to name letters and associate sounds to them. These are most probably the result of a deficit in the phonological component of language, impairing the ability to segment a spoken word into its underlying phonologic elements and link each letter to its corresponding sound (Shaywitz and Shaywitz 2005). As a result, the dyslexic reader experiences difficulty in decoding the word and then in identifying it, leading to problems in reading and spelling. Secondary consequences are problems in reading comprehension and reduced vocabulary and

background knowledge (Shaywitz and Shaywitz 2005). Poor phonological awareness is among the best early predictors of later reading disability, suggesting a causal role in the development of dyslexia (Peterson et al. 2007). Deficits in phonological awareness can be demonstrated in children of dyslexic parents already before kindergarten (Pennington and Lefly 2001; Snowling et al. 2003; Lyytinen et al. 2004). In adolescence or adulthood, most dyslexic individuals compensate for their reading disability through experience, but are still deficient in their phonological skills, and generally do not reach normal reading ability (Bruck 1992; Shaywitz et al. 1999).

Although the core deficit in dyslexia is most commonly believed to be in phonological processing, dyslexic individuals also show impairments in several correlated cognitive processes, such as verbal short-term memory and rapid naming (Vellutino et al. 2004). Verbal short-term memory is implicated in the transient storing of all the relevant representations, thus allowing grapheme-phoneme conversion and assembly of the phonemes. A deficit in rapid naming is connected with deficits in reading fluency and individuals with a double deficit, i.e., deficits in both phonological awareness and rapid naming, have been found to be more impaired on reading-related measures (Lovett et al. 2000). It has also been suggested that impairments in different cognitive processes lead to diverse subtypes of dyslexia. Based on the dual-route model of reading two types of dyslexia have been proposed: surface dyslexia that affects orthographic processing, and phonological dyslexia that is characterized by poor phonological decoding skills (Castles and Coltheart 1993). Generally, however, dyslexic children are impaired in both phonological and orthographic skills, and the validity of this model has been criticized.

In addition to the reading and language related deficits, dyslexic individuals often manifest also other behavioral symptoms, including sensory deficits in auditory and visual domains, impaired balance and motor control, and cerebellar dysfunction (Ramus 2003). Thus, in addition to the widely accepted phonological theory, there have been several alternative theories proposed for the development of dyslexia. Early studies of dyslexia focused on a simple visual processing theory, since it was observed that dyslexic individuals commonly made visual confusions between morphologically similar letters, especially those that have a symmetrical counterpart (such as b and d). However, this simple theory has long been invalidated, as it has been demonstrated that the reversal errors are restricted to one's own language, being linguistic rather than visual in nature (Vellutino 1979; Peterson et al. 2007). The visual sensory deficit has since been explained by the magnocellular theory, in which a generalized dysfunction of cells in the magnocellular pathway results in difficulties with rapidly processing visually presented stimuli (Stein and Walsh 1997). Many dyslexics also perform poorly on auditory tasks and show abnormal neurophysiological responses to auditory stimuli (McArthur and Bishop 2001). The auditory processing theory suggests that the phonological deficits in dyslexia are secondary to auditory deficits in the perception of short or rapidly varying sounds due to impairment in auditory temporal processing (Tallal 1980). The cerebellar theory of dyslexia attempts to explain the phonological deficits by a cerebellar dysfunction leading to impairment in motor, balance and automation skills. The visual, auditory, and motor deficits have recently been combined in a general magnocellular theory for dyslexia (Stein 2001). However, the visual,

auditory, and motor theories for dyslexia are abated by that these impairments seem not to be specific for dyslexia and there is little evidence for a causal relationship between sensorimotor dysfunction and reading impairment (Ramus 2003).

1.2.3 Comorbidity of dyslexia

Dyslexic individuals often show symptoms of related neurodevelopmental behavioral disorders, e.g., deficits in oral language acquisition, motor coordination deficits, visuospatial impairment, and attentional abilities. There exists frequent comorbidity between dyslexia and speech sound disorder (SSD), specific language impairment (SLI), attention deficit hyperactivity disorder (ADHD), dysgraphia (difficulties in writing), dyspraxia (motor coordination deficits), and dyscalculia (difficulties in mathematics). All these behavioral disorders manifest in learning problems at an early age, despite normal intelligence, that can affect the cognitive and emotional development of a child.

1.2.3.1 Speech sound disorder

SSD is a developmental disorder that is prevalent in childhood, with 15.6 % of pre-school age children and 3.8 % of 6 year-olds manifesting the disorder (Shriberg et al. 1999; Campbell et al. 2003). It is characterized by deficits in articulation and the cognitive representation of speech sounds (Stein et al. 2006). Children with SSD show also deficits in verbal short-term memory (Kenney et al. 2006). As in dyslexia, the core deficit is believed to be in phonological processing (Pennington 2006). Articulation is crucial in the early stages in learning to read, but not needed for skilled learning. Developmental problems in spoken language at a very early age and diagnosed SSD before school age have been shown to predict the later emergence of dyslexia in families at high risk (Lyytinen et al. 2003; Raitano et al. 2004). In effect, about 50 % of SSD probands will develop dyslexia (Stein et al. 2006).

1.2.3.2 Attention deficit hyperactivity disorder

ADHD, together with dyslexia, is the most prevalent neurobehavioral disorder in childhood, affecting 8-12 % of the general population (Faraone et al. 2003). The disorder manifests as inattention, hyperactivity, and/or impulsivity (Wilens et al. 2002). ADHD and dyslexia are highly comorbid, with estimates of co-occurrence between 15 and 40 % (Willcutt and Pennington 2000). Measures of ADHD and dyslexia exhibit at least some shared factors, as processing speed has shown to be impaired in both disorders (Pennington 2006).

1.2.4 Diagnosis of dyslexia

Dyslexia is usually diagnosed in primary school when parents or teachers become aware of that the dyslexic child has problems in reading. A family history of dyslexia as well as a history of early learning problems, e.g., delay in speaking and difficulties in

early letter identification, are strong risk factors for dyslexia and are often first evaluated (Grizzle and Simms 2005). Reading and spelling ability are assessed by measuring the central impairment in dyslexia, i.e., fluent and accurate word recognition, to evaluate if there are unexpected difficulties in reading and spelling contrasted by age, intelligence, and level of education. The commonly used criteria require a two standard deviations (SD) discrepancy between the observed reading ability and that expected on the basis of intelligence quotient (IQ) (Fisher and DeFries 2002). Commonly used exclusion criteria are a peripheral sensory impairment (e.g., deafness, poor vision), a neurological impairment, or other developmental disorders such as autism, ADHD, or mental retardation.

Even though the core deficit in dyslexia lies in phonological processing, many dyslexic individuals also show deficits in several other reading related components, making it difficult to precisely define the dyslexia phenotype. There is a variety of well-established psychometric tests to measure the reading and spelling ability as well as the various subcomponents of dyslexia (Francks et al. 2002b). As dyslexia is a quantitative disorder, clinical diagnosis usually derive from applying thresholds to psychometric measures that are normally distributed in the general population. Setting a diagnostic threshold can be difficult, as an arbitrary threshold of a continuous variable must be defined to classify the individual as dyslexic or not. In school-aged children, performance on both phonological and orthographic skills is usually assessed. Phonological awareness is tested by auditory-oral methods that do not involve the visual processing of print. The tasks may include phoneme deletion, or moving specific phonemes around within or between words. Phonological decoding is typically evaluated through oral reading of pronounceable pseudowords such as 'joop'. Orthographic coding can be measured through oral reading of words that violate standard letter-sound conventions (irregular words), such as 'yacht'. Another common measure is rapid recognition of correctly spelled words versus a phonologically identical non-word ('rain' vs. 'rane'). Rapid naming tasks measure the ability to rapidly retrieve names for items presented in a series, assessed by the time taken to name an array of letters, numbers, colors or objects. Verbal short-term memory is usually measured by reciting a string of orally presented numbers (digit span) or unrelated nonwords (nonword repetition). However, these are all arbitrary measures and even a well-designed test is likely to measure several of the cognitive processes, as they are connected and overlapping (Francks et al. 2002b).

Psychometric profiles can vary greatly among people with dyslexia and at different developmental stages (Fisher and DeFries 2002). The type of test administered depends on the age of the individual and also on the level of education. Adolescents and adults that have compensated for their deficit seem to have normal word recognition skills, but the underlying phonological deficits persist. These deficits can be measured with special tests developed for adults, measuring, e.g., spelling, reading rate, or phonological skills (Shaywitz et al. 1999). There are also oral-auditory phonological tests available assessing phoneme awareness and phonological skills in young children, before reading-age (Grizzle and Simms 2005). These test the child's knowledge of letter sounds, the ability to blend sounds into words, and the ability to name letters rapidly.

1.3 NEUROBIOLOGY OF DYSLEXIA

A traditional approach to study reading and reading disability has been to examine the consequences of focal brain lesions leading to acquired dyslexia (alexia). In the late 19th century, Dejerine suggested that a portion of the posterior brain region (including the angular gyrus and supramarginal gyrus in the inferior parietal cortex and the posterior part of the superior temporal gyrus) is critical for reading, and that damage to the left angular gyrus resulted in reading difficulties (reviewed by Richardson 1992; Shaywitz and Shaywitz 2005). Hinshelwood and Morgan, who reported the first cases of developmental dyslexia, suggested that the disorder in young dyslexic patients similarly resulted from abnormal brain development, particularly of the angular gyrus (Richardson 1992). The first pathological examination of the brain of a dyslexic boy by Drake in 1968 confirmed this initial hypothesis (reviewed by Habib 2000). Several malformations were observed in the cortical gyri of the left inferior parietal region, including ectopias in the outer (molecular) cortical layer. More recently, structural and functional neuroimaging studies of normal and impaired readers have provided insight into the neural systems involved in reading and dyslexia.

Magnetic resonance imaging (MRI) makes use of nuclear magnetic resonance to examine gross anatomical features of the brain. Diffusion tensor imaging (DTI) is an MRI technique that measures the restricted diffusion of water in tissue. It is predominantly used in studies of white matter, where the location, orientation, and anisotropy of the axonal tracts are measured. The architecture of the axons in parallel bundles and their myelin sheaths facilitate the diffusion of the water molecules preferentially along the direction of the axonal tracts. Such preferentially oriented diffusion is called anisotropic diffusion. The diffusion coefficient of a particular region depends on the number, orientation, and packing density of the axons. Voxel-based morphometry is another neuroimaging technique that allows investigation of focal differences in brain volume.

In functional brain imaging it is possible to examine brain function during performance of a cognitive task. Functional MRI (fMRI) and positron emission tomography (PET) measure metabolic activity and blood flow in the brain. Performance of a specific cognitive task activates neural systems in specific brain regions, leading to changes in metabolic activity, reflected by changes in cerebral blood flow and in cerebral utilization of substrates such as glucose. fMRI is commonly used in neurological studies of language disorders as it is noninvasive and safe, and hence ideal for studying children. It measures changes in blood oxygenation levels due to changes in activity, utilizing the different magnetic resonance signals of oxyhemoglobin and deoxyhemoglobin. In contrast, PET measures neuronal activity by use of radioisotopes and therefore can not be used in children, or repeatedly in adults. Electrophysiological techniques such as magnetoencephalography (MEG) and event related potential (ERP) measures allow examination of neural function at a much smaller time scale than fMRI or PET. However, they do not provide as much information about the location of the neural activity.

1.3.1 Neurobiology of reading

Mature reading is performed by a left hemisphere network of three important systems; an anterior system in the left inferior frontal cortex; a dorsal parietotemporal system involving the angular gyrus, supramarginal gyrus, and posterior portions of the superior temporal gyrus; and a ventral occipitotemporal system involving portions of the middle temporal gyrus and middle occipital gyrus (Shaywitz et al. 2002) (Figure 1, see also Figure 3A). These systems are responsible for mapping visual (orthographic) information onto auditory (phonological) and conceptual (semantic) representations (Turkeltaub et al. 2003). The phonological system can be functionally divided into two components: the dorsal parietotemporal component (the perisylvian region) and the anterior component (Schlaggar and McCandliss 2007). The perisylvian region is thought to function as an integrative component linking orthography to phonology. The left anterior component includes the inferior frontal gyrus and has been associated with speech production and phonological processing of words. The posterior reading system in the left occipito-temporal cortex includes the visual word form area involved in visual orthographic processing. It is critical for development of skilled reading and functions as an automatic, rapid word recognition system (Schlaggar and McCandliss 2007).

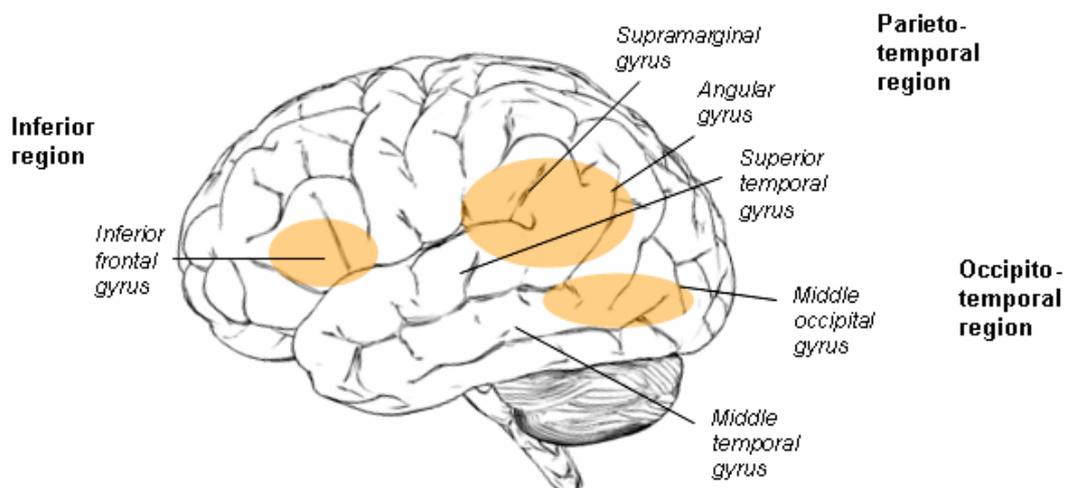


Figure 1. Left hemisphere neural systems involved in reading.

Learning to read is associated with two patterns of change in brain activity: increased activity in left middle temporal and inferior frontal gyri and decreased activity in right inferotemporal cortical areas (Turkeltaub et al. 2003). Activity in the left posterior superior temporal sulcus is associated with the maturation of the phonological processing abilities, which are crucial for fluent reading. Activity in the left ventral inferior frontal gyrus increases with reading ability and is related to phonological awareness and phonological naming ability (Turkeltaub et al. 2003). Brain activity in the anterior middle temporal gyrus also increases with reading ability. Later acquisition of a fluent reading skill is associated with engagement of the left inferotemporal word form area. Deficits in the development of these functional pathways may manifest in reading disability.

1.3.2 Neurobiology of reading disability

Early neurological studies of dyslexia attempted to identify a single marker that would aid in the diagnosis of dyslexia before children learn to read. Several differences have been indicated between dyslexic and normal readers, e.g., unusual symmetry of the cerebral hemispheres, cortical abnormalities, and different brain activation patterns and processing pathways in response to auditory and visual perception tasks. Although many areas of the brain are involved in reading, the most consistently disrupted regions in dyslexics are the left posterior parieto-temporal and occipito-temporal regions involved in phonological tasks. Thus, neurological studies support the cognitive theories of a phonological deficit in dyslexia. However, very rarely do measures of single anatomical structures distinguish dyslexics from controls; instead, dyslexic individuals mostly display a combination of impairments (Eckert 2004).

1.3.2.1 Anatomical studies

Galaburda and co-workers observed in postmortem studies abnormal accumulations of neurons (ectopias) in dyslexic brains, predominantly in the left hemisphere, including the perisylvian, temporo-occipital, temporoparietal, and frontal regions (Galaburda and Kemper 1979; Galaburda et al. 1985; Humphreys et al. 1990) (Figure 2A). These molecular layer ectopias and focal microgyri were suggested to result from deficits in neuronal migration. Ectopias are nests of neurons that have missed their target in the cortex during neuronal migration in brain development, and have therefore escaped into the molecular layer of the cortex, accompanied by mild disorganization of the adjacent cortical layers. Microgyri are more severe disturbances in neuronal migration where the organization of all layers of the cortex is severely affected. Moreover, Galaburda and others observed dyslamination of the cortical layers in the perisylvian language areas (Galaburda and Kemper 1979; Galaburda et al. 1985). Since these early postmortem studies, brain imaging and morphometry studies have found consistent anatomical differences between dyslexic and nonimpaired readers, namely in the planum temporale, inferior frontal gyrus, thalamus, corpus callosum, cerebellum, and white matter (Figure 2A and B).

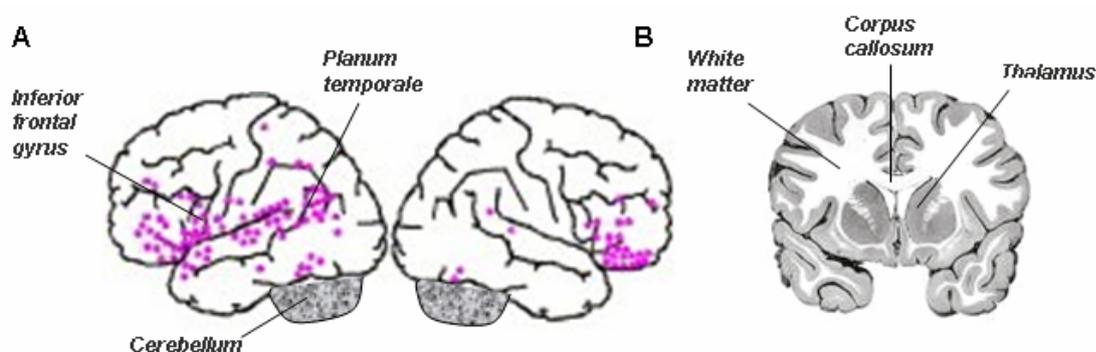


Figure 2. Brain regions showing anatomical differences in dyslexic subjects. (A) The distribution of cortical ectopias observed in dyslexic subjects is shown by red dots in the two brain hemispheres. (B) A coronal section of the human brain. Adapted from TRENDS in Neurosciences, 27; Ramus, Neurobiology of dyslexia: a reinterpretation of the data; 720-726. Copyright 2004, with permission from Elsevier.

The planum temporale within the posterior superior temporal gyrus forms the heart of the Wernicke's area, one of the most important functional areas for language in the perisylvian cortex. Therefore, it has been the most studied brain region in anatomical studies of dyslexia. Normally, the planum temporale is larger in the left than in the right hemisphere. In contrast, postmortem and imaging studies have observed unusual symmetry in the planum temporale of dyslexic subjects (Galaburda et al. 1985; Hynd and Semrud-Clikeman 1989). However, later studies have found exaggerated left asymmetry in dyslexic subjects or no asymmetry at all (Heiervang et al. 2000; Robichon et al. 2000b; Leonard et al. 2001; Eckert et al. 2003).

The inferior frontal gyrus functions in several different language tasks and plays a role in integrating different brain regions. Similarly to other language regions, it normally shows leftward asymmetry, whereas MRI studies of dyslexic adults have shown a larger right inferior frontal gyrus (Robichon et al. 2000b). A voxel-based morphometry study reported decreased gray matter in the left inferior frontal gyrus of dyslexic adults (Brown et al. 2001). The Broca's area within the inferior frontal gyrus has long been known as a critical region in language development, and it has been suggested that anomalous asymmetry relates to phonological segmentation (Robichon et al. 2000b).

The thalamus functions primary in transmission of sensory signals to the cortex. Dyslexic individuals show disproportionate number of small neurons in the left medial geniculate nucleus (MGN) (Galaburda et al. 1994), and disorganized and smaller cell bodies in the magnocellular layers of the lateral geniculate nucleus (LGN) (Livingstone et al. 1991). The anomalies in MGN have been hypothesized to be the cause of auditory deficits, and the ones in LGN the cause of the visual deficits, supporting the sensory deficit theories of dyslexia (Livingstone et al. 1991; Galaburda et al. 1994).

The corpus callosum is a central nerve bundle that allows communication between the right and left hemispheres of the brain. Some studies have found differences in the size, shape, and position of the corpus callosum in dyslexic individuals, resulting in communication deficits between the two hemispheres (Robichon and Habib 1998; Robichon et al. 2000a; von Plessen et al. 2002).

The cerebellum plays an important role in the integration of sensory perception and in motor control. Disrupted cerebellar pathways and/or primary cerebellar impairment have been proposed in the etiology of dyslexia (Nicolson et al. 2001), and reduced volume of left cerebellar gray matter has been reported in dyslexic individuals (Rae et al. 2002; Eckert et al. 2003).

White matter is composed of myelinated axons which connect various gray matter areas. DTI studies have shown bilateral differences between dyslexic and control individuals in white matter microstructure underlying the temporal-parietal areas (Klingberg et al. 2000; Deutsch et al. 2005). Reading and spelling performance correlated with anisotropy of white matter pathways in the left hemisphere, with lower anisotropy in the temporo-parietal region of dyslexic subjects. Lower anisotropy could reflect decreased myelination or number of axons, or structural disruption of the white

matter tracts. These results suggest differences in connectivity between cortical regions in dyslexic subjects (Klingberg et al. 2000; Deutsch et al. 2005).

The functional significance or the role in the etiology of dyslexia of many of the above mentioned anatomical regions is not yet clear. Functional brain imaging studies have attempted to resolve this by studying the differences in brain activation patterns between dyslexic and nonimpaired readers during reading- and language-related tasks.

1.3.2.2 Functional brain imaging studies

Functional brain imaging studies in dyslexic individuals have consistently shown impaired function of the left hemisphere posterior brain systems during reading. Specifically, dyslexic individuals show a deficiency within the phonological system in the temporo-parietal-occipital brain region involved in grapheme-phoneme conversion (Shaywitz et al. 1998; Paulesu et al. 2001; Shaywitz et al. 2002) (Figure 3). Also the cerebellum and thalamus have shown decreased activity in dyslexics as compared to controls during phonological tasks (Brunswick et al. 1999; McCrory et al. 2000). Differences in brain activation in the inferior frontal gyrus have also been suggested although these results have been inconsistent, as in some studies dyslexics show increased activation (Shaywitz et al. 1998; Brunswick et al. 1999; Georgiewa et al. 2002), and in some reduced activity has been observed (Corina et al. 2001; Ruff et al. 2002; Shaywitz et al. 2002). Shaywitz et al. (2003) proposed that dyslexic adults with a good reading accuracy and poor fluency activate the left inferior frontal gyrus, and individuals impaired in both components show decreased activity.

A functional brain imaging study of adult dyslexics from different cultures (English, French and Italian) showed same abnormal patterns of brain activation during reading and phonological tasks across languages, i.e., reduced activity in the left hemisphere (Paulesu et al. 2001) (Figure 3). The region showing most significant reduction in activation was the middle temporal gyrus, with marked decrease also in the inferior and superior temporal gyri and the middle occipital gyrus. Reduced activation in these regions has also been shown in MEG studies of Finnish dyslexics (Salmelin et al. 1996). These results suggest common neurological mechanisms and an underlying phonological deficit regardless of language, while the variation in prevalence could reflect difficulties specific to each language, when homogeneous diagnostic criteria are applied (Paulesu et al. 2001). However, neuroimaging studies in Chinese dyslexic individuals showed no deficit in the left temporo-parietal region (Siok et al. 2004). Instead, Chinese dyslexics displayed a functional disruption of the left middle frontal gyrus. This suggests that different writing systems, i.e., logographic vs. alphabetic, may lead to different neurofunctional organization patterns (Siok et al. 2004).

A common interpretation of the functional brain imaging results in dyslexia is that decreased occipitotemporal activity corresponds to deficits in word recognition (orthographic coding), decreased temporoparietal activity corresponds to difficulties in phonological processing, and increased activity in the inferior frontal gyrus relates to compensatory processes (Peterson et al. 2007). Dyslexic children compensate their impaired posterior systems by shifting to other ancillary sites, e.g., anterior sites in both

hemispheres such as the inferior frontal gyrus, as well as the right hemisphere analogue to the left occipito-temporal word form area (Shaywitz et al. 2002). The anterior sites are critical in articulation, and may therefore help the dyslexic child to develop an awareness of the sound structure of the word by forming the word with his lips, tongue, and vocal apparatus and thus allow the child to read, although more slowly and less efficiently than if the fast occipito-temporal word identification system were functioning (Shaywitz et al. 2002). The right hemisphere may represent sites that allow the poor reader to use other perceptual processes to compensate for the poor phonological skills (Shaywitz et al. 2002). Phonologically mediated reading intervention in dyslexic children has been shown to improve the disrupted function of brain regions associated with phonological processing, and produce compensatory activation in other brain areas, improving the reading fluency (Shaywitz et al. 2004).

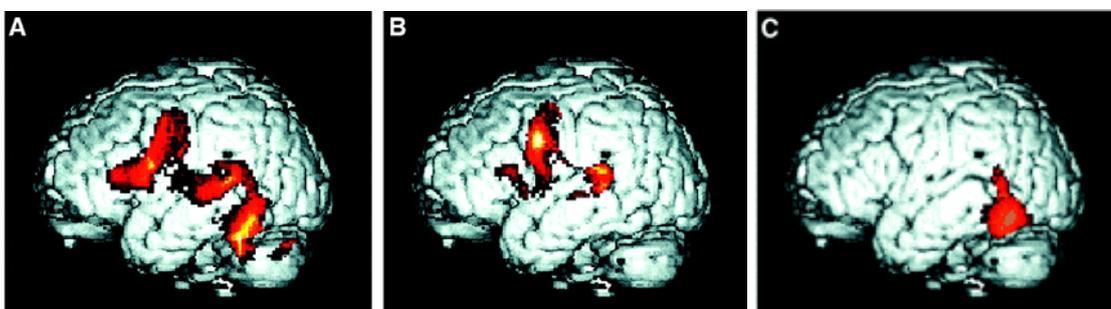


Figure 3. Brain areas activated during reading in normal (A) and dyslexic (B) individuals from three countries: UK, France, and Italy. The brain areas that were significantly more active in all normal compared to all dyslexic readers are shown in (C). Reprinted from *Science*, 291; Paulesu et al., Dyslexia: cultural diversity and biological unity; 2165-2167. Copyright 2001, with permission from AAAS.

1.4 GENETICS OF COMPLEX DISEASES

The completion of the human genome sequencing project in the beginning of the 21st century and the generation of high-density catalogues of common genetic variation in the human genome have offered considerable advances for the mapping of complex genetic traits. The SNP consortium has identified and mapped common variation in the human genome (Sachidanandam et al. 2001), and the International HapMap Consortium has genotyped these single nucleotide polymorphisms (SNPs) in different populations and created a haplotype map of SNPs that tend to be inherited together (International HapMap Consortium 2005). Moreover, high-throughput molecular genetic techniques have undergone a rapid development in the past decade, making it possible to perform large genome-wide scans at high density instead of relying on often weak a priori hypotheses.

The common diseases in the population, such as diabetes, cardiovascular disease, asthma, and the developmental and learning disorders, have a complex basis with several interacting genetic and environmental determinants contributing to the disease susceptibility. Each gene only contributes a small fraction to the overall heritability and allelic variants at multiple loci contribute to an increased risk. The relationship between genotype and phenotype is further complicated by genetic and allelic heterogeneity, i.e., there are different underlying susceptibility loci and alleles in different families. Moreover, in complex diseases there are often individuals carrying the risk variant but who do not manifest the trait, as well as affected individuals who do not carry the risk variant. These are cases of incomplete penetrance and phenocopies, respectively. Finally, some genes function differently depending on the parent-of-origin, known as imprinting. All these factors complicate the identification of genes that contribute to complex traits.

1.4.1 Genetic studies of complex diseases

The most commonly used approach to unravel the genetic basis of a complex trait is 'positional cloning'. The first step towards identifying the susceptibility gene is linkage analysis, where the goal is to establish statistically significant genome-wide evidence for linkage, followed by refinement of the chromosomal region by association. Finally, identification of the causal variant(s) and the etiological mechanism of the putative candidate gene provide final proof in its role in the disease under study.

In traditional mapping of genes for Mendelian disorders the usual first step involves the localization of a gene to a narrow genetic interval using parametric linkage analysis. Because of the strong relationship between genotype and phenotype, very narrow intervals can be defined. Identification of coding variants in affected individuals usually provides proof for the identity of the disease gene. Thus, many genes underlying Mendelian traits have been discovered in genetic mapping studies. However, monogenic diseases are rare, in fact it has been suggested that very few traits are truly monogenic, most are genetically complex (Nadeau 2001). The mapping of susceptibility genes for complex traits is complicated by their complex nature, severely limiting the power of traditional genetic linkage analysis which assumes single-gene inheritance and relies on precise specification of the transmission pattern, penetrance levels and phenocopy rates (Fisher and DeFries 2002). According to the common disease/common variant hypothesis many genetic variants that underlie common complex diseases are common, and therefore susceptible to detection (Reich and Lander 2001). However, there have been arguments against this hypothesis, proposing that the genetic contributions to complex diseases arise from many rare variants (Zwick et al. 2000). If the genetic variations influencing complex diseases are rare, the required sample sizes for identification of an effect are very large (Zondervan and Cardon 2004). In any case, as each variant contributes only with a small effect, large sample sizes are usually needed to identify the underlying susceptibility genes.

A prerequisite for successful genetic mapping is careful assessment of the phenotype. The choice of phenotype to be analyzed is not always straightforward. For Mendelian disorders, the individuals are usually classified as affected or unaffected. However, for complex traits, there is no clear definition of the phenotype. The diagnosis is often based on several different tests for multiple component phenotypes, and arbitrary thresholds are applied for phenotype classification. The phenotypic measures and diagnostic criteria used vary, leading to inconsistencies between studies. Simplifying a quantitative trait to a binary phenotype may also lead to loss of power (Fisher and DeFries 2002). Recently, many studies have started to use quantitative trait locus (QTL) based methods that investigate directly the effects of quantitative measures rather than analyzing a dichotomous arbitrary variable. Genetic heterogeneity can also seriously affect the power of a study, and may be reduced by limiting the analysis to precisely defined subgroups of the disease, or targeting families in isolated populations with a homogeneous genetic background (Lander and Schork 1994; Kere 2001).

1.4.2 Linkage analysis

Genetic linkage analysis uses genetic polymorphisms, usually microsatellites, to track inheritance of chromosomal regions within families. Microsatellites are short tandemly repeated sequences, usually of 2-4 nucleotides, and the number of copies differs between individuals. Two genetic loci that are close together on a chromosome tend to be inherited together in families, i.e., they are linked. When homologous chromosomes pair at meiosis, recombination may occur, separating loci that were previously together (Figure 4A). The probability of recombination, the recombination fraction θ , is a function of the distance between two linked loci. This genetic distance is expressed in centimorgans (cM), where 1 cM equals to a 1% chance of recombination at each meiosis, corresponding to approximately 1 Mb of DNA. In linkage analysis the recombination fraction between individual markers with known position and the disease locus is estimated. The overall likelihood on the two alternative assumptions that the loci are linked (recombination fraction = θ) or not (recombination fraction = 0.5) is calculated, and the ratio of these two likelihoods gives the odds of linkage. Linkage is reported as the logarithm of odds (LOD) score. The maximum LOD score is obtained by maximizing the estimate of θ . Linkage analysis can be made more powerful by use of multipoint analysis, where the location of a disease gene is estimated in combination with many linked loci simultaneously and the LOD score is maximized with respect to the map position.

In Mendelian disorders, where the mode of inheritance is known, standard parametric linkage analysis is used. A model for disease transmission must be specified, i.e., the frequency of the disease- and marker alleles, the mode of inheritance, and the penetrance of the disease. When studying complex diseases with several genetic components contributing to the trait and there is no clear mode of inheritance, model-free (non-parametric) methods should be used. They do not rely on assumptions of the mode of inheritance, allele frequencies, penetrance levels, or phenocopy rates. However, they require much larger datasets to yield sufficient power for the

identification of a susceptibility gene (Lander and Schork 1994). These methods exploit the fact that affected relatives display excess sharing of haplotypes identical by descent (IBD) in the region where the susceptibility gene is located. The simplest approach is to study affected sib-pairs, where linkage is demonstrated if the siblings share significantly more alleles IBD than would be expected by chance. To estimate the extent of allele sharing, maximum likelihood methods are applied. Alternative methods based on IBD sharing have been developed to analyze families with large numbers of affected relatives, as well as for analyzing QTLs. QTL approaches often have advantages over a qualitative approach as they directly exploit additional phenotypic information that is available from quantitative data (Fisher and DeFries 2002). Simple implementations of QTL mapping use regression analysis in sib-pairs to assess genotype-phenotype relationships, such as the Haseman-Elston and DeFries-Fulker methods (Haseman and Elston 1972; Cardon and Fulker 1994; Fisher and DeFries 2002). Multivariate linkage methods consider simultaneously each component in the context of the other components, allowing estimates for the effect of each trait (Marlow et al. 2003).

Genome-wide scans are the most thorough way of investigating genetic linkage. Polymorphic markers covering the whole genome are analyzed and IBD sharing is estimated by multipoint analysis using information of all markers on a chromosome. Single extended pedigrees or large samples of sib-pairs may be used. As multiple independently segregating genomic regions are analyzed, stringent thresholds should be adopted for declaring significant linkage to avoid false positive findings. Lander and Kruglyak (1995) proposed guidelines for considering results statistically significant or as suggestive evidence that need further proof. According to these criteria, LOD scores of >3.6 ($p\text{-value} < 4.9 \times 10^{-5}$) should be obtained in linkage analysis of complex traits to achieve a genome-wide significance of 5%. A finding should be considered suggestive if the probability of occurring by chance is one per genome-wide scan (corresponding to $\text{LOD} > 1.9$, $p\text{-value} < 1.7 \times 10^{-3}$). However, genome-wide scans of complex disorders rarely yield strong enough results to be considered significant according to the Lander and Kruglyak criteria (Altmuller et al. 2001). These criteria assume full information content, which is seldom achieved in genome-wide scans. An empirical approach for establishing rigorous thresholds for statistical significance is to use permutation testing on the exact family structures and markers employed in the study. Nevertheless, several results will still be false positives due to chance and multiple testing, so replication of the results on an independent sample set is crucial before establishing a genetic effect of a particular locus (Lander and Kruglyak 1995).

Because linkage focuses on families with recent ancestry, there are relatively few meioses and thus few recombinations. Therefore, disease loci identified by linkage are typically large, covering tens of cM. This corresponds to tens of Mb of genomic DNA, encompassing tens or even hundreds of genes. Moreover, the location of linkage peaks across different studies is usually highly variable (Roberts et al. 1999). Resolution of linkage to a finer scale in complex diseases is difficult, as phenocopies, incomplete penetrance, and heterogeneity all distort the linkage signal. These regions are therefore commonly refined in association studies using high-density linkage disequilibrium (LD) mapping of SNPs (Figure 4).

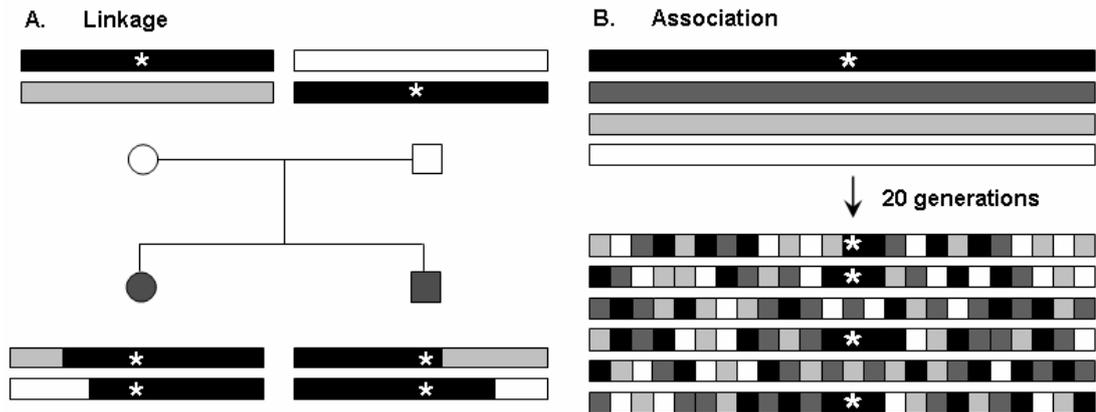


Figure 4. Linkage vs. association. Both linkage and association rely on the co-inheritance of adjacent DNA variants, which are separated primarily by recombination. Because linkage focuses on families of recent ancestry, in whom there have been relatively few opportunities for recombination to occur, disease loci that are identified by linkage will often be large (**A**). In contrast, association studies look at population-level recombination over many generations, so disease-associated regions are comparatively small (**B**). The disease locus is denoted by an asterisk. Adapted from Nature Reviews Genetics, 2; Cardon and Bell, Association study designs for complex diseases. Copyright 2001, with permission from Macmillan Publishers Ltd.

1.4.3 Association analysis

Genetic association studies aim to detect association between a polymorphism and a trait of interest. In contrast to linkage analysis where inheritance of markers is studied in families, association analysis tracks alleles associated with a trait across populations. Association analysis makes use of the existence of LD between markers at the population level. The most commonly used markers are SNPs, as they are abundant within the human genome. As association looks for historical recombination within populations across hundreds or thousands of generations, LD between two loci is maintained only if they are very close together (Figure 4B). Because LD decays rapidly over time and genetic distance, a much greater density of closely spaced markers than in linkage analysis is necessary (Cardon and Bell 2001). Other factors influencing LD are genetic drift, population growth and admixture, natural selection, recombination, and mutation (Ardlie et al. 2002). An observed association between a marker and a trait may be either direct, where the marker under study is the actual causal variant, or indirect, where the marker is in LD with the causal variant. In association studies of complex traits, the causal variants are usually not known and indirect studies are carried out.

Association analysis has greater statistical power than linkage analysis for finding genes with small effect in complex diseases (Cardon and Bell 2001). The power of association studies depends not only on the physical distance between a marker and a trait locus, but also on the contribution of the particular locus to the phenotype, i.e., its effect size, the allele frequencies for marker and trait loci, allelic and genetic heterogeneity, and the sample size (Cardon and Bell 2001). Larger sample sizes are needed when the effect sizes are weak, alleles rare, or LD incomplete. Greater power in

LD mapping is obtained by use of densely spaced markers and haplotype analysis, in which several adjacent markers in LD are considered simultaneously as haplotypes rather than individually. This allows the inference of likely historical crossover points, which localize the disease locus (Botstein and Risch 2003). Closely linked genetic markers are often transmitted as evolutionary conserved haplotype blocks (Gabriel et al. 2002). A haplotype block shows strong intermarker LD and limited haplotype diversity, whereas between the blocks there is little LD. It has been proposed that haplotype blocks are separated by recombination hot-spots, but chance probably plays an important role as well (Wall and Pritchard 2003). Although the extent of LD varies across genomic regions and populations, block boundaries are relatively consistent between populations (Conrad et al. 2006). The International HapMap Consortium has generated dense genome-wide SNP maps and has characterized the LD between the SNPs in different populations (International HapMap Consortium 2005). This has eased the association studies tremendously by limiting the number of markers to be typed to haplotype tagging SNPs; the minimum number of SNPs to be genotyped that retain as much as possible of the genetic variation (Carlson et al. 2004). When LD is high, evidence for association can be found with only a few tagging SNPs, however, refinement of the region and the precise localization of the genetic variant may be difficult. When LD within a genomic region is low, large numbers of SNPs at a high density are needed to identify a potential effect. However, when LD is low, haplotypes are useful in refining the SNP-phenotype association only if they help delineate rare allele frequencies or if there are significant interactions among the SNPs affecting the trait (Palmer and Cardon 2005). In addition, haplotypes may be important as different combinations of alleles in the same gene may have different effects on the protein product and on transcriptional regulation.

The simplest association study design is to compare marker allele frequencies at a genetic locus between affected and control individuals, or between disease chromosomes and control chromosomes. A problem in association studies using unrelated cases and controls is population stratification; subgroups in a population with different allele frequencies independent of disease, which can lead to false positive association results or failure to detect true effects. Therefore, careful matching of controls is required. Family-based association studies overcome the population stratification problem by providing an inherently matched control sample. The most commonly used test is the transmission disequilibrium test (TDT) (Spielman et al. 1993). The TDT compares the frequencies of transmitted vs. untransmitted alleles in affected offspring, by using the untransmitted parental alleles as controls. Various extensions to the TDT have been developed, such as for handling missing data, multiple siblings, or quantitative traits (Clayton and Jones 1999; Abecasis et al. 2000; Martin et al. 2000; Dudbridge 2003).

The large number of genotype-phenotype associations tested, i.e., several SNPs and haplotype combinations, and often also many phenotypes, creates many false-positives by chance. If associations are sought for at the genotype level, several genetic models (dominant, recessive, and additive) may be tested. Moreover, one may adjust for sex, disease subtype, age, parent-of-origin, or study interaction effects. Any of these tests may be entirely justified on the basis of prior biological hypotheses or in an effort to

replicate specific previous findings. However, the multiple testing will inflate the false positive rate, and must therefore be accounted for. The high degree of correlation between all tests can make determination of the extent of correction very difficult (Hattersley and McCarthy 2005). The standard Bonferroni correction is often over-conservative, as it assumes independence of all tests performed. Typically, many of the markers studied are not independent as they are in LD with each other. On the other hand, haplotype tagging SNPs are chosen to be as independent as possible, and therefore need a more stringent correction. The a priori probability of association should be accounted for, rather than the number of tests (Thomas and Clayton 2004). Bayesian methods allow calculation of a posterior probability of a true association when the prior probability of association is known, but they require also knowledge of the distribution and size of the underlying effects (Thomas and Clayton 2004). Nevertheless, for conclusive proof of an effect, the obtained association results should be replicated in independent sample sets, preferably even from different populations.

1.4.4 Chromosomal rearrangements

Balanced translocations and other chromosomal abnormalities co-segregating with a disorder are valuable tools for identifying the susceptibility gene. These rearrangements may truncate or inactivate the gene at the breakage site. The chromosomal breakpoints can be exactly mapped, in contrast to the wide intervals characteristic of linkage and even association studies. However, even in families where the breakpoint has been mapped, identification of the underlying susceptibility gene is not always straightforward. The rearrangement may influence regulatory elements, possibly affecting genes hundreds of kb away from the breakpoint by means of long-range position effects. Moreover, most chromosomal alterations do not have any phenotypic consequences. The effects of the rearrangement may also be restricted to the specific family carrying the rearrangement. Nevertheless, a chromosomal abnormality may offer a shortcut in finding a candidate susceptibility gene (Taipale et al. 2003; Hannula-Jouppi et al. 2005).

1.4.5 Candidate gene studies

Candidate genes in complex disorders are generally obtained by positional cloning, i.e., on the basis of their position obtained from genetic linkage or association studies. A candidate gene may also be suggested on the basis of its function, but for complex diseases the precise etiological background is usually not known and there are many possible candidates. Thus, selecting candidate genes purely based on function for screening is not straightforward in complex disorders.

When an association to a given phenotype is found and verified by replication, it is only the beginning of understanding its etiology and function. The QTLs identified are often large, containing numerous genes. Often the associated SNP is not the causal variant; rather it is in LD with it. To identify the underlying susceptibility gene and separate the

causal variants from normal human variation is a difficult task. The ultimate demonstration that a gene is responsible for a disease phenotype is the identification of mutations in affected subjects. In coding regions the functional consequences of missense, nonsense, and splicing polymorphisms can generally be easily assessed. However, in a complex disease the causal variants are often non-coding, in a possible regulatory region of the gene. Identification of the exact variants in the regulatory region may be difficult due to LD, and regulatory elements may be located just upstream of the gene, in introns, or more distal even hundreds of kb away. In addition to transcription, regulatory variants could affect the stability, splicing, localization and translation of mRNA. Functional studies of the candidate gene are required to determine the consequences of the causal variants.

The most conclusive evidence for a variant to be causal is to demonstrate its effect on the function of the encoded protein, preferably in animal models, or, if possible, by measurements in affected individuals. The effects on the phenotype may be difficult to prove when the variant is in a regulatory region. The putative causal variant may alter gene expression or the function of the protein product. The expression of the candidate gene and the function of the encoded protein should be thoroughly studied, and the tissue expression patterns and cellular distribution should be appropriate for the disorder under study. *In vitro* functional tests, e.g., reporter assays may be carried out to study the effect of the putative variant on gene expression. Animal models may be used to study similar phenotypes in other species. However, for many complex disorders, not the least for behavioral learning disorders, animal models are not directly comparable. The degree of evolutionary conservation is an important predictor of clinical significance, as it may highlight important regions intolerant to change, e.g., with important regulatory effects.

1.5 GENETICS OF DYSLEXIA

The familial transmission of developmental dyslexia has been recognized already in the beginning of the 20th century (reviewed by Temple 1997). Hallgren carried out the first large-scale family study on dyslexia in 1950, and proposed that transmission was due to an autosomal dominant gene. Since then, numerous family and twin studies have reported an increased risk of reading disability in relatives of dyslexic probands; ~40% in first-degree relatives (reviewed by Grigorenko 2001). In a large twin-study of dyslexia the concordance rate was significantly higher in monozygotic (68 %) than in dizygotic (38 %) twins (DeFries and Alarcón 1996). To estimate the proportion of phenotypic variation that is attributable to genetic effects, family and twin studies have investigated both the global phenotype of dyslexia as well as the specific cognitive components contributing to the disorder. Significant heritability has been observed for reading and the reading-related component processes, with estimates ranging from 44 to 87 % (DeFries et al. 1987; Wadsworth et al. 2000; Gayan and Olson 2001; Gayan and Olson 2003). Both shared and independent genetic effects as well as non-shared environmental influences affect the development of these reading-related skills (Gayan

and Olson 2003). Heritability estimates vary depending on the analyzed phenotype and due to differences in environmental characteristics between samples. Environmental factors that have been suggested to affect dyslexia susceptibility are home language and literacy environment, education and socioeconomic level (Grigorenko 2001). Moreover, a higher IQ and parental education level is associated with higher heritability (Knopik et al. 2002; Kremen et al. 2005). The precise inheritance mode for dyslexia has not been established. Segregation analysis using a qualitative phenotype has indicated evidence of a major gene mode of transmission (Pennington et al. 1991); while analyses using quantitative phenotypic measures indicate evidence of one or a few major genes and a polygenic component (Lewitter et al. 1980; Gilger et al. 1994; Wijisman et al. 2000). Nevertheless, although there are rare families in which dyslexia appears to be transmitted in an autosomal dominant fashion (Fagerheim et al. 1999; Nopola-Hemmi et al. 2001), the inheritance is generally complex with several genes of small effects and environmental factors contributing to the susceptibility.

Due to the complexity of the phenotype and the lack of a consensus definition, different genetic studies on dyslexia have used slightly different diagnostic and ascertainment criteria, complicating the comparison of results across studies. Conventionally dyslexia has been analyzed as a categorical trait, using a qualitative or quantitative definition of the phenotype. Quantitative strategies may provide a more powerful tool in genetic analysis, as they exploit all the available phenotypic information. However, quantitative measurements provide additional information only if they are stable across time and measurements, which may not be true in dyslexia if individuals of different ages are considered. Recently, linkage and association have moved to study several subcomponents of dyslexia and to look for convergence among the measures. To date, genetic linkage studies have identified nine dyslexia susceptibility loci (*DYX1-DYX9*) in the human genome listed by the HUGO Gene Nomenclature Committee (HGNC) (Table 1). Initial evidence for many of these loci comes from genome-wide linkage scans. Several genetic studies on dyslexia have also analyzed specific chromosomal regions based on previous findings or due to a potential candidate gene, i.e., targeted linkage studies.

Locus	Chromosome	Reference (first report and first replication)
<i>DYX1</i>	15q21	Smith et al. 1991; Grigorenko et al. 1997
<i>DYX2</i>	6p21-p22	Cardon et al. 1994; Grigorenko et al. 1997
<i>DYX3</i>	2p15-p12	Fagerheim et al. 1999, Fisher et al. 2002
<i>DYX4</i>	6q11-q12	Petryshen et al. 2001, Bates et al. 2007
<i>DYX5</i>	3p12-q13	Nopola-Hemmi et al. 2001, Fisher et al. 2002
<i>DYX6</i>	18p11.2	Fisher et al. 2002, Bates et al. 2007
<i>DYX7</i>	11p15.5	Hsiung et al. 2004, Fisher et al. 2002
<i>DYX8</i>	1p36-p34	Rabin et al. 1993, Grigorenko et al. 2001
<i>DYX9</i>	Xq27	Fisher et al. 2002, de Kovel et al. 2004

Table 1. Confirmed dyslexia susceptibility loci within the human genome.

1.5.1 *DYX1* on chromosome 15q21

Smith et al. (1983) performed the first linkage analysis for dyslexia by studying centromeric heteromorphisms on chromosome 15 in extended families with autosomal dominant inheritance. The subsequent linkage study using restriction fragment length polymorphisms (RFLPs) mapped the first susceptibility locus for dyslexia to 15q15-q21 (Smith et al. 1991). Grigorenko et al. (1997) observed significant linkage for single-word reading to D15S143 in 15q21 using single-point parametric analysis in six extended families from the US. Schulte-Körne et al. (1998) and Nöthen et al. (1999) reported multipoint non-parametric linkage (NPL) to the same marker and spelling disability in seven extended families from Germany. In a variance components analysis on 111 US families, Chapman et al. (2004) found suggestive evidence of linkage for phonological decoding and single-word reading to the same region. A genome-wide scan in an unselected sample of 403 Australian families found also support for linkage within 15q21 with a spelling phenotype (Bates et al. 2007). However, not all studies have replicated the evidence for linkage of dyslexia to chromosome 15 (Bisgaard et al. 1987; Rabin et al. 1993; Fisher et al. 2002; Raskind et al. 2005).

Morris et al. (2000) used a family-based LD-mapping approach and reported significant association to microsatellites within the 15q15 region in two sets of triads from the UK. As phospholipid metabolism had been suggested to be involved in the development of dyslexia (MacDonell et al. 2000), they screened two phospholipase genes within the associated region. However, neither contributed to the association signal (Morris et al. 2004). Marino et al. (2004) confirmed the association of dyslexia to the 15q region in an Italian sample set. Figure 5 summarizes the linkage and association results reported for the *DYX1* locus.

There is also evidence that the *DYX1* locus has pleiotropic effects, affecting also related developmental disorders with a shared etiology. Specifically, ADHD (Bakker et al. 2003; Loo et al. 2004) and SSD (Smith et al. 2005; Stein et al. 2006; Miscimarra et al. 2007) have shown evidence of linkage to the 15q14-q21 region.

1.5.1.1 *DYX1C1* – the first candidate gene for dyslexia

Nopola-Hemmi et al. (2000) described two families co-segregating balanced translocations and dyslexia; t(2;15)(q11;q21) and t(2;15)(p13; q22), respectively. The translocation breakpoints were mapped in both families to a narrow interval within the *DYX1* locus, 6-8 Mb apart from each other. Taipale et al. (2003) identified the first candidate gene for dyslexia, *DYX1C1* (dyslexia susceptibility 1 candidate 1; also known as *EKNI*, en keksi nimeä 1), by cloning the translocation breakpoint in the family co-segregating t(2;15)(q11;q21) and dyslexia. By sequencing, eight SNPs were identified within the *DYX1C1* region and genotyped in two independent Finnish case-control sample sets (totally 106 cases and 194 controls). The minor alleles of two polymorphisms; -3G→A (from here on referred to rs3743205 for clarity) in the translation initiation site disrupting three predicted transcription factor binding sites, and 1249G→T truncating the protein by the four last amino acids, were associated to

dyslexia both in single-marker ($P=0.002$ and $P=0.006$, respectively) and in haplotype analysis ($P=0.015$).

Since these initial findings, seven independent groups have attempted to replicate the association with *DYX1C1* and dyslexia. The results have, however, been inconclusive, with either negative or weak positive associations to the opposite alleles reported. Wigg et al. (2004) analyzed six polymorphisms within *DYX1C1*, including the two that were initially found associated to dyslexia (Taipale et al. 2003), in 148 Canadian families. The two SNPs showed modest evidence of association to the categorical trait of dyslexia, but only at the haplotype level and for the opposite alleles as reported by Taipale and co-workers. The major allele of rs3743205 was also associated to all quantitative traits analyzed. In addition, an intronic SNP (rs11629841) showed significant single-marker and haplotype association with dyslexia. Scerri et al. (2004) studied the eight variants identified by Taipale and others within *DYX1C1* in 264 nuclear families from the UK using quantitative trait association analysis. Although they concluded that *DYX1C1* was not associated with dyslexia in their sample set, 1249G→T showed nominally significant association with orthographic coding, both in single-marker and haplotype analysis. Brkanac et al. (2007) found also evidence of association of 1249G→T with dyslexia in their sample set of 191 multigenerational families from the US. However, these associations were again with the opposite allele than described by Taipale et al. (2003). Four studies using samples from the UK (Cope et al. 2005b), US (Meng et al. 2005a), and Italy (Bellini et al. 2005; Marino et al. 2005) did not find any evidence of association of the two SNPs within *DYX1C1* with dyslexia as a categorical trait or with any quantitative trait analyzed in single-marker or in haplotype analysis. In addition, Cope et al. (2005b) studied the intronic SNP (rs11629841) showing association in the Canadian samples (Wigg et al. 2004), but again, the results were negative.

The *DYX1C1* variants were also investigated in 100 Finnish families ascertained for autism, but no evidence for single-marker or haplotype association was observed (Ylisaukko-Oja et al. 2005).

1.5.2 *DYX2* on chromosome 6p21-p22

Smith et al. (1991) reported the first evidence for a gene on chromosome 6 that might influence reading deficits. Subsequently, Cardon et al. (1994) extended the study by analyzing an additional sample set from the Colorado twin study of reading disability (DeFries et al. 1987), and using targeted linkage analysis within the human leukocyte antigen (HLA) region of chromosome 6p, as elevated risk for autoimmune disease had been observed in relatives of dyslexic probands (Pennington et al. 1987; Hugdahl et al. 1990). Using a continuous measure of reading performance, a quantitative trait locus (QTL) was mapped to 6p21.3 in two independent sample sets from the US. Analysis of more severe deficits in reading performance provided even stronger evidence for linkage to this locus. The location of a QTL on chromosome 6p21.3 influencing dyslexia was further confirmed in subsequent linkage studies using independent sample

sets analyzing the various subcomponents of dyslexia. Grigorenko et al. (1997; 2000; 2003) analyzed a categorical definition for each subphenotype, and reported significant evidence for several reading-related components in their US sample set. Using a QTL approach, Gayan et al. (1999) and Fisher et al. (1999) also identified significant linkage for reading ability as well as several subcomponents that influence reading and language in samples from the US and UK, respectively. Fisher et al. (2002) performed the first QTL-based genome-wide scan for dyslexia in three independent sample sets, including the previously used samples from the UK (Fisher et al. 1999) and from the Colorado study (Cardon et al. 1994; Gayan et al. 1999), as well as additional new families from the UK. Although several loci within the human genome showed significant evidence of linkage, the *DYX2* was one of the most significant, in particular in the UK sample set (Fisher et al. 2002).

Kaplan et al (2002) performed a finer-scale linkage and association study in the US Colorado sample set (Cardon et al. 1994; Gayan et al. 1999). Consistent with the previously reported findings, they observed significant linkage to several reading-related measures. Quantitative association analysis pointed to a microsatellite marker within the *KIAA0319* gene on 6p22. Deffenbacher et al. (2004) used an extended set of the US Colorado sample (Cardon et al. 1994; Gayan et al. 1999; Kaplan et al. 2002), and by selecting for severity, significant linkage was detected to *DYX2* for all subphenotypes analyzed. They further refined the linkage region by genotyping a dense map of SNP markers over a 680 kb covering ten genes with a putative function in the central nervous system (CNS). Significant associations were obtained for two clusters of genes: *VMP/DCDC2/KAAG1* (vesicular membrane protein 1/doublecortin domain containing 2/kidney associated antigen 1) and *KIAA0319/TTRAP/THEM2* (*KIAA0319*/TRAF and TNF receptor associated protein/thioesterase superfamily member 2), in particular when selecting probands with extreme scores of reading. Francks et al. (2004) genotyped SNPs spanning the *KIAA0319/TTRAP/THEM2* locus in three sample sets; two from the UK and one from the US Colorado twin study (Fisher et al. 1999; Fisher and DeFries 2002). Quantitative association analysis identified a risk haplotype for several reading-related components spanning the *TTRAP* and *KIAA0319* genes in all three samples. Again, the association was strongest with more severely affected probands. Turic et al. (2003) genotyped microsatellites within the *DYX2* region in two independent sample sets from the UK. Association was found with dyslexia and several of its subphenotypes. Cope et al. (2005a) performed a high-density LD screen encompassing the two gene clusters *VMP/DCDC2/KAAG1* and *KIAA0319/TTRAP/THEM2*. First, a pooled case-control sample was analyzed, and only the significant SNPs were further genotyped in a family-based sample, all ascertained for severe dyslexia. Significant evidence of association with dyslexia was observed for several SNPs within *KIAA0319*.

Although there have been also many negative linkage findings to the *DYX2* locus on 6p22 (Field and Kaplan 1998; Schulte-Körne et al. 1998; Petryshen et al. 2000; Chapman et al. 2004; Raskind et al. 2005; Bates et al. 2007), it is the most consistently replicated locus for a complex cognitive trait. Figure 5 summarizes the reported linkage results for the *DYX2* locus. Moreover, the comorbid disorders SSD and ADHD have also been linked to the *DYX2* locus (Willcutt et al. 2002; Smith et al. 2005).

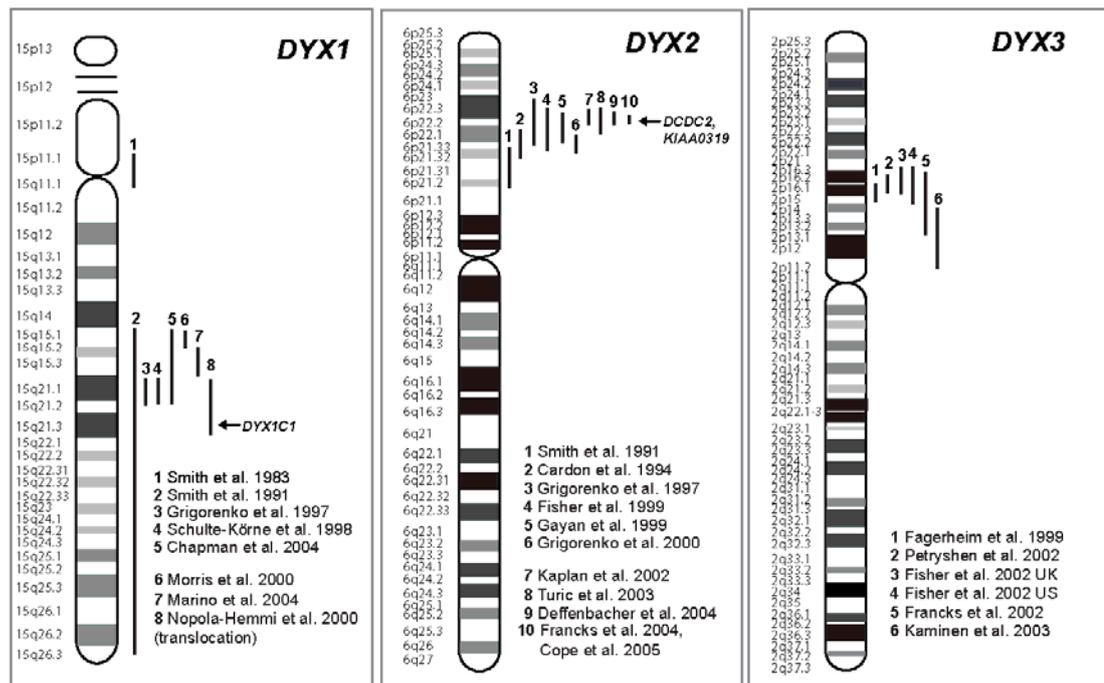


Figure 5. Reported linkage and association findings within *DYX1-DYX3*. Adapted from European Journal of Human Genetics, 14; Williams and O'Donovan, The genetics of developmental dyslexia. Copyright 2006, with permission from Macmillan Publishers Ltd.

1.5.3 *DYX3* on chromosome 2p15-p12

Four independent genome-wide scans using different analytical approaches have shown linkage of dyslexia to 2p12-p16. Fagerheim et al. (1999) studied a single extended Norwegian pedigree, in which inheritance was consistent with an autosomal dominant transmission. Parametric linkage analysis found significant evidence of linkage to 2p15-p16, which was supported by NPL analysis. Fisher et al. (2002) analyzed two large sets of nuclear families from UK and US using a quantitative non-parametric approach, and found significant single-point linkage results for orthographic coding at 2p16 in the UK sample and phoneme awareness at 2p13 in the US sample. Multi-point analysis showed evidence of linkage to 2p15 for phoneme awareness in the UK sample. Kaminen et al. (2003) analyzed a categorical phenotype for dyslexia in 11 extended Finnish families. Nonparametric and parametric linkage analysis using an autosomal dominant model peaked at 2p11. In their genome-wide scan on Australian families, Bates et al. (2007) identified linkage of reading and spelling to several loci in the human genome, among them the 2p15-p12 region. Petryshen et al. (2002) performed a targeted linkage study in Canadian families by genotyping seven microsatellite markers spanning the region on 2p15-p16 reported by Fagerheim et al. (1999). Multipoint variance component analysis of different reading-related measures yielded significant linkage for spelling. Francks et al. (2002a) genotyped microsatellites in the 2p12-p21 region and performed a quantitative sib-pair association study in the US twin-bases families (Fisher et al. 2002). Two loci at 2p21 and 2p12 yielded suggestive evidence of association for a range of reading-related measures. The most significant association

was found for word recognition and phonological coding at 2p21. Figure 5 summarizes the reported linkage results for the *DYX3* locus. As for the other dyslexia loci, there have been also negative reports to the *DYX3* locus. A variance component linkage analysis on 111 US families did not find any linkage to the *DYX3* region (Chapman et al. 2004; Raskind et al. 2005).

1.5.4 *DYX4 – DYX9*

1.5.4.1 DYX4 on chromosome 6q11-q12

Petryshen et al. (2001) reported suggestive evidence of linkage for phonological coding and spelling to the 6q11-q12 region in 96 Canadian families. The first independent support for this locus came from a genome-wide scan on 403 Australian families, where a spelling phenotype showed significant evidence for linkage (Bates et al. 2007).

1.5.4.2 DYX5 on chromosome 3p12-q13 and ROBO1

Nopola-Hemmi et al. (2001) identified linkage for dyslexia to 3p12-q13 in an extended Finnish family with an autosomal dominant inheritance pattern. The gene was identified by mapping the translocation breakpoint in an unrelated individual who co-segregated the t(3;8)(p12;q11) translocation and dyslexia (Hannula-Jouppi et al. 2005). The breakpoint disrupted an intron of the human orthologue of the *Drosophila* gene *roundabout*, *ROBO1*, on 3p12. Association analysis using SNPs in the extended Finnish family identified a dyslexia risk haplotype within *ROBO1*. The *DYX5* locus is also supported by QTL-based genome-wide scans for dyslexia, in which single- and multipoint analyses showed evidence of linkage in UK and US families (Fisher et al. 2002). Recently, a genome-wide scan on Australian families found also evidence for linkage of reading to the 3p12-q13 region (Bates et al. 2007). The *DYX5* locus is also pleiotropic for SSD (Stein et al. 2004).

1.5.4.3 DYX6 on chromosome 18p11.2

As noted already earlier, the first two QTL-based genome-wide scan for dyslexia identified several susceptibility loci throughout the human genome (Fisher et al. 2002). However, the strongest evidence for linkage was seen for 18p11.2 in two independent sample sets from the UK and US. Several reading-related measures showed linkage to this locus, although the most significant was seen for single-word reading. The results within 18p11.2 were further replicated in a third independent sample from the UK, where phoneme awareness showed the strongest evidence for linkage. Chapman et al. (2004) and Schumacher et al. (2006) attempted to replicate the findings within 18p11.2, but no linkage was observed to any of the quantitative phenotypes analyzed in their samples from the US and Germany, respectively. Nevertheless, Bates et al. (2007) reported the first independent replication of linkage to *DYX6* in a genome-wide scan on 403 Australian families.

1.5.4.4 *DYX7 on chromosome 11p15.5*

Hsiung et al. (2004) performed a targeted linkage analysis on 100 Canadian families ascertained for dyslexia. The study was based on the location of the *DRD4* (dopamine receptor D4) gene, which has been implicated in ADHD (Thapar et al. 2005). Significant evidence for linkage was observed to *DRD4* and to the proximal *HRAS* (Harvey-Ras) gene. LD analysis showed a trend of overtransmission of the ADHD risk allele in *DRD4*, although it did not reach significance. The *DYX7* locus is supported by the QTL-based genome-wide scan for dyslexia, in which single-point analysis showed evidence of linkage in the UK families (Fisher et al. 2002). However, in another genome-wide scan on Australian families, no support of linkage was found for this locus (Bates et al. 2007).

1.5.4.5 *DYX8 on chromosome 1p34-p36*

Rabin et al. (1993) reported suggestive linkage of dyslexia to 1p34-p36 in nine extended families of US origin. A targeted linkage study in eight extended dyslexic families from the US supported these initial findings (Grigorenko et al. 2001). Tzenova et al. (2004) confirmed the presence of a dyslexia susceptibility gene within the *DYX8* region. Both qualitative and quantitative linkage analysis in 100 Canadian families yielded significant evidence of linkage. In the genome-wide scan on Australian families, Bates et al. (2007) found suggestive evidence of linkage to the *DYX8* region. Froster et al. (1993) identified a balanced translocation $t(1;2)(p22;q31)$ co-segregating with dyslexia and delayed speech in a German family. However, the breakpoint on chromosome 1 resides on p22, being much more centromeric than *DYX8*. The *DYX8* locus seems to exert pleiotropic effects for speech and language, as shown in targeted linkage studies of SSD and SLI (Smith et al. 2005; Miscimarra et al. 2007).

1.5.4.6 *DYX9 on chromosome Xq27*

A genome-wide scan in a single extended family of Dutch origin yielded significant evidence for linkage of dyslexia to Xq27 (de Kovel et al. 2004). This locus showed also linkage in the QTL-based genome-wide scan in the UK sample (Fisher et al. 2002), and in the genome-wide scan on Australian families (Bates et al. 2007).

1.5.5 Other possible loci for dyslexia susceptibility

In addition to the nine confirmed dyslexia loci *DYX1-DYX9*, there is also evidence for linkage of dyslexia to chromosomes 2q22.3 (Raskind et al. 2005; Bates et al. 2007), 4p15 (Bates et al. 2007), 7q32 (Kaminen et al. 2003; Bates et al. 2007), 13q12 (Igo et al. 2006), and 17p13 (Bates et al. 2007). Moreover, the genetic loci identified in comorbid disorders may also be pleiotropic for dyslexia. Several overlapping linkage regions have already been identified for ADHD, speech and language disorders and dyslexia. The etiologies of all these neurodevelopmental behavioral disorders are complex with several genes and environmental factors contributing to the susceptibility, and they share symptoms and risk factors that partly overlap. Thus, there

may be shared genes with effects on several neurodevelopmental disorders, affecting reading, language, and learning in general.

Dyslexia and the inattention trait of ADHD show significant bivariate heritability (Willcutt et al. 2000). To analyze the shared genetic variance between ADHD and dyslexia, Gayan et al. (2005) performed a bivariate genome-wide linkage analysis in a sample set selected for reading disability but displaying also comorbid ADHD. Pleiotropic loci on 13q32, 14q32, and 20q11 were identified, as well as loci specific for each disorder. Moreover, ADHD families with comorbid dyslexia have shown linkage of a reading phenotype to 10q11, 16p12, 17q22 (Loo et al. 2004). However, not all dyslexia loci show linkage to ADHD, and vice versa, as there are expected to be independent genetic contributions to each trait as well.

In conclusion, there are many dyslexia loci, of which several have yet to be replicated in order to be confirmed. Thus, although some of these may be false positive findings, the number of dyslexia loci is likely to increase in the near future.

2 AIMS OF THE STUDY

The overall aim of the study was to map and identify candidate genes for developmental dyslexia, and more specifically to:

1. Map and identify the susceptibility gene(s) for *DYX3* on chromosome 2 using positional cloning (**Studies I and II**)
2. Map and identify the susceptibility gene(s) for *DYX2* on chromosome 6 using linkage disequilibrium mapping (**Study III**)
3. Verify the role of the first candidate gene for dyslexia, *DYX1C1*, as a true susceptibility gene using another population set (**Study IV**)
4. Identify yet a new dyslexia candidate gene on 15q21 by studying the translocation breakpoint t(2;15)(p12;q21) (**Study V**)

3 MATERIALS AND METHODS

3.1 SAMPLES

3.1.1 Finnish dyslexia cohort (I, II, V)

In total, the Finnish dyslexia cohort consisted of 19 Finnish families with 135 subjects (55 affected, 63 healthy, 17 phenotype unknown). Thirteen families were recruited from the Department of Pediatric Neurology at the Hospital for Children and Adolescents, University of Helsinki; three families were from the Association of Learning Disabled Individuals of Helsinki (HERO); and three families from the Central hospital of Central Finland, Jyväskylä. The inclusion criteria for the probands included a normal performance IQ, and remarkable deviation (≥ 2 years) in reading skills compared to chronological age. Children with an IQ < 85 or another developmental or neurological disorder were excluded. The diagnosis of dyslexia was determined by Finnish reading and spelling tests designed for children under 13 years of age and for adults. The adult reading assessment included oral text reading, pseudo- and non-word reading, and spelling to dictation. All subtests were analyzed for speed and accuracy. Neuropsychological tests were used to assess reading-related neurocognitive skills, i.e., phonological awareness, rapid automatized naming, and verbal short term memory. For a more detailed description of the diagnostic criteria and the ascertainment of phenotypes see (Nopola-Hemmi et al. 2001; Nopola-Hemmi et al. 2002). Genomic DNA was obtained from blood lymphocytes using a standard non-enzymatic extraction method (Lahiri and Nurnberger 1991).

3.1.1.1 *The t(2;15)(p12;q21) family (V)*

The family co-segregating the translocation $t(2;15)(p12;q21)$ and dyslexia was recruited from the Seinäjoki District Hospital, Finland. The father and all three children were carriers of the translocation, but only the girl had been diagnosed with dyslexia (Figure 6). The father had not been available for neuropsychological testing, although he had no history of reading and writing difficulties. All children were tested for dyslexia. The diagnosis of dyslexia was determined as described above. The proband performed normally in non-verbal psychological tests. In verbal skills, she had problems with phonological awareness, resulting in severe impairment in reading and writing. The boys had normal reading and language skills. For a more detailed description of the family, see (Nopola-Hemmi et al. 2000).

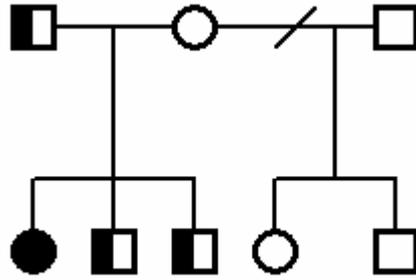


Figure 6. Pedigree of the t(2;15)(p12;q21) family. Black-and-white symbols denote unaffected carriers of the translocation, whereas the individual indicated with a solid black symbol co-segregates the translocation and dyslexia.

3.1.2 German dyslexia cohort (II, III, IV)

A total of 376 triads of German descent, including siblings, were recruited from the Departments of Child and Adolescent Psychiatry and Psychotherapy at the Universities of Marburg and Würzburg. Because clinical studies of dyslexia in Germany usually base sample selection on spelling disorder and the previous findings in this sample set rest on this criterion (Schulte-Körne et al. 1996; Schulte-Körne et al. 1998; Ziegler et al. 2005; Schumacher et al. 2006), the proband's spelling ability was used for inclusion criteria. As a spelling disorder cannot be reliably diagnosed at an early age, only those children attending a regular primary school who had reached at least the middle or second grade were included in the study. On the basis of an assumed correlation between the IQ and spelling of 0.4 (Schulte-Körne et al. 1996; Schulte-Körne et al. 2001), an expected spelling score was estimated. The child was diagnosed as dyslexic, if the discrepancy between the observed and expected spelling score was ≥ 1 SD. Spelling was measured using age-appropriate spelling tests (writing to dictation). Probands or siblings who showed symptoms of ADHD were excluded because the traits might overlap (Willcutt et al. 2002), and the symptoms of inattention and hyperactivity might influence the behavior of the child during the neuropsychological examinations. Additional exclusion criteria were a bilingual education; an IQ < 85 ; an uncorrected sensory hearing or vision disorder; or a psychiatric or neurological disorder influencing the development of reading and spelling ability. Quantitative measures of correlated phenotypic dimensions, i.e., spelling, word reading, phonological awareness, phonological decoding, orthographic processing, rapid naming (letters; numbers; colors and symbols) were analyzed in order to cover the phenotypic dimensions of dyslexia. For a detailed description of the diagnostic inclusion criteria and phenotypic measures, see (Schulte-Körne et al. 1996; Schulte-Körne et al. 2001; Ziegler et al. 2005; Schumacher et al. 2006). Genomic DNA was obtained from blood using a standard method (Miller et al. 1988).

3.1.3 US dyslexia cohort (V)

The US dyslexia cohort consisted of 13 Caucasian families of US origin with 60 subjects (22 affected, 29 healthy and 9 of unknown phenotype). The families were recruited and referred through schools, physicians, and community announcements at the Center for Clinical and Developmental Neuropsychology (CCDN) at the University of Georgia. All qualifying families had at least one proband between the ages of 8 and 12 years with significant reading problems and no history of neurological impairment, traumatic brain injury, psychiatric disorders, or severe pre- and/or perinatal complications. The test battery consisted of measures designed to assess intelligence, academic achievement, receptive and expressive language, phonological processing, memory, reading, spelling, visual-spatial ability, executive functioning, handedness, and social-emotional functioning. Genomic DNA was obtained from self-collected buccal swabs by a NaOH extraction method (Walker et al. 1999). Whole-genome amplification of the extracted DNA was performed by the Improved Primer Pre-amplification method (Hannelius et al. 2005).

3.1.4 US speech sound disorder cohort (V)

The SSD cohort consisted of 118 Caucasian families of US origin with 401 subjects (51 affected with both SSD and dyslexia, 118 affected with only SSD, 25 affected with only dyslexia, 198 unaffected with either SSD or dyslexia and eight of unknown phenotype). The proband child was identified through the clinical practices of speech-language pathologists in the Greater Cleveland area. Proband were enrolled in speech-language therapy for a moderate to severe speech sound production disorder. Children were also required to have normal hearing, as demonstrated by passing a pure-tone audiometric screening test; normal intelligence, defined as a performance IQ of at least 80; and a normal speech mechanism (adequate oral structures for producing speech sounds). An extensive battery of standardized speech sound production, receptive and expressive language, reading decoding and comprehension, spelling, oral-motor skills, memory, and phonological processing measures were administered to all probands and their siblings of four years age and older. Affected siblings met the same eligibility criteria as the probands. Affection status for parents was based on historical reports of speech-language problems and/or treatment. Parent(s) were interviewed concerning speech, language, developmental, and academic history for each child employing the child history questionnaire developed for the study. Family history information was collected on three generations of family members. Affection status for speech, language, reading, spelling, learning disabilities, ADHD, stuttering, and apraxia of speech as well as education and handedness information were recorded for each family member. Socioeconomic status was determined by the Hollingshead Four Factor Index of Social Class which employs the amount of formal education each parent has completed and the occupational level of each parent to compute a score for social class. To compare the SSD and dyslexia phenotypes, we attempted to classify individuals ascertained for SSD as affected or unaffected for reading disorder based on criteria employed in the US dyslexia cohort. Language, reading, spelling and phonological processing were measured in both cohorts, and in several instances the same test was

administered. The criteria for dyslexia in the SSD probands were a performance IQ >85 and a score of <1 SD below the mean in text reading or spelling, as well as on two or more of pseudoword reading, orthographic recognition, lexical decision accuracy, lexical decision speed, and/or segmentation. Genomic DNA was obtained from self-collected buccal swabs or blood draws.

3.2 GENETIC ANALYSIS

3.2.1 Genotyping (I – V)

3.2.1.1 Microsatellites (I – III)

Microsatellite markers were PCR-amplified, pooled, and electrophoresed on an automated capillary DNA sequencer (MegaBACE, Amersham Biosciences/GE). The alleles were visualized using the Genetic profiler v1.5, Genescan Analysis v2.1.1., or the Genotyper v2.0 software.

3.2.1.2 SNPs (I – V)

SNPs were genotyped using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. PCR assays and extension primers were designed using the SpectroDESIGNER software (Sequenom). Extension products were analyzed by a Mass ARRAY mass spectrometer (Bruker Daltonik) and peaks were identified using the SpectroTYPER RT 2.0 software (Sequenom). In study V, some SNPs were also genotyped using 5' exonuclease TaqMan Assay by Design or Assays in Demand, and the genotypes were assigned using the SDS 2.0 software. CEPH samples, negative controls, and replicates were included on each plate to assure consistency of the genotype calls. All genotypes were independently confirmed by two investigators blind to the disease status. To identify possible genotyping errors, data were checked for Mendelian consistency using Pedcheck (O'Connell and Weeks 1998), and unresolved inconsistencies were assigned as missing genotypes. Testing for Hardy Weinberg equilibrium proportions was done via a chi-squared goodness-of-fit test, using only the founders to eliminate the non-independence owing to family data. Intermarker LD was visualized and pairwise R^2 values were determined using Haploview v3.2 software (Barrett et al. 2005).

3.2.2 Linkage analysis (I)

Non-parametric multipoint linkage analysis was performed using the Genehunter 2.1 software (Kruglyak et al. 1996). The significance of the NPL score was evaluated by simulation.

3.2.3 Association analysis (I – V)

3.2.3.1 Haplotype pattern mining (I)

Haplotype pattern mining (HPM) was used to narrow down the observed linkage signal within the *DYX3* locus. HPM is a data-mining-based approach in which frequent haplotype patterns associated with a trait are sought (Toivonen et al. 2000). The haplotype patterns are sorted by the strength of their association to the trait, and a non-parametric model is used for localizing the underlying locus. HPM analyses case-control data using non-transmitted parental chromosomes as controls. The method is robust for mutations and missing data, as the haplotypes are allowed to contain gaps.

3.2.3.2 TDT (II – V)

TDTs were used to test for single marker as well as for haplotype associations (Spielman et al. 1993). The TDT compares the frequencies of transmitted vs. untransmitted alleles in affected offspring, by using the untransmitted parental alleles as controls. TDTHAP is an implementation of the TDT, that allows for missing data and phasing of uncertain haplotypes (Clayton and Jones 1999). Another implementation, QTDT, was used for analysis of quantitative component processes of the dyslexia phenotype (Abecasis et al. 2000). Various further extensions to the TDT were used to maximize power in the different datasets. The pedigree disequilibrium test (PDT) is best suitable to large pedigrees, such as the Finnish families (Martin et al. 2000). An extension of the PDT, PDTPHASE, was used to deal with missing data and for estimation of phase for uncertain haplotypes (Dudbridge 2003). The analysis of quantitative phenotypes was performed using QPDTPHASE, which is a quantitative trait implementation of the PDT. For haplotypes also global tests were performed, i.e., testing the effects for all haplotypes for a given set of SNPs instead of each haplotype independently. Statistical significance was estimated by permutation testing.

A variance component model developed for family-based association was used for single-marker analysis of quantitative traits in the two US cohorts. This method assesses association between a marker and phenotype, while simultaneously estimating residual and multifactorial (polygenic, familial, and marital) variance components. The method combines the original association method by George and Elston with a TDT-type analysis (George and Elston 1987; Spielman et al. 1993). To account for multiple tests when determining allelic association to a trait Sidak's correction was used (Sidak and Zbynek 1967).

3.2.4 DNA re-sequencing (I – V)

3.2.4.1 Genomic sequencing (II, III, V)

To characterize the LD in detail and to explore the genetic variation within the *VMP/DCDC2/KAAG1* and *MRPL19/C2ORF3* loci, the non-repetitive DNA segments

were PCR-amplified and re-sequenced by dye-terminator Sanger chemistry in dyslexic individuals carrying the respective risk alleles at each locus.

3.2.4.2 Mutation screening (I – V)

TACR1, *VMP*, *DCDC2*, *KAAG1*, *LRRTM4*, *CTNNA2*, *MRPL19*, *C2ORF3*, *CYP19A1* and *DYX1C1* were screened for coding polymorphisms by direct sequencing of all exons, exon/intron junctions, and the 5' and 3' untranslated regions (UTRs). In addition, the promoter sequences of *DYX1C1* and *CYP19A1* were analyzed. The human-specific primers were also used to sequence *MRPL19*, *C2ORF3* and *CYP19A1* orthologues in gorilla (*Gorilla gorilla*), chimpanzee (*Pan troglodytes*), pigmy chimpanzee (*Pan paniscus*) and orangutan (*Pongo pygmaeus*) DNA samples.

3.2.4.3 PCR and sequencing reactions (I – V)

All fragments were PCR-amplified and sequenced from both directions. Purified sequencing products were resolved using a MegaBACE 1000 instrument and MegaBACE long-read matrix, visualized using the Sequence Analyzer v3.0 software, and assembled and analyzed using the Pregap and Gap4 software (www.cbi.pku.edu.cn/tools/staden), comparing to the human database sequence (www.ncbi.nih.gov). Sequences were verified visually by two independent readers. In addition, a separate viewer compared each FASTA output from sequencing results to corresponding genomic sequences using Blast 2 sequences.

3.3 GENE IDENTIFICATION

3.3.1 Fluorescent *in situ* hybridization (V)

Fluorescent *in situ* hybridization (FISH) was used for mapping the t(2;15)(p12;q21) breakpoint in the dyslexic individual carrying the translocation. Metaphase spreads were obtained from EBV-transformed cell lines of the translocation carrier, and bacterial artificial chromosome (BAC) clones from chromosomes 2 and 15 were used as probes. Probes were fluorescently labeled by nick translation and FISH analyses were performed according to standard protocols.

3.3.2 Southern blot analysis (V)

Southern blot analysis was used to map the exact translocation breakpoint in t(2;15)(p12;q21). Genomic DNA from the dyslexic individual carrying the translocation and from an unrelated control were digested using restriction endonucleases, and subjected to electrophoresis and Southern hybridization as previously described (Taipale et al. 2003). PCR-amplified genomic fragments from non-repetitive regions of the chromosome 15 BAC clone RP11-108K3 were used as hybridization probes. PCR

and labeling reactions were performed as previously described (Hannula-Jouppi et al. 2005).

3.3.3 Gene predictions (II, V)

Putative genes/exons were *in silico* predicted using Genscan (genes.mit.edu/genscan.html) and GrailEXP (grail.lsd.ornl.gov/grailexp). The expression of the predicted genes/exons were tested by PCR on human cDNA libraries from leukocytes and fetal brain.

3.4 GENE EXPRESSION STUDIES

3.4.1 Transcript characterization (I – III, V)

The gene structures of *KAAG1*, *DCDC2*, *MRPL19*, *C2ORF3*, *FLJ13391* and *CYP19A1* were verified and improved by fully sequencing I.M.A.G.E. clones. The expression patterns were studied by PCR on human cDNA libraries from fetal brain and from leukocytes. PCR products were visualized by agarose gel electrophoresis and further sequenced. Rapid amplification of 5' complementary DNA ends (5'RACE) using Marathon-Ready cDNA from fetal brain tissue was performed to confirm the 5' end of *MRPL19*, *C2ORF3*, *KAAG1* and *DCDC2*.

3.4.2 Quantitative mRNA expression (II, V)

Ready-made TaqMan gene expression assays were used to study the expression levels of *MRPL19*, *C2ORF3*, *FLJ13391*, *CYP19A1*, *DCDC2*, *DYX1C1*, *ROBO1*, and *KIAA0319* in different areas of adult human brain (thalamus, hypothalamus, frontal-, occipital-, parietal-, temporal cortex, hippocampus, paracentral-, postcentral gyrus, and whole adult and fetal brain). For each tissue, three independent cDNA syntheses were performed. Quantitative real-time PCR was performed in quadruplets from each cDNA synthesis. Relative quantification of the data was performed using the comparative threshold cycle (Ct) method adjusting the Ct values to *18S* rRNA.

3.4.3 Northern blot analysis (I – III)

Northern blot analyses were performed to study the expression level, transcript size, and potential splice variants of *TACR1*, *DCDC2*, *KAAG1*, *MRPL19* and *C2ORF3* in different human tissues. cDNA probes were radioactively labeled, cleaned from unincorporated P³²-dCTP, and hybridized to commercial ready-made blots according to the manufacturer's recommendations.

3.4.4 Allele-specific expression (II)

Allele-specific mRNA expression levels of *MRPL19* and *C2ORF3* were assessed in individuals heterozygous for the risk haplotype and for a synonymous SNP in the two genes. Five Finnish dyslexic and six normal readers, and three German dyslexic and one normal reader were studied using a standard method (Pastinen et al. 2004). cDNA synthesis was performed from total RNA extracted from EBV-transformed lymphocyte cell lines. Peak heights were compared to genomic DNA and an allelic ratio was calculated for each sequence. The cDNA ratio values (unknown proportions) were normalized by dividing with the genomic values (1:1 proportion by definition). Data were pooled by genotype (risk haplotype heterozygotes vs. non-risk haplotype heterozygotes) to evaluate whether the normalized value differed from equal expression.

3.4.5 Electrophoretic mobility shift assay (IV, V)

Electrophoretic mobility shift assays (EMSA) were performed to investigate the protein binding affinities to different alleles in the promoters of *DYX1C1* and *CYP19A1*. DNA probes of 30 bp length for each allele studied were synthesized, annealed to form a duplex, and radiolabeled. The labeled probes were incubated with nuclear or whole-cell extracts from the neuroblastoma cell line SH-SY5Y. Specificity of the binding was confirmed by adding unlabelled oligonucleotide prior to addition of the labeled probe. To study the identity of the binding proteins, supershift assays using antibodies for TFII-I or Elk-1 were performed. The samples were electrophoresed on non-denaturing polyacrylamide gels and the radioactive pattern was visualized by autoradiography. Transcription factor binding sites for both alleles of each SNP were *in silico* predicted by the Transcription Element Search System (TESS).

3.5 FUNCTIONAL STUDIES (V)

3.5.1 Aromatase knock-out mice

To study the effects of aromatase during brain development, aromatase knock-out (ArKO) mice were generated through breeding heterozygous ArKO +/- mice (Fisher et al. 1998). The brains of embryos at day 17.5 (E17.5) and of five-month-old mice were dissected and postfixed. After fixation, brains were processed for paraffin (6 µm) sections.

3.5.1.1 Immunohistochemistry

Nissl staining was used to examine the histology of embryonic brains with light microscopy. For immunohistochemistry, the polyclonal rabbit anti-EGF and anti-NeuN, and the mouse monoclonal anti-parvalbumin, were used as primary antibodies.

The Cy3-conjugated anti-rabbit IgG and anti-mouse IgG were used as secondary antibodies. The number of parvalbumin-positive cells was counted on images in an area of $200 \times 200 \mu\text{m}$ in the somatosensory cortex in coronal sections. All pictures were location-matched between wild-type and ArKO mice.

3.5.1.2 Commissural measurements

Five two-month-old male ArKO mice and five wild-type littermates were used for commissural measurements. The commissures were visualized by cutting the fixed brain in half at the mid-sagittal plane, and staining the myelinated structures with gold chloride (Wahlsten et al. 2003). The areas of corpus callosum, hippocampal commissure, and anterior commissure were measured with the ImageJ program (Abramoff et al. 2004).

3.5.2 Process outgrowth of rat hippocampal neurons

To study the direct effect of aromatase in neurons, hippocampal neuronal cultures were prepared from the brains of E17 rat fetuses (Brewer and Cotman 1989). After 24 h culture *in vitro* cells were treated with only solvent; testosterone; testosterone and letrozole; testosterone, letrozole and β -estradiol; or β -estradiol. After three and four days *in vitro*, respectively, cells were fixed and immunostained with the neuronal marker TuJ1. Neurite outgrowth, tracing all processes, was measured using the Image-Pro Plus software.

3.6 EVOLUTIONARY ANALYSIS (II, V)

The *MRPL19*, *C2ORF3* and *CYP19A1* orthologues from chimpanzee (*Pan troglodytes*), pigmy chimpanzee (*Pan paniscus*), gorilla (*Gorilla gorilla*) and orangutan (*Pongo pygmaeus*) were screened for variations by direct sequencing with human-specific, intronic primers. Evolutionary analysis of the coding sequences were performed with a likelihood ratio test (LRT) using the codeml program of the paml3.15 package (Yang 1997). Mouse (*Mus musculus*) or dog (*Canis familiaris*) sequence was used as outgroup. Evolutionary comparison of the complete genomic region of *CYP19A1*, covering the full promoter as well as the coding sequence, was performed using the GenomeVISTA browser (pipeline.lbl.gov/cgi-bin/GenomeVista). The human sequence was aligned with the genomic sequences of dog (*Canis familiaris*), mouse (*Mus musculus*), opossum (*Monodelphis domestica*), and frog (*Xenopus tropicalis*).

4 RESULTS

4.1 GENETIC ANALYSIS

4.1.1 *DYX3* (I, II)

A previous genome-wide scan on eleven Finnish families identified linkage to a 40 cM region within the *DYX3* locus on 2p12 (Kaminen et al. 2003) (Figure 7A). NPL analysis pointed out to marker D2S2216 at 2p11 with an NPL score of 2.55 ($P=0.004$). Parametric analysis using an autosomal dominant model of inheritance gave a LOD score of 3.01 for marker D2S286 at 2p12. In order to refine the linkage region from the genome-wide scan, we performed four rounds of fine mapping in an extended Finnish sample set (Figure 7B-E).

4.1.1.1 *Linkage analysis (I)*

Eighteen microsatellite markers within the 40 cM candidate region from the genome-wide scan were successfully genotyped in the eleven Finnish families. The genotypes were analyzed together with the original genome-wide scan data for chromosome 2 (Kaminen et al. 2003). The candidate region was narrowed down to 15 cM from D2S2110 at 2p12 to D2S2181 at 2p11.1, with NPL scores >2 and nominal p-values <0.01 for all markers (Figure 7B). The highest NPL score was seen for the same marker as in the genome-wide scan; D2S2216 (NPL score 3.0; $P=0.001$). Pedigree-wise NPL scores were positive for only six of the eleven families. A modest NPL score of 0.91 ($P=0.1$) was also observed for D2S2352 located within the original *DYX3* linkage region at 2p15-p16 reported by Fagerheim et al. (1999).

4.1.1.2 *Haplotype pattern mining (I)*

HPM analysis using independent triads from the whole family data did not give any significant results. However, significant scores were obtained when selecting triads only from the six families contributing to the linkage peak on chromosome 2. Three neighboring markers, D2S2110, D2S286 and D2S2116, showed empirical p-values ≤ 0.01 for a common haplotype in dyslexic subjects (Figure 7C). The highest score was seen for D2S286 ($P<0.001$). In order to verify the observed association, six SNPs were genotyped within the D2S2110-D2S2116 interval. As no significant haplotypes could be obtained by HPM analysis of the microsatellites and SNPs combined, this region was considered unlikely to contribute to dyslexia. HPM analysis of eight additional microsatellite markers within the 15 cM region pointed out to a three-marker haplotype pattern D2S286-D2S2116-D2S169 (p-values of 0.001, 0.009 and 0.03, respectively) (Figure 7C). Therefore, this ~ 3.3 Mb region was focused on in the further fine mapping studies using high-density LD mapping.

4.1.1.3 Linkage disequilibrium mapping (II)

Linkage regions in complex disorders often remain large due to the lack of detailed informativeness in families. Therefore, we narrowed the 2p12 candidate region down further using SNPs and association analysis. Two rounds of LD mapping were performed in the eleven Finnish families analyzed in the original genome-wide scan (Figure 7D and E). In the second stage, eight additional Finnish families were genotyped. The genotype data were analyzed for single-marker and haplotype associations with dyslexia using TDT. Initially, the 5 Mb candidate region between markers D2S2109 and D2S329 was narrowed down to 157 kb by analyzing 24 SNPs (Figure 7D). Single-marker TDT yielded significant results for rs917235 (G, $P=0.0027$) and rs730148 (C, $P=0.0039$). Haplotype analysis showed the most significant results for a four-marker haplotype rs1859708-rs1986238-rs2010599-rs730148 (CCAC, $P=0.0039$). In the second mapping stage, marker density was increased to one every 8 kb over the 157 kb region (Figure 7E). This region covered the associated haplotypes from the previous round, as well as the only three genes in this gene-poor region; *FLJ13391* (TMEM, transmembrane protein 166), *MRPL19* (mitochondrial ribosomal protein L19), and *C2ORF3* (chromosome 2 open reading frame 3). Single-SNP analysis did not add any new information, whereas the most significant haplotype covered rs917235-rs714039 (GG, $P=0.0029$).

To confirm the observed association results, we genotyped an independent sample set consisting of 251 German families ascertained for spelling disorder. Totally 29 SNPs/DIPs from the two rounds of LD mapping were genotyped and overlapping risk haplotypes in the two populations were observed. In the first stage, a four-marker haplotype rs1859708-rs1986238-rs730148-rs721390 was over-transmitted to affected subjects (CCCC, $P=0.0036$) (Figure 7D). In the second round, the most significantly transmitted haplotype was rs917235-rs714039-rs6732511 (GGC, $P=0.0036$) (Figure 7E). As a joint analysis of the sample sets should give even more significant p-values in case of a true association (Skol et al. 2006), we performed a combined analysis of the two sample sets. Two overlapping three-marker haplotypes ($P=0.0049$ and 0.0013, respectively) covering totally 16.6 kb, delineated the region of association in both populations (Figure 7E).

As many studies of dyslexia have reported stronger positive associations with more severe phenotypes (Deffenbacher et al. 2004; Francks et al. 2004; Cope et al. 2005a), we re-analyzed the most significantly associated haplotype in the German sample set (rs917235-rs714039-rs6732511, GGC) by stratifying for severity. Detailed phenotype data were not available for the Finnish families. Probandes were classified according to their discrepancy between the observed and expected spelling scores. The odds ratio (OR) increased from 2.2 (95 % confidence interval, CI, 1.4–3.4, global $P=0.006$) for all probands to 5.2 (95 % CI 2.3–18.0, global $P=0.00005$) for the most severely affected individuals (displaying a discrepancy of ≥ 2.5 SD between the observed and expected spelling scores; 72 probands).

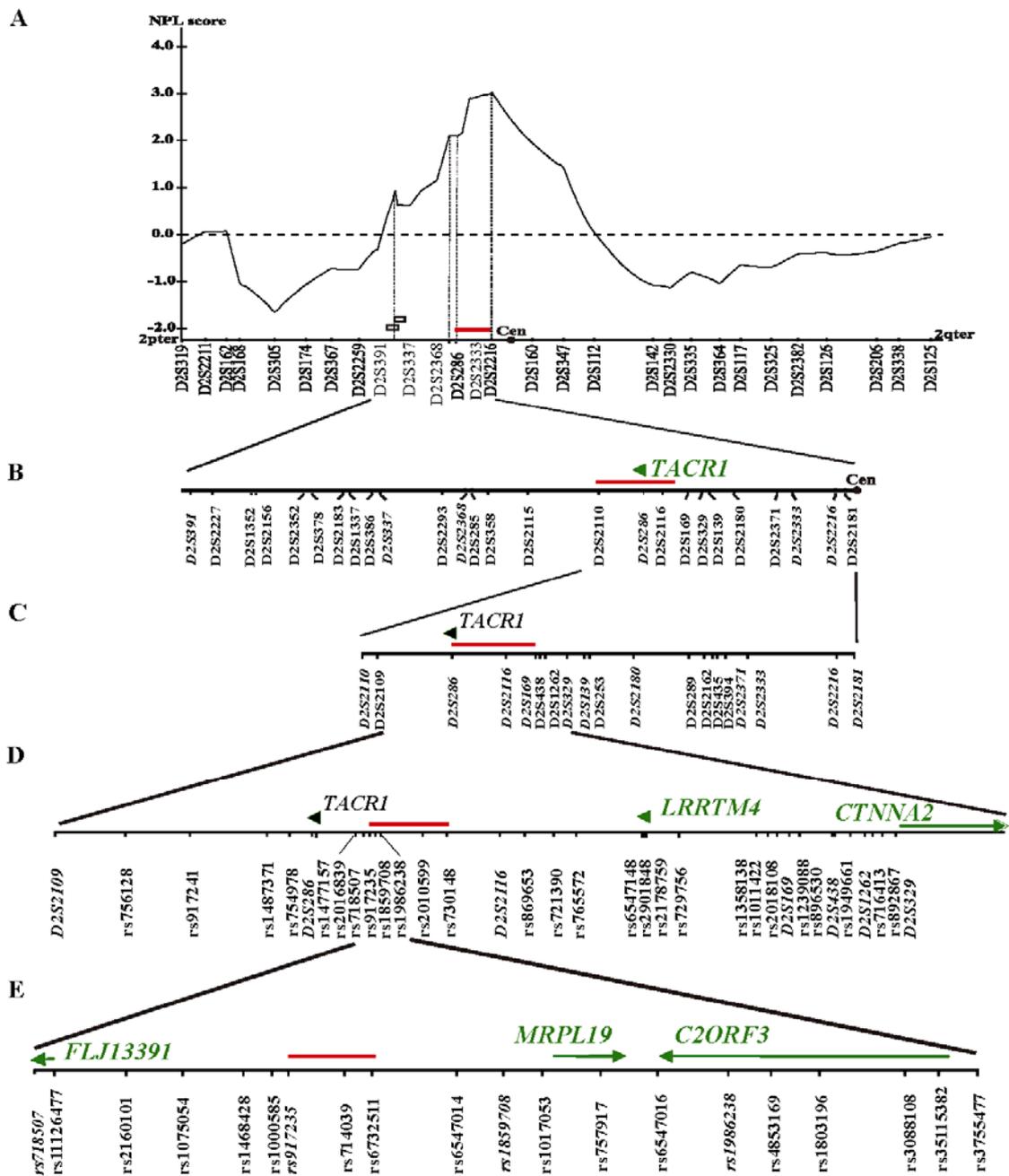


Figure 7. The mapping stages within the 2p12 dyslexia candidate region. **(A)** NPL peak from a previous genome-wide scan on eleven Finnish pedigrees. **(B)** The 40 cM candidate region from the first round of refinement using microsatellite markers and linkage- and HPM analysis (study I). **(C)** Additional microsatellites genotyped within a 15 cM region and analyzed by HPM (study II and unpublished results). **(D)** SNPs genotyped over a 5 Mb region in the third refinement round and analyzed using TDT (study II). **(E)** The last mapping stage over a 157 kb region, where additional SNPs were genotyped and analyzed using TDT. Markers used in a previous mapping stage are shown in italics. The significant linkage/association regions from each stage are highlighted by a red bar. The genomic extent and orientation of the candidate genes studied in each stage are shown with green arrows.

4.1.2 *DYX2* (III)

We studied the most replicated locus for dyslexia, *DYX2* on 6p22.1, in two sample sets from Germany ascertained for spelling disability. To identify LD within the *DYX2* region, we first genotyped 16 microsatellite markers encompassing a 24 Mb region between D6S289 and D6S1610. In the 137 triads analyzed, the most significant association was observed for D6S276 ($P=0.004$). D6S276 is located within the *DCDC2* (doublecortin domain containing 2) gene and in close proximity to *KAAG1* (kidney associated antigen 1) and *VMP* (vesicular membrane protein 1). To explore the genetic variation within this gene cluster, we re-sequenced the non-repetitive segments at the *VMP/DCDC2/KAAG1* locus including all coding and most intronic sequences, the putative promoters, and the 5' and 3' UTRs in two dyslexic individuals carrying the risk allele at D6S276, and in one non-carrier. To further refine the LD region and to determine its boundaries, we performed TDT analysis with 25 of the identified SNPs, encompassing all three genes with constant intermarker distances (Figure 8). Significant transmission disequilibrium was observed for rs793862 ($P=0.011$) located within *DCDC2*. A haplotype analysis using marker combinations of the most-significantly associated variants strengthened the observed association. Two two-marker haplotypes at rs28990373-D6S276 and rs793862-rs807701 were highly overtransmitted to affected probands ($P<0.0001$). To confirm the association results, we analyzed an independent sample set of German origin consisting of 239 triads ascertained for spelling disability. All markers that had shown significance in the initial single-marker- or haplotype analyses were genotyped. The haplotype A-C at rs793862-rs807701 was again overtransmitted to affected probands ($P=0.001$).

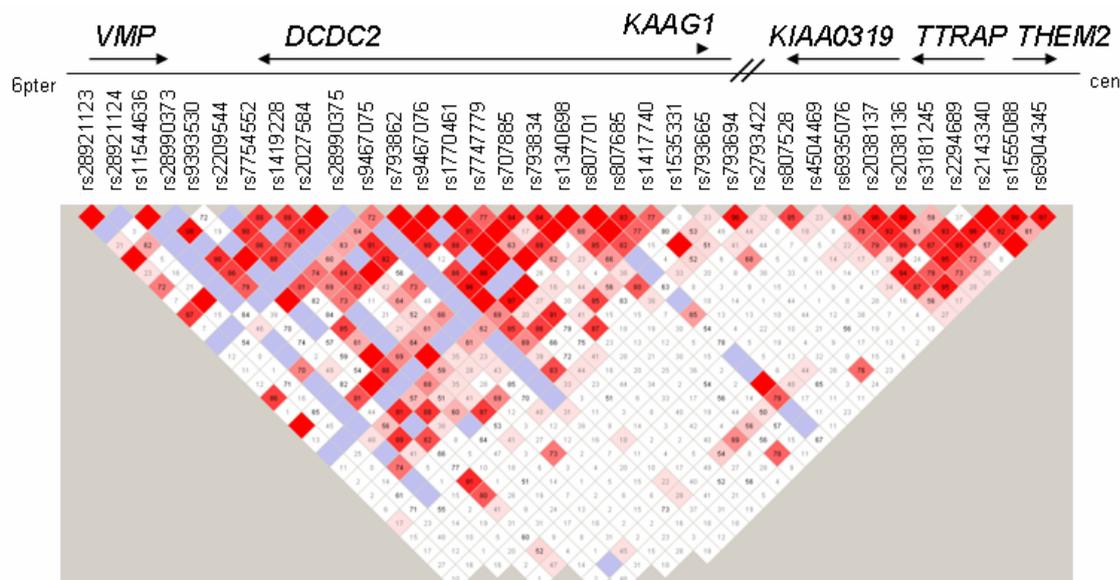


Figure 8. Genomic map over the 570 kb *DYX2* locus showing the two gene clusters, the SNPs genotyped, and the LD structure. Note that the map is not to scale; the *VMP/DCDC2/KAAG1* cluster covers ~200 kb, whereas the *KIAA0319/TTRAP/THEM2* covers ~150 kb, and the two clusters are separated by a ~220 kb gap (//).

The initial linkage findings to the *DYX2* locus were significantly enhanced when selecting for severity (Cardon et al. 1994; Gayan et al. 1999), and association studies within *DYX2* have reported positive association only in subsets of their sample with a more severe phenotype of dyslexia (Deffenbacher et al. 2004; Francks et al. 2004). Therefore, we stratified for families with more-severely affected children, and re-analyzed the three most significantly associated markers. For both markers located within *DCDC2*, the genotype relative risk (GRR) increased continuously with analysis of more-severe spelling disorder. In the pooled sample set, the risk haplotype at rs793862-rs807701 was associated with a GRR of 4.88 (95% CI 3.32– 7.15, $P < 0.0001$) for the most severely affected individuals (≥ 2.5 SD; 111 triads). In contrast, the GRR of the *VMP* variants did not show any increase with the application of more-severe dyslexia phenotypes.

None of the quantitative component processes studied (phonological awareness, phonological decoding, rapid naming, or orthographic coding) showed evidence of association in the initial sample set, not even when the analysis was restricted to more-severe phenotypes.

4.1.2.1 *The KIAA0319/TTRAP/THEM2 gene cluster*

Another gene cluster within the *DYX2* locus, *KIAA0319/TTRAP/THEM2* ~185 kb from *VMP/DCDC2/KAAG1*, has shown association with dyslexia in UK and Australian samples (Francks et al. 2004; Cope et al. 2005a; Luciano et al. 2007). To assess whether genetic variation in *KIAA0319/TTRAP/THEM2* would contribute to dyslexia in the German population, we genotyped SNPs in this gene cluster in the pooled sample set of 376 triads, including the most significant SNPs reported by Francks et al. (2004) and Cope et al. (2005a) (Figure 8). None of the SNPs showed association with spelling disability, and only one SNP (rs2294689) showed borderline significance in the most-severely affected children ($P = 0.046$). Our analysis had an average power of $>95\%$ for detection of an effect at the threshold level of 5%. Therefore, genetic variation at the *KIAA0319/TTRAP/THEM2* locus is most likely not of major importance for the development of dyslexia in the German population.

4.1.3 *DYX1C1 (IV)*

Given the previous inconsistent association findings within *DYX1C1* (dyslexia susceptibility 1 candidate 1), we aimed to explore the contribution of this locus to dyslexia in a large sample set of 366 German families ascertained for spelling disability. As the previous studies conducted on *DYX1C1* have not considered a putative influence of sex, which has been suggested as an important factor in dyslexia (Rutter et al. 2004; Harlaar et al. 2005), we extended our statistical analysis and divided our sample in female and male proband groups. An initial analysis, using the same SNPs as Taipale and co-workers (rs3743205 and 1249G→T), suggested a trend towards association at the haplotype level ($P = 0.08$, global $P = 0.73$ in the full sample set). Although the global p-value was not significant and we failed to find association at the single-marker level, we hypothesized that putative important SNPs and the

haplotype structure of *DYX1C1* gene had not been fully revealed. We therefore re-sequenced the whole coding region of *DYX1C1* as well as the corresponding 5' and 3' UTR regions in 10 patients and 10 controls of German descent, all randomly selected. This allowed detection of SNPs with a minor allele frequency of >10% with more than 98% confidence if the SNPs are in Hardy-Weinberg equilibrium (Gregorius 1980). Comparison of the analyzed gene regions to the public sequence revealed three previously unstudied SNPs, namely rs12899331, rs16787 and rs8043269, all located within the promoter region of *DYX1C1*. Variants rs12899331 and rs16787, rs3743205 and 1249G→T, and two additional polymorphic sites, rs3743204 (an intronic SNP) and rs600753 (572G→A), were selected for further analysis (Figure 9). Although no significant point-wise association was observed, haplotypes within *DYX1C1* were associated with spelling (global $P=0.04$). Analyzing both sexes separately revealed a strong effect in females, with the most significant results for the common three-marker haplotype at rs3743205-rs3743204-rs600753 (GGG), which carried an OR of 3.71 (95% CI 1.44–9.60).

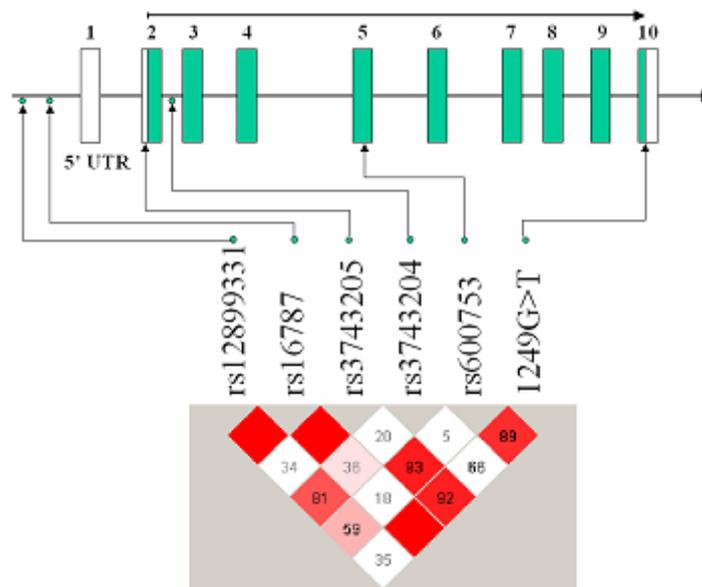


Figure 9. Genomic structure of the *DYX1C1* gene, the genotyped SNPs and the LD structure. Note that rs3743205 corresponds to -3G→A and rs600753 to 572G→A.

No significant haplotypes were observed in the male probands for the dichotomous trait of spelling, not even when stratifying for severity. However, when analyzing the quantitative subphenotypes of dyslexia, both the full sample set as well as the male probands showed significant association with phonological awareness and haplotypes within *DYX1C1*. A two-marker haplotype at rs3743205-rs3743204 (GG) was over-transmitted to affected subjects, with global p-values of 0.002 and 0.004 for the full sample set and for the males, respectively. In the female probands, the same haplotype at rs3743205-rs3743204-rs600753 (GGG) that was associated with spelling, showed association with short term memory ($P=0.005$, global $P=0.011$).

Our findings are supported by previous studies where the common alleles at rs3743205 and 1249G→T increased, although not statistically significantly, the risk for dyslexia or related subphenotypes (Scerri et al. 2004; Wigg et al. 2004; Cope et al. 2005b; Marino et al. 2005; Marino et al. 2007). To test for composite significance, we combined our results and the previous *DYX1C1* association findings which were obtained in samples with the same ethnic background (Skol et al. 2006). Because raw data were not available from the other studies, we performed a Fisher’s combined probability test using the haplotypic p-values that were available from the published studies. For the overall association, the p-value was 0.009, indicating that the same full-length or partial haplotype of *DYX1C1* is significantly associated with dyslexia in all studies (Table 2). Unfortunately, a combined sex-specific analysis using the different samples was not possible, because the raw data were not available.

Study	size	p-value	Haplotypes						
			rs12899331	rs16787	rs3743205	rs3743204	rs600753	1249G→T	1259C→C
Wigg	148	0.03	-	-	G	-	-	G	-
Marino	158	0.09	-	-	G	-	-	G	G
Cope	247	0.41	-	-	G	-	-	G	-
Our study	366	0.04	*	A	G	G	G	G	-
Pooled		0.009							

Table 2. Meta-analysis of the replication studies for *DYX1C1* association with dyslexia. Studies not included are: Taipale et al. (2003), as no information about the common alleles at rs3743205 and 1249G→T were reported; Scerri et al. (2004), as p-values were reported only for the significantly associated dyslexia subphenotypes; Meng et al. (2005a), as no haplotypic information was reported; Marino et al. (2007), as the same sample set was used in the previous study (Marino et al. 2005). For our study, the corrected global p-value for all probands is used. Size refers to the no. of triads used in each study. Asterisk denotes “any allele” and a minus sign indicates that the SNP is not genotyped in the specific material.

4.1.4 t(2;15)(p12;q21) (V)

Using FISH and Southern blot analysis, the translocation breakpoint in the individual co-segregating t(2;15)(p12;q21) and dyslexia was mapped to the complex promoter region of *CYP19A1* (cytochrome P450, family 19, subfamily A, polypeptide 1) gene. Specifically, the translocation localized between the promoter for skin, adipose tissue and fetal liver (I.4) and the promoter for fetal tissue (I.5), ~22 kb upstream from the brain-specific exon I.f and its respective promoter (Figure 10). *CYP19A1* encodes the

cytochrome P450 subfamily enzyme aromatase that catalyzes the formation of estrogen from androgens. In most mammals, it is expressed only in the gonads and in the brain, whereas in primates it is expressed in several other tissues as well (Bulun et al. 2003). *CYP19A1* transcripts in primates have numerous untranslated first exons in a tissue-specific fashion due to differential splicing as a consequence of the use of tissue-specific promoters. The chromosome 2 breakpoint mapped to an unremarkable region ~6.5 Mb centromeric from the 2p12 *DYX3* locus. This breakpoint region on 2p12 is very repeat-rich and contains no known genes. Furthermore, no new genes could be cloned from this region by gene prediction programs and PCR on human cDNA libraries. The gene desert stretches ~2 Mb on both sides on the breakpoint, suggesting that the chromosome 2p breakpoint is not relevant for the phenotype.

4.1.4.1 Association analysis SNPs within the *CYP19A1*

Two sample sets ascertained for dyslexia of Finnish and US origin, respectively, and one ascertained for SSD (from the US) were analyzed for association within the *CYP19A1* locus. Both single-SNP and haplotype PDTs yielded significant results for several SNPs within the *CYP19A1* gene. In both the US dyslexia and SSD populations the binary trait dyslexia showed the greatest evidence of point-wise SNP association with the T allele of rs11632903 ($P=0.025$ and $P=0.019$, respectively). Combining both populations using Fisher's combined probability test, increased the significance of the association ($P=0.0004$). rs11632903 is located 5.8 kb downstream from the untranslated brain-specific exon I.f (Figure 10). A risk haplotype spanning the brain specific exon I.f and its promoter was also observed in both US cohorts (rs11632903-rs1902586; TG; $P=0.023$ and rs1902586-2470176; GCAG; $P=0.023$, in the dyslexia and SSD cohorts, respectively). In the Finnish population there was evidence of transmission distortion to dyslexia affected offspring at the haplotype level only (rs8034835-rs2899472; GC; $P=0.039$). An overlapping four-marker haplotype was also identified in the US SSD cohort, spanning rs8034835-rs700518 within the coding region of *CYP19A1* (GCAG, $P=0.032$).

Several SNPs were associated with quantitative phenotypes in the US dyslexia and SSD cohorts using a variance components analysis. A random polygenic effect was the only variance component included in the model. Age was included in the baseline model as a covariate as it was found to have a significant effect in both populations. At each SNP and each trait an additive, dominant, and a recessive allele effect was tested. Quantitative measures were not available for the Finnish samples. Of the overtransmitted SNPs described above, significant association in both cohorts was observed with phonological decoding ($P<0.01$). In addition, the dyslexia cohort showed association with spelling and phonological awareness ($P<0.01$), and the US SSD cohort with phonological short-term memory and oral motor skills ($P<0.001$). Moreover, several SNPs within the coding region of *CYP19A1* demonstrated significant associations with a number of reading- and language-related measures after correction for multiple testing.

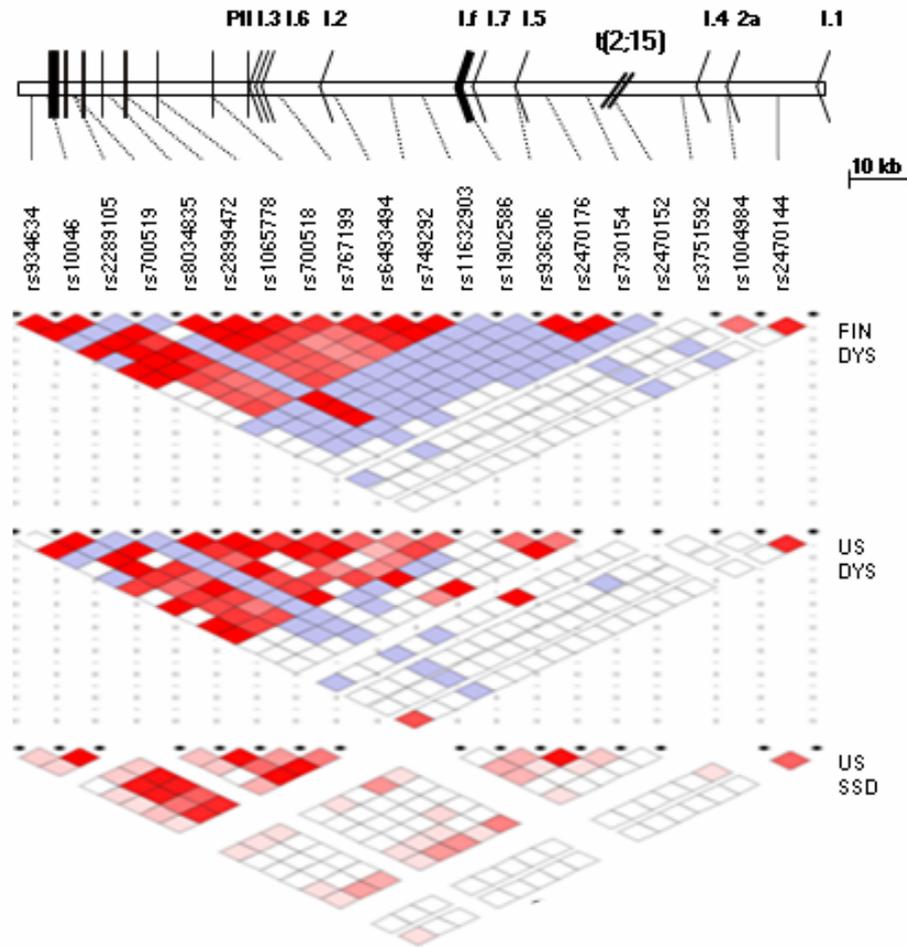


Figure 10. *CYP19A1* gene organization, the translocation breakpoint, the genotyped SNPs and the LD structure in the three cohorts analyzed. The coding exons are denoted with bars, promoter regions with arrowheads, and the t(2;15) indicates the breakpoint. The brain-specific exon I.f is highlighted with a thicker arrowhead. Note that in accordance with the chromosomal orientation, the gene is drawn from right (5') to left (3').

4.1.5 Causal variants in the dyslexia candidate genes (I – V)

4.1.5.1 *MRPL19*, *C2ORF3* and other positional candidates within *DYX3* (I, II)

As the identified risk haplotypes were located in an 80 kb intergenic region between the hypothetical gene *FLJ13391* and the *MRPL19* and *C2ORF3* genes, an extensive search for novel genes was performed. However, no additional genes could be identified using gene prediction programs and analyzing cDNA libraries. As *MRPL19* and *C2ORF3* are in high LD, and the LD extends to the risk haplotypes, we hypothesized that the risk haplotypes might lie in a putative regulatory region of the two genes. Therefore, we evaluated the allelic expression levels of *MRPL19* and *C2ORF3* in EBV-transformed lymphocyte cell lines of carriers and non-carriers of the risk haplotype. By comparing the peak height ratios in genomic DNA and cDNA, we observed a significant difference in the expression levels of *MRPL19* and *C2ORF3* for the two alleles of a

synonymous SNP in each of the two genes. In carriers of the risk haplotype, the more common allele was significantly less transcribed for both genes.

We hypothesized that there might be two separate mechanisms in the 2p12 region for the susceptibility of developing dyslexia, i.e., the risk haplotypes identified in the putative regulatory region and/or SNPs within the coding region of one or both of the two genes. Therefore, we sequenced all coding exons, the flanking sequences, and UTRs of *MRPL19* and *C2ORF3* in one affected individual from each of the 19 Finnish families. Two novel coding SNPs (cSNPs) were identified as heterozygous changes in both genes, respectively. We genotyped these as well as all previously known cSNPs reported in the dbSNP database, in the two sample sets of Finnish and German families. No over-transmissions could be observed to affected individuals, and the allele frequencies in affected and unaffected were approximately equal, suggesting that none of these variants was functionally relevant.

Our initial fine-mapping effort using microsatellite markers pointed to the marker D2S286. The *TACRI* (tachykinin receptor 1) gene encompasses the marker D2S286 and has a role in the CNS; in modulation of neuronal activity, inflammation, and mood (Derocq et al. 1996; De Felipe et al. 1998; Kramer et al. 1998). Direct sequencing of the full coding region of the gene revealed two already known SNPs in dyslexic subjects from the Finnish families. However, these were not associated to dyslexia when genotyping the full set of Finnish families. In addition to the *FLJ13391*, *MRPL19* and *C2ORF3* genes, *CTNNA2* (catenin alpha 2) and *LRRTM4* (leucine rich repeat transmembrane neuronal 4) are the only known genes, besides a cluster of pancreatic-specific genes, within a 5 Mb genomic region from *TACRI* to *CTNNA2* (see Figure 7 D and E). As these two genes are highly expressed in the human brain and thus represented also functional candidate genes for dyslexia, they were screened during the mapping process in dyslexic subjects of Finnish origin. However, no variants were detected in the coding exons or splice sites of either of them. Furthermore, TDT did not reveal any signs of association in the *LRRTM4* gene in the Finnish or the German sample set.

4.1.5.2 *DCDC2* (III)

The coding regions, flanking sequences, and the UTRs of *DCDC2* and *KAAG1* were sequenced in 47 dyslexic and 47 normal readers of German origin. Several variants in *DCDC2* were identified, including four amino acid substitutions. Of these, one SNP (rs2274305; Ser221Gly) seemed to be associated with dyslexia risk as it occurred more frequently in cases than in controls (64 vs. 50 %, respectively). However, it seemed unlikely that it would be a common risk allele, as it was not specific to the identified risk haplotype at rs793862-rs807701.

4.1.5.3 *DYX1C1* (IV)

To understand possible functional consequences of the critical new SNPs in the 5'UTR region, EMSAs were performed. Allele-specific differential retardation was detected for all three SNPs, suggesting functional effects on transcription or other regulatory

factor binding. We searched *in silico* information on suggested factors with altered binding properties. For all three SNPs, differences in the number of hits and also in the identity of the factors predicted to bind to the respective alleles were observed. For rs12899331, the binding site for the transcription factor Sp1 was altered. For rs16787 and rs3743205, binding sites of at least three transcription factors were disrupted, including GATA1, TFII-I, and Elk-1.

4.1.5.4 *CYP19A1* (V)

Three dyslexic subjects of US origin carrying the risk haplotype at the *CYP19A1* locus were sequenced over the full coding region of the gene, as well as over the brain-specific exon I.f. With the exception of the two genotyped variants (rs700519 and rs700518), no coding polymorphisms were identified. The SNPs identified in intronic sequence or in the 3'UTR did not seem to affect any known splice or regulatory sites. In addition to the already genotyped nonsynonymous SNP (rs700519), we sequenced the two remaining nonsynonymous SNPs existing in the public database at the time (rs2236722 and rs1803154; www.ncbi.nih.gov) in the Finnish and US dyslexia sample sets. Neither of these two SNPs could be identified in these families. As the dyslexia-associated SNPs clustered around the brain-specific promoter and its untranslated exon I.f, we searched for additional variation in this region by sequencing the non-repetitive sequence over a 20 kb region in one affected individual homozygous for the susceptibility haplotype. No additional variation than the already genotyped SNPs were found, which is not surprising as a promoter region is generally highly sensitive to sequence changes.

Among the dyslexia-associated SNPs, rs11632903 and rs1902586 flank the brain-specific promoter I.f of *CYP19A1*. As no additional variants were identified within this region, we hypothesized that these SNPs might have causative roles by affecting the binding of nuclear protein factors and thus the regulation of transcription of aromatase in the brain. *In silico* predictions of altered transcription factor binding for both SNPs indicated differences in the number of hits as well as in the identity of the predicted binding factors. In particular, the T allele of rs11632903 abolished TFII-I and Elk-I binding sites that were present for the C allele. To verify the predicted effects on transcription factor binding, we used each allele of both SNPs as probes in EMSA. Both rs11632903 and rs1902586 showed reduced binding to the dyslexia-associated alleles. TFII-I, but also Elk-1, bound weaker to the dyslexia-associated T allele of rs11632903 as verified by supershift assays with specific antibodies.

4.2 EXPRESSION ANALYSIS

4.2.1 Transcript characterization (I – III, V)

The gene structure and the exon-intron boundaries were verified in human fetal brain and lymphocyte cDNA libraries and I.M.A.G.E. clones. The expression levels and transcript sizes were also studied by Northern blots to human tissues (studies I – III). For *C2ORF3* and *TACR1*, novel splice variants were identified in addition to the ones existing in the public database. *C2ORF3* showed a weak expression across multiple human tissues, whereas Northern blot analysis of *TACR1* and *MRPL19* confirmed their ubiquitous and abundant expression. The expression pattern of *DCDC2* was investigated by Northern blot analysis to a panel of different regions of human brain, where it was expressed at an equal degree in all brain regions (amygdala, caudate nucleus, hippocampus, thalamus, whole brain), except for corpus callosum, where the expression was weak.

4.2.2 Correlation of mRNA expression across various regions of human brain (II, V)

We studied the expression levels of all dyslexia candidate genes, i.e., *MRPL19*, *C2ORF3*, *DCDC2*, *KIAA0319*, *DYX1C1*, *CYP19A1* and *ROBO1*, in nine different regions of adult human brain as well as in whole adult and fetal brain using quantitative real-time RT-PCR. In addition, the expression of *FLJ13391* was studied as it is located within the 2p12 candidate region. All dyslexia candidate genes were highly expressed in all areas of adult brain as well as in whole fetal brain, except *CYP19A1* which showed a more moderate expression level. High expression levels were observed in regions implicated in reading, such as the temporal-, parietal-, and frontal cortex. Thalamus and hypothalamus displayed also high levels of expression, especially for *CYP19A1*. More importantly, the pattern of expression across the different brain regions was correlated for the dyslexia candidate genes. In particular, *CYP19A1* showed a high correlation with *ROBO1* and *DYX1C1* ($R^2=0.72$ and $R^2=0.60$, respectively), and *C2ORF3* with *DCDC2* and *ROBO1* ($R^2=0.76$ and 0.69 , respectively). Different from the other dyslexia candidate genes, *KIAA0319*, and also to some extent *MRPL19*, showed a more even distribution of expression across the different brain regions, and thus their expression was correlated ($R^2=0.47$). In contrast, *FLJ13391*, which is not regarded as a dyslexia susceptibility gene, showed a completely different expression pattern than any of the dyslexia candidate genes.

4.3 FUNCTIONAL STUDIES (V)

4.3.1 Aromatase knock-out mice

To study the role of *CYP19A1* in brain development, we performed a detailed structural analysis of the brains of ArKO mice. Several signs of cortical disorganization were observed as compared to wild-type controls. The neuronal density in cortical areas was significantly increased at E17.5. ArKO E17.5 mice showed also an increased signal for epidermal growth factor (EGF), which plays an important role in the regulation of cell growth, proliferation and differentiation. Remarkably, even in five-month old ArKO mice, the cortical areas had an increased overall neuronal density as determined by the neuron-specific nuclear protein (NeuN) staining. Moreover, an increased number of parvalbumin-positive inhibitory interneurons and occasional cortical heterotopias were observed. The mid-sagittal areas of the anterior and hippocampal commissures as well as the corpus callosum were similar in size in ArKO and wild-type mice.

4.3.2 Hippocampal process outgrowth

To study the effects of aromatase function on undifferentiated neurons, we investigated process outgrowth of rat E17 hippocampal neurons in culture. Testosterone and estradiol-17 β both significantly promoted neurite outgrowth at day four in culture when compared to controls. The effect of testosterone was blocked by the aromatase inhibitor letrozole. However, letrozole did not block the effects of estradiol and testosterone together, indicating that aromatase-dependent conversion of testosterone to estradiol enhances neurite outgrowth in cultured hippocampal neurons.

4.4 EVOLUTIONARY ANALYSIS (II, V)

We looked for signs of recent selection in the *MRPL19*, *C2ORF3* and *CYP19A1* genes since the divergence from the orangutan and gorilla branches by sequencing the coding regions in four non-human primate species. *MRPL19* and *CYP19A1* were both highly conserved, with only a few amino acid substitutions identified. In contrast, several variants were identified in *C2ORF3* in all primates analyzed. The predicted *C2ORF3* proteins for pigmy chimpanzee, chimpanzee and gorilla differed by 1.0%, 1.5% and 1.9% of residues, respectively, as compared to the human homologue. The orangutan exons could not be amplified with the human-specific primers used, suggesting a too low sequence identity in the flanking intronic sequence.

We calculated the rate of synonymous (dS) and nonsynonymous substitutions (dN) in all species studied and applied LRTs to analyze the selection pressure $\omega=dN/dS$ during primate evolution. For *MRPL19* and *CYP19A1* the low information content drastically reduced the power and the LRTs were not reliable. For *C2ORF3*, low dN/dS values were observed in the primate lineages as compared to the out-group (dog). A two-ratio

model with a different ratio for the primates than for the out-group showed a significantly better fit to the data than a freely variable ratio, indicating a change in the selection pressure from dog to primates. Among the primates no significant heterogeneity could be observed, suggesting no change in the selection pressure during the evolution of *C2ORF3* from non-human primates to the human lineage. However, the lowest dN/dS ratios were observed in humans, where there were relatively more synonymous than nonsynonymous changes.

A comparative analysis of the complex promoter region of *CYP19A1* revealed extremely high conservation around the brain-specific promoter across a broad range of vertebrates. Moreover, in the untranslated brain-specific exon I.f, we identified one SNP that was different in humans compared to all the non-human species analyzed, including the sequences available in the public databases. *In silico* analysis of this SNP revealed a larger number of transcription factor binding sites for the human-specific variant, and an EMSA showed completely different retardation patterns for the human and non-human alleles.

5 DISCUSSION

5.1 GENETIC ANALYSIS OF DYSLEXIA

5.1.1 The *DYX3* locus and *MRPL19* and *C2ORF3*

A previous genome wide scan on eleven Finnish families identified linkage to the *DYX3* locus on 2p12 (Kaminen et al. 2003). We refined this initial linkage region to 157 kb in several rounds of fine-mapping in an extended set of Finnish families. The observed association results were replicated in an independent sample set of German origin, and a joint analysis of the two sample sets identified overlapping haplotypes co-segregating with dyslexia in an intergenic region between the *FLJ13391* and *MRPL19* and *C2ORF3* genes. Stratification analysis of more severely affected cases strengthened the associations further, significantly increasing the OR for the most common susceptibility haplotype. Since our extensive search for possible novel genes throughout the whole ~80 kb region between *FLJ13391* and *MRPL19* yielded no results, it is unlikely that the associated region would harbor a yet unidentified susceptibility gene. Instead, we hypothesized that the associated SNPs might be non-coding variants in a regulatory region of *MRPL19* and *C2ORF3*. In support of this hypothesis, our expression data showed that these two closely located genes are highly co-expressed in different brain regions. In addition, *MRPL19* and *C2ORF3* are in strong LD with each other belonging to a single haplotype block, and therefore it is plausible that they share a common regulatory region. The regulatory effect of the associated haplotypes was further established by allele-specific expression analysis where the expression levels of the two genes were significantly reduced in carriers of the risk haplotypes, as compared to non-carriers.

Because the two independent sample sets supported the association findings, there is strong evidence for the involvement of this locus in dyslexia. Other studies have found support for this region as well, although the linked/associated loci have been widely spread over the short arm of chromosome 2. The original linkage of dyslexia to the *DYX3* locus was identified in a single large family showing autosomal dominant transmission to 2p15-p16 (Fagerheim et al. 1999). Similar to our study, they used a categorical phenotype, but our region is 18 Mb centromeric to their reported locus. Using quantitative non-parametric linkage analysis, Fisher et al. (2002) identified linkage to a number of loci in the human genome, among them several loci on chromosome 2. In single-point analysis, they identified linkage to the same locus as we, at marker D2S286 ($P=0.0003$ for phonological awareness in the US families). A locus on 2p15 overlapping with the *DYX3* locus reported by Fagerheim et al., showed also linkage to phonological awareness ($P=0.001$) in the US families. In the UK sample set, linkage was more distal, with peaks on 2p16 and 2p25 for orthographic coding ($P=0.0007$ and $P=0.001$, respectively). Francks et al. (2002a) performed a quantitative sib-pair association study on the same set of 119 US twin-based families to refine the chromosome 2p12-p16 locus. Suggestive evidence of association was found to several markers and several phenotypes. Although weak, D2S286 showed again association

($P=0.049$ for orthographic choice), and D2S2114, which is located only 350 kb centromeric from our identified dyslexia locus, showed borderline association for all tested reading measures. However, the strongest results were seen for more distal markers at 2p21. Petryshen et al. (2002) analyzed a qualitative phenotype in Canadian families and identified linkage of spelling to 2p15-p16. However, they only genotyped markers within 2p15-p16, and therefore could not possibly identify a putative signal at the 2p12 locus. Recently, a genome-wide scan on Australian families identified a linkage peak at D2S2216 (Bates et al. 2007), the same marker showing the strongest linkage in our initial genome-wide scan.

Together, these reported findings suggest that there may be two susceptibility loci for dyslexia on chromosome 2p; the original linkage locus on 2p15 and our reported locus on 2p12, both supported in the large quantitative genome-wide scan (Fisher et al. 2002). Alternatively, there is only one locus with an inaccurate definition of its position by linkage. Linkage regions in complex disorders are often wide and often not completely overlapping between different studies, even though they may reflect the same susceptibility locus. As we used a categorical diagnosis of dyslexia both in the initial genome-wide scan and in further fine mapping studies, this locus seems to have a general effect on dyslexia, i.e., word reading and spelling. This conclusion is further supported by the observation of a stronger effect in the more severe cases from Germany. Even though Fisher et al. (2002) and Francks et al. (2002a) studied different quantitative processes of dyslexia, they found evidence of linkage and association to several of these components and approximately the same locus as we report here.

5.1.2 Strong evidence for *DCDC2* as a susceptibility gene for dyslexia

The *DYX2* locus on 6p22.1 is the most-replicated locus for dyslexia and is thus likely to contribute to common forms of the disorder across different populations. However, the exact genomic position and the genes contributing to this disorder have been unclear. At the linkage level, the effects of the two gene clusters in the region, *VMP/DCDC2/KAAG1* and *KIAA0319/TTRAP/THEM2*, have not been resolved. Association analyses have provided inconsistent results, with positive findings to either cluster in different studies. We identified significant association of markers within *DCDC2* with spelling disability in two independent sample sets from Germany, in particular in more severe phenotypes. The replication on an independent sample set as well as the increasing GRR with severity, provide compelling evidence for *DCDC2* as a susceptibility gene for dyslexia, contributing in particular to severe spelling disability. None of the quantitative component processes studied showed evidence of association in the initial sample set, not even when the analysis was restricted to more-severe phenotypes. Therefore, at the phenotypic level, no component process characterizes the genetic effect of the *DCDC2* locus in the German sample set more precisely than spelling disorder itself. Interestingly, the effect of severity was already reported in the initial linkage findings to the *DYX2* locus (Cardon et al. 1994; Gayan et al. 1999), and many association studies have also found stronger association to *DCDC2* or *KIAA0319* in more severe cases for dyslexia (Deffenbacher et al. 2004; Francks et al. 2004; Harold et al. 2006). Moreover, as a genome-wide scan for reading and spelling in an unselected

sample from the normal population failed to identify linkage to the *DYX2* locus (Bates et al. 2007), this locus seems to affect the lower tail of the distribution, contributing specifically to the development of dyslexia and not to general variation in reading ability.

The involvement of *DCDC2* is strongly supported by the many positive findings within this locus. Although Deffenbacher et al. (2004) found significant association for both gene clusters, the strongest association was found with SNPs located in the same genomic region within *DCDC2* as the association in the German sample sets. This region is inherited as a separate haplotype block. Meng et al. (2005b) refined their original association study by Kaplan et al. (2002) in the Colorado family sample by genotyping SNPs in the two gene clusters. Several SNPs within *DCDC2* showed significant evidence of association with spelling- and reading-related phenotypes. They also identified a large intronic deletion in intron 2 of *DCDC2* that encodes repeats of putative brain-related transcription factor binding sites (Meng et al. 2005b). Alleles of the deletion showed significant LD with several reading-related traits. However, in our German sample set, this deletion did not show any evidence for association with dyslexia or with the risk haplotype identified in this sample set (Ludwig et al. submitted).

However, there have been many negative reports for *DCDC2* as well. Most of these have identified association to the second gene cluster in the region; *KIAA0319/TTRAP/THEM2*, ~185 kb from the *VMP/DCDC2/KAAG1* locus. Cope et al. (2005a) performed a high-density screen encompassing the *VMP/DCDC2/KAAG1* locus, and did not detect any association with *DCDC2* in their UK case-control sample set. Instead, several reading phenotypes showed association with *KIAA0319*. However, as they did not select for disease severity, the power of their sample set may have been too small to detect the genetic effect of the *VMP/DCDC2/KAAG1* locus, especially with the use of a DNA-pooling approach in which no individual genotypes were obtained and haplotype analysis could not be performed. Francks et al. (2004) focused only on the gene cluster *KIAA0319/TTRAP/THEM2* and observed a significant association in *KIAA0319* and *TTRAP* in their UK and US sample, in particular in probands with severe reading disability. Harold et al. (2006) verified the role of *KIAA0319* in these UK sample sets, by genotyping also the most significant SNPs identified by us (III) and by Meng et al. (2005b) in *DCDC2*. Significant associations with several quantitative traits were observed within *DCDC2* in the family-based sample (Francks et al. 2004). However, in the more severe cases, only the intronic deletion identified by Meng et al. (2005b) showed nominal significance. In the case-control sample no significant associations within *DCDC2* were observed, consistent with the previous results in this set (Cope et al. 2005a). Harold et al. (2006) also analyzed additional SNPs within *KIAA0319*, and could localize the strongest association signal around the first exon, for several reading-related components. Statistical interactions between SNPs within the two genes were significant, indicating a possible epistatic effect between the two genes. Luciano et al. (2007) identified modest association signals within the *KIAA0319* and *TTRAP* genes and reading and spelling ability in an Australian sample set. However, they did not study SNPs within the *DCDC2* cluster.

There are several explanations for the inconsistent association results within *DYX2*. Sample ascertainment and phenotyping criteria have slightly varied between studies and languages. In our study, dyslexic individuals were selected on the basis of a discrepancy between spelling and IQ, whereas others selected dyslexic individuals on the basis of a discrepancy between word-reading ability and IQ (Deffenbacher et al. 2004; Francks et al. 2004; Cope et al. 2005a). Twin studies of dyslexia have indicated that the heritability of spelling deficits is higher than that of reading deficits (Stevenson et al. 1987). These are different phenotypes, but spelling and reading ability are strongly correlated (Schulte-Körne et al. 2007). Thus, different studies may have analyzed slightly different subsamples of the dyslexia/reading/spelling spectrum. *DCDC2* and *KIAA0319* may contribute to slightly different subphenotypes, thereby providing an explanation for the differing results within this locus. One might speculate that the *DCDC2* gene has more relevance for spelling ability, whereas *KIAA0319* has more relevance for word reading. However, Marlow et al. (2003) performed a multivariate linkage analysis in UK samples (Fisher et al. 2002), and showed that the QTL in *DYX2* influences multiple aspects of reading ability, not specific phenotypic subdimensions. Population-specific genetic heterogeneity within the *DYX2* region could also explain the different association findings in different populations. Nevertheless, both genes are probably involved in the development of dyslexia. As implied by Harold et al. (2006), there may be epistatic interactions between the two genes, contributing to the risk. However, these interactions should be verified by replication on independent sample sets. Collaborative efforts with large sample sizes are required for definite answers about the contribution of each of the two genes within this locus to the development of dyslexia.

5.1.3 Evidence for *DYX1C1* as a sex-specific risk factor for dyslexia

The *DYX1C1* gene on chromosome 15q21 was the first candidate gene proposed for dyslexia, based on cloning of a translocation breakpoint co-segregating with dyslexia and association of genetic variants in a Finnish cohort of dyslexic individuals (Taipale et al. 2003). Since then, seven replication studies have been published on *DYX1C1*. Support for *DYX1C1* as a risk factor for dyslexia came from three studies on families of European-Canadian, US, and Italian descent, in which SNPs and haplotypes within the *DYX1C1* gene were associated with dyslexia (Wigg et al. 2004; Brkanac et al. 2007) and with short-term memory (Marino et al. 2007). However, in two of these studies (Wigg et al. 2004; Brkanac et al. 2007), the observed association was to the opposite alleles as observed in the original study by Taipale et al. (2003). In four other dyslexia sample sets, mainly of European descent, nominal associations of *DYX1C1* SNPs were interpreted as negative (Scerri et al. 2004) or no associations with dyslexia were observed (Cope et al. 2005b; Marino et al. 2005; Meng et al. 2005a). The role of *DYX1C1* in dyslexia has thus been considered unsettled (Fisher and Francks 2006).

Therefore, we aimed to explore the contribution of *DYX1C1* to dyslexia in the German population and analyzed a large family-based sample of 366 triads. Adding the newly identified variants from a re-sequencing effort of *DYX1C1* disclosed a refined

haplotype structure, showing association with dyslexia. A detailed analysis revealed that the identified risk haplotype was associated with the dichotomous trait of spelling, in particular in females. When analyzing quantitative subcomponents for dyslexia, phonological awareness showed significant over-transmission to dyslexic subjects, in the full sample set as well as in males. In females, association was seen for short-term memory. Quantitative analysis of subphenotypes may have decreased the heterogeneity in our sample set, thereby providing more power to find an underlying effect. However, due to the multiple testing of several subphenotypes, as well as the relatively small sample size of females, these findings require independent replication before any definite conclusions about the sex-specific genetic effect of *DYX1C1* can be made. Nevertheless, to our knowledge, this study represents the first molecular genetic analysis on dyslexia considering an influence of sex on the risk conferred by specific genetic variants. Sex is an environmental factor that modifies both penetrance and expressivity of a disorder, and gene-environment interactions may have different effects on the same variant in men and women. Because differences in prevalence rates between females and males have been reported in epidemiological studies, an inclusion of sex in genetic analyses of dyslexia is warranted (Rutter et al. 2004; Harlaar et al. 2005). Interestingly, in a recently published genome-wide linkage study on autism (Szatmari et al. 2007), suggestive linkage was reported to 15q21-q25.3 that includes *DYX1C1*, and increased support for this region was obtained from families containing affected females. It may be hard to detect effects of genes for complex traits without separating the two sexes in analysis, and difficult to replicate in samples with different male-female ratios (Weiss et al. 2006). Most, if not all, dyslexia cohorts are enriched for male probands, and thus it may be difficult to replicate a female-specific effect unless very large sample sets are analyzed.

Several studies reported associations with the same alleles as Taipale et al. (2003), although in opposite directions, and were therefore interpreted as negative findings. These associations with the opposite alleles are not uncommon in complex diseases, where the interplay of multiple loci and environmental factors contribute to the susceptibility. Multiplicative or additive effects of multiple risk loci can produce associations with alleles in opposite directions due to differences in the correlation between the observed variant and other causal variants (Lin et al. 2007). LD patterns across the same gene may be highly variable across populations, even in populations of same ethnicity (Lin et al. 2007). Different populations may have different alleles in LD with the causal variant, as the causal variant may reside on different haplotypic backgrounds. Refinement of the block structure and using haplotype analysis aided in identification of the true risk haplotypes in our study. For example, Meng et al. (2005a) studied only the two originally identified SNPs, and perhaps therefore did not find any association within *DYX1C1*. In contrast, Wigg et al. (2004) studied additional intronic SNPs and found evidence of association with dyslexia. Even though Marino et al. (2005) found no evidence of association of the two original *DYX1C1* variants, in their subsequent study with more relaxed criteria for dyslexia they identified significant LD with short-term memory in both single-marker and haplotype analyses (Marino et al. 2007). Interestingly, also Wigg et al. (2004) found association with short-term memory. Marino et al. (2007) proposed that *DYX1C1* has more effect on short-term memory than on the categorical definition of dyslexia itself or other related phenotypes. Our results

are also in line of this hypothesis that *DYX1C1* affects a more defined subphenotype for dyslexia. However, it may also be that the two measures of short-term memory and phonological awareness limit the heterogeneity of the disorder, thus allowing for the identification of an effect. In fact, phonological awareness is perhaps the strongest indicator for dyslexia (Peterson et al. 2007), and problems in verbal short-term memory have been linked to the phonological processing deficits (Snowling 2000). Nonetheless, a meta-analysis of the comparable published results for *DYX1C1* showed that the haplotype structures across studies are consistent with a single, common risk haplotype in all the studied populations.

5.1.4 *CYP19A1* is associated with dyslexia and speech- and language disorders

The linkage peaks within the *DYX1* locus identified in different studies on dyslexia map approximately 8 Mb proximal to *DYX1C1* (Figure 11). Given the imprecision of genetic linkage for complex disorders, this is not unexpected. However, as the breakpoint in another translocation family co-segregating with dyslexia also mapped 6-8 Mb proximal from *DYX1C1* (Nopola-Hemmi et al. 2000), it seemed possible that there may be another susceptibility gene for dyslexia within the *DYX1* locus. The inconsistencies to replicate the *DYX1C1* association in some sample sets further suggested that *DYX1* might harbor another, more general gene for speech and language development and dyslexia. There is evidence for more than one susceptibility gene within other dyslexia loci as well. For the *DYX2* locus, there are already two susceptibility genes, as both *DCDC2* and *KIAA0319* contribute to dyslexia. Within *DYX3* there may be two separate dyslexia loci, as several studies have identified linkage to 2p15, in addition to our reported 2p12 locus.

We refined the location of the translocation breakpoint in the dyslexic individual co-segregating t(2;15)(p12;q21) and dyslexia to the complex promoter region of *CYP19A1*, ~4 Mb centromeric from *DYX1C1* (Figure 11). This suggested that the disruption of the regulatory region of *CYP19A1* might be the underlying cause of dyslexia in this individual. However, all three children and the father in the family carried the translocation, but only the female child had been diagnosed with dyslexia (see Figure 6). This could be due to reduced penetrance or variable expressivity affected by the effects of other genes or environmental factors. Sex-specific effects could also affect the variable expression, as already shown for *DYX1C1*, and was also demonstrated for the *DYX1* locus in general in a genome-wide linkage study on autism, where stronger linkage to the 15q21-q25.3 region was obtained in families containing affected females (Szatmari et al. 2007).

To study the role of *CYP19A1* in reading and speech- and language functions, we analyzed SNPs covering the whole genomic region of *CYP19A1* in three sets of families ascertained for dyslexia and SSD. These disorders are co-morbid and most likely share a common etiology (Pennington 2006). As in dyslexia, the deficit in SSD lies in phonological processing. Moreover, the *DYX1* locus has also been implicated in SSD (Figure 11). We observed significant associations for the binary traits of dyslexia

and SSD as well as for a range of reading-related quantitative measures within the *CYP19A1* locus, particularly in the two US cohorts. In the Finnish cohort, association was seen only at the haplotype level. However, as the results were stronger in the two US cohorts when analyzing the specific quantitative phenotypes, the power in the Finnish sample set may have also been increased if quantitative measures would have been available. In the Finnish set, a haplotype associated with dyslexia was observed across the coding region of the gene, whereas in the US dyslexia set the haplotype was located within the brain-specific promoter. In the US SSD cohort, overlapping haplotypes were observed for both these regions. These different haplotypes observed reflect the quite different patterns of LD observed across the *CYP19A1* region (see Figure 10). The Finnish sample set showed stronger LD across the gene than the US sets, which is not unexpected as Finland is a relatively homogeneous population (Peltonen et al. 1999), and therefore also the precise location of the association signal may be more difficult to resolve. The fact that slightly different SNPs show association in different sample sets and for different measures, may also be due to stochastic variation as a result of relatively small sample sizes and by sample specific patterns of missing genotype/phenotype data. In case of a true association, a combined analysis of all sample sets should be much more powerful (Skol et al. 2006). However, as the analyzed samples were ascertained for slightly different phenotypes, i.e., SSD and dyslexia, it was not desirable to pool the data. Instead, we combined the significant p-values for the common measures used in the two US cohorts by Fisher's combined method, and obtained stronger association results.

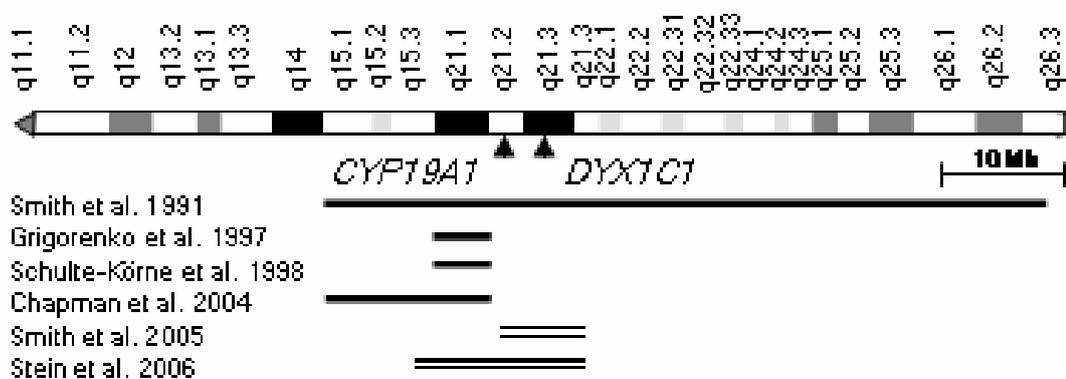


Figure 11. Chromosome 15q showing the relative positions of *CYP19A1*, *DYX1C1* and the linkage peaks in different studies of dyslexia (solid lines) and speech sound disorder (double lines).

Our results suggest that variation in *CYP19A1* may influence a broad phenotype of verbal skills, in particular phonological processing, and contribute to the susceptibility for both SSD and dyslexia. Genetic variation in *CYP19A1* has been associated with Alzheimer's disease (Iivonen et al. 2004; Huang and Poduslo 2006), and now we show that *CYP19A1* regulates also development of speech- and language functions early in life, as well as later reading skills.

5.1.5 General comments

In order for any genetic linkage or association to be valid, the results need to be replicated on an independent sample set to separate the false positives from the true associated variants. In general, if an association is found between the same markers and alleles and the same phenotype in an independent study, there is strong evidence for true association. However, replication studies in complex disorders have often yielded inconsistent results with different subphenotypes, markers, or even different alleles of the same marker often associated in the follow-up study. For complex disorders, there are multiple susceptibility genes with small to moderate individual effects, which may be increased or modified in gene-gene interactions and gene-environment interactions. There is also substantial genetic and phenotypic heterogeneity across and within populations. Some genes may have a major effect on the trait variance in the samples they have been reported, whereas in other sample sets they may have less effect. Thus, it is not surprising that different research groups have identified unique locations for linkage and association signals and have not always found support for loci reported by others. Moreover, the different studies are usually not directly comparable as there are differences in the phenotype definition, ascertainment, diagnostic criteria, marker selection, and the analysis methods. Therefore, lack of replication does not imply exclusion of a locus/gene. Rather, further validation may be required in even more sample sets, and functional studies of the candidate gene to validate its role in disease etiology. The sample sizes in genetic studies of complex diseases have often been too small, resulting in insufficient power to detect small to moderate effects, as well as imprecise or incorrect estimates of the magnitude of the observed effects. It is now generally agreed that for reliable association results large sample sizes, rigorous p-value thresholds, and replication in multiple independent sample sets, are required (Chanock et al. 2007).

The ascertainment method, the measures of the phenotype, and frequency and variance of the phenotype in the sample, all influence the genetic findings (Pennington 1997). A clearly defined phenotype classified by standard criteria should be used to reduce misclassification and to aid subsequent replication studies. However, due to the absence of a consensus definition for dyslexia, phenotype definitions have varied considerably across studies yielding increasing genetic heterogeneity, making it hard to interpret data across studies. The inclusion criteria may vary according to the number of affected individuals within the pedigree as well as the degree of severity, and different phenotypic measures have been used to assess the scores for the quantitative cognitive processes involved in reading and dyslexia. As there is substantial heterogeneity, it may be advantageous to divide according to severity or into subphenotypes to obtain a more homogeneous sample, instead of setting an arbitrary threshold of “affected” for a trait that is normally distributed within the general population. With several subphenotypes one may, however, lose power as the sample size gets smaller, and inflate the false positive rate due to multiple comparisons. However, whereas the probability for detection of the susceptibility locus may increase when analyzing more precise definitions, most of the dyslexia loci seem to affect reading in general. The correlation between the different components of reading and language are high and twin studies have shown that there are common as well as independent genetic factors (Gayan and

Olson 2003). Although different studies have found association/linkage to the same locus with different subcomponents, a multivariate genome-wide analysis in a large sample set suggested that each dyslexia locus has an impact on multiple traits (Marlow et al. 2003). The differing results observed may only reflect the different phenotypic measures used and heterogeneous samples. Some loci may contribute more strongly to one component than to another within specific families and populations. Even though precisely defined specific subphenotypes are analyzed, each subphenotype is a complex phenotype on its own, consisting of several independent processes. Different measures may target these subprocesses differently, leading to slightly different traits analyzed and thereby also to different results.

5.2 MOLECULAR MECHANISMS OF DYSLEXIA

Identification of the molecular mechanism for a susceptibility gene has traditionally involved extensive screening of the coding regions, UTRs, splice sites, and the putative promoter regions in immediate vicinity of the transcription start site. However, complex genetic diseases result more often from non-coding regulatory variants than from coding sequence variants (Knight 2005). To screen the entire genomic region of a gene in multiple subjects, as well as the often poorly defined promoter and other regulatory elements that may stretch over large distances, is a laborious task. Even if a putative causal variant is identified, interpreting the consequences of non-coding sequence variants is not straightforward, in contrast to nonsynonymous and splice-site alterations, as the complexity of gene regulatory regions is not yet fully understood.

So far, no coding variants to provide a simple functional explanation have been identified in any of the dyslexia candidate genes. The coding changes that have been found in the dyslexia susceptibility genes appear to have little functional relevance. Rather, the causal variants are likely to be in regulatory regions affecting the expression of these genes. For a subtle effect in the cognitive functions affecting the phonological skills it is not expected that the causative variant would be a deleterious coding change, rather some milder effect in a regulatory region. There is evidence for already four dyslexia susceptibility genes that the causal mechanism affects the regulation of the genes. Allele-specific expression analysis of *ROBO1* (Hannula-Jouppi et al. 2005) and *KIAA0319* (Paracchini et al. 2006) and the chromosome 2 candidate genes *C2ORF3* and *MRPL19* (II), have suggested that disrupted expression of one copy of the gene can predispose to dyslexia, maybe because of reduced amount of protein in the CNS. However, it should be noted that as the expression analyses were performed in lymphocytes, the effects are not necessary directly comparable to the expression in brain. Moreover, these patterns may not reflect the expression during fetal development, when the correct expression of dyslexia susceptibility genes is expected to be most critical. Allelic variation in gene expression is likely to be present for most genes, providing an additional mechanism of human variation and probably also for susceptibility to common diseases (Lo et al. 2003). Reduced amounts of functional

protein due to, e.g., impaired gene regulation, may well lead to developmental disorders.

The original risk alleles identified in *DYX1C1* were hypothesized to have a putative effect on the encoded protein (Taipale et al. 2003). However, as all replication studies have reported association to the opposite alleles than Taipale and co-workers, it is unlikely that these are the causal polymorphisms. Instead, they are probably in LD with the causal variants, which probably affect the expression and/or regulation of the gene. Our re-sequencing approach revealed variants within the *DYX1C1* promoter region that added new information to the gene-specific haplotype architecture and refined our association findings. The fact that we observed allele-specific differential binding in the EMSA with the associated SNPs in the promoter region suggest that these SNPs may affect the expression level of *DYX1C1*, providing a molecular mechanism for its role in the etiology of dyslexia. *In silico* searches suggested allele-specific differential binding of TFII-I and Elk-1. TFII-I is highly expressed in brain and is deleted in Williams-Beuren syndrome characterized, e.g., by unusual language functions (Roy 2001; Mervis and Berra 2007). Elk-1 has been shown to have a role in learning and memory in rats (Cammarota et al. 2000). Interestingly, the binding of these transcription factors was found to be reduced also to the dyslexia-associated alleles in the highly conserved brain-specific promoter of *CYP19A1*. A comparative analysis across several species highlighted the importance of the brain-specific promoter, with slight alterations in its sequence most likely leading to critical consequences in gene regulation. Thus, the SNPs associated with dyslexia and SSD might regulate the expression of *CYP19A1*. The fact that the translocation breakpoint co-segregating with dyslexia is located within the complex promoter region of *CYP19A1* further supports a regulatory effect as the causal mechanism.

Further functional studies are required to determine the underlying causative variants and their exact mechanism in the etiology of dyslexia. However, functional studies are not always straightforward and the findings should be interpreted carefully, as results from particular cell-lines and experimental settings studied may not be valid *in vivo*. To prove causality of a candidate gene requires the functional mutations or regulatory variants to be identified and to demonstrate that these mutations disrupt brain development in animal models. Finally, it should be shown that dyslexic individuals carrying these mutations show disruptions in brain development. An interaction is probably required between many gene variants and their protein products in the development of dyslexia. Dyslexia is a heterogeneous disorder, with many subtypes at the cognitive, neurobiological, and genetic level. Different individuals may have different genetic backgrounds and slightly different causal pathways for the disorder. Certain combinations of alleles may engender high susceptibility for the disorder, and some variants may be of high severity. These genes or alleles are, however, not really “genes for dyslexia”, or for reading or language (Pennington 1997). They will probably be involved in other developmental disorders as well. Our quantitative expression analysis showed wide expression of all the dyslexia susceptibility genes across different regions in the human brain, including regions implicated in reading. However, the expression was not restricted to language areas. Moreover, their expression is not restricted to the CNS either; rather, most of them are ubiquitously expressed.

5.3 NEUROBIOLOGICAL MECHANISMS OF DYSLEXIA

The brain is a highly complex system that requires a complex network of events to develop correctly. Small disruptions in this circuitry may result in a neurodevelopmental disorder, such as dyslexia. Reading is a complex task and deficits may arise from several of the associated cognitive processes. The core deficit in dyslexia in the phonological system involves a network of cortical areas (Figure 12). The observed structural anomalies reported in dyslexia are located within these regions (Figure 12A), and functional brain imaging studies show that the areas involved in phonological processing show abnormal activation in dyslexics. The anomalies and activation patterns vary between individual dyslexics. The differences may reflect variation in the symptoms and subcomponents, as well as the variation in the underlying susceptibility genes. Dyslexic individuals often compensate for their disability, thereby activating other neural circuits for reading. There may be differences in these compensatory mechanisms employed. However, the compensatory mechanisms are most likely not as efficient, and thus the problems in fluent reading often persist throughout their lives.

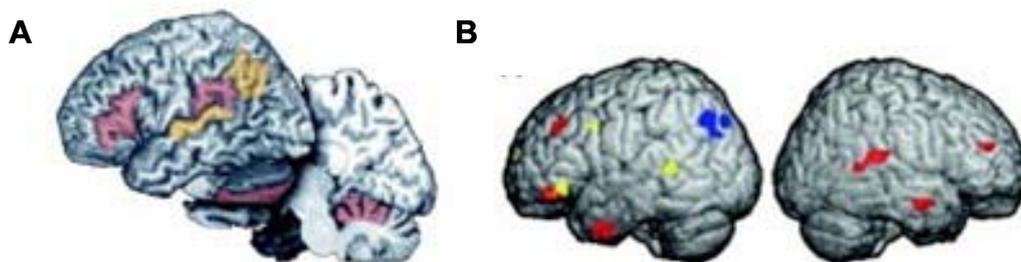


Figure 12. Neurobiological findings in dyslexia. (A) Brain areas activated in oral language tasks that exhibit structural differences from controls in studies of dyslexia. Areas in pink are supported by several published studies, areas in yellow by only one. (B) Brain areas activated during performance of the main phonological skills impaired in dyslexia: phonological awareness (yellow), rapid naming (red), and verbal short-term memory (blue). Reprinted from *TRENDS in Neurosciences*, 27; Ramus, *Neurobiology of dyslexia: a reinterpretation of the data*; 720-726. Copyright 2004, with permission from Elsevier.

5.3.1 Neuronal migration and dyslexia

Several lines of evidence support neuronal migration deficits and the resulting cortical abnormalities (ectopias and microgyri), as the primary cause of dyslexia. Anatomical studies of dyslexic brains show cortical abnormalities, functional brain imaging studies show that these areas are involved in phonological processing and show abnormal activation in dyslexics, mouse and rat models with ectopias show learning deficits, and finally, many of the dyslexia candidate genes are involved in neuronal migration (Ramus 2004).

Galaburda and co-workers proposed already in the 1980's that ectopias and neuronal migration would be the causal mechanism for developmental dyslexia (Galaburda et al.

1985). However, as there was no attempt to replicate these findings, they have been considered inconclusive until recently, when dyslexia susceptibility genes have been identified and their function is beginning to be revealed (Ramus 2006). Five of the seven susceptibility genes reported for dyslexia so far have been implicated in brain development. *Dyx1c1*, *Dcdc2*, and *Kiaa0319* are involved in cortical neuronal migration during development, as indicated by RNAi *in utero* in rat (Meng et al. 2005b; Paracchini et al. 2006; Wang et al. 2006). Interestingly, disruption of *Dyx1c1 in utero* results in brain malformations in adult rats similar to those seen in postmortem brains of dyslexics (Rosen et al. 2007). The effects of *Dcdc2* and *Kiaa0319* have not been studied in adult rats, so no comparisons to the dyslexic neocortical phenotype can be made. *Robo1* is implicated in axonal guidance across the commissures as well as in neuronal migration (Andrews et al. 2006). *Robo1* knockout mice show cortical disorganization characterized by increased neuronal density and commissural defects due to altered axonal pathfinding. Interestingly, ArKO mice showed similar cortical disorganization and occasional ectopias, indicating an important role of also *Cyp19a1* in brain development, perhaps as well in neuronal migration. Estrogen is actively synthesized in the brain by the enzyme aromatase (*CYP19A1*). Estrogen regulates several processes that are necessary for the proper development of brain structures and connections, including neuronal migration, survival and death, and synaptic plasticity (Beyer 1999). The putative role of *MRPL19* and/or *C2ORF3* in human brain development has not been studied so far. If they are found to cause similar phenotypes in brain organization as the other dyslexia susceptibility genes, their role in the development of dyslexia is reinforced.

Evidence from studies on mice and rats further support the hypothesis that dyslexia results from abnormal neuronal migration resulting in cortical disorganization. Induction of microgyri in rats leads to local changes in cortico-cortical connectivity and changes of connectivity across corpus callosum (Galaburda et al. 2006). Reduced corticocortical connectivity has been reported in dyslexic individuals as well (Paulesu et al. 1996; Klingberg et al. 2000). It has been hypothesized that the cortical abnormalities observed in dyslexia lead to a disconnection between the different language systems, resulting in reduction of activity in the brain areas involved in phonological processing (Paulesu et al. 2001). Furthermore, mice and rats with spontaneous or induced ectopias and microgyri display a variety of learning and memory deficits (Denenberg et al. 1991; Rosen et al. 1995; Boehm et al. 1996; Hyde et al. 2000). The location of the cortical disruption influences the specific type of learning deficit exhibited (Hyde et al. 2001). Focal cortical malformations are also associated with deficits in rapid auditory processing in rodents (Fitch et al. 1994; Peiffer et al. 2004). The auditory deficits are more marked in young than in adult animals; compatible with that the auditory deficits in dyslexic are mostly apparent in children only, as the condition improves with age due to compensatory mechanisms (Galaburda et al. 2006). Auditory deficits may therefore often be undetectable at the time of diagnosis of dyslexia. In others, auditory deficits could remain and the phonological problems improve, leading to a diagnosis other than dyslexia (Galaburda et al. 2006). Interestingly, RNAi disruption of *Dyx1c1* in rats results also in auditory processing deficits and impairments in spatial learning (Threlkeld et al. 2007). The important roles of estrogen in learning and memory have also been demonstrated by studies on rats and

non-human primates, although so far these studies have been performed only on ovariectomized, hormone-treated animals (Korol 2004; Hao et al. 2006).

How can a general neuronal migration disorder produce such a specific phenotype as dyslexia, rather than, e.g., mental retardation or general learning disability? The dyslexia susceptibility genes are probably involved in multiple functions during development. A subtle effect in a mild quantitative trait like reading disability would not be expected to result from a devastating mutation of a protein critical for neuronal development; rather a small change in its expression level. Such change could affect the efficiency of neuronal migration. The deficit may be defined only to certain cortical areas, such as those that are necessary for phonological processing, at a specific time point in development. However, the susceptibility genes may lead to more severe developmental phenotypes if their function is disrupted more radically, resulting in large-scale brain malformations. In addition, it is likely that they may confer risk to other, co-morbid learning disorders as well, depending on the precise disruption as well as the effect of other genes.

5.3.2 A neuronal network model for dyslexia

Generation of the cerebral cortex is a complex and highly coordinated process. In the developing brain, cortical neurons are generated in the ventricular and subventricular zones. Postmitotic neurons then migrate along processes of radial glia to form the six layers of the cerebral neocortex. The layers are formed in an “inside-out” manner from deepest to most superficial, so that each newly generated cohort of neurons must migrate past the previously formed neurons. Appropriate neuronal migration and positioning during development is essential for proper brain function and the construction of a functional synaptic circuitry.

The developmental pathways involved in neuronal migration and axon growth are dependent on coordinated changes in cell adhesion and cytoskeletal reorganization. Neuronal migration is achieved through a rearrangement of cytoskeletal components in response to extracellular cues, mediated by intracellular signaling pathways (Ayala et al. 2007). Galaburda et al. (2006) proposed a possible molecular network for the dyslexia susceptibility genes in neuronal migration (Figure 13). The role of *ROBO1* orthologues in axonal guidance is well established. In addition, mouse *Robo1* guides tangential migration of cortical interneurons (Andrews et al. 2006), and maybe also radial migration (Ayala et al. 2007). *KIAA0319* encodes an integral membrane protein which may serve as an adhesion molecule due to its homology to PKD1 which is involved in cell adhesion (Paracchini et al. 2006). RNAi studies have suggested that *Kiaa0319* could mediate the necessary adhesion between neurons and radial glia in radial neuronal migration (Paracchini et al. 2006). These two transmembrane proteins, *ROBO1* and *KIAA0319*, may relay signals to the intracellular signaling pathways, eventually leading to rearrangement of the cytoskeleton and neuronal movement. *DCDC2* contains two doublecortin homology domains that were originally identified in the doublecortin gene *DCX*. *DCX* encodes a cytoplasmic protein that directs neuronal

migration by regulating the organization and stability of microtubules, and is mutated in human X-linked lissencephaly and double cortex syndrome (Reiner et al. 1993; des Portes et al. 1998; Bai et al. 2003). Thus, *DCDC2* may be involved in modulating changes in cytoskeletal dynamic processes involved in neuronal migration. Microtubules provide stability to the growing neurites and are crucial for the association of the centrosome to the nucleus during nucleokinesis. Thus, it is clear that proper regulation of microtubule dynamics is necessary for neuronal migration (Ayala et al. 2007). The cellular function of *DYX1C1* is unknown, but it contains three tetratricopeptide repeat domains, which are known to be involved in protein-protein interactions (Taipale et al. 2003). The tetratricopeptide repeat domains of *Dyx1c1* are necessary and sufficient for normal neural migration (Wang et al. 2006). *DYX1C1* is rapidly upregulated in injury (Linkai et al. submitted), and may thus be involved in dynamical cellular processes.

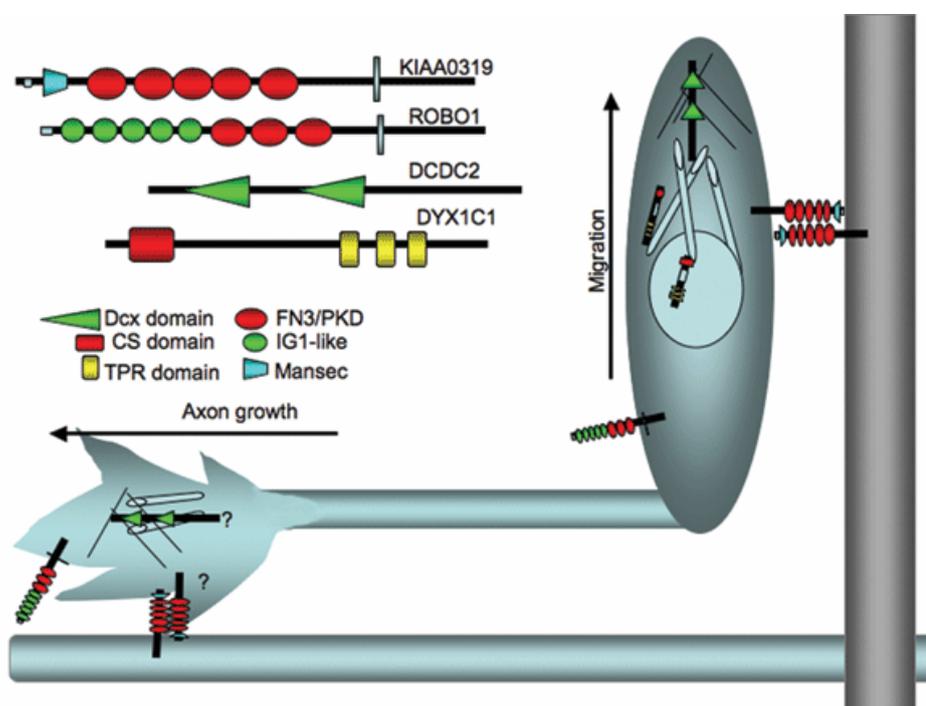


Figure 13. Protein domains and possible functions of the dyslexia candidate genes during neuronal migration and axonal pathfinding. MRPL19, C2ORF3, and CYP19A1 are missing from the model. Reprinted from Nature Neuroscience, 9; Galaburda et al., From genes to behavior in developmental dyslexia; 1213-1217. Copyright 2006, with permission from Macmillan Publishers Ltd.

The domain structure or cellular localization of *C2ORF3* is still unknown. However, in our quantitative expression analysis, *C2ORF3* was highly co-expressed with the other dyslexia susceptibility genes, including *ROBO1*. Interestingly, microarray data indicate that *C2ORF3* is highly co-expressed with *SRGAP1* (bioinformatics.ubc.ca/tmm). *SRGAP1* is a SLIT-ROBO1 Rho GTPase-activating protein that interacts with an intracellular domain of *ROBO1* and is involved in the downstream signaling cascade upon binding of the SLIT1/2 ligand to the *ROBO1* transmembrane receptor (Wong et al. 2001). The Rho GTPases play important roles in regulating the actin cytoskeleton and have been implicated in axon guidance, neurite extension, and neuronal migration

(Govek et al. 2005). Mutations in these signaling pathways have been reported in human neurological disorders, emphasizing their importance in the development and proper function of the nervous system (Govek et al. 2005). However, functional studies are needed to reveal the cellular localization and possible role of C2ORF3 in neuronal development.

Most of the energy requirement for cellular growth, differentiation, and migration is met by ATP produced by mitochondria in oxidative phosphorylation. *MRPL19* is a highly conserved gene that may have a central role in ribosome biogenesis and mitochondrial protein synthesis. Minor changes in the protein, leading to marginally impaired energy metabolism may have developmental consequences in critical tissues, such as impaired neuronal migration during development. Many of the mitochondrial ribosomal proteins encoded in the nucleus have been associated with several neurological disorders, such as deafness (O'Brien et al. 2005), in accordance to the fact that energy production is critical in the active brain.

The neuronal expression of *CYP19A1* localizes to the cell soma, axons, dendrites, synaptic boutons and the growth-cones (Naftolin et al. 1996; Beyer and Hutchison 1997). Estrogen regulates the transcription of cytoskeletal proteins that are required for neurite growth (Beyer 1999). In addition, estrogens have more direct, rapid-acting roles on neural activity by modulating neural signaling transduction pathways, e.g., by altering the properties of extracellular receptors (Cornil et al. 2006). Our results support an important regulatory role for local aromatase activity and estrogen production in brain development, as inhibition of aromatase activity significantly decreased neurite outgrowth and branching in rat hippocampal cultures. An effect of estradiol on neurite outgrowth has been shown also in cortical neurons (Brinton et al. 1997). Thus, aromatase may have an important regulatory role in the neuronal network involved in dyslexia.

5.4 SEX-SPECIFIC EFFECTS OF DYSLEXIA

Our findings within the *DYX1* locus may offer insights into previous reports that dyslexia and SSD are more common in boys than in girls, a phenomenon that has remained unexplained to date. Although it has been suggested that dyslexia affects males and females nearly equally (Rumsey 1992), recent studies have indicated that there is an increased risk in males, and at least before 8 years of age, severe risk is more heritable in males (Rutter et al. 2004; Harlaar et al. 2005).

The different genetic effects of *DYX1C1* that we observed in males and females, leading to a slightly different phenotype, are not surprising as several male-female differences have been seen in brain anatomy, chemistry and function (Cahill 2006). Many factors, such as hormonal influences and interaction with sex-linked genes, may affect the sex-specific expression and regulation. Several susceptibility genes for common complex traits act differently in the two sexes, even when the trait

distributions don't differ by sex (Weiss et al. 2006), and there is evidence for sex-specific linkages for psychiatric traits such as autism (Stone et al. 2004; Lamb et al. 2005; Szatmari et al. 2007).

Although we did not observe a sex-specific effect within the *CYP19A1* locus, it does not imply that such an effect does not exist. The overall genetic effect of *CYP19A1* in the development of dyslexia and speech- and language disorders is likely to be small, and therefore much larger sample sizes may be needed for the identification of specific effects. In association of *CYP19A1* variants with Alzheimer's disease, no sex-specific effects for the associated SNPs were observed either (Iivonen et al. 2004; Huang and Poduslo 2006), despite a significant difference between the sexes in disease pathology (Barnes et al. 2005). Nevertheless, aromatase plays a decisive role in development and in the sexual dimorphism of the brain (Colciago et al. 2005). Sexually dimorphic phenotypes are evident in ArKO mice, in which males show gender-specific loss of dopaminergic neurons in the medial preoptic area and arcuate regions of the hypothalamus, as well as gender-specific obsessive compulsive disorder (Hill et al. 2004; Hill et al. 2007). Rats with cortical microgyri develop similar anomalies in the MGN as observed in dyslexic brains, i.e., more small and fewer large neurons (Herman et al. 1997; Peiffer et al. 2002a; Peiffer et al. 2002b). In accordance with the hypothesis that a disruption of MGN has an impact on the auditory system, these rats are impaired in an auditory discrimination task. Interestingly, only male rats and mice with microgyri or ectopias were initially found to have an impairment in the auditory system (Herman et al. 1997). Later, it was observed that the impairment was dependent on hormonal stimuli, as female rats that were injected with testosterone during gestation showed disruption of MGN and impaired auditory function (Rosen et al. 1999). Interestingly, androgens have been suggested to have effects on early human brain development as well, with excess levels impairing auditory temporal processing and possibly also phonological processing in the left hemisphere (Beech and Beauvois 2006). Aromatase shows high expression in thalamus, and the enzyme activity has suggested to be higher in males (Lephart et al. 2001). Proper aromatase function could play an important regulatory role of the hormonal environment, and enlighten, at least in part, the sex-specific effects observed in dyslexia. Developmental differences in males and females could explain why more males are generally affected by developmental disorders, including dyslexia and SSD (Galaburda et al. 2006). Plasticity in cortical development may vary with age and sex, resulting in individual differences in the behavioral phenotype.

5.5 EVOLUTION OF SPEECH, LANGUAGE AND READING

Although only humans can read, all genes implicated in dyslexia so far are conserved across species. While slight alterations in these genes may increase the risk for dyslexia, they are not "genes for reading" and did not evolve for reading (Fisher 2006). Rather, they are involved in important functions of brain development, such as neuronal migration as implicated by RNAi studies in rats and studies of brain morphology in KO

mice. However, subtle changes in these neurodevelopmental genes may have been involved in the evolution of higher-order cognitive skills in humans. Humans are distinguishable from other species in the use of vocal speech, reading and written language. There are, however, similarities in the development of human language and birdsong that also requires a vocal-learning ability (Doupe and Kuhl 1999). Interestingly, aromatase is involved in the control of vocalization for songbird singing and in the communication of teleost fish (Forlano et al. 2006). Our results show that *CYP19A1* also controls the development of related but higher-order skills in humans, such as speech, language, and reading. Comparative analysis of the *CYP19A1* genomic region revealed high levels of conservation within the brain-specific promoter region across a wide range of vertebrates. The brain-specific promoter is highly conserved even in the songbird (Ramachandran et al. 1999), highlighting its important regulatory function. Remarkably, we identified a human-specific SNP within the brain-specific exon I.f that showed completely different patterns of protein binding than the non-human allele. This SNP could therefore have a functional role for altered aromatase regulation and/or expression in the human brain.

MRPL19 is also a highly conserved gene with only a few nucleotide changes identified between the coding region of human and non-human primates. *MRPL19* may have a central role in ribosome biogenesis and mitochondrial protein synthesis across a wide range of species. Thus, it is not surprising that it is under strong evolutionary constraint, although proteins in mammalian mitochondrial ribosomes are evolving more rapidly than cytoplasmic ribosomal proteins (O'Brien 2002). In contrast, rather high levels of variation were identified between *C2ORF3* and its primate orthologues. The non-primate lineage showed comparatively higher *dN/dS* ratios, suggesting that *C2ORF3* may be under functional constraint due to an important function in the brain acquired during primate evolution. The fact that the human lineage showed the lowest *dN/dS* ratios may indicate an even more specialized and constrained function in humans. However, a test for heterogeneity among the primate species revealed no evidence of change in the selection pressure during primate evolution of *C2ORF3*. The relatively low *dN/dS* ratios estimated for *C2ORF3* are consistent with previous reports of low *dN/dS* ratios for nervous system genes (Duret and Mouchiroud 2000). However, the stringent definition of adaptive evolution ($\omega > 1$) in estimations of selection pressure may be misleading for many genes expressed in brain, as low *dN/dS* values may mask signs of adaptive evolution. Nevertheless, an accelerated rate of protein evolution in primates, especially in the human lineage, has been shown for a subset of genes important for nervous system development and function (Dorus et al. 2004; Khaitovich et al. 2005). Positive selection during recent human evolution has been suggested for the highly conserved *FOXP2*, a gene implicated in language development and that is mutated in a severe form of speech and language disorder (Lai et al. 2001; Enard et al. 2002). Interestingly, orthologues of *FOXP2* are also involved in the vocal control of birdsong (Teramitsu and White 2006), similarly to *CYP19A1*. The selection pressure of *ROBO1* has been found to be different between the human, chimpanzee and gorilla branches as compared to the orangutan (Hannula-Jouppi et al. 2005), although *ROBO1* has since been proposed to be a slowly evolving gene due to the large excess of silent changes in each primate lineage (Fisher and Francks 2006). Within *DYX1C1*, several coding changes between human and non-human primates have been identified,

but no tests for deviation of a normal evolutionary rate have been performed (Taipale et al. 2003). The evolution of *KIAA0319* and *DCDC2* have so far not been studied, although also these genes exist in very similar versions in other species, such as in rodents.

6 CONCLUSIONS AND FUTURE PERSPECTIVES

When I started this thesis project in year 2002, no candidate genes for dyslexia had been published. Since then, there has been a remarkable expansion in the field, with currently seven susceptibility genes proposed; *DYX1C1*, *ROBO1*, *DCDC2*, *KIAA0319*, *MRPL19*, *C2ORF3* and *CYP19A1*. The number is likely to grow even more, as there are still several replicated linkage regions within the human genome where no candidate gene has yet been identified.

Reading is a complex cognitive process involving a large network of factors, and consequently also multiple genes of small to moderate effects. The encoded protein products are involved in brain development through a complicated network of interactions required for proper function. Slight alterations in this network may lead to subtle deficits such as reading disability. The dyslexia susceptibility genes are involved in common developmental processes such as neuronal migration and axon guidance, suggesting an underlying causal mechanism in dyslexia. Slight alterations in the expression of the dyslexia susceptibility genes may alter the neuronal migration producing subtle cortical malformations affecting cognitive processes critical for learning and reading.

The major challenge is to identify the precise causal variants and distinguish them from neutral variants, and to identify the underlying etiology for the disorder. The improvement of genotyping and statistical methods and enlargement of cohort sizes in large collaborative studies have begun to give more insight into the genetic basis of complex disorders. The availability of high-density SNP maps and high-throughput genotyping with decreasing costs has made whole-genome association studies technically and practically feasible. In the future, complete re-sequencing of the entire genome in cases and controls may replace indirect SNP genotyping approaches for the identification of susceptibility genes, but so far this is only feasible for single individuals. Recently, a large European Consortium, NEURODYS (www.neurodys.com), was formed, integrating dyslexia research from various disciplines. Large dyslexia cohorts from 10 different countries are genotyped to identify new candidate genes and replicate the existing findings. The identified risk-conferring genes will also be analyzed for gene-gene and gene-environment interactions, as well as gene-specific contributions to a variety of neurobiological correlates of dyslexia. These efforts will allow for a deeper understanding of the etiology of dyslexia.

Dyslexia is the best known of the neurodevelopmental disorders, with regard to genetics as well as neuropsychology and neurobiology. Understanding the complex etiology of dyslexia will greatly aid in understanding also other learning and developmental deficits. Several of these disorders are comorbid with shared effects contributing to the phenotype. The dyslexia susceptibility genes thus make also good candidates for other neurodevelopmental disorders.

The ultimate goal for dyslexia research is more effective and earlier diagnosis and treatment. Despite the small effect sizes of the dyslexia susceptibility genes, understanding the molecular pathogenesis and the complex neural network behind it will have large effects on diagnosis and intervention. Early identification and intervention of children at risk may substantially improve future reading acquisition and even prevent reading difficulties in some cases. Written language is the basis of education systems throughout the world, and failure to learn to read fluently is a severe handicap with lifelong socioeconomic, mental health, and life satisfaction repercussions.

7 ACKNOWLEDGEMENTS

This work was carried out at the Department of Biosciences and Nutrition, Karolinska Institutet, during the years 2002 to 2007. A number of people have contributed to the completion of this thesis and I would like to express my sincere gratitude to all of you.

First and foremost I would like to thank my supervisor Professor *Juha Kere* for inviting me to come to Stockholm and for convincing me that this is The Project! I have never regretted it! In addition to your scientific vision, I truly admire your continuous enthusiasm, optimism and good mood! I am really lucky to have had you as my supervisor.

Special thanks also to *Myriam Peyrard-Janvid*, my fantastic co-supervisor! Thank you for letting me work independently enough and for all your advice, help, and support through these years, always when I have needed it! Both in science and in personal life.

Many thanks to our great dyslexia team! So many nice discussions and moments we've shared at work, at conferences... Especially big thanks for the exceptional laboratory work of *Ingegerd Fransson*, and to *Marco Zucchelli*, for your invaluable statistical analyses and patiently explaining to me (at least trying to!) the secrets of statistics - and sometimes also of real life. Thank you *Isabel Tapia-Páez* and *Hans Matsson* for your essential contribution to the dyslexia team, I really appreciate your help and support - in many matters! Thank you *Kristiina Tammimies* for your always smiling attitude! Thanks also to my dyslexisters in Finland, *Satu Massinen* (my personal advisor!), *Nina Kaminen-Ahola* and *Katariina Hannula-Jouppi*. Thank you *Päivi Onkamo* for the initial analyses and *Katja Kivinen* for your bioinformatics help, and for correcting my scientific Finnish - it was worth a lot, literally!

I would like to thank all co-authors for excellent collaboration and for enabling this thesis. In particular, I would like to thank the German team, especially *Faten Dahdouh*, *Johannes Schumacher* and *Markus Nöthen*, and the Cleveland team, especially *Lara Sucheston* and *Sudha Iyengar*. *Jan-Åke Gustafsson*, thank you for our collaboration but also for providing excellent research facilities for our group as the chair of the department. I also want to thank *Heikki Lyytinen*, *Jaana Nopola-Hemmi*, *Gerd Schulte-Körne*, *Barbara Lewis*, *Jeff Gilger* and *George Hynd* for sample collection. And of course, big thanks for all families and individuals who participated in this study - and thus made it possible!

For the excellent genotyping and co-operation at MAF, thank you *Cecilia Lindgren*, *Kristina Duvefelt*, *Astrid Lindstedt*, *Päivi Kiviluoma*, *Ville Mäkelä* and *Linda Berglind*.

I also want to thank the administrative and service personnel at the department for their help during these years, especially *Kristina Bergholm* and *Mairon Sandin* for your assistance in practical matters.

The past and present members of the JKE group - *Anna, Astrid, Cecilia, Christina, Erik, Erja, Hanna, Hannele L, Hans, Harriet, Hong, Ingegerd, Isabel, Katja, Kristiina, Kristina, Linda, Marco, Marie, Mauro, Myriam, Per, Petteri, Pu, Päivi K, Päivi O, Rasko, Sara, Tao, Tiina, Ulf* and *Ville* - thank you all!!! I have really enjoyed my time with you both in and outside of work. Special thanks to our first (party!) team, *Harriet, Petteri, Päivi O* and *Ulf*, you made me feel at home when I had just moved to Stockholm! *Erja*, thank you for great friendship and all the wonderful and crazy time we had here! Tjejmiddag-klubben - *Anna, Astrid, Christina, Hanna, Kristiina, Linda* and *Sara* - thank you for excellent food and wine, nice discussions and regular updates on the gossip!

All the members of the other groups at the department and Novum for friendly chats and all the nice time together at diverse parties and biosci pubs during these years... *Alberto, Aleksi, Amena, Ann, Antonio, Christophoros, Ivan, Jan-Phillipp, Joëlle, Juhana, Kirsten, Patrick, Paula, Tassos, Tomas...* and many, many more.

I am grateful to all my *friends* and *relatives* for your support. Especially, I would like to thank my friends during this time in Stockholm for making life also outside work really enjoyable! Thanks *Ivan* and *Joëlle* for your nice parties, the cheese, skiing, and great friendship! Thanks *Pauliina* and *Tassos* for lots of laughs, and of course for RuisRock! *Amena* and *Peter*, thank you for your relaxed attitude and for refreshing my Finnish. Thank you *Sanna*, for being a true friend all the time here, *Milla, Monika* and *Sari* for the wild evenings out, you really made me feel me like a V.I.P.! *Karoliina, Henri* and *Vappu*, you stayed in Stockholm only for a short while, but we spent so many fun moments together!

My deepest gratitude to my family. My sister and dear friend *Hannele* who has always understood me better than anyone. *Tuomas*, for putting up with the pranks of Flickorna Anthoni. My dear parents, I can never thank you enough. *Äiti*, the best mother of all, taking care of me, being always there when I have needed it. *Isi*, preparing me to be here where I am now, it wasn't always easy, but your invaluable opinions, encouragement, and support have helped me all the way.

Finally and most of all, thank you *Vilho* for your endless support and understanding, for always believing in me, for your love!



Stockholm, November 2007

This thesis project was financially supported by Karolinska Institutet Center for Biotechnology, the Swedish Research Council, the Bank of Sweden Tercentenary Foundation, the Academy of Finland, the Sigrid Jusélius Foundation, the Päivikki and Sakari Sohlberg Foundation, and the Centennial Foundation of Helsingin Sanomat.

8 REFERENCES

- Abecasis GR, Cardon LR, Cookson WO (2000) A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 66:279-292
- Abramoff M, Magelhaes P, Ram S (2004) Image Processing with Image J. *J Biophotonics International* 11:36-42
- Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M (2001) Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet* 69:936-950
- Andrews W, Liapi A, Plachez C, Camurri L, Zhang J, Mori S, Murakami F, Parnavelas JG, Sundaresan V, Richards LJ (2006) Robo1 regulates the development of major axon tracts and interneuron migration in the forebrain. *Development* 133:2243-2252
- Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3:299-309
- Ayala R, Shu T, Tsai LH (2007) Trekking across the brain: the journey of neuronal migration. *Cell* 128:29-43
- Bai J, Ramos RL, Ackman JB, Thomas AM, Lee RV, LoTurco JJ (2003) RNAi reveals doublecortin is required for radial migration in rat neocortex. *Nat Neurosci* 6:1277-1283
- Bakker SC, van der Meulen EM, Buitelaar JK, Sandkuijl LA, Pauls DL, Monsuur AJ, van 't Slot R, Minderaa RB, Gunning WB, Pearson PL, Sinke RJ (2003) A whole-genome scan in 164 Dutch sib pairs with attention-deficit/hyperactivity disorder: suggestive evidence for linkage on chromosomes 7p and 15q. *Am J Hum Genet* 72:1251-1260
- Barnes LL, Wilson RS, Bienias JL, Schneider JA, Evans DA, Bennett DA (2005) Sex differences in the clinical manifestations of Alzheimer disease pathology. *Arch Gen Psychiatry* 62:685-691
- Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263-265
- Bates TC, Luciano M, Castles A, Coltheart M, Wright MJ, Martin NG (2007) Replication of reported linkages for dyslexia and spelling and suggestive evidence for novel regions on chromosomes 4 and 17. *Eur J Hum Genet* 15:194-203
- Beech JR, Beauvois MW (2006) Early experience of sex hormones as a predictor of reading, phonology, and auditory perception. *Brain Lang* 96:49-58
- Bellini G, Bravaccio C, Calamoneri F, Donatella Cocuzza M, Fiorillo P, Gagliano A, Mazzone D, del Giudice EM, Scuccimarra G, Militerni R, Pascotto A (2005) No evidence for association between dyslexia and DYX1C1 functional variants in a group of children and adolescents from Southern Italy. *J Mol Neurosci* 27:311-314
- Beyer C (1999) Estrogen and the developing mammalian brain. *Anat Embryol (Berl)* 199:379-390

- Beyer C, Hutchison JB (1997) Androgens stimulate the morphological maturation of embryonic hypothalamic aromatase-immunoreactive neurons in the mouse. *Brain Res Dev Brain Res* 98:74-81
- Bisgaard ML, Eiberg H, Moller N, Niebuhr E, Mohr J (1987) Dyslexia and chromosome 15 heteromorphism: negative lod score in a Danish material. *Clin Genet* 32:118-119
- Boehm GW, Sherman GF, Rosen GD, Galaburda AM, Denenberg VH (1996) Neocortical ectopias in BXS^B mice: effects upon reference and working memory systems. *Cereb Cortex* 6:696-700
- Botstein D, Risch N (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 33 Suppl:228-237
- Brewer GJ, Cotman CW (1989) Survival and growth of hippocampal neurons in defined medium at low density: advantages of a sandwich culture technique or low oxygen. *Brain Res* 494:65-74
- Brinton RD, Tran J, Proffitt P, Montoya M (1997) 17 beta-Estradiol enhances the outgrowth and survival of neocortical neurons in culture. *Neurochem Res* 22:1339-1351
- Brkanac Z, Chapman NH, Matsushita MM, Chun L, Nielsen K, Cochrane E, Berninger VW, Wijsman EM, Raskind WH (2007) Evaluation of candidate genes for DYX1 and DYX2 in families with dyslexia. *Am J Med Genet B Neuropsychiatr Genet* 144:556-560
- 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
- Bruck M (1992) Persistence of dyslexics' phonological awareness deficits. *Dev Psychol* 28:874-886
- 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 (Pt 10):1901-1917
- Bulun SE, Sebastian S, Takayama K, Suzuki T, Sasano H, Shozu M (2003) The human CYP19 (aromatase P450) gene: update on physiologic roles and genomic organization of promoters. *J Steroid Biochem Mol Biol* 86:219-224
- Cahill L (2006) Why sex matters for neuroscience. *Nat Rev Neurosci* 7:477-484
- Cammarota M, Bevilacqua LR, Ardenghi P, Paratcha G, Levi de Stein M, Izquierdo I, Medina JH (2000) Learning-associated activation of nuclear MAPK, CREB and Elk-1, along with Fos production, in the rat hippocampus after a one-trial avoidance learning: abolition by NMDA receptor blockade. *Brain Res Mol Brain Res* 76:36-46
- Campbell TF, Dollaghan CA, Rockette HE, Paradise JL, Feldman HM, Shriberg LD, Sabo DL, Kurs-Lasky M (2003) Risk factors for speech delay of unknown origin in 3-year-old children. *Child Dev* 74:346-357

- Cardon LR, Bell JI (2001) Association study designs for complex diseases. *Nat Rev Genet* 2:91-99
- Cardon LR, Fulker DW (1994) The power of interval mapping of quantitative trait loci, using selected sib pairs. *Am J Hum Genet* 55:825-833
- 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
- Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 74:106-120
- Castles A, Coltheart M (1993) Varieties of developmental dyslexia. *Cognition* 47:149-180
- Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, et al. (2007) Replicating genotype-phenotype associations. *Nature* 447:655-660
- Chapman NH, Igo RP, Thomson JB, Matsushita M, Brkanac Z, Holzman T, Berninger VW, Wijsman EM, Raskind WH (2004) Linkage analyses of four regions previously implicated in dyslexia: confirmation of a locus on chromosome 15q. *Am J Med Genet B (Neuropsychiatr Genet)* 131:67-75
- Clayton D, Jones H (1999) Transmission/disequilibrium tests for extended marker haplotypes. *Am J Hum Genet* 65:1161-1169
- Colciago A, Celotti F, Pravettoni A, Mornati O, Martini L, Negri-Cesi P (2005) Dimorphic expression of testosterone metabolizing enzymes in the hypothalamic area of developing rats. *Brain Res Dev Brain Res* 155:107-116
- Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA, Pritchard JK (2006) A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet* 38:1251-1260
- Cope N, Harold D, Hill G, Moskvina V, Stevenson J, Holmans P, Owen MJ, O'Donovan MC, Williams J (2005a) Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *Am J Hum Genet* 76:581-591
- Cope NA, Hill G, van den Bree M, Harold D, Moskvina V, Green EK, Owen MJ, Williams J, O'Donovan MC (2005b) No support for association between dyslexia susceptibility 1 candidate 1 and developmental dyslexia. *Mol Psychiatry* 10:237-238
- Corina DP, Richards TL, Serafini S, Richards AL, Steury K, Abbott RD, Echelard DR, Maravilla KR, Berninger VW (2001) fMRI auditory language differences between dyslexic and able reading children. *Neuroreport* 12:1195-1201
- Cornil CA, Ball GF, Balthazard J (2006) Functional significance of the rapid regulation of brain estrogen action: where do the estrogens come from? *Brain Res* 1126:2-26
- De Felipe C, Herrero JF, O'Brien JA, Palmer JA, Doyle CA, Smith AJ, Laird JM, Belmonte C, Cervero F, Hunt SP (1998) Altered nociception, analgesia and aggression in mice lacking the receptor for substance P. *Nature* 392:394-397

- de Kovel CG, Hol FA, Heister JG, Willemen JJ, Sandkuijl LA, Franke B, Padberg GW (2004) Genomewide scan identifies susceptibility locus for dyslexia on Xq27 in an extended Dutch family. *J Med Genet* 41:652-657
- 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. *Hum Genet* 115:128-138
- DeFries JC, Alarcón M (1996) Genetics of specific reading disability. *Ment Retard Dev Disabil Res Rev* 2:39-47
- DeFries JC, Fulker DW, LaBuda MC (1987) Evidence for a genetic aetiology in reading disability of twins. *Nature* 329:537-539
- 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 Res* 562:98-104
- Derocq JM, Segui M, Blazy C, Emonds-Alt X, Le Fur G, Brelire JC, Casellas P (1996) Effect of substance P on cytokine production by human astrocytic cells and blood mononuclear cells: characterization of novel tachykinin receptor antagonists. *FEBS Lett* 399:321-325
- des Portes V, Pinard JM, Billuart P, Vinet MC, Koulakoff A, Carrie A, Gelot A, Dupuis E, Motte J, Berwald-Netter Y, Catala M, Kahn A, Beldjord C, Chelly J (1998) A novel CNS gene required for neuronal migration and involved in X-linked subcortical laminar heterotopia and lissencephaly syndrome. *Cell* 92:51-61
- Deutsch GK, Dougherty RF, Bammer R, Siok WT, Gabrieli JD, Wandell B (2005) Children's reading performance is correlated with white matter structure measured by diffusion tensor imaging. *Cortex* 41:354-363
- Dorus S, Vallender EJ, Evans PD, Anderson JR, Gilbert SL, Mahowald M, Wyckoff GJ, Malcom CM, Lahn BT (2004) Accelerated evolution of nervous system genes in the origin of Homo sapiens. *Cell* 119:1027-1040
- Doupe AJ, Kuhl PK (1999) Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* 22:567-631
- Dudbridge F (2003) Pedigree disequilibrium tests for multilocus haplotypes. *Genet Epidemiol* 25:115-121
- Duret L, Mouchiroud D (2000) Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol Biol Evol* 17:68-74
- Eckert M (2004) Neuroanatomical markers for dyslexia: a review of dyslexia structural imaging studies. *Neuroscientist* 10:362-371
- Eckert MA, Leonard CM, Richards TL, Aylward EH, Thomson J, Berninger VW (2003) Anatomical correlates of dyslexia: frontal and cerebellar findings. *Brain* 126:482-494
- Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S (2002) Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418:869-872

- Fagerheim T, Raeymaekers P, Tonnessen FE, Pedersen M, Tranebjaerg L, Lubs HA (1999) A new gene (DYX3) for dyslexia is located on chromosome 2. *J Med Genet* 36:664-669
- Faraone SV, Sergeant J, Gillberg C, Biederman J (2003) The worldwide prevalence of ADHD: is it an American condition? *World Psychiatry* 2:104-113
- Field LL, Kaplan BJ (1998) Absence of linkage of phonological coding dyslexia to chromosome 6p23-p21.3 in a large family data set. *Am J Hum Genet* 63:1448-1456
- Fisher CR, Graves KH, Parlow AF, Simpson ER (1998) Characterization of mice deficient in aromatase (ArKO) because of targeted disruption of the *cyp19* gene. *Proc Natl Acad Sci U S A* 95:6965-6970
- Fisher SE (2006) Tangled webs: tracing the connections between genes and cognition. *Cognition* 101:270-297
- Fisher SE, DeFries JC (2002) Developmental dyslexia: genetic dissection of a complex cognitive trait. *Nat Rev Neurosci* 3:767-780
- Fisher SE, Francks C (2006) Genes, cognition and dyslexia: learning to read the genome. *Trends Cogn Sci* 10:250-257
- Fisher SE, Francks C, Marlow AJ, MacPhie IL, Newbury DF, Cardon LR, Ishikawa-Brush Y, Richardson AJ, Talcott JB, Gayan 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. *Nat Genet* 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. *Am J Hum Genet* 64:146-156
- Fitch RH, Tallal P, Brown CP, Galaburda AM, Rosen GD (1994) Induced microgyria and auditory temporal processing in rats: a model for language impairment? *Cereb Cortex* 4:260-270
- Forlano PM, Schlinger BA, Bass AH (2006) Brain aromatase: New lessons from non-mammalian model systems. *Front Neuroendocrinol* 27:247-274
- Francks C, Fisher SE, Olson RK, Pennington BF, Smith SD, DeFries JC, Monaco AP (2002a) Fine mapping of the chromosome 2p12-16 dyslexia susceptibility locus: quantitative association analysis and positional candidate genes SEMA4F and OTX1. *Psychiatr Genet* 12:35-41
- Francks C, MacPhie IL, Monaco AP (2002b) The genetic basis of dyslexia. *Lancet Neurol* 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. *Am J Hum Genet* 75:1046-1058

- Froster U, Schulte-Korne G, Hebebrand J, Remschmidt H (1993) Cosegregation of balanced translocation (1;2) with retarded speech development and dyslexia. *Lancet* 342:178-179
- 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 AM, Kemper TL (1979) Cytoarchitectonic abnormalities in developmental dyslexia: a case study. *Ann Neurol* 6:94-100
- Galaburda AM, LoTurco J, Ramus F, Fitch RH, Rosen GD (2006) From genes to behavior in developmental dyslexia. *Nat Neurosci* 9:1213-1217
- Galaburda AM, Menard MT, Rosen GD (1994) Evidence for aberrant auditory anatomy in developmental dyslexia. *Proc Natl Acad Sci U S A* 91:8010-8013
- Galaburda AM, Sherman GF, Rosen GD, Aboitiz F, Geschwind N (1985) Developmental dyslexia: four consecutive patients with cortical anomalies. *Ann Neurol* 18:222-233
- Gayan J, Olson RK (2001) Genetic and environmental influences on orthographic and phonological skills in children with reading disabilities. *Dev Neuropsychol* 20:483-507
- Gayan J, Olson RK (2003) Genetic and environmental influences on individual differences in printed word recognition. *J Exp Child Psychol* 84:97-123
- Gayan J, Smith SD, Cherny SS, Cardon LR, Fulker DW, Brower AM, Olson RK, Pennington BF, DeFries JC (1999) Quantitative-trait locus for specific language and reading deficits on chromosome 6p. *Am J Hum Genet* 64:157-164
- Gayan J, Willcutt EG, Fisher SE, Francks C, Cardon LR, Olson RK, Pennington BF, Smith SD, Monaco AP, DeFries JC, Rhee SH, Hewitt JK, Corley RP, Chhabildas N, Hulslander J, Friedman MC, Budhiraja N, Boada R, Oglie JS, Tunick RA, Chhabildas NA (2005) Bivariate linkage scan for reading disability and attention-deficit/hyperactivity disorder localizes pleiotropic loci. *J Child Psychol Psychiatry* 46:1045-1056
- George VT, Elston RC (1987) Testing the association between polymorphic markers and quantitative traits in pedigrees. *Genet Epidemiol* 4:193-201
- Georgiewa P, Rzanny R, Gaser C, Gerhard UJ, Vieweg U, Freesmeyer D, Mentzel HJ, Kaiser WA, Blanz B (2002) Phonological processing in dyslexic children: a study combining functional imaging and event related potentials. *Neurosci Lett* 318:5-8
- Gilger JW, Borecki IB, DeFries JC, Pennington BF (1994) Commingling and segregation analysis of reading performance in families of normal reading probands. *Behav Genet* 24:345-355
- Govek EE, Newey SE, Van Aelst L (2005) The role of the Rho GTPases in neuronal development. *Genes Dev* 19:1-49
- Gregorius HR (1980) The probability of losing an allele when diploid genotypes are sampled. *Biometrics* 36:643-652

- Grigorenko EL (2001) Developmental dyslexia: an update on genes, brains, and environments. *J Child Psychol Psychiatry* 42:91-125
- Grigorenko EL, Wood FB, Golovyan L, Meyer M, Romano C, Pauls D (2003) Continuing the search for dyslexia genes on 6p. *Am J Med Genet B (Neuropsychiatr Genet)* 118:89-98
- Grigorenko EL, Wood FB, Meyer MS, Hart LA, Speed WC, Shuster A, Pauls DL (1997) Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. *Am J Hum Genet* 60:27-39
- Grigorenko EL, Wood FB, Meyer MS, Pauls DL (2000) Chromosome 6p influences on different dyslexia-related cognitive processes: further confirmation. *American Journal of Human Genetics*:715-723
- Grigorenko EL, Wood FB, Meyer MS, Pauls JE, Hart LA, Pauls DL (2001) Linkage studies suggest a possible locus for developmental dyslexia on chromosome 1p. *Am J Med Genet* 105:120-129
- Grizzle KL, Simms MD (2005) Early language development and language learning disabilities. *Pediatr Rev* 26:274-283
- Habib M (2000) The neurological basis of developmental dyslexia: an overview and working hypothesis. *Brain* 123 Pt 12:2373-2399
- Hannelius U, Lindgren CM, Melen E, Malmberg A, von Döbeln U, Kere J (2005) Phenylketonuria screening registry as a resource for population genetic studies. *J Med Genet* 42:e60
- Hannula-Jouppi K, Kaminen-Ahola N, Taipale M, Eklund R, Nopola-Hemmi J, Kaariainen H, Kere J (2005) The axon guidance receptor gene *ROBO1* is a candidate gene for developmental dyslexia. *PLoS Genet* 1:e50
- Hao J, Rapp PR, Leffler AE, Leffler SR, Janssen WG, Lou W, McKay H, Roberts JA, Wearne SL, Hof PR, Morrison JH (2006) Estrogen alters spine number and morphology in prefrontal cortex of aged female rhesus monkeys. *J Neurosci* 26:2571-2578
- Harlaar N, Spinath FM, Dale PS, Plomin R (2005) Genetic influences on early word recognition abilities and disabilities: a study of 7-year-old twins. *J Child Psychol Psychiatry* 46:373-384
- Harold D, Paracchini S, Scerri T, Dennis M, Cope N, Hill G, Moskvina V, Walter J, Richardson AJ, Owen MJ, Stein JF, Green ED, O'Donovan MC, Williams J, Monaco AP (2006) Further evidence that the *KIAA0319* gene confers susceptibility to developmental dyslexia. *Mol Psychiatry* 11:1085-1091, 1061
- Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 2:3-19
- Hattersley AT, McCarthy MI (2005) What makes a good genetic association study? *Lancet* 366:1315-1323
- Heiervang E, Hugdahl K, Steinmetz H, Inge Smievoll A, Stevenson J, Lund A, Ersland L, Lundervold A (2000) Planum temporale, planum parietale and dichotic listening in dyslexia. *Neuropsychologia* 38:1704-1713

- Herman AE, Galaburda AM, Fitch RH, Carter AR, Rosen GD (1997) Cerebral microgyria, thalamic cell size and auditory temporal processing in male and female rats. *Cereb Cortex* 7:453-464
- Hill RA, McInnes KJ, Gong EC, Jones ME, Simpson ER, Boon WC (2007) Estrogen deficient male mice develop compulsive behavior. *Biol Psychiatry* 61:359-366
- Hill RA, Pompolo S, Jones ME, Simpson ER, Boon WC (2004) Estrogen deficiency leads to apoptosis in dopaminergic neurons in the medial preoptic area and arcuate nucleus of male mice. *Mol Cell Neurosci* 27:466-476
- Hsiung GY, Kaplan BJ, Petryshen TL, Lu S, Field LL, Stein CM, Schick JH, Gerry Taylor H, Shriberg LD, Millard C, Kundtz-Kluge A, Russo K, Minich N, Hansen A, Freebairn LA, Elston RC, Lewis BA, Iyengar SK (2004) A dyslexia susceptibility locus (DYX7) linked to dopamine D4 receptor (DRD4) region on chromosome 11p15.5. *Am J Med Genet* 125B:112-119
- Huang R, Poduslo SE (2006) CYP19 haplotypes increase risk for Alzheimer's disease. *J Med Genet* 43:e42
- Hugdahl K, Synnevag B, Satz P (1990) Immune and autoimmune diseases in dyslexic children. *Neuropsychologia* 28:673-679
- Humphreys P, Kaufmann WE, Galaburda AM (1990) Developmental dyslexia in women: neuropathological findings in three patients. *Ann Neurol* 28:727-738
- 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 BXS^B mice. *Dev Psychobiol* 39:286-300
- Hyde LA, Sherman GF, Hoplight BJ, Denenberg VH (2000) Working memory deficits in BXS^B mice with neocortical ectopias. *Physiol Behav* 70:1-5
- Hynd GW, Semrud-Clikeman M (1989) Dyslexia and brain morphology. *Psychol Bull* 106:447-482
- Igo RP, Jr., Chapman NH, Berninger VW, Matsushita M, Brkanac Z, Rothstein JH, Holzman T, Nielsen K, Raskind WH, Wijsman EM (2006) Genomewide scan for real-word reading subphenotypes of dyslexia: novel chromosome 13 locus and genetic complexity. *Am J Med Genet B Neuropsychiatr Genet* 141:15-27
- Iivonen S, Corder E, Lehtovirta M, Helisalml S, Mannermaa A, Vepsalainen S, Hanninen T, Soininen H, Hiltunen M (2004) Polymorphisms in the CYP19 gene confer increased risk for Alzheimer disease. *Neurology* 62:1170-1176
- International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437:1299-1320
- Kaminen N, Hannula-Jouppi K, Kestila 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. *J Med Genet* 40:340-345
- Kaplan DE, Gayan J, Ahn J, Won TW, 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. *Am J Hum Genet* 70:1287-1298

- Kenney MK, Barac-Cikoja D, Finnegan K, Jeffries N, Ludlow CL (2006) Speech perception and short-term memory deficits in persistent developmental speech disorder. *Brain Lang* 96:178-190
- Kere J (2001) Human population genetics: lessons from Finland. *Annu Rev Genomics Hum Genet* 2:103-128
- Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Paabo S (2005) Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* 309:1850-1854
- Klingberg T, Hedehus M, Temple E, Salz T, Gabrieli JD, Moseley ME, Poldrack RA (2000) Microstructure of temporo-parietal white matter as a basis for reading ability: evidence from diffusion tensor magnetic resonance imaging. *Neuron* 25:493-500
- Knight JC (2005) Regulatory polymorphisms underlying complex disease traits. *J Mol Med* 83:97-109
- Knopik VS, Smith SD, Cardon L, Pennington B, Gayan J, Olson RK, DeFries JC (2002) Differential genetic etiology of reading component processes as a function of IQ. *Behav Genet* 32:181-198
- Korol DL (2004) Role of estrogen in balancing contributions from multiple memory systems. *Neurobiol Learn Mem* 82:309-323
- Kramer MS, Cutler N, Feighner J, Shrivastava R, Carman J, Sramek JJ, Reines SA, et al. (1998) Distinct mechanism for antidepressant activity by blockade of central substance P receptors. *Science* 281:1640-1645
- Kremen WS, Jacobson KC, Xian H, Eisen SA, Waterman B, Toomey R, Neale MC, Tsuang MT, Lyons MJ (2005) Heritability of word recognition in middle-aged men varies as a function of parental education. *Behav Genet* 35:417-433
- Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58:1347-1363
- Lahiri DK, Nurnberger JI, Jr. (1991) A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res* 19:5444
- Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413:519-523
- Lamb JA, Barnby G, Bonora E, Sykes N, Bacchelli E, Blasi F, Maestrini E, Broxholme J, Tzenova J, Weeks D, Bailey AJ, Monaco AP (2005) Analysis of IMGSAC autism susceptibility loci: evidence for sex limited and parent of origin specific effects. *J Med Genet* 42:132-137
- Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 11:241-247
- Lander ES, Schork NJ (1994) Genetic dissection of complex traits. *Science* 265:2037-2048
- Landerl K, Wimmer H, Frith U (1997) The impact of orthographic consistency on dyslexia: a German-English comparison. *Cognition* 63:315-334

- Leinonen S, Muller K, Leppänen P, Aro M, Ahonen T, Lyytinen H (2001) Heterogeneity in adult dyslexic readers: Relating processing skills to the speed and accuracy of oral text reading. *Read Write Interdisc J* 14:265-296
- Leonard CM, Eckert MA, Lombardino LJ, Oakland T, Kranzler J, Mohr CM, King WM, Freeman A (2001) Anatomical risk factors for phonological dyslexia. *Cereb Cortex* 11:148-157
- Lephart ED, Lund TD, Horvath TL (2001) Brain androgen and progesterone metabolizing enzymes: biosynthesis, distribution and function. *Brain Res Brain Res Rev* 37:25-37
- Lewitter FI, DeFries JC, Elston RC (1980) Genetic models of reading disability. *Behav Genet* 10:9-30
- Lin PI, Vance JM, Pericak-Vance MA, Martin ER (2007) No gene is an island: the flip-flop phenomenon. *Am J Hum Genet* 80:531-538
- Linkai R, Kaminen-Ahola N, Tatlisumak T, Elomaa O, Kere J, Lindsberg P (Submitted) Expression of *Dyx1c1* following cerebral ischaemia and heat shock - relation to heat shock proteins 70 and 90
- Livingstone MS, Rosen GD, Drislane FW, Galaburda AM (1991) Physiological and anatomical evidence for a magnocellular defect in developmental dyslexia. *Proc Natl Acad Sci U S A* 88:7943-7947
- Lo HS, Wang Z, Hu Y, Yang HH, Gere S, Buetow KH, Lee MP (2003) Allelic variation in gene expression is common in the human genome. *Genome Res* 13:1855-1862
- 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. *Mol Psychiatry* 9:485-493
- Lovett MW, Steinbach KA, Frijters JC (2000) Remediating the core deficits of developmental reading disability: a double-deficit perspective. *J Learn Disabil* 33:334-358
- Luciano M, Lind PA, Duffy DL, Castles A, Wright MJ, Montgomery GW, Martin NG, Bates TC (2007) A Haplotype Spanning KIAA0319 and TTRAP Is Associated with Normal Variation in Reading and Spelling Ability. *Biol Psychiatry* 62:811-817
- Ludwig KU, Schumacher J, Schulte-Körne G, König IR, Warnke A, Plume E, Anthoni H, Peyrard-Janvid M, Meng H, Ziegler A, Remschmidt H, Kere J, Gruen JR, Müller-Myhsok B, Nöthen MM, Hoffmann P (Submitted) Investigation of the *DCDC2* intron 2 deletion/compound short tandem repeat polymorphism in a large German dyslexia sample
- Lyytinen H, Aro M, Eklund K, Erskine J, Guttorm T, Laakso ML, Leppänen PH, Lyytinen P, Poikkeus AM, Torppa M (2004) The development of children at familial risk for dyslexia: birth to early school age. *Ann Dyslexia* 54:184-220
- Lyytinen H, Leinonen S, Nikula M, Aro M, Leiwo M (1995) In search of the core features of dyslexia: observations concerning dyslexia in the highly orthographically regular Finnish language. In: Berninger V (ed) *The varieties of*

orthographic knowledge II: relationships to phonology, reading, and writing. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 177-204

- Lyytinen P, Eklund K, Lyytinen H (2003) The play and language behavior of mothers with and without dyslexia and its association to their toddlers' language development. *J Learn Disabil* 36:74-86
- MacDonell LE, Skinner FK, Ward PE, Glen AI, Glen AC, Macdonald DJ, Boyle RM, Horrobin DF (2000) Increased levels of cytosolic phospholipase A2 in dyslexics. *Prostaglandins Leukot Essent Fatty Acids* 63:37-39
- Marino C, Citterio A, Giorda R, Facchetti A, Menozzi G, Vanzin L, Lorusso ML, Nobile M, Molteni M (2007) Association of short-term memory with a variant within *DYX1C1* in developmental dyslexia. *Genes Brain Behav* 6:640-646
- Marino C, Giorda R, Luisa Lorusso M, 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. *Eur J Hum Genet* 13:491-499
- Marino C, Giorda R, Vanzin L, Nobile M, Lorusso ML, Baschirotto C, Riva L, Molteni M, Battaglia M (2004) A locus on 15q15-15qter influences dyslexia: further support from a transmission/disequilibrium study in an Italian speaking population. *J Med Genet* 41:42-46
- Marlow AJ, Fisher SE, Francks C, MacPhie IL, Cherny SS, Richardson AJ, Talcott JB, Stein JF, Monaco AP, Cardon LR (2003) Use of multivariate linkage analysis for dissection of a complex cognitive trait. *Am J Hum Genet* 72:561-570
- Martin ER, Monks SA, Warren LL, Kaplan NL (2000) A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am J Hum Genet* 67:146-154
- McArthur GM, Bishop DV (2001) Auditory perceptual processing in people with reading and oral language impairments: current issues and recommendations. *Dyslexia* 7:150-170
- McCrary E, Frith U, Brunswick N, Price C (2000) Abnormal functional activation during a simple word repetition task: A PET study of adult dyslexics. *J Cogn Neurosci* 12:753-762
- Meng H, Hager K, Held M, Page GP, Olson RK, Pennington BF, DeFries JC, Smith SD, Gruen JR (2005a) TDT-association analysis of *EKN1* and dyslexia in a Colorado twin cohort. *Hum Genet* 118:87-90
- 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 (2005b) *DCDC2* is associated with reading disability and modulates neuronal development in the brain. *Proc Natl Acad Sci U S A* 102:17053-17058
- Mervis CB, Berra AM (2007) Language and communicative development in Williams syndrome. *Ment Retard Dev Disabil Res Rev* 13:3-15
- Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215

- Miscimarra L, Stein C, Millard C, Kluge A, Cartier K, Freebairn L, Hansen A, Shriberg L, Taylor HG, Lewis B, Iyengar SK (2007) Further Evidence of Pleiotropy Influencing Speech and Language: Analysis of the DYX8 Region. *Hum Hered* 63:47-58
- 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. *Am J Med Genet B (Neuropsychiatr Genet)* 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. *Hum Mol Genet* 9:843-848
- Nadeau JH (2001) Modifier genes in mice and humans. *Nat Rev Genet* 2:165-174
- Naftolin F, Horvath TL, Jakab RL, Leranath C, Harada N, Balthazart J (1996) Aromatase immunoreactivity in axon terminals of the vertebrate brain. An immunocytochemical study on quail, rat, monkey and human tissues. *Neuroendocrinology* 63:149-155
- Nicolson RI, Fawcett AJ, Dean P (2001) Developmental dyslexia: the cerebellar deficit hypothesis. *Trends Neurosci* 24:508-511
- 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. *J Med Genet* 38:658-664
- Nopola-Hemmi J, Myllyluoma B, Voutilainen A, Leinonen S, Kere J, Ahonen T (2002) Familial dyslexia: neurocognitive and genetic correlation in a large Finnish family. *Dev Med Child Neurol* 44:580-586
- Nopola-Hemmi J, Taipale M, Haltia T, Lehesjoki AE, Voutilainen A, Kere J (2000) Two translocations of chromosome 15q associated with dyslexia. *J Med Genet* 37:771-775
- 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. *Eur Child Adolesc Psychiatry* 8 Suppl 3:56-59
- O'Brien TW (2002) Evolution of a protein-rich mitochondrial ribosome: implications for human genetic disease. *Gene* 286:73-79
- O'Brien TW, O'Brien BJ, Norman RA (2005) Nuclear MRP genes and mitochondrial disease. *Gene* 354:147-151
- O'Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63:259-266
- Palmer LJ, Cardon LR (2005) Shaking the tree: mapping complex disease genes with linkage disequilibrium. *Lancet* 366:1223-1234
- Paracchini S, Thomas A, Castro S, Lai C, Paramasivam M, Wang Y, Keating BJ, Taylor JM, Hacking DF, Scerri T, Francks C, Richardson AJ, Wade-Martins R, Stein JF, Knight JC, Copp AJ, Loturco J, Monaco AP (2006) The chromosome

6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. *Hum Mol Genet* 15:1659-1666

- Pastinen T, Sladek R, Gurd S, Sammak A, Ge B, Lepage P, Lavergne K, Villeneuve A, Gaudin T, Brandstrom H, Beck A, Verner A, Kingsley J, Harmsen E, Labuda D, Morgan K, Vohl MC, Naumova AK, Sinnott D, Hudson TJ (2004) A survey of genetic and epigenetic variation affecting human gene expression. *Physiol Genomics* 16:184-193
- Paulesu E, Demonet JF, 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 RS, Frith CD (1996) Is developmental dyslexia a disconnection syndrome? Evidence from PET scanning. *Brain* 119 (Pt 1):143-157
- Peiffer AM, Friedman JT, Rosen GD, Fitch RH (2004) Impaired gap detection in juvenile microgyric rats. *Brain Res Dev Brain Res* 152:93-98
- Peiffer AM, Rosen GD, Fitch RH (2002a) Rapid auditory processing and MGN morphology in microgyric rats reared in varied acoustic environments. *Brain Res Dev Brain Res* 138:187-193
- Peiffer AM, Rosen GD, Fitch RH (2002b) Sex differences in rapid auditory processing deficits in ectopic BXSB/MpJ mice. *Neuroreport* 13:2277-2280
- Peltonen L, Jalanko A, Varilo T (1999) Molecular genetics of the Finnish disease heritage. *Hum Mol Genet* 8:1913-1923
- Pennington BF (1997) Using genetics to dissect cognition. *Am J Hum Genet* 60:13-16
- Pennington BF (2006) From single to multiple deficit models of developmental disorders. *Cognition* 101:385-413
- Pennington BF, Gilger JW, Pauls D, Smith SA, Smith SD, DeFries JC (1991) Evidence for major gene transmission of developmental dyslexia. *JAMA* 266:1527-1534
- Pennington BF, Lefly DL (2001) Early reading development in children at family risk for dyslexia. *Child Dev* 72:816-833
- Pennington BF, Smith SD, Kimberling WJ, Green PA, Haith MM (1987) Left-handedness and immune disorders in familial dyslexics. *Arch Neurol* 44:634-639
- Peterson RL, McGrath LM, Smith SD, Pennington BF (2007) Neuropsychology and genetics of speech, language, and literacy disorders. *Pediatr Clin North Am* 54:543-561
- Petryshen TL, Kaplan BJ, Fu Liu M, de French NS, Tobias R, Hughes ML, Field LL (2001) Evidence for a susceptibility locus on chromosome 6q influencing phonological coding dyslexia. *Am J Med Genet* 105:507-517
- Petryshen TL, Kaplan BJ, Hughes ML, Tzenova J, Field LL (2002) Supportive evidence for the DYX3 dyslexia susceptibility gene in Canadian families. *J Med Genet* 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. *Am J Hum Genet* 66:708-714
- Rabin M, Wen XL, Hepburn M, Lubs HA, Feldman E, Duara R (1993) Suggestive linkage of developmental dyslexia to chromosome 1p34-p36. *Lancet* 342:178
- Rae C, Harasty JA, Dzendrowskyj TE, Talcott JB, Simpson JM, Blamire AM, Dixon RM, Lee MA, Thompson CH, Styles P, Richardson AJ, Stein JF (2002) Cerebellar morphology in developmental dyslexia. *Neuropsychologia* 40:1285-1292
- Raitano NA, Pennington BF, Tunick RA, Boada R, Shriberg LD (2004) Pre-literacy skills of subgroups of children with speech sound disorders. *J Child Psychol Psychiatry* 45:821-835
- Ramachandran B, Schlinger BA, Arnold AP, Campagnoni AT (1999) Zebra finch aromatase gene expression is regulated in the brain through an alternate promoter. *Gene* 240:209-216
- Ramus F (2003) Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? *Curr Opin Neurobiol* 13:212-218
- Ramus F (2004) Neurobiology of dyslexia: a reinterpretation of the data. *Trends Neurosci* 27:720-726
- Ramus F (2006) Genes, brain, and cognition: a roadmap for the cognitive scientist. *Cognition* 101:247-269
- 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. *Mol Psychiatry* 10:699-711
- Reich DE, Lander ES (2001) On the allelic spectrum of human disease. *Trends Genet* 17:502-510
- Reiner O, Carrozzo R, Shen Y, Wehnert M, Faustinella F, Dobyns WB, Caskey CT, Ledbetter DH (1993) Isolation of a Miller-Dieker lissencephaly gene containing G protein beta-subunit-like repeats. *Nature* 364:717-721
- Richardson SO (1992) Historical perspectives on dyslexia. *J Learn Disabil* 25:40-47; discussion 48-52
- Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS (1999) Replication of linkage studies of complex traits: an examination of variation in location estimates. *Am J Hum Genet* 65:876-884
- Robichon F, Bouchard P, Demonet J, Habib M (2000a) Developmental dyslexia: re-evaluation of the corpus callosum in male adults. *Eur Neurol* 43:233-237
- Robichon F, Habib M (1998) Abnormal callosal morphology in male adult dyslexics: relationships to handedness and phonological abilities. *Brain Lang* 62:127-146
- Robichon F, Levrier O, Farnarier P, Habib M (2000b) Developmental dyslexia: atypical cortical asymmetries and functional significance. *Eur J Neurol* 7:35-46

- Rosen GD, Bai J, Wang Y, Fiondella CG, Threlkeld SW, Loturco JJ, Galaburda AM (2007) Disruption of Neuronal Migration by RNAi of *Dyx1c1* Results in Neocortical and Hippocampal Malformations. *Cereb Cortex*
- Rosen GD, Herman AE, Galaburda AM (1999) Sex differences in the effects of early neocortical injury on neuronal size distribution of the medial geniculate nucleus in the rat are mediated by perinatal gonadal steroids. *Cereb Cortex* 9:27-34
- Rosen GD, Waters NS, Galaburda AM, Denenberg VH (1995) Behavioral consequences of neonatal injury of the neocortex. *Brain Res* 681:177-189
- Roy AL (2001) Biochemistry and biology of the inducible multifunctional transcription factor TFII-I. *Gene* 274:1-13
- Ruff S, Cardebat D, Marie N, Demonet JF (2002) Enhanced response of the left frontal cortex to slowed down speech in dyslexia: an fMRI study. *Neuroreport* 13:1285-1289
- Rumsey JM (1992) The biology of developmental dyslexia. *JAMA* 268:912-915
- Rutter M, Caspi A, Fergusson D, Horwood LJ, Goodman R, Maughan B, Moffitt TE, Meltzer H, Carroll J (2004) Sex differences in developmental reading disability: new findings from 4 epidemiological studies. *JAMA* 291:2007-2012
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928-933
- Salmelin R, Service E, Kiesila P, Uutela K, Salonen O (1996) Impaired visual word processing in dyslexia revealed with magnetoencephalography. *Ann Neurol* 40:157-162
- 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. *J Med Genet* 41:853-857
- Schlaggar BL, McCandliss BD (2007) Development of neural systems for reading. *Annu Rev Neurosci* 30:475-503
- Schulte-Körne G, Deimel W, Remschmidt H (2001) [Diagnosis of reading and spelling disorder]. *Z Kinder Jugendpsychiatr Psychother* 29:113-116
- 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. *Am J Hum Genet* 63:279-282
- Schulte-Körne G, Ziegler A, Deimel W, Schumacher J, Plume E, Bachmann C, Kleinsang A, Propping P, Nöthen MM, Warnke A, Remschmidt H, König IR (2007) Interrelationship and familiarity of dyslexia related quantitative measures. *Ann Hum Genet* 71:160-175
- Schulte-Körne G, Ziegler A, Deimel W, Schumacher J, Plume E, Bachmann C, Kleinsang A, Propping P, Nöthen MM, Warnke A, Remschmidt H, König IR, Müller K, Gutenbrunner C (1996) Familial aggregation of spelling disability. *J Child Psychol Psychiatry* 37:817-822

- Schumacher J, Konig IR, Plume E, Propping P, Warnke A, Manthey M, Duell M, Kleinsang A, Reipsilber D, Preis M, Remschmidt H, Ziegler A, Nothen MM, Schulte-Korne G (2006) Linkage analyses of chromosomal region 18p11-q12 in dyslexia. *J Neural Transm* 113:417-423
- Shaywitz BA, Shaywitz SE, Blachman BA, Pugh KR, Fulbright RK, Skudlarski P, Mencl WE, Constable RT, Holahan JM, Marchione KE, Fletcher JM, Lyon GR, Gore JC (2004) Development of left occipitotemporal systems for skilled reading in children after a phonologically- based intervention. *Biol Psychiatry* 55:926-933
- 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. *Biol Psychiatry* 52:101-110
- 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 ability. *N Engl J Med* 326:145-150
- Shaywitz SE, Fletcher JM, Holahan JM, Shneider AE, Marchione KE, Stuebing KK, Francis DJ, Pugh KR, Shaywitz BA (1999) Persistence of dyslexia: the Connecticut Longitudinal Study at adolescence. *Pediatrics* 104:1351-1359
- Shaywitz SE, Shaywitz BA (2005) Dyslexia (specific reading disability). *Biol Psychiatry* 57:1301-1309
- Shaywitz SE, Shaywitz BA, Fletcher JM, Escobar MD (1990) Prevalence of reading disability in boys and girls. Results of the Connecticut Longitudinal Study. *JAMA* 264:998-1002
- Shaywitz SE, Shaywitz BA, Fulbright RK, Skudlarski P, Mencl WE, Constable RT, Pugh KR, Holahan JM, Marchione KE, Fletcher JM, Lyon GR, Gore JC (2003) Neural systems for compensation and persistence: young adult outcome of childhood reading disability. *Biol Psychiatry* 54:25-33
- Shaywitz SE, Shaywitz BA, Pugh KR, Fulbright RK, Constable RT, Mencl WE, Shankweiler DP, Liberman AM, Skudlarski P, Fletcher JM, Katz L, Marchione KE, Lacadie C, Gatenby C, Gore JC (1998) Functional disruption in the organization of the brain for reading in dyslexia. *Proc Natl Acad Sci U S A* 95:2636-2641
- Shriberg LD, Tomblin JB, McSweeney JL (1999) Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *J Speech Lang Hear Res* 42:1461-1481
- Sidak, Zbynek (1967) Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *J Amer Statistical Assoc* 62:626-633
- Siok WT, Perfetti CA, Jin Z, Tan LH (2004) Biological abnormality of impaired reading is constrained by culture. *Nature* 431:71-76
- Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 38:209-213
- Smith SD, Kimberling WJ, Pennington BF (1991) Screening for multiple genes influencing dyslexia. *Read Writ Interdisc J* 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, Boada R, Shriberg LD (2005) Linkage of speech sound disorder to reading disability loci. *J Child Psychol Psychiatry* 46:1057-1066
- Snowling MJ (2000) Dyslexia. Blackwell, Oxford, UK
- Snowling MJ (2001) From language to reading and dyslexia. *Dyslexia* 7:37-46
- Snowling MJ, Gallagher A, Frith U (2003) Family risk of dyslexia is continuous: individual differences in the precursors of reading skill. *Child Dev* 74:358-373
- Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506-516
- Stein CM, Millard C, Kluge A, Miscimarra LE, Cartier KC, Freebairn LA, Hansen AJ, Shriberg LD, Taylor HG, Lewis BA, Iyengar SK (2006) Speech sound disorder influenced by a locus in 15q14 region. *Behav Genet* 36:858-868
- Stein CM, Schick JH, Gerry Taylor H, Shriberg LD, Millard C, Kundtz-Kluge A, Russo K, Minich N, Hansen A, Freebairn LA, Elston RC, Lewis BA, Iyengar SK (2004) Pleiotropic effects of a chromosome 3 locus on speech-sound disorder and reading. *Am J Hum Genet* 74:283-297
- Stein J (2001) The magnocellular theory of developmental dyslexia. *Dyslexia* 7:12-36
- Stein J, Walsh V (1997) To see but not to read; the magnocellular theory of dyslexia. *Trends Neurosci* 20:147-152
- Stevenson J, Graham P, Fredman G, McLoughlin V (1987) A twin study of genetic influences on reading and spelling ability and disability. *J Child Psychol Psychiatry* 28:229-247
- Stone JL, Merriman B, Cantor RM, Yonan AL, Gilliam TC, Geschwind DH, Nelson SF (2004) Evidence for sex-specific risk alleles in autism spectrum disorder. *Am J Hum Genet* 75:1117-1123
- Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, Vincent JB, et al. (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 39:319-328
- 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. *Proc Natl Acad Sci U S A* 100:11553-11558
- Tallal P (1980) Auditory temporal perception, phonics, and reading disabilities in children. *Brain Lang* 9:182-198
- Temple CM (1997) Reading disorders. In: *Developmental Clinical Neuropsychology*. Psychology press, East Sussex, UK, pp 162-223

- Teramitsu I, White SA (2006) FoxP2 regulation during undirected singing in adult songbirds. *J Neurosci* 26:7390-7394
- Thapar A, O'Donovan M, Owen MJ (2005) The genetics of attention deficit hyperactivity disorder. *Hum Mol Genet* 14 Spec No. 2:R275-282
- Thomas DC, Clayton DG (2004) Betting odds and genetic associations. *J Natl Cancer Inst* 96:421-423
- Threlkeld SW, McClure MM, Bai J, Wang Y, Loturco JJ, Rosen GD, Fitch RH (2007) Developmental disruptions and behavioral impairments in rats following in utero RNAi of *Dyx1c1*. *Brain Res Bull* 71:508-514
- Toivonen HT, Onkamo P, Vasko K, Ollikainen V, Sevon P, Mannila H, Herr M, Kere J (2000) Data mining applied to linkage disequilibrium mapping. *Am J Hum Genet* 67:133-145
- 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, Van Den Bree M, 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. *Mol Psychiatry* 8:176-185
- Turkeltaub PE, Gareau L, Flowers DL, Zeffiro TA, Eden GF (2003) Development of neural mechanisms for reading. *Nat Neurosci* 6:767-773
- 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. *Am J Med Genet B Neuropsychiatr Genet* 127:117-124
- Vellutino FR (1979) The validity of perceptual deficit explanations of reading disability: a reply to Fletcher and Satz. *J Learn Disabil* 12:160-167
- Vellutino FR, Fletcher JM, Snowling MJ, Scanlon DM (2004) Specific reading disability (dyslexia): what have we learned in the past four decades? *J Child Psychol Psychiatry* 45:2-40
- von Plessen K, Lundervold A, Duta N, Heiervang E, Klauschen F, Smievoll AI, Ersland L, Hugdahl K (2002) Less developed corpus callosum in dyslexic subjects--a structural MRI study. *Neuropsychologia* 40:1035-1044
- Wadsworth SJ, Olson RK, Pennington BF, DeFries JC (2000) Differential genetic etiology of reading disability as a function of IQ. *J Learn Disabil* 33:192-199
- Wahlsten D, Colbourne F, Pleus R (2003) A robust, efficient and flexible method for staining myelinated axons in blocks of brain tissue. *J Neurosci Methods* 123:207-214
- Walker AH, Najarian D, White DL, Jaffe JF, Kanetsky PA, Rebbeck TR (1999) Collection of genomic DNA by buccal swabs for polymerase chain reaction-based biomarker assays. *Environ Health Perspect* 107:517-520
- Wall JD, Pritchard JK (2003) Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet* 4:587-597

- Wang Y, Paramasivam M, Thomas A, Bai J, Kaminen-Ahola N, Kere J, Voskuil J, Rosen GD, Galaburda AM, Loturco JJ (2006) DYX1C1 functions in neuronal migration in developing neocortex. *Neuroscience* 143:515-522
- Weiss LA, Pan L, Abney M, Ober C (2006) The sex-specific genetic architecture of quantitative traits in humans. *Nat Genet* 38:218-222
- 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. *Mol Psychiatry* 9:1111-1121
- Wijsman EM, Peterson D, Leutenegger AL, Thomson JB, Goddard KA, Hsu L, Berninger VW, Raskind WH (2000) Segregation analysis of phenotypic components of learning disabilities. I. Nonword memory and digit span. *Am J Hum Genet* 67:631-646
- Wilens TE, Biederman J, Spencer TJ (2002) Attention deficit/hyperactivity disorder across the lifespan. *Annu Rev Med* 53:113-131
- Willcutt EG, Pennington BF (2000) Comorbidity of reading disability and attention-deficit/hyperactivity disorder: differences by gender and subtype. *J Learn Disabil* 33:179-191
- Willcutt EG, Pennington BF, DeFries JC (2000) Twin study of the etiology of comorbidity between reading disability and attention-deficit/hyperactivity disorder. *Am J Med Genet* 96:293-301
- Willcutt EG, Pennington BF, Smith SD, Cardon LR, Gayan J, Knopik VS, Olson RK, DeFries JC (2002) Quantitative trait locus for reading disability on chromosome 6p is pleiotropic for attention-deficit/hyperactivity disorder. *Am J Med Genet* 114:260-268
- Wong K, Ren XR, Huang YZ, Xie Y, Liu G, Saito H, Tang H, Wen L, Brady-Kalnay SM, Mei L, Wu JY, Xiong WC, Rao Y (2001) Signal transduction in neuronal migration: roles of GTPase activating proteins and the small GTPase Cdc42 in the Slit-Robo pathway. *Cell* 107:209-221
- Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555-556
- Ylisaukko-Oja T, Peyrard-Janvid M, Lindgren CM, Rehnstrom K, Vanhala R, Peltonen L, Jarvela I, Kere J (2005) Family-based association study of DYX1C1 variants in autism. *Eur J Hum Genet* 13:127-130
- Ziegler A, König IR, Deimel W, Plume E, Nöthen MM, Propping P, Kleensang A, Müller-Myhsok B, Warnke A, Remschmidt H, Schulte-Körne G (2005) Developmental dyslexia--recurrence risk estimates from a german bi-center study using the single proband sib pair design. *Hum Hered* 59:136-143
- Ziegler JC, Perry C, Ma-Wyatt A, Ladner D, Schulte-Körne G (2003) Developmental dyslexia in different languages: language-specific or universal? *J Exp Child Psychol* 86:169-193
- Zoccolotti P, De Luca M, Di Pace E, Judica A, Orlandi M, Spinelli D (1999) Markers of developmental surface dyslexia in a language (Italian) with high grapheme-phoneme correspondence. *Appl Psycholinguist* 20:191-216

Zondervan KT, Cardon LR (2004) The complex interplay among factors that influence allelic association. *Nat Rev Genet* 5:89-100

Zwick ME, Cutler DJ, Chakravarti A (2000) Patterns of genetic variation in Mendelian and complex traits. *Annu Rev Genomics Hum Genet* 1:387-407