Molecular genetic analyses in developmental dyslexia & related endophenotypes

Dissertation

zur Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität zu Bonn

vorgelegt von

Kerstin Urte Ludwig

aus Dresden

Bonn 2009

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

Die vorliegende Arbeit wurde am Institut für Humangenetik der Rheinischen Friedrich-Wilhelms-Universität zu Bonn angefertigt.

1. Gutachter: Prof. Dr. Markus M. Nöthen

2. Gutachter: Prof. Dr. Michael Hoch

Tag der Promotion: 16. November 2010

Erscheinungsjahr: 2010

Im Gedenken an OStR Heinz Lorenz

TABLE OF CONTENTS

TABLE OF CONTENTS I

ABBREVIATIONS IV

1. INTRODUCTION 1

2. BASIC PRINCIPLES 3

2.1 Developmental dyslexia 3

2.1.1 Clinical classification 3

2.1.2 Neurocognitive theories 4

2.1.3 Neurobiological studies 4

2.1.4 Dyslexia and related endophenotypes 6

2.1.5 Therapy and remediation 8

2.2 Identification of causal genes in human disorders 8

2.2.1 Genetic variability 9

2.2.2 Linkage and association analysis 11

2.2.3 Genotyping technologies 13

Genotyping based on BeadArray Technology 13

Genotyping based on MassExtend Reaction 15

2.3 Genetics of dyslexia 16

2.3.1 Formal genetics 17

2.3.2 Dyslexia susceptibility loci 17

2.3.3 Dyslexia candidate genes 19

2.4 The NeuroDys Consortium 22

2.5 Scope of the doctoral thesis 23

3. MATERIAL AND METHODS 24

3.1 Devices 24

3.2 Chemicals and reagents 25

3.3 Solutions 26

3.4 Commercial systems 27

3.5 Bioinformatic tools 28

3.5.1 Software 28

3.5.2 Databases 28

3.6 Study probands 29

3.6.1 German dyslexia (DYS-) sample 29

Diagnosis of dyslexia and inclusion criteria 29

Clinical assessment of related endophenotypes 30

3.6.2 Probands of the NeuroDys sample 33

3.6.3 Control individuals 34

3.7 Protocols 35

3.7.1 Preparation of nucleic acids 35

Isolation methods 35

Determination of concentration and quality 36

Generation of DNA pools 37

3.7.2 Processing of nucleic acids 37

Polymerase chain reaction 38

Agarose gel electrophoresis 39

3.7.3 Genotyping of DNA samples 39

Genotyping based on BeadArray Technology (Illumina®) 39

Genotyping based on MassExtend Reaction (Sequenom®) 43

3.7.4 Sequencing of DNA fragments 46

3.7.5 Expression analysis 48

Reverse transcription from mRNA to cDNA 48

Qualitative expression analysis in cDNA panels 49

Quantitative Real-Time PCR analysis 49

3.8 Statistical analysis 51

3.8.1 General concepts 51

Quality control 51

Correction for multiple testing 52

3.8.2 Association studies of qualitative traits 53

Case-control studies 53

Family-based association tests 54

3.8.3 Analysis of quantitative endophenotypes 55

4. RESULTS 56

4.1 Candidate gene approach 56

4.1.1 Investigation of genes within the DYX2 locus 56

An intronic deletion in *DCDC2* as causal variant 56

Genetic interaction between *KIAA0319* and *DCDC2* 57

4.1.2 Investigation of *GRIN2B* and short-term memory 58

4.1.3 Investigation of *LRRTM1* and human handedness 61

4.2 Genome-wide approach 62

4.2.1 Association analysis of dyslexia as qualitative trait 63

4.2.2 Genome-wide analysis of dyslexia-related endophenotypes 67

Analysis of event-related potentials 67

Analysis of arithmetical skills 70

5. DISCUSSION 75

5.1 Genetic factors for dyslexia as qualitative trait 75

5.2 Quantitative measures of dyslexia-related endophenotypes 83

6. SUMMARY 96

7. OUTLOOK 98

8. REFERENCES 100

9. LIST OF PUBLICATIONS 119

10. ATTACHMENTS VIII

ABBREVIATIONS

A	Adenine
aa	Amino acid
ABI	Applied Biosystems
ADHD	Attention-deficiency / hyperactivity disorder
AFE	Allele frequency estimate
ANOVA	Analysis of variance
bp	Base pairs
C	Cytosine
°C	Degree Celcius
cDNA	Copy deoxyribonucleic acid
Chr.	Chromosome
CI	Confidence interval
cm	Centimeter
CMH	Cochran-Mantel-Haenszel test
CNV	Copy number variation
Co.	Corporation
conc.	Concentration / concentrated
corr.	Corrected
CR	Call rate
ct	Cycle threshold
Cyc	Cyclophilin
Da	Dalton
DCDC2	Doublecortin domain containing 2
ddNTP	Dideoxyribonucleic triphosphate
DEPC	Diethyl pyrocarbonate
dest.	Distilled
DNA	Deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
dNTP	Deoxyribonucleic triphosphate
DSM	Diagnostic and Statistical Manual of Mental Disorders
DYS	Developmental dyslexia
DYX	Dyslexia susceptibility locus
DYX1C1	Dyslexia susceptibility 1 candidate 1
EBV	Eppstein-Barr virus
EC	Exact calculation

EEG	Electroencephalography
e.g.	For example
ERP	Event-related potential
EST	Expressed sequence tag
et al.	Et alii
EtBr	Ethidium bromide
EtOH	Ethanol
EU	European Union
F	Forward
Fig.	Figure
fMRI	Functional magnetic resonance imaging
FRET	Fluorescence resonance energy transfer
G	Guanine
g	Gram
g	Gravitational force
GRIN2B	N-methyl-D-aspartate subunit 2B
GWAS	Genome-wide association study
h	Hour
HGNC	HUGO Gene Nomenclature Committee
HNR	Heinz-Nixdorf recall
HUGO	Human Genome Organization
HWE	Hardy-Weinberg-equilibrium
Hz	Hertz (international unit for frequency)
ICD	International classification of diseases
i.e.	That is
ID	Sample identification number
Inc.	Incorporation
IQ	Intelligence quotient
IPS	Intraparietal sulcus
kb	Kilobase pairs
LD	Linkage disequilibrium
LOD	Logarithmic odds ratio
LRRTM1	Leucine-rich repeat transmembrane neuronal 1 gene
M	Molar
MAF	Minor allele frequency

MALDI-ToF-MS	Matrix assisted laser desorption / ionization time-of-flight mass spectrometry
MARS	Munich Antidepressant Response Signature
mat.	Maternal
max.	Maximum
Mb	Megabase pairs
MDS	Multi-dimensional-scaling
mg	Milligram
min	Minute
min.	Minimum
ml	Milliliter
mM	Millimolar
mm	Millimeter
MMN	Mismatch negativity
mRNA	Messenger ribonucleic acid
ms	Millisecond
MTC	Multiple tissue cDNA panel
MYO18B	Myosin 18B
µg	Microgram
µl	Microliter
µm	Micrometer
n	Number
n.a.	Not available
NC	Number comparison
NCBI	National Center of Biotechnology Information
ng	Nanogram
NGFN	Nationales Genomforschungsnetz (Germany)
nl	Nanoliter
nm	Nanometer
NMDA	N-methyl-D-aspartate
nom.	Nominal
NT	Non-transmission
OD	Optical density
OMIM	Online Mendelian Inheritance in Men
o.n.	Over night
OR	Odds ratio
pat.	Paternal
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction

pmol	Picomol
QTDT	Quantitative transmission disequilibrium test
R	Reverse
RNA	Ribonucleic acid
ROBO1	Roundabout 1
rpm	Rounds per minute
RR	Relative risk
RT	Room temperature
RT-PCR	Reverse transcription polymerase chain reaction
SAP	Shrimp alcaline phosphatase
SBE	Single-base extension
sd	Standard deviation
SD	Degree of severity
sec	Second
SLC2A3	Solute carrier family 2 member 3
SNP	Single nucleotide polymorphism
STR	Short tandem repeat
T	Thymidine
T	Transmission
Tab.	Table
TBE	Tris borate EDTA buffer
TEDS	Twin Early Development Study
TDT	Transmission disequilibrium test
tSNP	Tagging SNP
U	Unit
UEP	Unextended primer
UK	United Kingdom
US	United States (of America)
UTR	Untranslated region
vs.	Versus

1. INTRODUCTION

Each human being is unique. The inter-individual differences can be attributed to genetic variants which, besides environmental factors, play an important role in the development of human traits. In this context, the term 'genetic variability' refers to differences in the genetic information which is encoded in the deoxyribonucleic acid (DNA) of each individual. Its entity, also denoted as 'genome', comprises a sequential order of molecular bases. Substitutions at single-base level, so-called single nucleotide polymorphisms (SNPs), account for about 80% of the genetic variants in humans (Wang *et al.* 1998; HapMap Consortium 2003; 2005) and occur in about one out of 300 bases (Li & Sadler 1991; Wang *et al.* 1998; Cargill *et al.* 1999; HapMap Consortium 2003; 2005). Structural variants such as insertions or deletions, duplications, translocations or complex rearrangements provide with at a lower frequency. However, as they usually comprise larger numbers of bases, often in the range from one kilobase (kb) to several megabases (Mb), they are also of relevance (Feuk *et al.* 2006; Redon *et al.* 2006).

Most of the genetic variability does not have grave consequences at the phenotypic level, but some of the changes in DNA sequence confer risk for human diseases. These variations are of special interest in the field of human molecular genetics. In monogenic disorders, single mutations directly cause a pathogenic phenotype. These diseases are inherited according to Mendelian laws, however, their prevalence in the general population is very low (less than 1 / 1,000). In contrast, multifactorial heritable or so-called 'genetically complex' disorders appear at relatively high frequencies, in the range of several percent. In these common diseases, several mutations in the genome contribute to a certain disease phenotype. Hereby, the specific risk provided by each of these variants is variable.

Among this group of common diseases is dyslexia, a specific neurodevelopmental disorder (Shaywitz 1998; Shaywitz & Shaywitz 2005). Dyslexia is characterized by severe impairments in learning to read and to spell which are unexpected with regards to general intelligence, level of education and visual or auditory skills (Schulte-Körne *et al.* 2001b). Affected individuals often suffer from secondary symptoms such as attention-deficiency / hyperactivity disorder or depression, and show high rates of scholar dropouts and unemployment (Shaywitz 1998). Given its prevalence rate of 5 – 12% in the general population (Shaywitz *et al.* 1990; Katusic *et al.* 2001), dyslexia belongs to the most common of the neurodevelopmental disorders, thus also representing a social burden for the economic systems in developed countries.

Although a familial clustering of dyslexia has been observed already a hundred years ago (Hinshelwood 1907; Stephenson 1907), the genetic studies on dyslexia have so far identified a

limited number of candidate genes which only explain a small fraction of the dyslexia cases in the population (Williams & O'Donovan 2006; Schumacher *et al.* 2007). Hence, one main goal of research is to identify new genes and to further characterize genes that might contribute to dyslexia susceptibility. Recent advances in high-throughput technologies, international joint efforts and the inputs from functional studies increase the chance of identifying true candidate genes.

The cognitively complex nature of the dyslexia phenotype allows the genetic analysis of dyslexia as categorically defined disorder, but also offers the possibility to correlate genetic variants with particular quantitative measures that enable reading and spelling skills in humans. A more precise knowledge of the genetic causes of dyslexia will increase our understanding of the biological mechanisms that underlie reading and spelling. It will also shed a light on more general aspects of the evolutionary events that contributed to the development of brain architecture, human cognition and language. The correlation of specific genetic variants with particular cognitive processes will help to precisely determine the reason for the particular impairment in each dyslexic child and, thus, increases the chance for successful individual remediation and therapy.

2. BASIC PRINCIPLES

2.1 Developmental dyslexia

Developmental dyslexia, also referred to as 'reading and spelling disorder' or 'dyslexia', is one of the most common neurodevelopmental disorders and belongs to the family of learning disabilities. It is a severe and specific impairment in the acquisition of reading and spelling skills, which is unexpected in relation to other cognitive abilities (Schulte-Körne *et al.* 2001b). Dyslexia affects about 5 - 12% of school-aged children (Shaywitz *et al.* 1990; Katusic *et al.* 2001) and occurs in all languages, however, the specific prevalence rate might vary with language-specific characteristics and the application of different diagnostic criteria (Paulesu *et al.* 2001; Ziegler *et al.* 2003).

2.1.1 Clinical classification

Dyslexia (OMIM 600202) has been included in the *International Classification of Diseases* (ICD-10; (WHO 1993)) and the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV; (APA 1994)). The ICD-10 describes dyslexia as "a disorder manifested by difficulties in learning to read despite conventional instruction, adequate intelligence and sociocultural opportunity". To diagnose a child with dyslexia, according to ICD-10 criteria, the child's performance in reading has to be below the level that would be expected based on the child's age, intelligence quotient (IQ) and schooling. This discrepancy score can alternatively also base on the children's spelling ability, which is used in most German studies (Schulte-Körne & Remschmidt 2003). Importantly, dyslexia has to be distinguished from general learning impairments that are not caused by a developmental deficit such as, for instance (e.g.), inadequate schooling or neurological disorders.

Longitudinal studies have shown that dyslexia involves an extremely stable developmental disorder that, in contrast to general popular opinion, does not disappear with adolescence (Shaywitz *et al.* 1999). Compared to what would be expected according to their level of intelligence, affected individuals lifelong suffer from psychosocial consequences such as a lower educational level and higher rates of unemployment (Strehlow *et al.* 1992; Maughan 1995; Bruck 1998). In childhood, about 20% of dyslexic children present with symptoms of attention-deficiency / hyperactivity disorder (ADHD; (August & Garfinkel 1990; Purvis & Tannock 1997; Shaywitz 1998; Willcutt *et al.* 2000; Kaplan *et al.* 2001)). During their later life, affected individuals often develop depressive disorders and disorders of social behaviors (Frauenheim & Heckerl 1983; Naylor *et al.* 1990; Schulte-Körne *et al.* 2001b).

2.1.2 Neurocognitive theories

A number of theories have been suggested that try to explain the impairments in reading and writing in dyslexic individuals. The most prominent theory, which is supported by most of the investigators in the field, is the phonological theory (Liberman *et al.* 1989; Ramus *et al.* 2003). It suggests the presence of a general deficit in the processing of phonological aspects of language. In order to learn to read and to spell, children have to develop a phonological awareness of how language is structured, and that spoken language can be segmented into smaller elements (= phonemes) that correlate to a defined set of symbols (= graphemes). It has been shown that this awareness is impaired in dyslexic individuals (Bradley & Bryant 1978; Wagner & Torgesen 1987; Bruck 1992; Torgesen 1995; Shaywitz 2003), and that a training of phonological awareness improves the process of learning to read (Castles & Coltheart 2004). However, also individuals affected with dyslexia but intact phonological abilities have been reported (Castles & Coltheart 1996; Valdois *et al.* 2003). This illustrates that dyslexia cannot always be explained by the phonological deficit theory alone, but that at least some subgroups presenting with other deficits do exist. Several alternative hypotheses such as the 'rapid auditory processing deficit theory' or the 'magnocellular theory' have been suggested, and arguments for each of the hypotheses have been critically reviewed elsewhere (Ramus *et al.* 2003).

2.1.3 Neurobiological studies

First insights into the neurobiological basis of dyslexia were given by anatomical studies performed on postmortem brain specimen obtained from dyslexic individuals (Galaburda & Kemper 1979; Galaburda *et al.* 1985). Subtle cortical anomalies such as nests of neurons and focal microgyria were identified, which located to cortical regions in the left hemisphere of the brain (Shaywitz & Shaywitz 2008).

Static structural measurements in human brain have identified differences between dyslexic and non-impaired readers in three brain regions: the left tempo-parietal, left frontal and the left occipito-temporal region (Fig. 1; (Eliez *et al.* 2000; Klingberg *et al.* 2000; Brown *et al.* 2001)). Subsequently, functional imaging studies have started to examine which brain functions are triggered during the performance of particular cognitive tasks, e.g. reading, and which of them are altered in dyslexic individuals. The use of functional magnetic resonance imaging (fMRI), for instance, is non-invasive and can be used repeatedly, which makes it an ideal tool for

(Shaywitz & Shaywitz 2005). FMRI measurements

based on the fact that the neural system gets activated in specific brain regions when performing a particular cognitive task. This activation requires energy and can therefore be monitored by measuring changes in brain metabolic activity (e.g. changes in the cerebral blood

flow and the utilization of metabolic substrates such as glucose). In a fMRI study involving the reading of pseudowords (= non-words), significant differences in brain activation patterns were found between dyslexic and non-impaired children: the non-impaired children showed greater activation in left hemisphere sites such as the parieto-temporal and middle temporal / middle occipito-temporal region (Fig. 1; (Shaywitz *et al.* 2002)). These findings were in concordance with results from other brain imaging studies that also provided evidence for a failure of the left hemisphere posterior brain systems in dyslexic individuals, in reading tasks (e.g. (Horwitz *et al.* 1998; Brunswick *et al.* 1999; Temple *et al.* 2000; Paulesu *et al.* 2001; Shaywitz *et al.* 2002)) and in non-reading visual processing tasks (Eden *et al.* 1996; Demb *et al.* 1998).

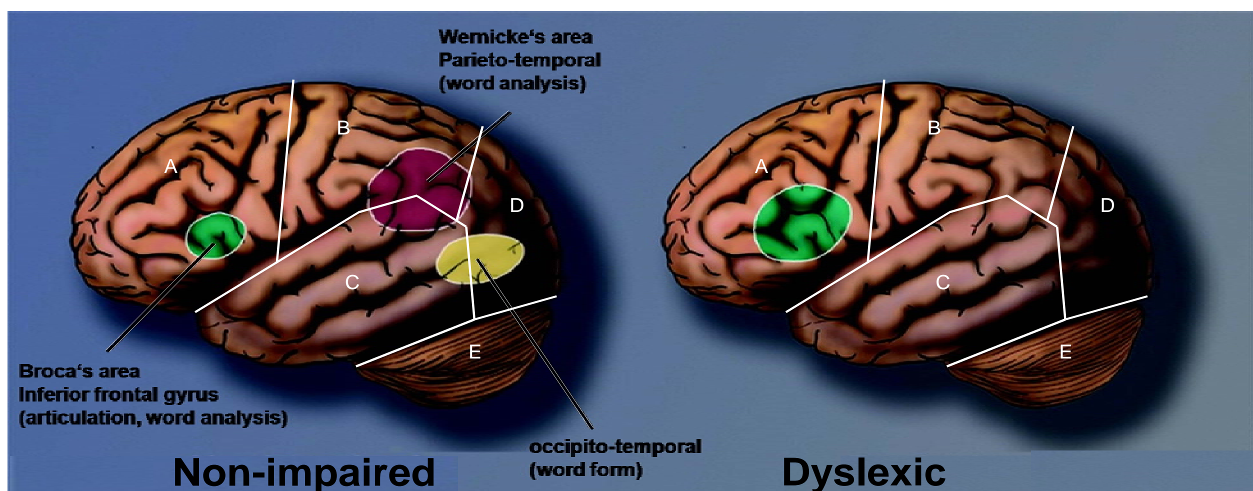


Fig. 1: Brain regions involved in normal and impaired reading. The left hemisphere of a human brain is schematically represented, with the main brain regions separated by white lines (A – frontal lobe, B – temporal lobe, C – parietal lobe, D – occipital lobe, E – cerebellum). On the left side, the three neural systems for reading are indicated by their activation patterns in non-impaired readers. On the right side, the 'neural signature for dyslexia' is shown. In dyslexic readers, no activation is seen in the two posterior systems (yellow, violet in the non-impaired readers), while the anterior system is overactivated (indicated by the enlarged green Broca's area). This is suggested to be the compensatory mechanism in impaired readers. Figure modified (Shaywitz & Shaywitz 2008).

Some other functional methods such as positron emission tomography (Paulesu *et al.* 1996; McCrory *et al.* 2005; Dufor *et al.* 2007) and magnetoencephalography (Salmelin *et al.* 2000; Wehner *et al.* 2007) have also contributed to our understanding of cognitive processes in dyslexia. Results from these studies support the aforementioned fMRI findings and further strengthen the hypothesis that disruptions of (i) the parieto-temporal and (ii) the occipito-temporal left hemisphere reading systems are the neurobiological reasons why dyslexic children fail to develop adequate reading skills. The pattern of underactivation in these two posterior regions of the left hemisphere is now denoted as the 'neural signature for dyslexia' (Shaywitz & Shaywitz 2008). However, there is evidence that the brain of impaired readers develops some kind of compensatory mechanisms that seem to involve areas around the

inferior frontal gyrus in both hemispheres, as well as the right occipito-temporal region (Shaywitz & Shaywitz 2005).

2.1.4 Dyslexia and related endophenotypes

On the cognitive level, dyslexia is a compound disorder which is characterized by the disturbance in at least one of the specific cognitive processes that enable reading and spelling. Cognitive components involved in reading and spelling include visual processing, phonological awareness, verbal short-term memory, phonological decoding, orthographic processing and rapid naming (Tab. 1). Each of these different cognitive dimensions eases the separate skills of reading and spelling, however, it is assumed that any single of these cognitive impairments is neither necessary nor sufficient to cause the disorder (Plomin *et al.* 1997). Dyslexic individuals each provide with a particular combination of some of these disturbances, resulting in an individual, partial representation of the general, entire dyslexia cognitive spectrum. Therefore, the analysis of single neuropsychological measures in dyslexia probands as quantitative traits can be used to further dissect the complex dyslexia phenotype. A detailed description of these dyslexia-related endophenotypes is given in Tab. 1.

Tab. 1: Endophenotypes of the dyslexia cognitive spectrum.

Component	Description
Visual processing	Reading requires the magnocellular system to respond to stimuli of low spatial frequency and low contrast. An impaired perception of these stimuli has been found repeatedly in dyslexic individuals. The exact nature of this deficiency is not yet clear.
Phonological awareness	Phonemes are the smallest meaningful sound units of spoken words. The ability to perceive, segment and manipulate them is crucial to construct an acoustic speech flow. The capacity for phonological awareness describes an oral language skill that is required to learn the letter-sound correspondences that are characteristic for the respective language. It is often tested through a phoneme deletion task.
Verbal short-term memory	Various aspects of memory are required for reading and spelling. For many known words, no dissection into phonemes is needed anymore. Instead, they are recalled directly from memory.
Phonological decoding	Phonological coding describes the ability to first put together the phonemes and then verbally express words which have never been previously read or heard. The skill demonstrates an understanding of letter-sound correspondences. Impaired individuals have difficulties in the reading of pseudowords.
Orthographic processing	Orthographic coding refers to the process of recognizing a word by its holistic form. It is measured by a pseudohomophone task where an orally presented word has to be compared with a visual representation of two phonologically indistinguishable words, of which only one is correctly spelled (e.g. "Wachstum" and "Waxtum").
Rapid naming	Rapid naming is a measure of the speed of processing and describes the temporal aspect of the cognitive events related to reading and spelling. This skill is often associated with reading fluency.

Table adapted (McGrath *et al.* 2006; Schumacher *et al.* 2007).

In addition to the cognitive components that are directly related to reading and writing skills as described in Tab. 1, there are additional subdimensions that are expected to correlate with certain aspects of the written language system. The left-right asymmetrical function of the brain,

for instance, is a conserved feature of vertebrate central nervous systems (Rogers & Andrew 2002) and has been suggested to underlie many aspects of behavior, cognition and emotion in humans (Hughdal & Davidson 2003). One measure of changes in brain asymmetry is the preferred usage of one hand in daily practice, i.e. left- or right-handedness. Additionally, a serious deficit in the acquisition of numerical abilities and calculation skills often co-occurs in dyslexic children (Dirks *et al.* 2008). Although the reason therefore still remains elusive, there is substantial evidence that arithmetic and literacy skills depend on similar cognitive factors (Hecht *et al.* 2001; Geary & Hoard 2002).

Electrophysiological measurements are another approach to assess some of the cognitive endophenotypes of dyslexia (Shaywitz & Shaywitz 2008). Among these, the mismatch negativity (MMN) is considered to be the neural correlate of speech perception and has been found impaired in dyslexic children (Schulte-Körne *et al.* 1998a). The MMN can be derived from electroencephalography (EEG) measurements and is generated by the automatic response of the brain to any change in auditory stimulation. The MMN is a negative curve, obtained by subtracting an event-related potential (ERP) to a standard stimulus (e.g. /da/) from the ERP to a certain deviant stimulus (e.g /ba/). It has been shown that MMN measurements present with both, a strong heritability and high retest reliability, respectively (Hall *et al.* 2006). The speech MMN is also altered in dyslexic adults (Schulte-Körne *et al.* 2001a), which provides further evidence that it is an excellent candidate trait marker for an important perceptual process in dyslexia. The generation of MMN data is presented in Fig. 2.

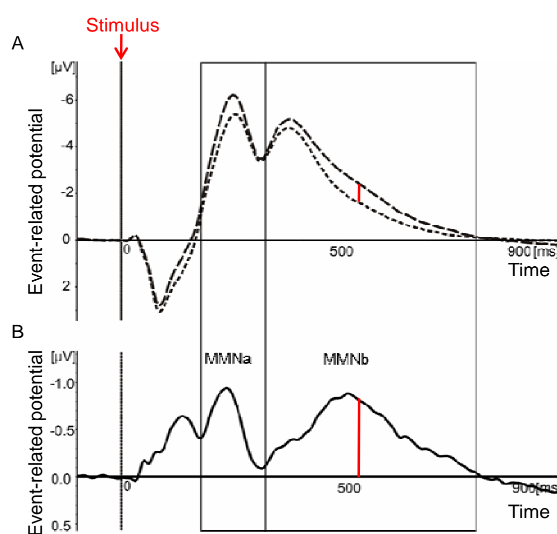


Fig. 2: Event-related potentials (ERPs) and mismatch negativity (MMN) curves. On the x-axis, the time course of the EEG measurement is indicated, and the event-related action potential is shown on the y-axis. At time-point 0, the standard or the deviant stimulus is applied. (A) Grand average of the standard (dotted line) and deviant (dashed line) curves. The red bar schematically represents the distance used for calculation of the mismatch values. (B) Over the entire time course, this results in a negativity curve. Dyslexic individuals show attenuated MMN values in particular in the time windows 180 – 300 ms (MMNa) and 300 – 710 ms (MMNb). Figure modified (Roeske *et al.* 2009).

2.1.5 Therapy and remediation

Therapeutic approaches to treat reading and spelling disability can be split into two main methods, namely prevention and intervention programs. Prevention is based on the hypothesis that a successful treatment in childhood should start as early as possible. These programs address at-risk children who show first behavioral aspects that can be seen as correlates of a subsequent development of dyslexia. First studies investigating the outcome of pre-school screenings have provided evidence that weak reading and spelling abilities in 2nd grade of school can be precisely predicted in kindergarten-age (Jansen *et al.* 2002; Maurer *et al.* 2009). Prevention methods have been shown to be of efficient benefit to the children if they are administered correctly and in an intensity required by the particular child (Clay 1985; Torgesen *et al.* 1999; Denton & Mathes 2003; Tunmer & Chapman 2003; Vellutino *et al.* 2003).
In contrast to prevention programs, intervention programs are administered to school-aged children who have already been diagnosed with dyslexia. One distinguishes between pure linguistic approaches, which try to remediate the impairments on the phonological level, e.g. by training grapheme-phoneme correlations (Alexander & Slinger-Constant 2004), and the second group of neurobiological-based therapy approaches. The latter aim at improving the functional performance of affected neuronal systems by taking advantage of the neuronal plasticity of the human brain (Eden & Moats 2002). Although first studies indicate that, following intervention, many dyslexic children improved their reading and also demonstrated an increase in activation of at least some of the neural systems required for reading (Temple *et al.* 2000; Temple *et al.* 2003; Eden *et al.* 2004; Shaywitz *et al.* 2004), there is still a considerable number of children who are resistant to nowadays' treatment. These children represent a challenge for future identification of other factors required for an optimal therapeutic outcome (Alexander & Slinger-Constant 2004; Rüsseler 2006).

2.2 Identification of causal genes in human disorders

The sequence of the human nuclear genome has been unraveled by a joint effort of an international consortium, the Human Genome Project (HGP). In 2001, the first draft assembly of the human haploid genome was presented to the public (Lander *et al.* 2001; Venter *et al*. 2001). Up to today, most of the remaining missing parts of the DNA sequence have been filled in, and a general reference sequence for the human genome is publically available. Current estimates suggest that about 22,000 genes are encoded in the human DNA, however, they are expected to only account for 1 - 2% of the entire DNA sequence (Lander *et al.* 2001; Levine & Tjian 2003). Although the function of the remaining parts is not fully understood yet, it is assumed that a

large fraction is involved in regulation of gene activity and might be responsible for structural aspects of the cell.

2.2.1 Genetic variability

The DNA sequence is characterized by a considerable, individual variability. Currently, it is assumed that about 98% of the human genome is similar between two randomly compared individuals. Given an average sequence length of the human genome of about 3.2 billion base pairs, this suggests that the variable part of the genomic information still comprises several million bases (Shianna & Willard 2006). This human genetic variability contributes to inter-individual differences, together with external factors such as environmental influences (Li & Sadler 1991; Sachidanandam *et al.* 2001; WTCCC 2007). In most cases, genetic variants emerge from errors that occur during replication of DNA or the correction of DNA damages (Cooper & Krawczak 1993). Most of the genetic variations do not have grave consequences on phenotypic level. In this case, variants are referred to as 'polymorphisms'. However, a part of them are directly causing for or contributing to human diseases. In case of monogenic (Mendelian) disorders, these causal variants are described as 'mutations'. In genetically complex, multifactorial diseases, in which different genetic variations interact with environmental factors, the border between 'mutations' and 'polymorphisms' is less clear cut. Polymorphisms, although they do not contribute to disease risk, are of considerable value for biomedical research, as they can be used as genetic markers in the search for the genetic factors underlying human diseases.

The genetic architecture of traits

All human traits, regardless of being diseases or phenotypic traits such as height or eye color, are at least partially genetically determined. Thereby, the phenotypic manifestation of a given variant mainly depends on its genetic effect (penetrance) (Hirschhorn & Daly 2005). Most of the monogenic disorders are caused by rare mutations with high genetic penetrance, making mutation carriers directly develop the phenotype. In contrast, common disorders are genetically complex and interact with externtal factors. Here, the penetrance of the genetic variants is low. An overview of the correlation of allele frequencies and penetrance is given in Fig. 3.

Initially, it was expected that most of the complex disorders would underlie the 'common variant – common disease' hypothesis (Lander 1996; Cargill *et al.* 1999; Chakravarti 1999). It predicts that the genetic risk for common diseases will often be related to disease-predisposing alleles with relatively high frequencies (Reich & Lander 2001). This has been proven to be true for at least some examples such as the *APOE* ε4 allele in Alzheimer's disease (Corder *et al.* 1993) or the PPARγ Pro12Ala variant in type II diabetes (Altshuler *et al.* 2000). However, in the last

years, a second competing theory has arisen. The 'rare variant – common disease' hypothesis suggests that multiple rare variants, which mostly occur *de novo*, are responsible for disease susceptibility (Pritchard 2001). As shown in Fig. 3, the identification of these rare variants will require large samples and, most efficiently, whole-genome sequencing.

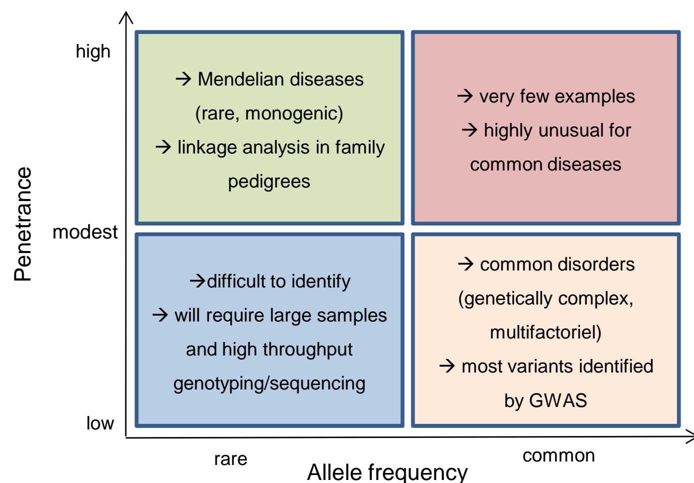


Fig. 3: The architecture of human traits. Genetic variants and their contribution to human disease susceptibility are indicated with respect to allele frequencies (x-axis) and the penetrance of these variants (y-axis). Own illustration.

<u>Single nucleotide polymorphisms</u>

The most abundant type of genetic sequence variants, representing about 80% of genomic variability, is the exchange of single bases (single nucleotide polymorphisms, SNPs; (Wang *et al.* 1998; HapMap Consortium 2003; 2005)). Up to now, at least 10 to 11 million SNPs that occur with a minor allele frequency (MAF) of at least 1% have been identified in the human genome (HapMap Consortium 2005). Information on these is accessible in publically available databases such as the National Center for Biotechnology Information (NCBI) or HapMap. SNPs are abundant in both, coding and non-coding regions of the genome. Depending on their location, they can have different impacts at the phenotypic level: SNPs in protein-coding regions of a gene may either be synonymous, yielding the same amino acid (aa), or non-synonymous. The latter situation results in an amino acid change which might alter the structure and / or the function of the respective encoded protein. In contrast, SNPs in regulatory regions, such as introns or the promoter-regions up- and downstream of a gene, might alter its expression and, subsequently, can influence the amount of the protein that is generated.

Besides SNPs, other types of variable elements include gained (insertion) or lost (deletion) genetic material, different forms of repetitive elements such as e.g. microsatellites (short tandem repeats (STRs) with 1 to 9 bp motifs) and copy number variants (CNVs). Although they show lower frequencies in the genome compared to SNPs, they comprise longer DNA sequences and are thus also of interest (Feuk *et al.* 2006; Redon *et al.* 2006).

The haplotype block structure of the human genome

Given the large number of SNPs, it is rarely possible to genotype all of them in individual samples. Genetic research therefore takes advantage of the fact that the human genome is organized in so-called 'haplotype blocks'. As shown by systematical investigations, recombinations do not occur at random positions in the genome, but are specifically restricted to certain loci (Jeffreys *et al.* 2001; HapMap Consortium 2003; 2005; Myers *et al.* 2005). In consequence, SNPs located between two recombinational hot spots are rarely separated from each other during meitotic recombinational events. They are dependent from one another and therefore provide with the same genetic information. This observation has been implemented as concept of linkage disequilibrium (LD). The number of SNPs that are needed to be analyzed in each individual to capture its genetic variability can now be dramatically decreased, as it is possible to design tagging SNPs (tSNPs) for each haplotype block (Johnson *et al.* 2001; Becker 2007; Steemers & Gunderson 2007). To determine the extent to which two SNPs are linked, the measures 'r^2' and 'D'' have been implemented, which both can take values from 0 (no LD) to 1 (perfect / complete LD; (HapMap Consortium 2005; Steemers & Gunderson 2007)). Two SNPs are in 'perfect' LD if $r^2 = 1$, which indicates that the alleles of the two SNPs always show the same state and provide with equal frequencies. A 'complete' LD is shown by $D' = 1$, which represents that the two SNPs have not been separated by recombination (Skipper *et al.* 2005).

LD is not only observed between two adjacent SNPs but can also include a large number of SNPs, which makes some LD blocks extend over several hundred base pairs. Based on this organization of the human genome, it is currently estimated that about 300,000 to 1,000,000 tSNPs are sufficient to capture the majority of genetic variation within a population (Gabriel, 2002). This knowledge has facilitated the design of genetic experiments that aim at investigating the genetic variants which underlie human disorders.

2.2.2 Linkage and association analysis

In linkage analyses, highly polymorphic microsatellite markers are used to track the inheritance of different chromosomal regions within families. Large pedigrees are screened for a joint appearance of a specific genetic variant and a disease. In systematic linkage scans, the entire genome is evenly covered with about 400 microsatellites. Genotype data for each of the investigated markers are analyzed with respect to linkage by using either parametric or non-parametric approaches. Parametric linkage analyses require precise knowledge about the genetic model underlying the trait of interest, such as mode of inheritance (recessive, dominant) and penetrance. This approach is often used in monogenic disorders where formal genetic studies have suggested a specific genetic model, e.g. cystic fibrosis (Zielenski & Tsui 1995). In

contrast, non-parametric approaches are performed without any prior knowledge about the underlying genetic model. This makes them a valid statistical tool for the analysis of complex disorders, in which genetic heterogeneity and the additional impact of environmental factors make it more difficult to find markers that co-segregate with the trait of interest. As only a few variants contributing to complex diseases have been conclusively identified by linkage analyses so far, it has become clear that the application of this approach in the analysis of complex traits is limited (Altmüller *et al*. 2001).

Another systematical approach for the identification of disease loci is the method of association analyses, which search for population-wide correlations between a phenotype and specific allelic variants of SNPs (Cardon & Bell 2001). Association analyses do not require extended family pedigrees, however, they are based on a large number of unrelated individuals in so-called case-control studies. Here, allele or genotype frequencies are compared between groups of affected (cases) and unaffected individuals (controls) of ethnically similar populations. In contrast to linkage analyses, association studies provide an increased resolution: while linkage regions are often several megabases (Mb) in size (and thus contain a large number of potential candidate genes), an associated region generally maps to some thousand to hundred thousand base pairs.

Association studies can also be applied as family-based designs using familial trios, which comprise at least one affected child and its parents. This type of analysis additionally offers the possibility to investigate parent-of-origin (= imprinting) effects. In genomic imprinting, paternal and maternal alleles of a gene have different levels of activity and, in some cases, one of the alleles is even completely silenced (Reik & Walter 2001). The underlying molecular mechanisms involve epigenetic events such as methylation or histone modifications (Delaval & Feil 2004), and imprinted genes have already been shown to be involved in cognition and behavior (Isles & Wilkinson 2000).

Apart from genotyping each DNA of large case or control samples individually, association studies can also be performed on DNA pools. In this approach, large numbers of DNA samples (cases or controls) are pipetted in one pooled sample which is subsequently genotyped. This method allows for an estimation of genome-wide allele frequencies in large samples, with the data pointing towards genomic regions that contain markers with differences in allele frequencies between cases and controls. Although the absolute values are not necessarily correct at the quantitative level, they do suggest markers that should be taken into subsequent individual genotyping. Pooling is a cost-efficient method to detect general allele frequency

differences between cases and controls, and its utility as first screen in order to detect genetic susceptibility variants has already been demonstrated (Kirov *et al.* 2008; Shifman *et al.* 2008).

2.2.3 Genotyping technologies

Given the complexity of the human genome and the large number of samples needed for genetic analyses in complex disorders, robust high-throughput genotyping methods are required. During the last years, advances in the field of technology have led to the development of systematical methods that allow for simultaneous analysis of some ten up to several hundreds of thousands markers in one assay, often with parallelizing also the number of individuals included.

Genotyping based on BeadArray Technology

The most established platforms for parallelized analysis of thousands of markers are provided by Affymetrix and Illumina. The Affymetrix system is based on randomly chosen SNPs that are analyzed after an oligonucleotide-based amplification of the DNA (Matsuzaki *et al.* 2004). In contrast, the BeadArray technology of Illumina is based on a whole-genome amplification without a prior polymerase chain reaction (PCR), and SNPs included in this method are tSNPs that were chosen to capture as much of the genetic variation within each individual as possible (Gunderson *et al.* 2005). The quality of data generated by Illumina technology has been shown to be superior to the Affymetrix data in terms of reproducibility and success rate (Suarez *et al.* 2005; Kim *et al.* 2009).

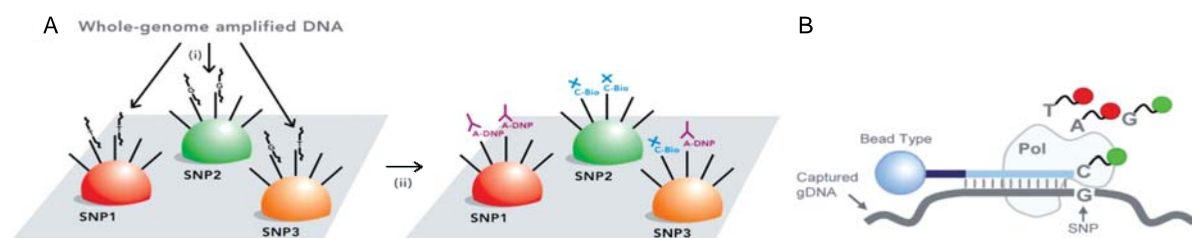


Fig. 4: Schematic representation of Illumina Infinium-II assay. (A) Overview of the basic principle of the BeadArray technology. (i) Whole-genome amplified DNA fragments hybridize to their complementary oligonucleotides that are immobilized on beads. After single-base extension, (ii) an immunohistochemical staining is performed to provide fluorescent allele-specific signals. (B) Single-base extension. Following successful hybridization, the oligonucleotide attached to the bead is extended of one base in accordance with the particular allele of the DNA fragment. Hereby, the nucleotide is modified so as to provide the basis for the subsequent staining. Figure adapted (Steemers & Gunderson 2007).

As shown in Fig. 4, the principle of Illumina's BeadArray is a direct hybridization of amplified DNA fragments to sequence-specific oligonucleotides that are 50 bases long and end directly adjacent to the target SNP. The oligonucleotides for each SNP are linked to micro beads of 3 µm size, which are immobilized on a BeadChip (Steemers *et al.* 2006). Following the sequence-

specific hybridization, an enzymatically catalyzed single-base extension (SBE) with labeled nucleotides takes place (Steemers & Gunderson 2007). During this reaction, one single nucleotide is added to the oligonucleotide that is attached to the bead, depending on the allele of the hybridized strand. The labeled nucleotides are subsequently stained based on immuno-histochemical reactions using Cy5 (green, C/G) and Cy3 (red, A/T) fluorochromes. Thus, the Illumina system is a two-color system which is not able to analyze A/T and C/G SNPs. However, according to information provided by the company, these SNPs only account for about 17% of the genomic variants in the human genome, and most of them can be replaced by other SNPs in high LD (Steemers & Gunderson 2007). To read the genotype calls, BeadChips are scanned using a confocal laser scanner (BeadArray Reader), and data are analyzed using the BeadStudio software (Fig. 5).

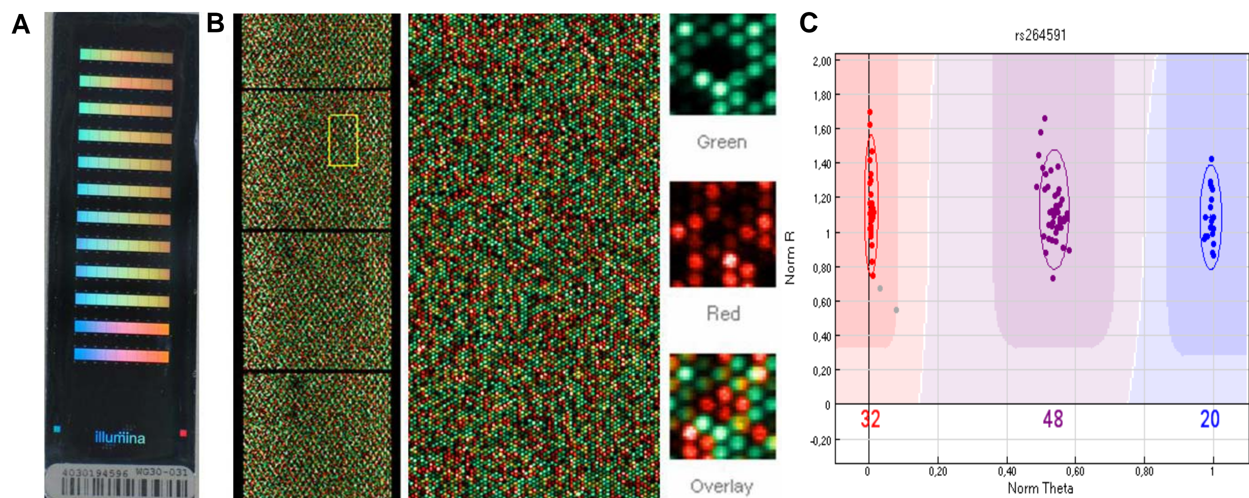


Fig. 5: HumanHap300K and the corresponding analysis by BeadScan / BeadStudio software. (A) Illumina's HumanHap300K BeadChip. (B) Scanning process. One section is read (left). The yellow square indicates the area that is enlarged in the middle figure. On the right side, single beads located within the selected area are shown for the two different laser channels (red, green), and an overlay is presented at the bottom. Beads that appear yellow in the overlay image represent heterozygous genotypes. (C) Example of a cluster plot as shown by BeadStudio analysis software. For SNP rs264591, the total number of 100 individuals is split into groups of homozygous AA (n = 32), heterozygous AB (n = 48) and homozygous BB (n = 20) individuals.

The SNP content is designed by the company itself, based on respectively current HapMap data. As the BeadArray technology has been further optimized during the last years, the number of SNPs that can be analyzed per person on one BeadChip has increased from 300,000 (HumanHap300) to around 1 million (Human1M-DUO) till date. In the last years, genome-wide association studies using Illumina BeadArray technology have identified a large number of disease-associated loci (Duerr *et al.* 2006; Winkelmann *et al.* 2007; Hillmer *et al.* 2008; Birnbaum *et al.* 2009).

Genotyping based on MassExtend Reaction

For candidate gene-, finemapping- or replication studies, it is often required to focus on a selected number of SNPs (i.e. 10 – 100) in an increased number of individuals. Herefore, genotyping based on MassExtend reactions and MALDI-ToF-MS (matrix assisted laser desorption / ionization time of flight mass spectrometry) by Sequenom is one state-of-the-art technology (Tang *et al.* 1995; Haff & Smirnov 1997). The iPLEX Gold MassEXTEND platform enables parallel analysis of up to 40 SNPs in one single assay (plex).

An initial multiplex PCR step with standard primers specific for each SNP of the plex is required. The primers generate amplicons of 80 – 120 bp in size. To prevent residual nucleotides from being incorporated during the subsequent single-base extension (SBE) reaction, they are enzymatically inactivated using shrimp alcaline phosphatase (SAP; (Oeth *et al.* 2005)). For each of the target SNPs, SBE takes place during a second PCR. In this reaction, mass-modified dNTPs and SNP-specific extension primers (unextended primers (UEP)) are used. As molecules with high masses are more difficult to accelerate in the final MS measurement, UEP primers are used in four groups of different concentrations (= 'primer adjustment'). During the SBE reaction, primers anneal directly adjacent to the SNP of interest and are elongated by one nucleotide, according to the allele present in the respective strand (Haff & Smirnov 1997; Ross *et al.* 1998). This extension reaction yields a mixture of extended and unextended products of different lengths and, accordingly, of different masses (Blondal *et al.* 2003).

As the final analysis takes place in an electrical field, cationic molecules need to be removed from the assay. Therefore, the mix is purified by an ion-exchange resin (Oeth *et al.* 2005). The purified mix is subsequently spotted on a SpectroChip. Its matrix consists of an organic acid (3-hydroxypicolin acid) that is available in 100 – 1,000 fold excess as compared to the spotted material. Analysis of the reaction takes place in a mass spectrometer under vacuum (Fig. 6).

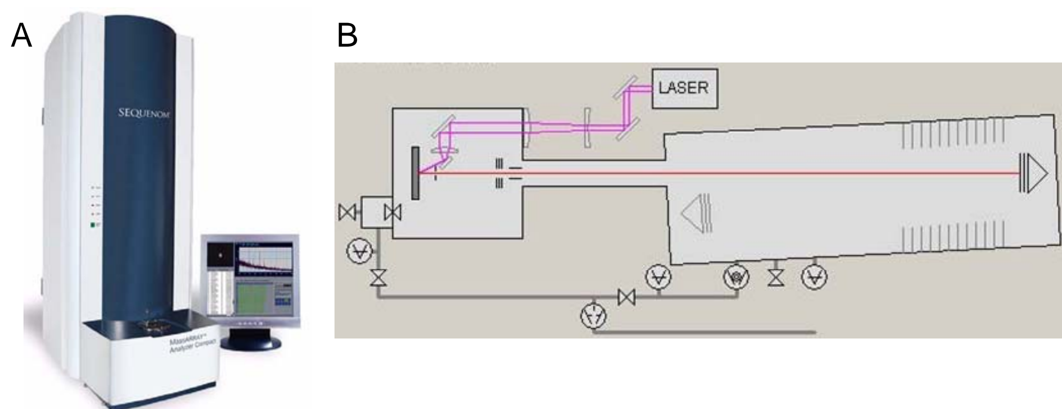


Fig. 6: The principle of MALDI-ToF-MS. (A) Mass spectrometer by Sequenom. (B) Schematic representation of the working principle. Analyte molecules are ionized by a laser impulse and extracted from the matrix. With different speed, they fly in an electric field until they reach the detector. The time of flight correlates with the molecule's mass.

A short laser impulse on the sample provides the energy required by the matrix to ionize the analyte molecules, and to let them detach from the matrix (desorption / ionization). The ionized molecules are accelerated in the electrical field that is provided in the vacuum flight channel of the mass spectrometer (Cotter 1992; Ragoussis *et al.* 2006). Depending on their respective masses, molecules show different flight times until they reach the detector: the larger the mass, the longer the time of flight. The previously performed primer adjustment allows for a good signal-to-noise ratio.

The Sequenom technology enables accurate measurements of molecules with masses in the range from 4,500 to 9,000 Da (Ragoussis *et al.* 2006). By using the system's analysis software (Typer 3.4 and 4), the time-of-flight measurements are converted into the respective masses of the molecules, and the resulting genotypes can be determined. A graphical representation of such SNP analysis is depicted in Fig. 7.

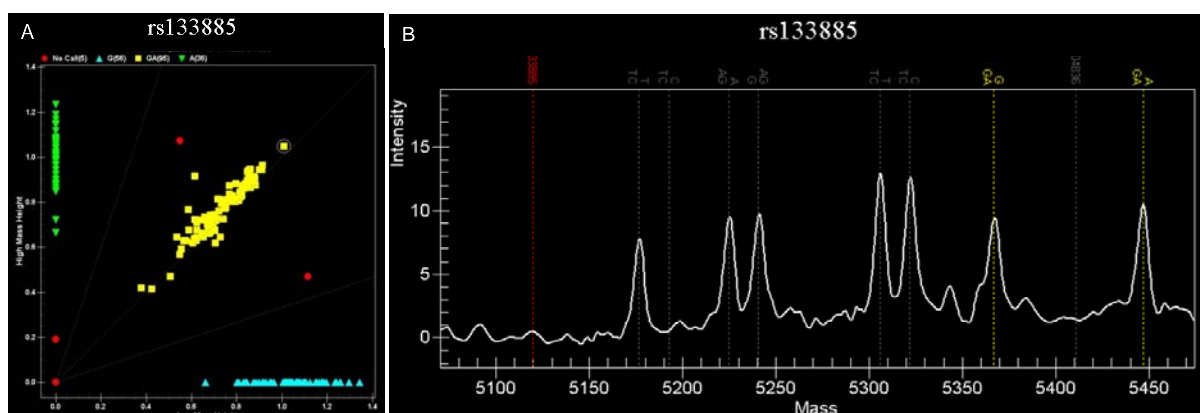


Fig. 7: Graphical representation of SNP genotyping using MassExtend technology. The analysis is depicted for SNP rs133885. (A) Cluster plot. All individuals of one analysis are represented with their individual genotypes. Individuals carrying the homozygous "GG" genotype are shown by blue triangles, while homozygous "AA" samples are depicted in green triangles. Heterozygous "AG" carriers are clustered in the middle (yellow squares). (B) Mass spectrum. The spectrum represents the masses that have been measured for one individual (heterozygous "AG", encircled in white in (A)). The unextended primer, with a mass of 5120 Da, is marked in red. No peak is visible at this mass, indicating that the UEP primers have been entirely converted into analytes. Along the x-axis, different grey-labeled masses are depicted. They correspond to mass peaks generated by other SNPs in the multiplex reaction. At 5367 Da, which corresponds to the expected mass for the G allele of rs133885, a distinct peak is indicated by a yellow line. The mass corresponding to the A allele (5448 Da) is also detected, indicating that the corresponding DNA sample has a heterozygous AG genotype.

2.3 Genetics of dyslexia

The dyslexia phenotype was first described in 1895 (Hinshelwood 1895). Only some years later, it was suggested that dyslexia clusters in families (Hinshelwood 1907; Stephenson 1907), and this observation was confirmed in many subsequent studies (Hallgren 1950; Walker & Cole 1965; Owen *et al.* 1971; Rutter & Yule 1975). Since then, numerous studies have been conducted to assess its epidemiology and to identify genetic loci and candidate genes that confer susceptibility for reading and spelling disability.

2.3.1 Formal genetics

Most of what is known about the epidemiology of dyslexia comes from family-, twin- and adoption studies. A child with an affected parent has a risk of 40 - 60% to develop dyslexia, with the risk being further increased when other family members are also affected (Stephenson 1907; Hallgren 1950; Stevenson 1991; Olson *et al.* 1994; Schulte-Körne *et al.* 1996; Ziegler *et al.* 2005). Compared to the prevalence rate in the general population, first degree relatives of affected individuals have a 4 to 10fold increased risk, and this number is even higher when stricter diagnostic criteria are applied (Ziegler *et al.* 2005).

Twin studies allow for an estimation of the contribution of both, environmental and genetic factors, to disease development. For dyslexia, it has been shown that concordance rates in monozygotic twins are higher than in dizygotic twins, indicating that the trait is highly heritable (DeFries *et al.* 1987; Stevenson *et al.* 1987; Olson 2002; Plomin & Kovas 2005). The proportion of inherited factors involved is estimated to range between 40 and 80%, with the highest numbers being reported for the phenotypic subdimensions word reading (up to 58%) and spelling (70%; (Gayán & Olson 2001; Olson 2002; Plomin & Kovas 2005)). The effect of shared and non-shared environmental factors has been suggested to be low for word reading, however, it is substantially higher for correlated traits such as phonological awareness (Gayán & Olson 2001).

Recent evidence points to a two-fold increase of risk for male children (Shaywitz *et al.* 1990; Rutter *et al.* 2004). This sex ratio has been suggested to be influenced by severity, IQ and assessed cognitive profiles (Olson 2002). Two studies investigated whether sex has an influence on dyslexia heritability: although results from the Colorado Twin Study showed similar heritability between the sexes (Wadsworth *et al.* 2000; Hawke *et al.* 2006), the London Twins Early Development Study (TEDS) found a higher heritability for boys compared to girls (Harlaar *et al.* 2005). However, it still remains elusive whether the latter findings are attributable to sex-specific genetic factors or environmental influences.

2.3.2 Dyslexia susceptibility loci

Linkage studies for dyslexia have suggested nine chromosomal regions (DYX, dyslexia susceptibility loci) which are expected to harbor dyslexia candidate genes. They are listed as DYX1 to DYX9 by the HUGO Gene Nomenclature Committee (HGNC) and are described in Tab. 2. In addition, it was tried to correlate quantitative endophenotypes with one or more of these general linkage regions, but there is only little support from independent studies so far (Williams & O'Donovan 2006; Schumacher *et al.* 2007). Nevertheless, there is an impressive consistency of linkage findings for dyslexia as compared to other neuropsychiatric disorders.

This is especially true for the dyslexia susceptibility loci on 1p34-p36 (DYX8), 6p21-p22 (DYX2), 15q21 (DYX1) and 18q11 (DYX6), as they have received support from at least two large family samples. A summary of the linkage findings for dyslexia is given in Tab. 2.

Tab. 2: Dyslexia susceptibility loci DYX1 to DYX9.

Locus	Sample description	Qualitative trait or endophenotype	References
DYX1 (15q21)	9 multiplex families, USA 6 multiplex families, USA 7 families, Germany 90 families, USA	Reading Single-word reading Spelling Single-word reading	(Smith *et al.* 1983) (Grigorenko & Chang 1997) (Schulte-Körne *et al.* 1998b) (Chapman *et al.* 2004)
DYX2 (6p21-6p22)	19 multiplex families, USA 82 families, UK 79 families, USA 89 families, UK 119 families, USA 104 families, USA 8 multiplex families, USA	Dyslexia Orthographic and phonological processes Orthographic and phonological processes Phonological decoding Phonological decoding Orthographic and phonological processes Single-word reading, phoneme awareness	(Cardon *et al.* 1994) (Fisher *et al.* 1999) (Gayán *et al.* 1999) (Fisher *et al.* 2002) (Fisher *et al.* 2002) (Kaplan *et al.* 2002) (Grigorenko *et al.* 2003)
DYX3 (2p15 – 2p16)	1 multiplex family, Norway 89 families, UK 119 families, USA 96 families, Canada 11 multiplex families, Finland	Dyslexia Orthographic choice Phonological awareness Phonological coding, spelling Dyslexia	(Fagerheim *et al.* 1999) (Fisher *et al.* 2002) (Fisher *et al.* 2002) (Petryshen *et al.* 2002) (Kaminen *et al.* 2003)
DYX4 (6q11-6q12)	96 families, Canada	Phonological coding, spelling	(Petryshen *et al.* 2001)
DYX5 (3p12-3q13)	1 multiplex family, Finland	Dyslexia	(Nopola-Hemmi *et al.* 2001)
DYX6 (18p11)	89 families, UK 119 families, USA 84 families, UK	Single-word reading Single-word reading Phoneme awareness	All (Fisher *et al.* 2002)
DYX7 (11p15)	100 families, Canada	Dyslexia	(Hsiung *et al.* 2004)
DYX8 (1p34-1p36)	9 families, USA 8 multiplex families, USA 100 families, Canada	Dyslexia Single-word reading, phonological decoding Spelling, phonological coding	(Rabin *et al.* 1993) (Grigorenko *et al.* 2001) (Tzenova *et al.* 2004)
DYX9 (Xq26-Xq27)	1 multiplex family, Netherlands 89 families, UK	Dyslexia Single-word reading	(de Kovel *et al.* 2004) (Fisher *et al.* 2002)

Table modified (Schumacher et al. 2007). Linkage of dyslexia has also been reported to other genomic regions, however, the findings still lack replication in independent samples. These loci include linkage on 13q12 for word reading (Igo et al. 2006), 2q22 for phonological decoding efficiency (Raskind et al. 2005) and 4p12 and 12p for non-word repetition (Brkanac et al. 2008).

Different studies have also been conducted to assess pleiotropic effects on dyslexia and ADHD, two frequent comorbid childhood disorders. One study investigating dyslexia families with ADHD symptoms identified linkage regions on 14q32, 13q32 and 20q11 (Gayán *et al.* 2005), and a second study involving families with ADHD provided evidence for linkage with reading ability on 10q11, 16p12 and 17q22 (Loo *et al.* 2004). Although the nature of those linkage findings still remains elusive, it is likely that these loci harbor candidate genes which could at

least partially explain the comorbidity observed between the two frequent childhood disorders dyslexia and ADHD.

2.3.3 Dyslexia candidate genes

Given the relatively consistent linkage findings for dyslexia, there was initially great hope that at least one candidate gene would be identified in each of the susceptibility loci. Based on finemapping studies and breakpoint mapping in translocation families, candidate genes have been suggested for four of the linkage regions, and even for those, findings are not consistent across studies. Currently, there are six candidate genes discussed in the field.

DYX1: *DYX1C1*

DYX1C1 (*dyslexia susceptibility 1 candidate 1*, chr. 15: 53,497,246…53,587,724 bp, NCBI build 36) has been identified by breakpoint mapping in a Finnish two-generation family. Here, a translocation t(2;15)(q11;q21) was linked to reading-associated problems in four family members (Nopola-Hemmi *et al.* 2000; Taipale *et al.* 2003). Studies assessing whether *DYX1C1* also explains the general linkage findings at DYX1 provided some evidence for association (Taipale *et al.* 2003; Bates *et al.* 2009). *DYX1C1* is expressed in many tissues including those of the central nervous system (Taipale *et al.* 2003), and *in vivo* RNA interference (RNAi) studies in rats indicate that its rodent homologue is involved in neuronal migration (LoTurco *et al.* 2006; Wang *et al.* 2006). Interestingly, the final neuronal outcome in these rodent brains resembles the picture observed in human postmortem dyslexic brains (Galaburda *et al.* 2006). Very recently, it has been shown that *DYX1C1* might be involved in the regulation of estrogens, implicating effects of hormonal pathways in dyslexia (Massinen *et al.* 2009). However, replication of the initial genetic findings failed in a number of independent association studies carried out to date (Scerri *et al.* 2004; Bellini *et al.* 2005; Cope *et al.* 2005b; Marino *et al.* 2005; Meng *et al.* 2005a).

DYX2: *DCDC2* and *KIAA0319*

Most consistent evidence for an involvement in dyslexia has been described for two genes located within DYX2, namely *DCDC2* (*doublecortin domain containing protein 2*, chr. 6: 24,279,962…24,491,499 bp) and *KIAA0319* (chr. 6: 24,652,311…24,754,362 bp). Both genes map within a region of 500 kb on chromosome 6p22, and each of them is located in a cluster of three genes (*VMP*/*DCDC2*/*KAAG1* and *KIAA0319*/*TTRAP*/*THEM2*). First evidence for association of these two gene clusters with dyslexia was obtained by LD mapping in the linkage region of DYX2, in a set of US-American families (Deffenbacher *et al.* 2004). Two subsequent studies provided evidence for *DCDC2* to be the causative gene (Meng *et al.* 2005b; Schumacher *et al.* 2006). This, however, could not be replicated in two UK samples (Cope *et al.* 2005a; Harold *et al.*

2006). In terms of causality, Schumacher and colleagues postulated a two-marker haplotype located in intron 7 of *DCDC2* to confer dyslexia risk (Schumacher *et al.* 2006). In contrast, Meng and colleagues suggested an intronic deletion of 2,445 bp in *DCDC2*, comprising putative brain-associated transcription factor binding sites, to be the causal variant (Meng *et al.* 2005b). Functional evidence for either of the theories has not yet been presented. *DCDC2* belongs to the family of doublecortin domain containing genes (Reiner *et al.* 2006). In mice, it has been shown that two functioning copies of two members of this family, *Dcx* and *Dclk*, are required for axonal growth and for neuronal migration in the cerebral cortex (Deuel *et al.* 2006; Koizumi *et al.* 2006). The comparison of both, biochemical and cellular properties of *Dcdc2*, with *Dclk* and *Dcx* suggests shared functional features (Coquelle *et al.* 2006) and, indeed, it has been shown that an RNAi-mediated local loss of *Dcdc2* in rodents leads to an interruption of normal neuronal migration in the neocortex (Meng *et al.* 2005b).

Evidence for the second candidate gene at this locus, *KIAA0319*, has been presented in two independent UK studies (Francks *et al.* 2004; Cope *et al.* 2005a). There was no replication of these findings in two studies including samples from Germany (Schumacher *et al.* 2006) and the US (Meng *et al.* 2005b). Notably, variants in *KIAA0319* have been found to be associated with reading skills in the general UK population (Paracchini *et al.* 2008). It has been suggested that the causal variant, localized near of or within exon 1, causes a reduced expression of *KIAA0319* in lymphoblastoid and neuronal cell lines (Paracchini *et al.* 2006; Dennis *et al.* 2009). Similar to *Dcdc2*, a downregulation of the rodent homologue *Kiaa0319* induced by RNAi results in disturbed neuronal migration (Paracchini *et al.* 2006).

DYX3: *MRPL19* and *C2orf3*

MRPL19 (*mitochondrial ribosomal protein L19*, chr. 2: 75,727,417…75,751,387 bp) and *C2orf3* (chr. 2: 75,742,802…75,791,830 bp) have been identified in a LD mapping study of the 2p12 dyslexia candidate region, in Finnish families and a German sample (Anthoni *et al.* 2007). A two-marker haplotype was found to be overtransmitted to affected probands, and it was shown that the findings were significantly stronger when the samples were stratified for severity (Anthoni *et al.* 2007). The associated markers map to a region of 16.6 kb which contains *MRPL19* and *C2orf3*. Functional data for both genes are not yet available, however, *MRPL19* and *C2orf3* have found to be strongly co-expressed in various tissues of the adult human brain, and this expression pattern correlated with those found for the other candidate genes (Anthoni *et al.* 2007). *MRPL19* is a highly conserved gene and has been suggested to act in mitochondrial energy metabolism (Kenmochi *et al.* 2001). This might be in accordance with the fact that energy production is critical in the active brain for the performance of cognitive processes (Anthoni *et al.* 2007).

DYX5: *ROBO1*

ROBO1 (*roundabout 1*, chr. 3: 78;729,080…79,721,751 bp) was first identified by breakpoint mapping in one dyslexic Finnish individual carrying a t(3;8)(p12;q11) translocation (Hannula-Jouppi *et al.* 2005). The breakpoint disrupted the coding region of *ROBO1*, however, the evidence is still weak as one sibling of the affected individual also presented with the translocation but did not show any dyslexic symptoms. Thanks to its identification in axon-patterning mutants in *Drosophila melanogaster*, the role of *ROBO1* in neuronal development is well understood (Kidd *et al.* 1999; Andrews *et al.* 2006), and it has been shown that the gene and its vertebral homologues are important in axon growth, pathfinding, branching and interneuron migration (Yuan *et al.* 1999; Erskine *et al.* 2000; Andrews *et al.* 2006).

In summary, the present data suggest that the proteins encoded by *DYX1C1*, *KIAA0319*, *ROBO1* and *DCDC2* may be functionally linked in pathways involved in neuronal migration and axon growth. They provide first insights into the molecular processes underlying dyslexia, however, for none of the genes a direct functional link of specific genetic variants to particular processes in the developing and mature human brain has yet been established.
The current state of knowledge regarding the chromosomal regions harboring dyslexia susceptibility loci and candidate genes, respectively, is summarized in Fig. 8.

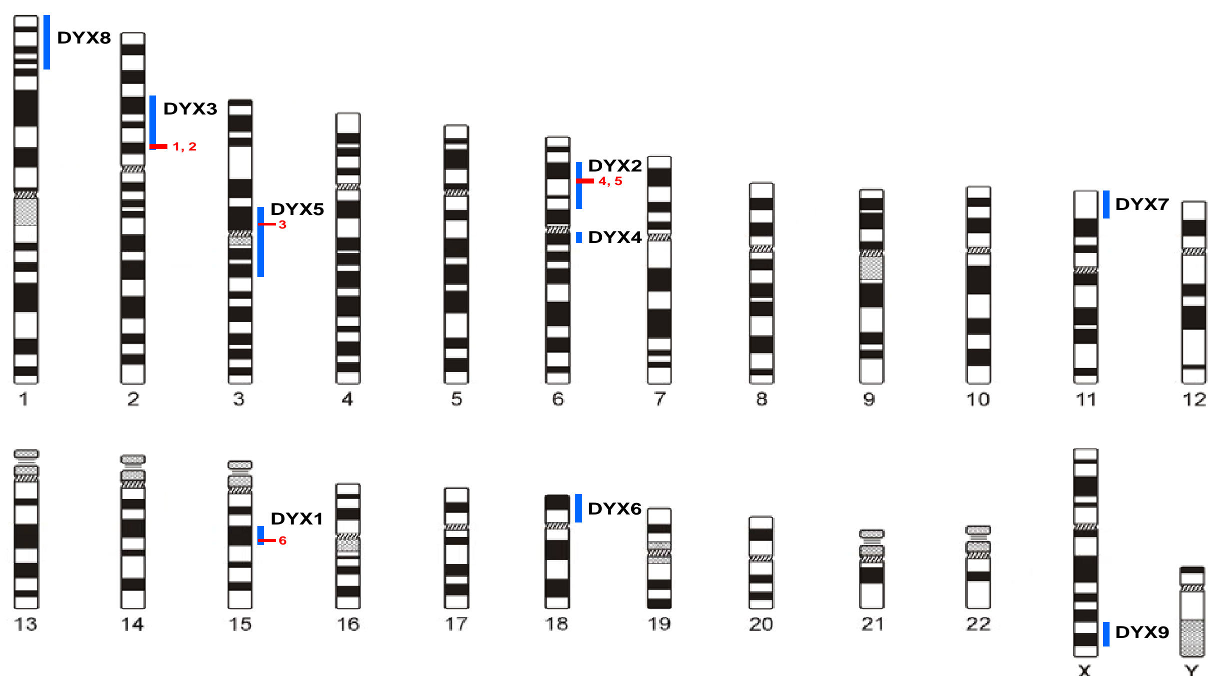


Fig. 8: Dyslexia susceptibility loci and candidate genes. The chromosomal regions which have been reported for dyslexia in linkage studies are shown in blue vertical lines (DYX1-9). Horizontal lines represent dyslexia candidate genes: 1 – *MRPL19*, 2 – *C2orf3*, 3 - *ROBO1*, 4 - *DCDC2*, 5 - *KIAA0319*, 6 - *DYX1C1*. Figure modified (Schulte-Körne *et al.* 2007a).

2.4 The NeuroDys Consortium

According to the prevalence rate of 5 to 12% (see section 2.1), it is estimated that approximately 2,500,000 children in the European Union (EU) are affected with dyslexia. Given this high number and the social problems that co-occur with the disorder, dyslexia also constitutes a substantial economic problem for the societies within the EU.

In 2006, the 'NeuroDys' consortium started as a multicentre, multidisciplinary project with the main goals to (i) identify the genetic factors that confer dyslexia susceptibility and (ii) investigate the biological basis of dyslexia at different levels. The envisaged approaches involved genetic studies as well as the analysis of the impact of environmental factors and the integration of different levels of neurosciences (structural and functional brain studies). The coordinative European effort comprised four molecular genetic laboratories and statistical departments (Bonn / Munich, Cardiff, Oxford, Stockholm), three neuroimaging laboratories (Maastricht, Salzburg, Zürich) and four electrophysiological laboratories (Budapest, Jyväskylä, Munich, Toulouse). Additional expertise was included from the field of phenotypic dimensions (Paris) and environmental interactions (London, Jyväskylä).

One major focus of the NeuroDys project was the collection of a large, well characterized European dyslexia sample in which standardized inclusion criteria and quantitative measures were applied similarly in all participating countries. The sample collected over a period of 2.5 years consists of dyslexia individuals and unaffected controls from France, Hungary, England, Sweden, Austria, Switzerland, the Netherlands and Germany. At the end of 2008, the European NeuroDys sample consisted of about 1,500 cases and 2,200 controls.

2.5 Scope of the doctoral thesis

Although formal genetic studies suggest a large genetic contribution to dyslexia susceptibility, the number of candidate genes identified to date is still limited (Williams & O'Donovan 2006; Anthoni *et al.* 2007; Schumacher *et al.* 2007). The first aim of the present thesis was to identify further genes that confer risk for the development of dyslexia as categorically defined trait. Additionally, it was intended to decompose the dyslexia cognitive spectrum into its single components and to perform genetic analyses on these dyslexia-related, quantitative endophenotypes. This was expected to increase the chance to identify genetic variants that contribute to particular cognitive processes which might be disturbed in subgroups of dyslexia patients.

In a candidate gene approach, available literature and databases were to be screened for genes that have been suggested to contribute to (i) cognitive abilities in the general population or (ii) comorbid traits such as attention-deficiency / hyperactivity disorder. In addition, the well replicated dyslexia susceptibility locus DYX2 was to be further analyzed in order to explore the causal variants that could explain the associations found for the two candidate genes *DCDC2* and *KIAA0319*. Using Sequenom technology, the respective hypotheses were analyzed in the German dyslexia sample comprising about 400 parent-child trios. Findings from these studies have already been published (Ludwig *et al.* 2008a; Ludwig *et al.* 2008b; Ludwig *et al.* 2009a; Ludwig *et al.* 2009b).

In a second part, a first genome-wide association analysis (GWAS) for dyslexia was to be carried out in order to identify new genomic loci. Using Illumina's BeadArray technology, 200 German dyslexia cases were assessed on a genome-wide level, and the frequencies of genetic variants were compared to German controls. Subsequently, data from a large European consortium (NeuroDys) were to be integrated, in order to increase the power of the study. Additionally, within the dyslexic cases only, the allelic contribution to quantitative endophenotypes was analyzed. Associated markers from either of the two approaches were to be reanalyzed in independent sample sets and, if replicated, were to be followed-up to dissect their functional mechanism. Results from these studies have also been published (Roeske *et al.* 2009) or will be available soon (Ludwig *et al.* in preparation).

3. MATERIAL AND METHODS

3.1 Devices

Autoclaves - Systec D-150, Systec GmbH
- Varioklav® 135 S Dampfsterilisator, H+P Labortechnik GmbH

Cell culture device - Neubauer improved counting chamber, BRAND AG

Centrifuges - Megafuge 1.0 R; Biofuges *stratus/ fresco/ pico*, Heraeus GmbH
- Concentrator 5301, Eppendorf AG
- Vacuum centrifuge VR-mini, Thermo Fisher Scientific Inc.

Concentration measurement device
- NanoDrop® ND-1000 Spectrophotometer, Peqlab Biotechnologie GmbH

DNA storage system - 2D CYPHER™ system (1.2 ml tubes, cluster racks, SmartScan solo, SmartScan™ 96), Thermo Fisher Scientific Inc.

Drying chamber - T 20 P, Heraeus GmbH

Electrophoresis chambers - WIDE MINI-SUB® CELL GT, Bio-Rad Laboratories GmbH
- Sub-Cell Model 96, Bio-Rad Laboratories GmbH

Gel documentation - Gel Doc™ XR System, Bio-Rad Laboratories GmbH

Genotyping systems - BeadStation 500GX System, Illumina® Inc.
- Infinium Plus Whole-Genome Genotyping Starter-Manual Option Package, Illumina® Inc.
- MassARRAY™ Nanodispenser, SAMSUNG Techwin Co. Ltd. for Sequenom®
- MassARRAY™ Compact Analyzer, Bruker Daltonics Inc. for Sequenom®

Isolation of nucleic acids - Chemagen AG stand 2x12, Chemagen AG

Mix and stir devices - Thermomixer comfort, Eppendorf AG
- Vortex Genie 2, Scientific Industries Inc.
- ARE Heating Magnetic Stirrer, VELP® Scientifica Srl.

pH meter - inoLab pH 720, Wissenschaftlich-Technische Werkstätten GmbH

Pipettes - Research® variable pipette set (0.1-2.5 µl, 0.5-10 µl, 2.0-20 µl, 10-100 µl, 20-200 µl, 100-1000 µl, 500-5000 µl), Eppendorf GmbH
- Eight channels pipettes 0.5-10 µl, 10-100 µl, Eppendorf GmbH
- Transferpette®-8/-12 electronic, BRAND GmbH & Co.KG
- Finnpipette® 16 channels, VWR International GmbH
- Multipette® plus, Eppendorf AG
- Transferpette® electronic, BRAND GmbH & Co. KG

Pipette robot systems - Biomek® Laboratory Automation Workstations NX MC and NX S8G, Beckman Coulter GmbH

Pipette support - Cell Mate II, Matrix Technologies Corporation, Thermo Fisher Scientific Inc.

Power supply - PowerPac Power Supplies, Bio-Rad Laboratories GmbH

Scales - TE3102S / TE3135-DS, Sartorius AG

Sequencing device - 3130xl Genetic Analyzer, Applied Biosystem Deutschland GmbH

Shaking devices - REAX 2 / TITRAMAX 101 / UNIMAX 1010, all Heidolph Instruments GmbH & Co. KG

Sterile hood - HERA*safe*, Heraeus GmbH

Sterilization oven - UT 6000, Heraeus GmbH

Thermal cycler - PTC-200 and PTC-100, MJ Research Inc.
- ABI Prism®7900HT Fast real-Time PCR System, Applied Biosystems Deutschland GmbH (TaqMan)

Vacuum systems - MZ 2C Membran Vacuum Pump, Vacuubrand GmbH
- Heto Vacuum centrifuge, Thermo Fisher Scientific Inc.

3.2 Chemicals and reagents

If possible, high quality reagents and chemicals 'pro analysis' were used.

- Acetic acid ($C_2H_4O_2$), Merck KGaA
- Agarose low EEO (Agarose Standard), AppliChem GmbH
- AmpliTaq DNA polymerase [5 U/µl], Applied Biosystems Deutschland GmbH
- 5 x Big Dye Terminator 3.1 sequencing buffer, Applied Biosystems Deutschland GmbH
- BioTherm DNA polymerase [5 U/µl], GeneCraft®, Ares Bioscience GmbH
- Boric acid, Invitrogen Co.
- Bromphenol blue, Sigma-Aldrich Co.
- Diethyl pyrocarbonate ($C_6H_{10}O_5$) (DEPC), Sigma-Aldrich Co.
- Dimethyl sulfoxide (C_2H_6SO) (DMSO), Sigma-Aldrich Co.
- dNTPs [10 mM], GeneCraft®, Ares Bioscience GmbH
- Ethanol absolut (C_2H_5OH) (EtOH) (100%), AppliChem GmbH
- Ethidium bromide ($C_{21}H_{20}N_3Br$) (EtBr) (1%), Merck KGaA
- Fetal calf serum (FCS), Biochrom AG
- FicollTM 400, Amersham Pharmacia Biotech Inc.
- Formamide (CH_3NO), AppliChem GmbH
- Fungizone (250 µg/ml amphotericin B, 205 µg/ml sodium deoxycholate), Invitrogen Co.
- HotStar Taq DNA polymerase [5 U/µl], Qiagen GmbH
- Hydrochloric acid (HCl) (32%), Merck KGaA

- Isopropanole (C_3H_8O), AppliChem GmbH
- L-Glutamine [200 mM] (100 x), Invitrogen Co.
- Magnesium chloride ($MgCl_2$) [50 mM], GeneCraft®, Ares Bioscience GmbH
- Magnesium chloride ($MgCl_2$) [25 mM], Qiagen GmbH
- Magnesium sulfate ($MgSO_4$), Merck KGaA
- ß-Mercaptoethanol (C_2H_6OS) (ß-ME), Serva Electrophoresis GmbH
- Potassic hydrogen carbonate ($KHCO_3$), Merck KGaA
- PCR buffer containing $MgCl_2$ [15 mM] (10 x), Qiagen GmbH
- PCR buffer without $MgCl_2$ (10 x), GeneCraft®, Ares Bioscience GmbH
- Penicillin/Streptomycin (Pen/Strep) (100 x), Invitrogen Co.
- Phenol (C_6H_6O), Sigma-Aldrich Co.
- Phosphate buffered saline (PBS) (10 x), Invitrogen Co.
- 3130 POP7- polymer, Applied Biosystems Deutschland GmbH
- RPMI 1640 (2.0 g/l $NaHCO_3$, without L-Glutamine), Biochrom AG
- Sodium acetate (CH_3COONa), Merck KGaA
- Sodium chloride (NaCl), Merck KGaA
- Sodium hydroxide (NaOH), Merck KGaA
- Sodium hypochlorite (NaOCl) (13%), AppliChem GmbH
- Titriplex III ($C_{10}H_{14}N_2Na_2O_8 * 2 H_2O$) (EDTA), Merck KgaA
- Tris, ICN Biomedicals GmbH
- Tris-BASE ($NH_2C(CH_2OH)_3$) (Trizma), Sigma-Aldrich Co.
- Ultrasol F, MEDICENT GmbH

3.3 Solutions

Until mentioned otherwise, solutions, dilutions and buffers were prepared using H_2O dest.

Loading buffer (bromphenole blue buffer):

10 ml 10 x TBE, 10 ml 0.1% bromphenole blue, 40 ml 20% Ficoll, added up to 100 ml

EDTA [1 mM], 0.1% DEPC:

500 µl 0.5 M EDTA, added up to 250 ml, 250 µl DEPC

Ethanol (EtOH, 85%, 70%):

850 ml (700 ml), added up to 1 l with H_2O

Formamide [95%]/ 1 mM EDTA:

9.5 ml 100% formamide, 480 µl H_2O, 20 µl EDTA

Sodium acetate [3 M], pH 5.2:

123 g sodium acetate, added up to 500 ml with H_2O

Sodium hydroxide [0.1 M]:

1 g NaOH and 250 µl DEPC, added up to 250 ml with H_2O, shaking o.n. at 37°C

0.4% sodium hypochlorite:

15.4 ml NaOCl, 484.6 ml H_2O

10 x TBE buffer:

0.01 M EDTA pH 8.4 with NaOH [5 M], 1 M Tris, 0.9 M boric acid

Tris-Cl [10 mM], pH 8.0:

0.6 g Tris-BASE, 500 ml H_2O

Tris-EDTA (TE^{-4}):

0.1 mM EDTA, 10 mM Tris-Cl pH 8.0

3.4 Commercial systems

- AMPure Kit, Agencourt Bioscience Co.
- Big Dye terminator cycle sequencing kit v3.1, Applied Biosystems Deutschland GmbH
- Blood & Cell Culture DNA Kit, Qiagen GmbH
- Chemagic DNA Blood Kit, Chemagen AG
- CleanSEQ Kit, Agencourt Bioscience Corporation
- Custom TaqMan® Gene Expression Assay *"SLC2A3-1ex2"*, Applied Biosystems Deutschland GmbH
- DNA Ladder 100 bp and 1 kb, GeneCraft®, Ares Bioscience GmbH
- DNA Ladder 100 bp and 1 kb, New England Biolabs Inc.
- GenElute™ Agarose Spin Columns, Sigma-Aldrich Co.
- GFX PCR DNA and Gel Purification Kit, Amersham Biosciences GmbH
- Infinium-II Whole-Genome Genotyping Kit, Illumina® Inc., including BeadChips HumanHap300v1.1, HumanHap550v3 and HumanHap1M-DUO v3.0
- iPLEX™ Gold Reagent Kit, Sequenom® GmbH
- Micron YM-100 centrifugal units, Millipore Co.
- Multi tissue cDNA panels Human I and Fetal I, Clontech Laboratories Inc.
- Oragene™ DNA self collection kit, DNA Genotek Inc.
- Proteinase K, Qiagen GmbH
- QIAprep® Spin Miniprep Kit, Qiagen GmbH
- Ready Reaction Mix 3.1, Applied Biosystems Deutschland GmbH
- RNeasy Micro Kit, Qiagen GmbH
- RNase-free DNase Set, Qiagen GmbH
- SpectroCHIP® Arrays and Clean Resin Kit, Sequenom® GmbH
- Super Script III First-Strand Synthesis SuperMix, Invitrogen Co.
- TaqMan® Universal PCR Master Mix, No AmpErase® UNG, Applied Biosystems Deutschland GmbH
- TaqMan® Endogenous Control Assay Human Cyc (Cyclophilin, "4326316E"), Applied Biosystems Deutschland GmbH

3.5 Bioinformatic tools

3.5.1 Software

Applied Biosystems - File Builder 3.0
- GeneScan 3.1.2
- SDS 2.2.2
- 3130xl Data Collection v3.0

Beckman-Coulter - Biomek® Software 3.2
Biocomputing Platforms - BC/Gene v2.5.5
Bio-Rad - Quantity One®
Chip Bioinformatics Tools - http://snpper.chip.org/
Conor McCarthy - Chromas Lite Version 2.0
DNASTAR Inc. - SeqMan II Version 5.0
HapMap - HaploView 4.0
Illumina - BeadScan v3.1
- BeadStudio v3.2, including genotyping module 3.3.4

Paul Stothard (The Sequence Manipulation Site)
- http://bioinformatics.org/sms/index.html

Peqlab Biotechnologie - NanoDrop® ND-100 v3.3.0
Primer3 - http://frodo.wi.mit.edu/
Restriction mapping - http://www.restrictionmapper.org/
Sequenom - Assay Design 3.1
- Typer v3.4 and v4.0
- SpectroPoint
- RT-Workstation 3.3 and FLEXcontrol

Thermo Fisher Scientific Inc.
- ABgene 2D CYPHER™ Pilot Databases
- ABgene SmartScan 96
- ABgene SmartScan Solo

3.5.2 Databases

CEPH human genome diversity project
- ftp://ftp.cephb.fr/hgdp_supp1/

Database of genomic variants
- http://projects.tcag.ca/variation/

Ensembl - http://www.ensembl.org/index.html
GENEVAR database - http://www.sanger.ac.uk/humgen/genevar/
HapMap - http://www.hapmap.org/
Imprinting catalogue - http://igc.otago.ac.nz/home.html
mRNA-by-SNP browser - http://www.sph.umich.edu/csg/liang/asthma/
Myers Lab - http://labs.med.miami.edu/myers/LFuN/data.html
NCBI - http://www.ncbi.nlm.nih.gov

SwissProt - http://beta.uniprot.org/uniprot
UCSC Genome Bioinformatics
- http://genome.ucsc.edu/
Whole brain atlas - http://www.med.harvard.edu/AANLIB/home.html

3.6 Study probands

From 2000 to 2005, dyslexia probands were recruited at the Departments of Child and Adolescent Psychiatry and Psychotherapy at the Universities of Marburg and Würzburg. The resulting German sample of dyslexia probands (German DYS-sample) was used as basis for the present thesis. From 2006 to 2009, sample recruitment strategies were unified on a European level within the NeuroDys consortium (NeuroDys sample).

3.6.1 German dyslexia (DYS-) sample

Children considered as potential dyslexia probands were referred to the clinics by parents, teachers, special educators or health professionals on the basis of a prior diagnosis of dyslexia or observed difficulties in learning to read and to spell.

Diagnosis of dyslexia and inclusion criteria

The proband's spelling ability was used as the diagnosis criterion. Spelling was measured using an age-appropriate German spelling-test (writing to dictation; (Brähler *et al.* 2002)). The subject's intelligence quotient (IQ) was assessed with one of two 'Culture Fair Tests' (CFT-1 or CFT-20), depending on the age of the proband (Weiß & Osterland 1997; Weiß 1998). The two measures of 'spelling' and 'IQ' were then used to calculate an observed spelling score, based on an assumed correlation of 0.4 between the proband's IQ and spelling ability. Children were classified as 'affected' when there was a discrepancy of at least one standard deviation (sd) between the observed spelling score and the one expected based on the child's IQ and age (= spelling discrepancy score).

Families were excluded if the proband or a sibling showed symptoms of attention-deficiency / hyperactivity disorder (ADHD), according to a standardized clinical interview with the proband's mother (Unnewehr *et al.* 1998). Families were also excluded in cases where the proband had experienced a bilingual education or presented with an intelligence quotient < 85, an uncorrected disorder of peripheral hearing or vision, or a psychiatric or neurological disorder with a possible impact on the development of reading and spelling ability.

Written informed consent was obtained from all participating individuals or their parents if the proband was aged 12 years or younger.

In total, the German DYS-sample consisted of 400 probands and their parents, all of German origin. If available, also siblings of the affected children were included. In the case of two probands, only the mother participated in the study. Probands were recruited at an age between 8 and 19 (mean = 11.99, sd = 2.30). Based on the spelling discrepancy score, the sample was divided into four subgroups of different degrees of severity (SD). The exact composition of the German DYS-sample is presented in Tab. 3.

Tab. 3: The German DYS-sample.

	SD ≥ 1.0	SD ≥ 1.5	SD ≥ 2.0	SD ≥ 2.5
total	400	367	250	116
male	*289*	*273*	*188*	*93*
female	*111*	*94*	*62*	*23*

The total number of probands is given for the entire sample (SD ≥ 1.0) and the different severity groups. The respective numbers of male and female individuals are shown in italics. Data for siblings are not shown, as they are not part of the main German DYS-sample. SD - degree of severity.

Clinical assessment of related endophenotypes

Following the initial diagnosis of dyslexia and inclusion in the sample, probands were assessed using a wide battery of psychometric tests. The tests are based on the measures presented in Tab. 1. and they allow for discrimination between specific dyslexia-related endophenotypes. Additional tests targeted more abstract phenotypes (handedness, event-related potentials, arithmetic abilities) to extend the endophenotypic spectrum. Children were trained with some practice items prior to starting the test, in order to ensure that the respective task was fully understood. During the trials, no feedback was given. A more detailed description of the phenotypic measures and test batteries has been published (Schulte-Körne *et al.* 2001b).

Word reading and phonological decoding

The ability to fluently read words was assessed using the standardized "Salzburger Lese- und Rechtschreibtest" (Landerl *et al.* 1997) in cases where the probands attended school from 2nd to 4th grade. Children at or above 5th grade underwent a non-standardized reading test, which consisted of a list of 48 words (Schulte-Körne 2001). For both versions of the test, the number of words read correctly in one minute was used as quantitative measure.

A similar test set up was used to address the proband's ability of phonological decoding. In contrast to the word reading task, children were now presented a list of 48 pseudowords (non-words, e.g. "sesa", "mume"). Again, the quantitative measure was made up of the number of pseudowords that were read correctly in one minute.

Phonological awareness

Children from 2nd to 4th grade were administered three tests targeting (i) *phoneme deletion*, (ii) *phoneme segmentation* and (iii) *phoneme reversal*. Each of the tests was presented aurally, while the probands had to respond orally.

In the *phoneme deletion* test, children were asked to repeat the presented word without the first phoneme (e.g. 'Ball' without first phoneme /b/, n = 15). To assess *phoneme segmentation*, children had to split a pseudoword (n = 10) into its phonemes. The *phoneme reversal* test consisted of an exchange of phonemes: children were instructed to switch the first two phonemes of a known, real word (e.g. 'Leder' was to become 'elder', n = 15). For children attending grade 5 or above, the tasks were similar but contained more complex items, in order to avoid ceiling effects due to low task difficulty. For each of the two age groups (2nd to 4th grade / 5th grade and above), the outcomes of all three tests were combined. The average was used as measure of phonological awareness.

Orthographic processing

The ability of orthographic processing was measured by using real words (e.g. "Wachstum") and their respective pseudohomophone ("Waxtum"). Hereby, the children's ability to discriminate between correct and incorrect spelling was targeted.

Each word was presented to the children via headphones. Thereafter, the correctly spelled word or its pseudohomophone appeared on the screen, and subjects were asked to decide whether the word was orthographically correct or not (n = 35).

Rapid naming

The tasks assessing rapid naming were developed based on a previous study (Denckla & Rudel 1974). Four different trials were conducted, and each of them contained a different series of items which were presented printed on a sheet of paper. Children were asked to name them as quickly and accurately as possible:

1. Rapid naming of numbers: one-digit numbers (7, 2, 9, 6, 4)
2. Rapid naming of letters: single vowels or consonants (p, s, a, o, d)
3. Rapid naming of objects: line drawings of common objects, in color (scissors, candle, comb, key, clock)
4. Rapid naming of colors: circles of five different colors (red, green, brown, blue, black)

The respective raw scores of the naming of numbers and letters were used as single measures, as both of them are reflecting speed and fluency. The single raw scores of naming objects and colors were joined into a combined score. This measure represents a purer measure of simple naming speed as it is not influenced by letter- or number knowledge (Meyer *et al.* 1998).

Verbal short-term memory

To measure phonological short-term memory, a standardized digit span test was used (German adaptation of the WISC-R-test; (Tewes 1983)). Probands were read increasingly long series of numbers. The series had to be repeated in both directions, namely forwards and backwards. While the forward digit span reflects pure storage and recall in short-term memory, the backward digit span requires the processing of information and thus presents with additional aspects of working memory.

Basic mathematical abilities

The assessment of the proband's mathematical abilities was performed based on a previously published test battery (Landerl *et al.* 2004). The test consisted of three different trials that addressed (i) number comparison, (ii) addition and (iii) multiplication, respectively.
Number comparison (NC) was assessed as follows: Randomly arranged dots from one to nine were presented to the child on the left part of a computer screen, and a written number was shown on the right half of the screen. Children had to decide whether the number of dots corresponds to the written number or not (n = 32). To measure mental arithmetic abilities, simple additions and multiplications (n = 24 each) were presented in two separate blocks. Numbers included in the addition task reached from one to 19, single-digit numbers (two to nine) were used for multiplication. Calculations were presented on a computer screen such as "6 + 2 = 8", and children were asked to decide as quickly as possible and without making mistakes, whether the result of the addition / multiplication was correct or not. Results of the two subtests were combined, and the corresponding variable was termed 'exact calculation' (EC). Based on the measures for NC and EC, a previous principal component analysis had generated a combined arithmetical variable, the basic mathematical factor (BMF), which represents a general measure of mathematical skills (Schulte-Körne *et al.* 2007b).

Handedness

For each proband, the preferred use of the left or the right hand in common day practice was assessed using a self-report handedness questionnaire. It provided extensive information on different aspects of the preferred use of hands and was transformed into a quantitative measure (Schulte-Körne *et al.* 1998a). For all tests that required button press, children were asked to perform the tasks with the hand they had declared their preferred one in the questionnaire.

Electroencephalography (EEG) measurements

The neuropsychological functions of the probands were assessed and analyzed by experienced staff at each of the two clinical centers. All devices, materials and methods used for these data

are in place at the two sites. Therefore, the clinical assessment procedure will only be described in brief.

For each proband, an electroencephalography (EEG) was recorded during a passive oddball paradigm with synthetic syllables (1700 standards /da/ and 300 deviants /ba/). Stimulus length was 240 ms, with an inter-stimulus interval set at 740 ms. The EEG was recorded at 250 Hz sampling. Signals were averaged into epochs of 1100 ms, including a prestimulus baseline of 100 ms. An overview of the single electrode positions is given in Fig. 9.

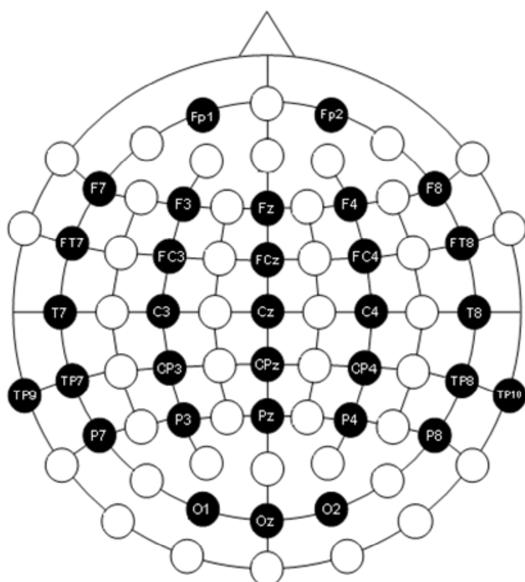


Fig. 9: Electrode arrangements as used for the probands. The head of a child is shown from the top-view, with the position of the nose indicated. Electrodes are arranged based on an extended version of the international 10/20 system (Jurcak *et al.* 2007). Apart from the 29 electrodes that were used for the main analysis, additional electrodes were attached below the subjects' right and left eyes, in order to measure ocular movements (electrooculogram), and at the right mastoid. The reference electrode was attached at the left mastoid.

Based on the EEG data, the mismatch negativity (MMN) was calculated as the difference between the averages of deviant and standard trials (see section 2.1.4). For the two components of the MMN, namely MMNa and MMNb (Fig. 2), the mean of the nine fronto-central electrodes F3, Fz, F4, FC3, FCz, FC4, C3, Cz and C4 was taken into further analysis.

3.6.2 Probands of the NeuroDys sample

NeuroDys criteria for sample collection were as follows: Children had to attend school in grade 3 or 4, resulting in an age of 8 to 12 years. Potential probands were either selected based on results from classroom-tests (e.g. Munich, Salzburg) or were referred to the clinics by practioners (e.g. Paris), but were excluded from participation in the study if they experienced a bilingual education or presented with symptoms of ADHD. Children had to perform a one-minute reading test whose result was used to classify the respective child as affected proband (reading score 1.25 sd below the mean of the age) or as unaffected control (reading score 0.85 above the mean). Regardless of being included in case or control group, subjects were further phenotypically characterized. The subsequent test batteries assessed the proband's spelling ability (exclusion of controls if ranked ≤ 20% of general mean), intelligence quotient (exclusion if

IQ ≤ 85) and most of the endophenotypes presented in section 3.6.1. An overview of the NeuroDys sample including sample sizes, country and the experiments in which single samples were used is given at a later stage, after description of the controls and the experimental setups (Tab. 5 , section 3.7.3).

3.6.3 Control individuals

The external control individuals used in this work have not been assessed with respect to reading and spelling performance (= population-based controls). Thus, it is assumed that about 5 - 12% of the control individuals also present with symptoms of dyslexia, according to the prevalence rate in the German population.

Heinz-Nixdorf recall study controls

The Heinz-Nixdorf recall (HNR) study has been performed at the University Hospital of Essen, and was initiated in order to ascertain a large longitudinal sample addressing population-based incidences of cardio-vascular diseases (http://www.recall-studie.uni-essen.de/). The entire sample comprised about 4,500 men and women aged 45 to 75 years. Probands included in this thesis were of German descent and have been randomly chosen among the entire HNR-cohort.

Munich Antidepressant Response Signature controls

The control sample collected within the Munich Antidepressant Response Signature (MARS) study (http://www.mpipsykl.mpg.de/research/) contained 550 healthy individuals, aged 18 to 75 years, who lived in the Munich area and were randomly chosen among all inhabitants officially registered in the city. Probands had passed a telephone interview to exclude known neurological, medical or psychiatric diagnoses. For this study, only probands with German ancestry were used.

Controls from the 'Nationales Genomforschungs Netz' (Germany)

In a project funded by the German 'Nationales Genomforschungs Netz' (NGFN), DNA samples were collected for genetic studies in psychiatric disorders. Besides probands affected with any of the psychiatric disorders of interest, also adult control samples were recruited in cooperation with the Central Institute of Mental Health in Mannheim. They were classified among the controls after they had passed extensive clinical examinations to exclude possible psychiatric disorders. Samples included in this thesis were randomly chosen among the about 1,000 individuals available.

Hair donor controls

For validation of expression findings, 37 adult control RNA samples derived from blood were used. They were taken from a sample of 121 male hair donor controls that have been recruited at the Institute of Human Genetics in Bonn for a study investigating hairless disorders. The choice of individuals was genotype-driven. Individuals were aged between 20 and 40 years, and had given 35 ml of blood that was used for isolation of both, DNA and RNA. As gene expression levels vary during day time, sampling had only been conducted between breakfast and lunch. An overview of the samples used in this thesis is given in Tab. 4.

Tab. 4: Control samples used in this thesis.

	HNR	MARS	NGFN	Hair donor
total	383	543	685	37
male	*224*	*246*	*286*	*37*
female	*159*	*297*	*399*	-

3.7 Protocols

3.7.1 Preparation of nucleic acids

The structural difference between double-stranded genomic DNA and single-stranded mRNA has fundamental impact on their respective stability. DNA is little susceptible to spontaneous degradation by catalytic hydrolysis and can be stored from -20°C up to room temperature (RT). Working with DNA does not require highly pure reagents, however, to avoid DNA contamination or degradation, pipette tips and reaction tubes should be autoclaved and free of DNases. In contrast, RNAs are easy targets for spontaneous degradation by RNAses, which appear ubiquitously, are very stable and renature quickly after autoclaving (Sela *et al.* 1957). Even a partial degradation of RNA molecules has profound impact on RNA experiments such as expression analysis. A continuous working on ice is required, and RNA samples have to be stored at -80°C. It is important to ensure that all devices, chemicals and the lab bench are free of RNAses. H_2O was either RNase-free as provided in the commercial kits, or has been individually prepared by adding 0.1% DEPC, shaking o.n. at 37°C and subsequent autoclaving. DEPC is known to be a strong inhibitor of RNases (Fedorcsak & Ehrenberg 1966).

Isolation methods

Isolation from proband samples

Probands participating in the genetic studies provided either a blood or a saliva sample for extraction of genomic DNA. As sample recruitment was performed over a period of several years, different techniques were applied to extract DNA. For the manual isolation of DNA from

peripheral blood, 10 ml of sample were used as input for the salting-out method (Miller *et al.* 1988). Its principle is a first lysis of the cells, a subsequent digestion of proteins by Proteinase K and a final precipitation of the DNA by ispropanol. For the automatic procedure, the 'chemagic DNA Blood Kit special' based on the Chemagen AG Stand 2x12 system was used. Here, DNA extraction is medicated by an interaction between magnetic particles and the negatively charged DNA molecules. The input consisted of 5 – 10 ml peripheral blood. While the manual isolation yielded about 100 µg genomic DNA, the automatic isolation method provided with 150 - 300 µg. For saliva samples, the Oragene™ DNA Self-Collection Kit was used. Two ml of spit saliva were required for an extraction of about 50 µg genomic DNA. Although the output of DNA isolated from blood is generally of better quality and higher yield than from saliva, the latter method is non-invasive and is therefore particularly suitable when working with children.

Isolation from Eppstein-Barr virus transformed cell lines

At the stage of recruitment of the German DYS-sample, blood serum was taken to generate lymphoblastoid cell lines. Immortalization was performed by transforming the cells with Eppstein-Barr virus (EBV), which is an effective procedure for inducing long-term growth of human B-lymphocytes. Cells were maintained in the presence of RPMI 1640 medium which was spiked with growth factors contained within the fetal calf serum. Penicillin, streptomycin and amphotericin B were added for protecting the cell culture from bacteria and fungi. The cell lines were cultivated in cell culture flasks at 37°C in the presence of 5% CO_2. Three times a week, the medium was changed or filled up, depending on the cell's nutrient uptake rate which was indicated by a color change of the medium.
Isolation of DNA was performed with the Blood & Cell Culture DNA Mini Kit (Qiagen), and RNA was extracted using the RNeasy Micro Kit (Qiagen). Both isolation methods were performed according to the protocols provided by the manufacturer. The input consisted of 5×10^6 cells for DNA, and of about 1.8×10^6 cells for RNA extraction, respectively. The number of cells was determined using a Neubauer counting chamber. To remove any possible genomic contamination that could interfere with the experiments, RNA was cleaned up using the protocols 'DNase Digestion of RNA before RNA Cleanup' and 'RNA Cleanup and Concentration' as described in the RNeasy Micro Kit manual.

Determination of concentration and quality

Quantity and quality for both, RNA and DNA, were determined based on photometrical (UV) measurements at wavelengths of 230, 260 and 280 nm. Measurements were performed using a NanoDrop ND-1000 spectrophotometer which, compared to other standard devices, presents with the advantage of a very low sample volume (1 µl) required for accurate measurements.

Using the software of the NanoDrop ND-1000 device, an easy switch between RNA and DNA measurement parameters is possible.

The maximal absorption wavelength for nucleic acids is 260 nm. Based on the Beer-Lambert law which predicts a linear correlation between absorbance and concentration, the optical density (OD) at 260 nm is used for calculation of sample concentration. For DNA, one unit (1.0) of OD corresponds to 50 ng/µl, while for RNA, one unit of OD is considered to be equivalent to 40 ng/µl. For analysis of quality, the samples' absorbance values at 230 nm and 280 nm were determined. Impurities such as phenol or aromatic compounds absorb at 230 nm, while proteins show their absorption maximum at 280 nm. Thus, calculation of the ratios at (i) 260 nm / 230 nm and (ii) 260 nm / 280 nm allow for an accurate prediction of the samples' purity. For high quality nucleic acid samples, the ratio at 260 nm / 280 nm should be around 1.8 for DNA and 2.1 for RNA samples.

Extracted DNA was subsequently prepared in two different working dilutions (100 ng/µl, 20 ng/µl) which were used for the standard experiments. Tubes containing 100 ng/µl were stored at -20°C, 20 ng/µl dilutions were stored at 4°C, in order not to freeze and thaw the samples too often. All working dilutions were prepared in 2 ml ABgene tubes and were managed using a 2D-barcode system. Remaining DNA was stored as 'stock'-DNA at -80°C.

Generation of DNA pools

A genome-wide DNA pooling setup was used for the analysis of three case-control combinations from Central Europe, the UK and Finland, respectively. The numbers of samples included in each pool are provided in Tab. 5.

For preparation of the Central European case and control pools, only suitable DNA samples were used (successful prior PCR amplification, no whole-genome amplified samples and high-quality spectrometric ratios). Samples were diluted in plates to 5 ng/µl and, after equilibration o.n., DNA concentrations were measured and adjusted to the target concentration. The individual DNA samples were then combined in two pools (case or control). Finally, pools had to be concentrated to 50 ng/µl, which was performed using Micron YM-100 centrifugal units.

The preparation of the DNA pools from Finland and the UK was performed in partner laboratories using similar protocols.

3.7.2 Processing of nucleic acids

Design of primers for DNA / cDNA amplification was performed either using the publically available online program Primer3 (primers for PCR, sequencing and expression analysis), or software tools provided by Sequenom® (primers for genotyping using MassExtend reaction).

For Primer3, the genomic DNA / cDNA sequence was retrieved from online databases (Ensembl, UCSC), and standard parameters were administered.

Primers were synthesized at Metabion (Martinsried, Germany) in standard quality. Primers for MassExtend reactions were additionally MALDI-ToF-checked. Primers and probes required for quantitative Real-Time PCR analysis (TaqMan assays) were designed and generated by Applied Biosystems (ABI, Darmstadt). All primers and their corresponding sequences are presented in Attachment I.

Polymerase chain reaction

The polymerase chain reaction (PCR) enables the exponential amplification of specific DNA segments that normally just appear in one or some few copies (Mullis & Faloona 1987). The basic principle of a PCR is the cyclic change of different temperatures. An initial step of 90 – 95°C (denaturation) is followed by a step of 45 - 65°C (annealing of primers, temperature is primer-dependent) and a final elongation step of 72°C (DNA synthesis, mediated by a thermally stable Taq polymerase). Depending on the structure of the primers, the length of the DNA region of interest and the experimental question, the number of cycles and the duration of each of the above mentioned steps can be varied in order to obtain a successful, specific amplification. As template for PCR reactions, genomic DNA as well as reversely transcribed cDNA can be used.

A standard PCR assay was performed in a total volume of 25 µl, as follows:

	Volume	Final concentration
PCR mix	15.3 µl H_2O dest. 2.5 µl 10 x PCR buffer with $MgCl_2$ [15 mM] 4.0 µl nucleotide mix [1.25 mM each] 1 µl F-primer [10 pmol/µl] 1 µl R-primer [10 pmol/µl] 0.2 µl Taq DNA polymerase [5 U/µl]	 1 x buffer with 1.5 mM $MgCl_2$ 0.2 mM of each dNTP 0.4 pmol/µl F-primer 0.4 pmol/µl R-primer 0.04 U/µl Taq DNA polymerase
DNA	1.0 µl ([20 ng/µl], for genomic DNA)	20 ng
Final volume	25.0 µl	

With exception of the template, all components of the PCR assay were mixed together in a mastermix. Twenty-four µl of the mix were distributed into 0.2 ml PCR reaction tubes, and 1.0 µl of nucleic acid template was added individually. After brief vortexing and a centrifugation step, the PCR reaction was carried out in a thermal cycler. A standard protocol for PCR reactions was as follows:

	5 min	95°C	Initial denaturation
35 cycles	30 sec	95°C	Denaturation
	30 sec	60°C	Annealing of primers
	1 min	72°C	Elongation of primers
	5 min	72°C	Final elongation

Attachment I provides information on all primers and, if applicable, particular modifications in the assay composition and / or the PCR protocol.

Agarose gel electrophoresis

PCR amplicons were verified on agarose gels. Agarose is a polysaccharide which, after solidifying, forms a three-dimensional network that allows for migration of DNA molecules. Briefly, if an anode is attached to the system, negatively charged PCR products migrate through the agarose pores based on the effect of molecular sieving. Thereby, the fragments show different migration speeds that depend on the following parameters: the applied voltage, the size of the DNA fragments and the concentration of agarose. To make the migrated products visible, ethidium bromide (EtBr) was added. During the migration process, EtBr incorporates into the DNA double helix and provides fluorescent shiny bands upon excitation with ultra-violet light (366 nm).

In general, gels of 1.5% agarose concentration were used for the analysis of 200 – 500 bp long fragments. In order to later make the PCR products visible, 6 µl of 1% EtBr were added. Five µl of PCR product were mixed with 1 µl of 6 x loading buffer, and applied to the wells of the solidified gel. 1 x TBE was used as running buffer. Two µl of a length standard, chosen among a 100 bp and a 1 kb ladder with respect to the size of the PCR product, were added in at least one slot to compare the position of the bands with the length standard. This allows for analysis of the size of the PCR product.

3.7.3 Genotyping of DNA samples

Genotyping based on BeadArray Technology (Illumina®)

Whole-genome genotyping was performed using the Illumina BeadArray genotyping system (see section 2.3.4). The Infinium-II Whole-Genome Genotyping Kits for single sample analysis on HumanHap300v1.1 and HumanHap550v3, respectively, were used for individual genotyping. The High Density Gemini assay (HumanHap1M-DUO, two samples per chip) was applied for genotyping of the DNA pools (see section 3.7.1).

The HumanHap300v1.1 was the first BeadChip for genome-wide analysis provided by Illumina. It contained 317,503 tSNPs that were spread over the entire human genome. In genomic regions containing genes, conserved elements and the area of the Major Histocompatibility Complex (chr. 6), the SNP density was increased. 7,300 of these SNPs were non-synonymous. The average intermarker distance on the HumanHap300v1.1 BeadChip was 9.27 kb. The HumanHap550v3 was introduced as second BeadChip on the market and provided a set of 561,466 tSNPs. Although the general HapMap-based chip design remained the same,

mitochondrial SNPs, SNPs on the Y-chromosome and SNPs within known CNV regions were added. The overlap between this new chip and the HumanHap300v1.1 was 307,795 SNPs. Given the increased number of markers on the HumanHap550v3, the average intermarker distance decreased to 5.3 kb.

In the following section, the protocol of an Illumina Infinium-II assay is described as applied for HumanHap300v1.1 and HumanHap550v3. For the HumanHap1M-DUO chips (HD Gemini assay), the protocol differed slightly. The respective modifications of the protocol are given in (*parentheses*). Both assays were conducted as described in the respective protocols by the manufacturer. Therefore, only a brief description of the most important steps is given. Prior to processing, all required reagents were mixed well and centrifuged briefly at 1,300 rpm. In parallel to each step, a tracking worksheet was filled in with information about each step, the lot numbers of the reagents and incubation times, in order to monitor possible processing errors.

<u>1. Whole genome amplification:</u>

- In a 96 well plate, 15 µl (*8 µl*) of DNA sample (50 ng/µl) are mixed with an equal amount of 0.1 M NaOH and incubated at RT for 10 min.
- 270 µl (135 µl) of MP1 (neutralization mix) and 300 µl (150 µl) of AMM (amplification mix) are added, the plate is sealed with a cap mat and inverted at least 10 times to mix contents, before pulse centrifugation.
- The mix is incubated in a hybridization oven for 20 to 24 hours at 37°C.

<u>2. Fragmentation and precipitation</u>

- The plate is removed from the oven and briefly centrifuged at 600 rpm.
- Each well containing sample is split into four (two) wells (150 µl per well).
- 50 µl of fragmentation mix (FRG) is added to each well, and the plate is sealed.
- The plate is vortexed for 1 min at 1,600 rpm and centrifuged at 600 rpm, followed by an incubation step for 1 h at 37°C.
- The plate is centrifuged (600 rpm), and 100 µl of precipitation agent (PA1) is added.
- After sealing, the plate is vortexed (1 min at 1,600 rpm) and centrifuged at 600 rpm.
- An incubation step for 5 min at 37°C is performed, followed by another centrifugation for 1 min at 600 rpm.
- To precipitate the fragmented DNA, 300 µl of isopropanol (100%) are added to each well. The plate is sealed with a cap mat, and inverted at least 10 times.
- The mix is incubated in the fridge (4°C) for 30 min, and then centrifuged at 4,000 rpm for 40 min.
- Immediately thereafter, the cap mat is removed. The supernatant is decanted by quickly inverting the plate. Blue pellets, containing the fragmented and precipitated DNA samples, remain stuck at the bottom of the wells.
- To entirely remove the solutions, the plate is tapped firmly on absorbing paper towels for 1 min. The inverted plate is dried at room temperature for 1 h.

3. Resuspension

- 45 µl of RA1 (resuspension, hybridization and wash solution) are pipetted into each well that contains sample.
- The plate is heat sealed with foil and incubated in the hybridization oven for 1 h at 48°C.
- The plate is vortexed at 1,800 rpm for 1 min and pulse centrifuged at 1,300 rpm.

4. Denaturation of samples and hybridization

- Samples are denatured for 20 min at 95°C.
- After centrifugation at 1,300 rpm, the heat-sealed foil is carefully removed.
- The four (*two*) wells containing the same sample are reunified in one well and the entire volume is concentrated at the bottom of each well by centrifugation for 1 min at 1,300 rpm.

The subsequent hybridization procedure differs slightly between Infinium-II and HD Gemini assay. Therefore, it is described in two separate parts:
a) Infinium-II assay:

Preparation of the BeadChips for subsequent hybridization:

- The DNA samples are hybridized to Illumina BeadChips via glass back plates. Prior to use, they have to be cleaned with 70% EtOH and completely dried.
- Two wash dishes are prepared: one is filled with 200 ml EtOH (100%), the second one with 200 ml PB1 (wash solution).
- The hybridization chambers (each providing space for up to four chips) are prepared, such as that 200 µl of PB2 is pipetted into the reservoirs.
- BeadChips are unpacked one after the other and submerged into 100% EtOH, using the respective wash dish and a chip holder.
- By moving up and down (10 x), chips are washed. This step is repeated twice after 5 and 10 min.
- The chip holder is transferred into the wash dish containing PB1. Again, after moving up and down, the washing is repeated twice after 2.5 and 5 min.
- BeadChips are assembled into 'Flow Through Chambers'. Therefore, the chips are dried by centriguation for 1 min at 1,300 rpm and assembled using spacers, glass back plates and two metal clamps.
- The 'Flow Through Chambers' are transferred into the 'Chamber Rack' on the TeFlow system.
- Just before loading the samples on the chips, the BeadChips are prepared by pipetting 150 µl 100% formamide into the 'Flow Through Chambers', followed by two rounds of 150 µl RA1.
- 160 µl of each DNA sample is now transferred onto the chip.
- After 1 min, the DNA has spread over the entire surface of the BeadChip. 'The Flow Through Chambers' are now removed from the 'Chamber Rack' and put into the prepared hybridization chambers.

b) HD Gemini assay:

- *BeadChips are placed into a 'Hyb Chamber insert'. The two sample sides are separated from one another via a foil that is coated onto the chip.*
- *84 µl of each DNA sample is transferred into the appropriate BeadChip inlet port.*
- *'Hyb Chamber inserts' containing the BeadChips are placed in the prepared hybridization chambers.*

Finally, for both assays, the chambers are well closed and incubated in the hybridization oven for 16 to 24 h at 48°C (o.n.). Hereby, the rocker is set at shaking level 5. As preparation for the next day, the coating agent XC4 has to be prepared by adding 330 ml 100% EtOH and intensive shaking.

Assembly of 'Flow Through Chambers' after hybridization (HD Gemini assay only):

- *BeadChips are removed from the hybridization oven and 'Hyb Chamber inserts'.*
- *The foil stuck to the chips is pulled off in a rapid motion. Immediately thereafter, the chip is submerged in a wash dish containing 200 ml of PB1.*
- *After all BeadChips have been treated likewise, the chip holder is moved up and down for 10 times.*
- *One after the other, the BeadChips are now assembled into Flow Through Chambers while they are submerged in PB1, using the alignment fixture. The assembly is performed according to the protocol for Infinium-II.*

5. Single-base extension, staining and signal amplification

A solution of 95% formamide/1 mM EDTA is freshly prepared, and all required downstream reagents are centrifuged for 5 min at 4,000 rpm. The Chamber Rack is preheated to 44°C.

- The hybridization chambers are removed from the oven (for Infinium-II assay only), and the 'Flow Through Chambers' are placed into the 'Chamber Rack'.
- 150 µl of RA1 is pipetted on each 'Flow Through Chamber'. This step is repeated five times.
- With inter-incubation times of 10 minutes, XC1 and XC2 (450 µl each) are added one after the other.
- 200 µl of the two color extension mix (TEM) are pipetted to each 'Flow Through Chamber', followed by an incubation step for 15 min.
- 450 µl of 95% formamide/1 mM EDTA are added. After 1 min, the step is repeated and the mix is further incubated (5 min).
- The 'Chamber Rack' is cooled down to the temperature marked on the LTM (labeling two color mastermix) tube.
- Meanwhile, 450 µl of XC3 are added to the 'Flow Through Chambers'. This step also is repeated, and the last incubation is left until the Rack reaches its set temperature.
- 250 µl of LTM are added and incubated for 10 min.
- 450 µl of XC3 are added, followed by an incubation for 1 min
- Again, 450 µl of XC3 are added and incubated for 5 min.

Subsequently, a sandwich-like procedure of signal labeling and amplification is performed two times. Each step consists of:

- Addition of 450 µl of anti-stain two color mastermix (ATM), incubation for 10 min.
- Addition of 450 µl of XC3, incubation for 1 min.
- Addition of 450 µl of XC3, incubation for 5 min.
- Addition of 250 µl of LTM and incubation for 10 min.
- Addition of 450 µl of XC3, incubation for 1 min.
- Addition of 450 µl of XC3, incubation for 5 min.

In a next step, the BeadChips are prepared for the scanning procedure. Herefore, a wash dish is filled with 310 ml PB1.

- 'Flow Through Chambers' are removed from the Chamber Racks and disassembled.
- The BeadChips are placed into the chip holder and submerged into PB1. After moving up and down for at least 10 times, the chips are incubated in PB1 for 5 min.
- A second wash dish is filled with 330 ml of XC4. The chip holder with the chips is transferred into this dish and moved up and down at least 10 times.
- After incubation for 5 min, the chip holder is smoothly removed from the dish and placed horizontally on the work bench, with the hybridized surface of the BeadChip on top.
- BeadChips are placed on a rack and dried under vacuum for 50 – 55 min.

6. Imaging of BeadChips and data analysis

- BeadChips are introduced into the BeadArray reader and scanned for about 3 hours using the BeadScan software.
- Hereby, the system assigns the corresponding decoding data to each chip via its barcode.
- For analysis, the raw data are used as input for the BeadStudio software, which transforms them into genotypes according to the manifest files of the respective chip type.

In total, 200 German dyslexia cases were individually genotyped on HumanHap300v1.1. Two sets of controls (383 HNR controls, 543 MARS controls) were analyzed using the HumanHap550v3, and the overlapping SNP content (n = 307,795 SNPs) was statistically tested for association. In a second, independent step, DNA pools of dyslexia cases and controls were genotyped using the Illumina HumanHap1M-DUO.

Genotyping based on MassExtend Reaction (Sequenom®)

The Sequenom iPLEX Gold assay together with MassExtend reaction and MALDI-ToF technology was used for genotyping of a limited set of markers in a large number of individuals. For selected SNPs (under use of their rs-number), genomic data were exported from databases using the SNPper-application (CHIP bioinformatics). These data included flanking sequences, chromosomal positions and genetic locations. The orientation of the flanking sequence was checked in terms of identical orientation as provided in the official database dbSNP. In case of discrepancy, the sequence was converted into its reverse complement using the online 'Sequence Manipulation Site'. Next, the final data set was processed using the online tool PreXTEND, which is provided by Sequenom's RealSNP browser. This program validates SNPs and their flanks in terms of uniqueness in the genome, and designs primers for amplicons of 80 bp to 120 bp length. The PreXTEND output file was used as input for the software Assay Design 3.1, which arranges the primers into multiplex reactions (maximal number of SNPs per plex = 40) and designs the corresponding oligonucleotides (standard and UEP primers). For assay design, standard parameters of the

SBE-method were applied. Information on all SNPs, plexes and primers used in the experiments are given in Attachment I.

Individual DNAs were arranged in 4 x 96-well format using the ABgene system. The samples were diluted to 10 ng/µl, using the Beckman NX-S8G robot. For quality control, three controls (negative, positive and double DNA samples) were added on each plate. The dilution plates were combined into one 384-well plate by transferring 1 µl of each sample (= 10 ng) with the Beckman NX-MC. After centrifugation, the DNA was dried o.n. at RT.

In each plex, the standard primers were diluted to a final concentration of 500 nM each. For the UEP primers, mass groups were generated using the primer adjustment tool of the software Typer 3.4. Based on these results, the respective amount of each primer within a group was added to the primer mix (7 µM, 9.3 µM, 11.6 µM and 14 µM). Primer groups are also given in Attachment I.

In the following sections, the protocol for a representative Sequenom® iPLEX Gold reaction is described for one 384-well plate. As pipette robot systems were used, an overhang was included for preparation of each mix. If more plates were used, the volumes were adjusted accordingly. All reagents were mixed and centrifuged prior to their use.

Protocol for a standard PCR mix:

	Volume per reaction (1x)	**Final volume per plate (480x)**
PCR mix	2.7 µl H_2O dest. 0.63 µl 10 x PCR buffer Qiagen, $MgCl_2$ [15 mM] 0.25 µl dNTP mix Nextec [10 mM each] 0.325 µl $MgCl_2$ [25 mM] 1.0 µl PCR-primer mix [500 nM each] 0.1 µl Taq DNA polymerase [5 U/µl]	1,296 µl H_2O dest. 300 µl 10 x PCR buffer Qiagen, $MgCl_2$ [15 mM] 120 µl dNTP mix Nextec [10 mM each] 156 µl $MgCl_2$ [25 mM] 480 µl PCR-Primer mix [500 nM each] 48 µl Taq DNA polymerase [5 U/µl]
DNA	10 ng (dried)	10 ng per well
Final volume	5.0 µl	2,400 µl

Twenty-five µl of the PCR mix were manually pipetted into each well of a 96-well plate. Using the Biomex NC-MX, 5 µl of the standard PCR mix were then transferred to each well of the 384-well plate containing the dried DNA samples. After sealing with an adhesive foil, the plate was vortexed and briefly centrifuged. The plate was put in a thermal cycler, and the following cycling parameters were used:

	15 min	95°C	Initial denaturation
45 cycles	20 sec 30 sec 1 min	95°C 56°C 72°C	Denaturation Annealing of primers Elongation of primers
	3 min	72°C	Final elongation

After the PCR was finished, three samples were checked on an agarose gel. Residual dNTPs had to be neutralized by dephosphorylation so that they could not be incorporated during the

following reactions. For this purpose, a digestion step using shrimp alcaline phosphatase (SAP) was performed.

Protocol for SAP digestion:

	Volume per reaction (1x)	Final volume per plate (480x)
SAP mix	1.53 µl H_2O dest. 0.3 µl SAP enzyme 0.17 µl 10 x SAP buffer	734.4 µl H_2O dest. 144 µl SAP enzyme 81.6 µl 10 x SAP buffer
Final volume	2.0 µl	960 µl

For distribution of the SAP mix into the plate containing the standard PCR mix, the Biomek NX-MC was used. Ten µl of SAP mix were first manually pipetted into each well of a 96-well plate, and the robot system dispensed 2 µl into each well of the 384-well plate. The reaction protocol for SAP digestion was as follows:

40 min	37°C	SAP digestion
5 min	85°C	SAP inactivation

In the subsequent extension PCR reaction, the UEP primers were elongated for one single base and thereby generated allele-specific analyte molecules.

Protocol for extension reaction:

	Volume per reaction (1x)	Final volume per plate (490x)
UEP mix	0.619 µl H_2O dest. 0.2 µl 10 x iPLEX Gold buffer 0.2 µl Termination mix 0.94 µl UEP primer mix 0.041 µl iPLEX Gold enzyme (sequenase)	303.31 µl H_2O dest. 98 µl 10 x iPLEX Gold buffer 98 µl Termination mix 460.6 µl UEP primer mix 20.09 µl iPLEX Gold enzyme (sequenase)
DNA	-	-
Final volume	2.0 µl	980 µl

The distribution of 2 µl of the reaction mix into each well of the 384-well plate was performed similar to the distribution of the SAP mix. Cycler conditions were as follows:

<table>
<tr><td colspan="2"></td><td>30 sec</td><td>94°C</td><td>Initial denaturation</td></tr>
<tr><td rowspan="3">45 cycles</td><td>1 x</td><td>5 sec</td><td>94°C</td><td>Denaturation</td></tr>
<tr><td rowspan="2">5 cycles</td><td>5 sec</td><td>52°C</td><td>Primer annealing</td></tr>
<tr><td>5 sec</td><td>80°C</td><td>Primer extension</td></tr>
<tr><td colspan="2"></td><td>3 min</td><td>72°C</td><td>Final extension</td></tr>
</table>

In a next step, the reaction mix was purified using an ion exchange resin. For this purpose, 16 µl of H_2O and 6 mg of dried resin had to be given to each well of the plate. This was performed using a separate 384-well plate into which the resin was added. The 384-well plate containing the analyte mix was inverted and placed on the resin plate. Thereafter, turning both plates resulted in the resin falling out of the matrix and into the analyte mix. After sealing, the plate was mounted into a plate inverting device, and was continuously shaked for 5 min. Finally, the plate was centrifuged for 5 min at 4,000 rpm. The UEP analytes were then spotted on a

SpectroCHIP (Sequenom), using the Nanodispenser and the SpectroPoint software. The chip was entered into the Sequenom analyzer and measured using the MassArray RT workstation software 3.3. Data were exported into Typer 3.4 / 4.0, respectively, and automatically generated genotype calls were manually checked.

In this thesis, genotyping by MassExtend reaction was used for individual genotyping in the replication steps and for validation of DNA pools. For the latter application, data analysis differed slightly. Here, the ratios of the area below the allele-specific peaks are used as estimator for allele frequencies (Bansal *et al.* 2002; Shifman *et al.* 2008).

Tab. 5 summarizes the use of the different technologies throughout this work. It also provides information on the single samples that were analyzed in each of the steps.

Tab. 5: Samples and technologies included in the different stages of the GWAS.

Country	Initial GWAS[1]		Replication study[2]		Pooling analysis[1]		Replication after pooling[2]	
	cases	controls	cases	controls	cases	controls	cases	controls
Austria	-	-	170	147	116	181	186	208
Germany NeuroDys	-	-	108	194	104	188	207	214
Germany DYS-sample	200	383 (HNR) 543 (MARS)	200	685 (NGFN)	196	400 (NGFN)	200	685 (NGFN)
Switzerland	-	-	*26*	*43*	*25*	*40*	*27*	*44*
Hungary	-	-	*5*	*39*	-	-	*78*	*154*
Netherlands	-	-	115	106	100	103	153	174
France	-	-	*92*	-	-	-	161	204*
UK Cardiff	385	1406	209	268	426	219	187	219
UK Oxford			328	288			327	359
Finland	-	-	156	189	286	321	-	-
Sum	**585**	**2326**	**1409**	**1959**	**1253**	**1052**	**1526**	**2261**

All individuals in subsequent stages are part of the previous ones (except for pooling, where DNA quality is crucial and samples had been excluded). *In italics:* samples with less than 100 individuals in one group. [1] - Illumina technology, [2] - Sequenom technology. GWAS – genome-wide association study (individual genotyping), * - Controls are a mixture of both, population-based and non-dyslexic individuals.

3.7.4 Sequencing of DNA fragments

Sequencing allows for precise determination of the successive order of nucleotides. Possible applications include e.g. the confirmation of the correct identity of an amplified DNA fragment, the confirmation of ambiguous genotypes for particular individuals or the identification of novel mutations by sequencing coding genomic regions. Using a successfully amplified DNA fragment as template, sequencing was performed as follows, in four consecutive steps.

1. AmPure purification after PCR

A PCR product was first purified from residual components in the PCR reaction, using the AmPure Kit (Agencourt). It is based on the „Solid Phase Reversible Immobilization“ (SPRI™) technology (DeAngelis *et al.* 1995) where, under particular buffer conditions, DNA binds to paramagnetic beads and can be separated from other molecules in the solution via attachment

to a magnetic plate. After some washing steps, the addition of a solvent allows the DNA fragments to detach from the beads.

Protocol for AmPure purification:

1. Per volume of the PCR reaction, 1.8 x volume of AmPure buffer, containing the magnetic beads in solution, is added to the PCR reaction.
2. Mixing by pipetting up and down (10 times).
3. The mixture is placed onto the magnetic plate, for 5 – 10 min.
4. The supernatant is discarded.
5. 200 µl of 70% EtOH is pipetted into the reaction tubes.
6. After incubation for 30 sec, the supernatant is removed and discarded. Steps 5 and 6 are to be repeated for a second time.
7. The beads, with the purified PCR products attached, are air-dried for 20 – 30 min.
8. 40 µl of TE^{-4} are added, and PCR fragments are dissolved by mixing up and down.

2. Cycle sequencing reaction

The purified PCR products were next used as template for the subsequent cycle sequencing reaction, in which dideoxynucleotides (ddNTPs) are incorporated into an extending PCR fragment according to the principle implemented by Sanger (Sanger *et al.* 1977). This technique is based on the PCR principle, with the exception that the extension is performed in only one direction (forward F or reverse R). The reaction mixture consists of normal dNTPs and fluorescently labeled ddNTPs which compete for incorporation in the elongating strand. As soon as a ddNTP is incorporated, the fragment is not further extended, and a fluorescently labeled PCR fragment of particular size is obtained. Hereby, the endstanding color label is complementary to the particular position within the DNA fragment. As thousands of fragments are formed during the cycle sequencing reaction, it is statistically ensured that ddNTPs are incorporated at different positions in different fragments, resulting in a mixture of fluorescently labeled fragments of different size. The set up of a standard cycle sequencing reaction was as follows:

	Volume per reaction
Cycle sequencing mix	13.75 µl H_2O dest. 3.75 µl 5 x Big Dye Terminator v3.1 sequencing buffer 0.5 µl Big Dye - ready reaction mix v3.1 1.0 µl Primer F (or R) [3.2 pmol/µl]
Template	1.0 µl Purified PCR product [approx. 10 ng PCR product]
Final volume	20.0 µl

	1 min	96°C	Initial denaturation
25 cycles	10 sec 5 sec 4 min	96°C 50°C 60°C	Denaturation Annealing of primers Elongation of primers

3. CleanSEQ purification after cycle sequencing

Prior to sequencing analysis, the cycle sequencing products were purified in order to remove impurities that interfere with the fluorescent signal during capillary electrophoresis. Similar to the AmPure purification previously described, a purification system based on magnetic beads (CleanSEQ) was used.

Protocol for CleanSEQ purification:

1. 10 µl of CleanSEQ (with resuspended magnetic beads) are added to each reaction.
2. 62 µl of 85% EtOH are added, and the reaction is mixed by pipetting up and down.
3. The reaction tubes are placed onto the magnetic plate.
4. After incubation for 3 min, the supernatant is removed and discarded.
5. 100 µl of 85% EtOH are added to the beads, and the mixture is incubated for 30 sec.
6. The supernatant of the washing step is removed and discarded.
7. The reaction is air-dried for 10 min.
8. 40 µl of H_2O are added to the wells.
9. In order to fully resuspend the fragments, the mixture is incubated for 5 min.

4. Capillary electrophoresis and data analysis

The resuspended cycle sequencing product was subjected to sequencing electrophoresis using the automatic 16-capillary sequencing device 3130xl (Applied Biosystems). A standard capillary length of 36 cm allows for the separation of DNA fragments up to 700 bp in length. Ten µl of purified product were used as template, and the fragments contained in the mix were electrophoretically separated. At the end of the capillary, a laser beam excitates the fluorescent molecules, and a detector registers the fluorescent signals of subsequently arriving fragments. Thereby, an electropherogram is generated. The results of the sequencing electrophoresis were analyzed using the software "Chromas Lite 2.0" and "SeqMan II".

3.7.5 Expression analysis

Functional assays were conducted either using commercial cDNA panels of multiple human tissues, or based on RNAs that have been extracted from EBV-transformed cell lines or were available from blood donor controls. While cDNA panels can be used directly, RNA has to be first transcribed into cDNA.

Reverse transcription from mRNA to cDNA

Total RNA from cell culture or lymphocytes was reversely described into cDNA using 'Super Script III First-Strand Synthesis SuperMix (Invitrogen)', according to the protocol provided by the manufacturer. In principle, oligo(dT)-molecules that bind specifically to the poly(A)-tail of mRNAs are used as anchor primer, and a retroviral reverse transcriptase uses single-stranded RNA as template to synthesize a complementary DNA (cDNA) strand. RNA samples that

showed an insufficient concentration for reverse transcription were up-concentrated using a centrifugation step under vacuum.

Protocol:

1. In order to denature secondary structures of RNA, a premix is generated as follows, and heated to 65°C:

	Volume per reaction
Premix 1	1.0 µl Oligo(dT) [50 µM] 1.0 µl dNTP-mix [10 mM] filled up to 10 µl using H_2O [RNase-free]
mRNA	1,000 ng
Final volume	10.0 µl

2. The reaction mix is incubated on ice for 1 min.
3. The premix is filled up with the following reagents:

	Volume per reaction
Premix 2	2.0 µl 10 x RT-buffer 4.0 µl $MgCl_2$ [25 mM] 2.0 µl DTT [0.1 mM] 1.0 µl RNase OUT [40 U/µl] 1.0 µl SuperScript™ III RT enzyme [200 U/µl]
Premix 1	10 µl
Final volume	20.0 µl

4. For cDNA synthesis, the reaction mix is placed at 50°C for 50 min.
5. Incubating the mix at 85°C takes place for 5 min, in order to stop cDNA synthesis.
6. 1.0 µl RNase H is added, for degradation of residual RNA.

Finally, cDNA is stored at -20°C or, if used immediately, can be kept on ice.

Qualitative expression analysis in cDNA panels

To assess whether a gene transcript is present in a tissue of interest, commercially available multiple tissue cDNA panels of both, fetal and adult human tissues, have been analyzed. Here, analysis of cDNA panels has been applied for expression analysis of two different transcripts of *MYO18B*. The analysis of cDNA panels using PCR only provides semi-quantitative information on whether the gene is expressed in a particular tissue or not.

Quantitative Real-Time PCR analysis

To address relative expression levels, a quantitative PCR was performed as Real-Time PCR, using the TaqMan device (Applied Biosystems). Such a quantitative method is considered to be a highly sensitive approach to detect transcripts and to estimate their total abundance in the sample of interest (Livak *et al.* 1995; Nolan *et al.* 2006). After each cycle of a specific PCR, the number of amplified products is registered ('real time'). By comparing the amplification rate of a ubiquitously expressed, known reference gene with the one of the transcript of interest, one can draw conclusions on the present amount of template (cDNA). Hereby, the choice of the reference gene is of crucial importance (Bustin 2002).

The quantification is enabled by a hybridization probe that is specifically designed for the transcript of interest. For multi-gene exons, this probe is designed over exon boundaries to ensure selective analysis of cDNA. At its 5′ end, the hybridization probe is coupled to a reporter fluorescent dye, and the 3′ end carries a non-fluorescent quencher. Based on these molecules, the assay works according to the fluorescence-resonance-energy-transfer (FRET) technology (Cardullo *et al.* 1988), whose principle is explained in Figure 10: Schematic representation of the FRET method. .

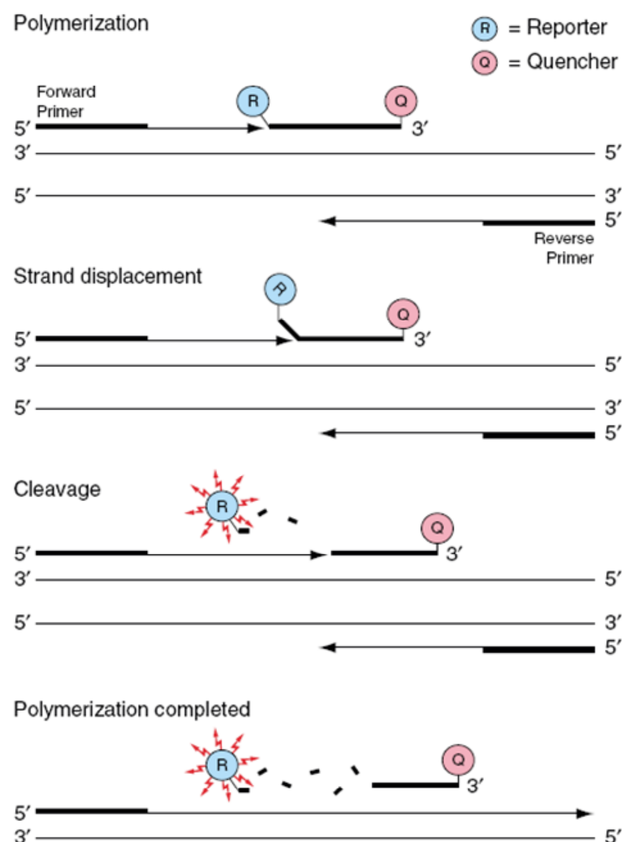


Figure 10: Schematic representation of the FRET method. Hybridization of the probe occurs during the annealing step of each cycle. As the probe has a higher melting temperature compared to both primers, this allows an earlier binding. After excitation, the reporter emits light of a specific wave length. As long as both molecules are located close to one another, the light emission is transferred on the non-fluorescent quencher that absorbs the energy. During the extension step of the PCR, the Taq polymerase, by using its 5′-3′ exonuclease activity, fragments the probe. Reporter and quencher are separated, and the light emitted by the reporter can be detected by the TaqMan device. The intensity of the fluorescence signal is proportional to the number of cDNA-strands. Taken from Applied Biosystems TaqMan Universal PCR Master Mix Protocol.

For quantification of the amount of template, the cycle threshold (ct-) value is calculated. This value represents the specific cycle, in which the amplification of the target fragment changes from steady into exponential phase. The ct-value is automatically determined via a threshold which is the same for all samples.

The PCR reaction assay was mixed together as follows:

	Volume per reaction
PCR mix	10.0 µl 2 x TaqMan Universal PCR Master Mix 1.0 µl 20 x Assay mix (specific for target transcript) 1.0 µl 20 x Assay mix (Endogenous Control Cyc, “4326316E”) add H_2O dest. to 20 µl
cDNA template	1.0 µl of cDNA
Final volume	20.0 µl

An exact pipetting is a crucial requirement for this experiment, as smallest deviations or contaminations might lead to non-reproducible, wrong results (Ding & Cantor 2003). For the analysis, the used standard amplification program had the following settings:

	10 min	95°C	Denaturation
40 cycles	15 sec	95°C	Denaturation
	1 min	60°C	Annealing of primers and extension reaction

In the present thesis, functional analysis using TaqMan was performed for the determination of allele-specific expression of *SLC2A3*. The corresponding sequence of the RNA and positions of the exon-spanning primers are given in Attachment II. Three independent RT reactions were performed, with each of them yielding 21 µl. The reference gene *cyclophilin* (TaqMan assay 4326316E) was measured in the same reaction. For the assay, 1 µl of cDNA was used as template. Measurements were performed in 384-well format, in triplicates for each sample.

3.8 Statistical analysis

Large-scale genotyping as performed here requires robust quality control and subsequent statistical analysis of the results, in order to provide information on the genetic correlation between the tested markers and the trait of interest. Statistical analyses presented within this thesis have been performed in cooperation with the Department of Statistical Genetics at the Max-Planck-Institute of Psychiatry in Munich.

3.8.1 General concepts

Quality control

Quality control (QC) refers to the measures applied to ensure high quality datasets. They include (i) technical controls (e.g re-sequencing of individuals for given alleles, comparison of double samples) and (ii) bioinformatic approaches, which address genotyping biases such as wrong clustering or genotyping errors due to low sample quality. The use of strict QC criteria helps to generate clean, powerful datasets and decreases the risk of false-positive results.

Call rates

For each sample and each SNP that is included in a study, call rates are calculated in order to assess the quality of the genotyping and / or the samples. One distinguishes between a call rate per sample (CR_{sample}) and a call rate per SNP (CR_{SNP}). Both types of call rates provide a tool to estimate the quality of an assay or a sample. In the presented studies, a call rate of 98% was set as inclusion threshold.

Hardy-Weinberg-equilibrium

In an ideal population, frequencies of alleles and genotypes remain constant over thousands of generations. This concept is described as Hardy-Weinberg-equilibrium (HWE; (Hardy 1908; Weinberg 1908)). Although such an ideal population does not exist due to evolution processes and the genomic haplotype structure, the HWE can be used to detect allele-specific imbalances in analyzed samples. A large deviation from HWE can provide a hint on possible genotyping errors, but can also point towards a biological process of interest. Therefore, HWE was only assessed in control samples. Markers were excluded from statistical analysis if they presented with a strong deviation from HWE ($P \leq 0.00001$).

Minor allele frequency

Only SNPs with a minor allele frequency (MAF) higher than 5% were analyzed. Lower MAFs were not considered due to power issues: To identify a genetic effect mediated by a SNP with a low allele frequency, a very large number of samples would be required to obtain sufficient information on allelic and phenotypic correlation in cases and controls.

Mendelian errors

In family-based data sets, the analyses of Mendelian transmission patterns provide an additional tool to check for allelic inconsistencies that are derived from genotyping errors or from wrong parent-child trios. For all SNPs that passed the QC measures above, Mendelian inheritance was checked. In case of genotype inconsistencies, the respective calls were either zeroed (if this inconsistency only occurred once per SNP or parent-child trio), or the SNP / family was completely taken out of the data set in case of several inconsistencies.

Correction for multiple testing

In large data sets, some of the markers will show statistically significant results simply by chance. Such results represent false-positives and introduce interpretation errors in studies. Therefore, the number of statistical hypotheses has to be taken into account (Balding *et al.* 2007), a procedure known as ‘correction for multiple testing’. The most conservative method is the correction according to Bonferroni, who postulated that in the presence of n independent tests (hypotheses), the nominal significance level should be adjusted accordingly (Bonferroni 1937). That is, for $P = 0.05$, the significance level for n tests would then be $P_{Bonf} = 0.05/n$.

Alternative concepts for adequate correction have been suggested (Balding 2006; Dudbridge & Gusnanto 2008; Rice *et al.* 2008), e.g. that the interconnection of markers (e.g. SNPs in high LD) and traits (e.g. closely related endophenotypes) should be taken into account (Westfall & Young 1993; Nyholt 2004). Notably, the level of significance is also influenced by the fact whether there

is a prior statistical hypothesis or not. In a hypothesis-free study (e.g. GWAS), *P*-values are calculated two-sided. In contrast, replication studies are hypothesis-driven, thus those *P*-values are calculated one-sided.

3.8.2 Association studies of qualitative traits

Case-control studies

Case-control studies are performed either on genome-wide level, or in restricted data sets (e.g. replication experiments, candidate gene studies). The main difference hereby is the number of markers that are investigated. In both approaches, allele or genotype frequencies are compared between affected and unaffected individuals using different genetic models. Given two alleles A and B, the 'allelic model' compares the differences in allele frequencies, i.e. A vs. B. The 'genotypic model' refers to the comparison of AA vs. AB vs. BB. Also recessive / dominant mechanisms are possible in biological traits. These carrier-models compare frequencies between AA/AB and BB (carrier-A) or AA vs. AB/BB (carrier-B; (Ziegler & König 2006)).

For the analysis of the German GWAS case-control data, the genotypic model as implemented in the Armitage-trend test (ATT) was applied. *P*-values were combined with corresponding UK data in a meta-analysis, using logistic regression. This method allows for correlation of binary traits (e.g. case-control) with predictors (e.g. genotypes) under consideration of covariates, such as country of origin. The inclusion of covariates provides the possibility to correct *P*-values for population stratification or country-specific effects. Apart from country of origin, also the first four axes of variation of a multi-dimensional scaling analysis, which models ancestry differences explicitly, were included as covariates.

Similar to logistic regression, the Cochran-Mantel-Haenszel (CMH) test also allows for the inclusion of covariates. However, this test provides the opportunity to calculate the genetic effects for the single data sets alone and, subsequently, combines these results. The CMH is therefore particularly suitable for the combined analysis of different samples, and has been used for the analysis of our replication data.

Pooled case-control samples, which were run in at least four replicates each, were analyzed based on normalized green and red fluorescence data. For each SNP, the ratio between green and red signal intensity indicates the ratio between the two alleles. Mean allele frequency estimates (AFEs) were calculated for each pool, by combining the results from the single replicates. For each of the analyzed SNPs, AFE values in the control pools were compared to HapMap data (Kirov *et al.* 2006), and SNPs showing the 10% worst correlations (i.e. 5% in each direction) were excluded. SNPs were further excluded if they presented with a MAF ≤ 0.05 or

varied extremely between the replicates (coefficient of variation > 0.5). AFEs were compared between case and control pools for each country separately, as previously suggested (Sham *et al.* 2002), and *P*-values of pools from the distinct populations were finally combined using Fisher's Product method. Top SNPs were identified based on an association (or the trend of it) that was in the same allelic direction in all three populations. Additionally, if the SNP was also available in the initial GWAS, it was also required to show the same direction in this individual analysis.

<u>Odds ratio and relative risk</u>

The comparison of allele frequencies between cases and controls permits an estimation of the effect range which a given marker contributes to disease development. This estimation can be given as odds ratio or as relative risk. The odds ratio indicates whether in the presence of a given allele, the disease appears more often (OR > 1) or less often (OR < 1) than predicted by the general prevalence rate. Thus, the odds ratio is a measure to demonstrate the contribution of an allele to the disease of interest (Ziegler & König 2006) and is generally referred to the risk allele. In contrast, the relative risk describes the probability for an individual to develop the disease in the presence of a given, associated allele within a specific time window (Thomas 2004).

<u>Interaction and haplotype analyses</u>

Statistical interactions are present if the genetic effect of a certain allele / haplotype is changed in presence of a specific allele / haplotype at a second locus. For this analysis, data sets are conditioned on the first SNP and are statistically reanalyzed.

Haplotype analyses are performed to investigate the causality of a given marker. In situations where the causative variant is located on a haplotype background, lower *P*-values than in the single marker analysis are obtained. Haplotypes were assessed for nominally associated SNPs, or by using a sliding-window approach. Here, two to five consecutive SNPs surrounding an associated marker were included in the statistical analysis. Interaction and haplotype analyses can also be performed in data sets including families and / or quantitative traits.

Family-based association tests

In contrast to case-control studies, family-based data do not present with the issue of population stratification, as the non-transmitted alleles within a family are used as control alleles (Balding *et al.* 2007). The concept that disease-associated alleles are unequally transmitted from parents to their affected offspring has been introduced in 1993 (Spielman *et al.* 1993). In such a transmission disequilibrium test (TDT), it is checked whether a particular allele of a given SNP is transmitted significantly more often from parents to their affected child than would be expected based on unbiased Mendelian inheritance. In case of 'no association'

between the investigated marker and the trait, each of the two parental alleles has the same chance to be transmitted to the child. However, if an allele is associated with an increased risk for development of the disease, it will be transmitted more often than expected by chance (Thomas 2004).

Additionally, the availability of information on parental alleles allows for the detection of possible parent-of-origin effects (imprinting). Herefore, statistical analysis was restricted to maternally or paternally derived alleles. In situations where an imprinting effect was suggested, the phenotypic distribution in children to whom a certain allele was transmitted maternally or paternally, respectively, was compared to the phenotypic distribution in children to whom it was not.

In this thesis, family-based association tests have been conducted for all studies that involved the German DYS-sample, which consisted of 400 parent-child trios.

3.8.3 Analysis of quantitative endophenotypes

The endophenotypic data available in the German DYS-sample were used to correlate the children's performances in each of these quantitative measures with genotypic information. In case of the GWAS data, this type of analysis does not require controls and was performed via an analysis of variance (ANOVA). In ANOVAs, means of quantitative endophenotypes are compared between the different genotypic groups, and different genetic models are taken into consideration (see section 3.8.2).

To perform family-based statistical analysis for the endophenotype measures, an adaption of the TDT was used. The quantitative transmission disequilibrium test (QTDT) allows for the analysis of quantitative measures in parent-child trios and has, for some of the experimental questions, advantages over the analysis of binary traits (Ziegler & König 2006). Similar to the TDT, parent-of-origin effects and their phenotypic consequences were also assessed using the quantitative measures available.

In situations where a SNP was found to be associated with a quantitative measure, it was analyzed how much of the phenotypic variability of this trait could be attributed to the given variant. This estimation of a variant's effect size was performed using the r^2 goodness-of-fit measure.

4. RESULTS

4.1 Candidate gene approach

Literature was screened for genes or proteins that (i) have been suggested to be involved in human cognitive performance, or have been found associated with (ii) single cognitive processes related to reading and writing, or (iii) a dyslexia comorbid disorder. Genes of interest were then assessed in databases in order to identify additional functional evidence. Candidate genes identified by this approach were next analyzed in terms of their genetic contribution to dyslexia or some of the related endophenotypes in the German DYS-sample. The genetic variability at the candidate loci was tried to be captured using haplotype tagging SNPs (tSNPs).

For already published dyslexia candidate genes, we performed replication studies to either confirm previous findings, or we attempted to identify the causal variants at these loci. As single studies within this thesis were conducted at different points in time, the number of included samples varied between 396 and 400 parent-child trios. Also, the separate analysis of subgroups, comprising individuals of different affection grades, contributes to differences in sample numbers between studies.

4.1.1 Investigation of genes within the DYX2 locus

An intronic deletion in *DCDC2* as causal variant

Meng and colleagues (2005) reported a deletion region of 2,445 bp, located in intron 2 of *DCDC2*, to be causative for dyslexia in an US sample (Meng *et al.* 2005b). The deletion harbors a compound short tandem repeat (STR) polymorphism (GenBank accession no. BV677278) that is composed of variable copy numbers of $(GAGAGGAAGGAAA)_n$ and $(GGAA)_n$ repeat units.
A total of 1188 individuals of the German DYS-sample were included in our analysis. The experimental design was similar to the original study (Attachment III). In brief, one forward and two reverse primers were included in one PCR reaction, yielding products of 525 bp for non-deleted and 215 bp for deleted alleles. The non-deleted product was subsequently sequenced to determine STR alleles. Eight individuals were homozygous for the deletion and 184 individuals provided with a heterozygous state. Sequencing of the STR revealed nine different alleles in the German DYS-sample (Tab. 6). One individual failed sequencing analysis, resulting in 2374 alleles that were used for statistical calculations.

Tab. 6: Distribution of the *DCDC2* intron 2 deletion / compound STR polymorphism alleles.

Compound STR, GenBank accession number BV677278							Allele frequency*	TDT results	
Alleles	Repeat unit 1	Repeat unit 2	SNP	Repeat unit 3	Repeat unit 4	Repeat unit 5	n = 1582 alleles	T / NT	*P*-value
1	(GAGAGGAAGGAAA)2	(GGAA)7		(GGAA)2	(GGAA)4	(GGGA)2	0.606	150 / 175	0.165
2	(GAGAGGAAGGAAA)1	(GGAA)9	DelGAAA	(GGAA)0	(GGAA)4	(GGGA)2	-	-	-
3	(GAGAGGAAGGAAA)1	(GGAA)6		(GGAA)2	(GGAA)4	(GGGA)2	0.055	37 / 31	0.467
4	(GAGAGGAAGGAAA)2	(GGAA)6		(GGAA)2	(GGAA)4	(GGGA)2	0.106	72 / 61	0.340
5	(GAGAGGAAGGAAA)2	(GGAA)8		(GGAA)2	(GGAA)4	(GGGA)2	0.043	32 / 26	0.430
6	(GAGAGGAAGGAAA)2	(GGAA)8		(GGAA)2	(GGAA)3	(GGGA)2	0.048	31 / 28	0.696
7	(GAGAGGAAGGAAA)2	(GGAA)8		(GGAA)1	(GGAA)4	(GGGA)2	-	-	-
8	(GAGAGGAAGGAAA)2	(GGAA)7	DelGAAA	(GGAA)0	(GGAA)4	(GGGA)2	-	-	-
9	(GAGAGGAAGGAAA)1	(GGAA)7		(GGAA)2	(GGAA)4	(GGGA)2	0.008	3 / 7	0.200
10	(GAGAGGAAGGAAA)2	(GGAA)4		(GGAA)2	(GGAA)4	(GGGA)2	0.043	32 / 24	0.284
19	(GAGAGGAAGGAAA)2	(GGAA)9		(GGAA)2	(GGAA)3	(GGGA)2	0.003	3 / 1	0.306
20	(GAGAGGAAGGAAA)2	(GGAA)9		(GGAA)2	(GGAA)4	(GGGA)2	0.004	2 / 4	0.410
Deletion	x	x	x	x	x	x	0.086	52 / 57	0.632

The STR consists of five differently structured repeat units which differ in their respective numbers between the alleles. Between repeat units 2 and 3, a GAAA-deletion is present in two of the alleles (2, 8). In the deletion allele, all repeat units are deleted (x). Alleles suggested by Meng and colleagues which could not be identified in the German DYS-sample are marked (-). * - Frequency among parents only, TDT – transmission disequilibrium test, T/NT – ratio of transmitted (T) and non-transmitted (NT) alleles.

In Tab. 6, the deletion and the different alleles of the STR that were observed in the German DYS-sample are summarized. Allele frequencies in the general population were estimated based on the parental alleles. With a frequency of 60.6%, allele 1 was the most common allele observed in the German population. In addition, two other STR alleles (3, 4) and the deletion itself also represented common alleles with frequencies > 5%. Three rare alleles (9, 19, 20) were also identified.

Subsequently, a transmission disequilibrium test (TDT) was performed for the alleles observed in the German DYS-sample. As shown in Tab. 6, this analysis did not yield any significant associations with dyslexia as qualitative trait. The lowest *P*-value was observed for the common allele 1, with $P = 0.165$. Combining the rare alleles (frequency < 5%) and the deletion, as proposed in the original study, was also not significant ($P = 0.227$). Applying the endophenotype measures ‘reading’ and ‘spelling’ did not improve upon the results (data not shown).

Genetic interaction between *KIAA0319* and *DCDC2*

In 2006, an association of dyslexia with seven SNPs located in and around exon 1 of *KIAA0319* was reported in two UK samples (Harold *et al.* 2006). In the same study, evidence was found for a genetic interaction between two of these SNPs (rs4504469, rs761100) and the two-marker haplotype of *DCDC2,* previously postulated by our group ([rs793862-rs807701 (A–C)]; (Schumacher *et al.* 2006)). Replication of these findings was now attempted in the German DYS-sample, using parent-child trios with a severely affected child ($SD \geq 2$, $n = 244$) in order to increase power. Six of the seven SNPs were included in the assay, as rs4504469 had already

been genotyped previously, in the Schumacher *et al.* (2006) study, where it had not provided any significant results.

The results for the remaining six SNPs analyzed in the current study are presented in Tab. 7.

Tab. 7: Results of the TDT for SNPs within *KIAA0319* and interaction analysis with *DCDC2*.

Marker in *KIAA0319*		Single marker TDT analysis by affection grade		Interaction analysis			
				DCDC2 risk haplotype [rs793862-rs807701 (A-C)]			
SNP-ID	Position[a]	SD ≥ 2 (n = 244)	SD ≥ 2.5 (n = 114)	dyslexia	spelling	word reading	phonological decoding
rs4504469[b, c]	24,696,863	-	-	0.0553	0.2334	0.2382	0.6269
rs2179515	24,736,182	0.5430	0.9156				
rs761100[c]	24,740,621	0.3055	0.4859	0.3567	0.0912	**0.0351**	0.3437
rs7766230	24,741,408	0.4828	0.8997				
rs17491230	24,753,676	0.0579	0.3757				
rs1555090	24,756,086	0.6346	1.0				
rs3212236	24,756,434	0.4794	1.0				

Transmission disequilibrium test (TDT) results are presented for six of the seven *KIAA0319* markers as rs4504469 has already been analyzed in Schumacher *et al.* (2006). Interaction analyses were performed applying the allelic model and are restricted to the two *KIAA0319* markers with significant interactions in the original study. Results are presented for dyslexia as qualitative trait (dyslexia) and the three endophenotypes most comparable to the ascertainment criteria in the UK data set. *P*-values are in bold if significant. [a] - Position according to dbSNP129, [b] - SNP previously included in Schumacher *et al.*, [c] – SNPs showing significant interaction *P*-values in Harold *et al.*

None of the six markers in *KIAA0319* showed significant association with dyslexia or any of the endophenotypes in the TDT, neither for the entire data set nor when restricting the analysis to the most severely affected patients (SD ≥ 2.5; Tab. 7). The lowest *P*-value observed was *P* = 0.0579 for rs17491230.

Testing for interactions between SNPs in *KIAA0319* and the *DCDC2* risk haplotype was only performed for *KIAA0319*-SNPs rs4504469 and rs761100, as these combinations were postulated as significant by Harold and colleagues (2006). For dyslexia as qualitative trait, no significant result was obtained (Tab. 7). However, a trend of association was observed for the risk haplotype and rs4504469, with *P* = 0.053. Analysis of the related endophenotypes revealed a nominally significant interaction for the quantitative endophenotype 'word reading' (rs761100; *P* = 0.0351).

4.1.2 Investigation of *GRIN2B* and short-term memory

Variations in the gene coding for the N-methyl-D-aspartate receptor subunit 2B (*GRIN2B*) have been suggested to be involved in memory-related aspects of human cognition (de Quervain & Papassotiropoulos 2006) and attention-deficiency / hyperactivity disorder (ADHD; (Dorval *et al.* 2007)). Additional evidence was provided by the UCSC genome browser which maps *GRIN2B* within a linkage region of the dyslexia-related endophenotype 'phonological memory' (Brkanac *et al.* 2008). *GRIN2B* was therefore analyzed for dyslexia *per se* and the cognitive endophenotype 'short-term memory'.

397 parent-child trios were included in the analysis. Given the large genomic size and complex linkage disequilibrium (LD) structure of the *GRIN2B* locus, the region of interest was evenly covered using an intermarker distance of 20 kb, including 100 kb up- and 50 kb downstream of the gene. The density of SNPs within introns 2 and 3 of *GRIN2B* was increased, as the positive findings from the two candidate gene studies overlapped in these regions. Also, it was attempted to include the associated SNPs reported in the original studies, which was successful for all SNPs except rs1805474. In total, 66 SNPs were included in the assay, and genotyping was successful for 61 SNPs (Attachment IV).

Sixty-one SNPs were first tested for association with dyslexia as qualitative trait using TDT analysis. Only one SNP showed a nominally significant *P*-value in the overall German DYS-sample (n = 397, *P* = 0.013 for rs933614). Stratifying the sample for severity or haplotype analyses did not provide any evidence for a genetic contribution of SNPs at this locus to dyslexia (data not shown). In a next step, the quantitative phenotype 'short-term memory' was analyzed by applying the quantitative transmission disequilibrium test (QTDT). Four significant association signals were obtained (rs1012586, rs2268119, rs2216128, rs2192973), with the lowest nominal *P*-value observed being *P* = 0.0243 for rs2268119 (see Attachment IV). We subsequently stratified our sample for severity, and analyzed the four nominally significant markers in differently affected groups. Results for the four SNPs showing significant *P*-values for association with short-term memory in the QTDT are given in Tab. 8.

Tab. 8: QTDT results and effect sizes for short-term memory.

Marker		**all (n = 397)**		**SD ≥ 1.5 (n = 365)**		**SD ≥ 2.0 (n = 249)**	
SNP-ID	Position[a]	*P*-value	Effect size	*P*-value	Effect size	*P*-value	Effect size
rs1012586	13,746,899	**0.0401**	0.0058	**0.0411**	0.0068	0.0587	0.0171
rs2268119	13,763,901	**0.0243**	0.0064	**0.0289**	0.0073	0.1879	0.0068
rs2216128	13,774,281	**0.0406**	0.0064	**0.0158**	0.0142	0.0536	0.0253
rs2192973	13,787,822	**0.0381**	0.0080	**0.0150**	0.0156	0.0683	0.0205

The quantitative transmission disequilibrium test (QTDT) was performed across different severity groups (all (SD ≥ 1.0), SD ≥ 1.5, SD ≥ 2.0). *P*-values (bold if ≤ 0.05) are given for nominally significant SNPs. Effect sizes are presented for each of the SNPs for each of the severity groups, according to the r^2 goodness-of-fit measure.

Tab. 8 is structured with respect to the subgroups of the German DYS-sample which were constructed based on the affection grades of the probands. Although group size decreased with higher affection status, *P*-values for the four SNPs remained significant or borderline significant. All four SNPs are located in intron 3 of *GRIN2B* and only show moderate LD with another. In order to investigate how much of the phenotypic measures can be explained by each of the four SNPs, we calculated effect sizes across different severity groups. As shown in Tab. 8, effect sizes range from about 1% in the overall group to about 2% in more severely affected individuals (SD ≥ 2.0).

To further investigate the genetic model underlying the given association, the four SNPs were analyzed with respect to a possible biased parental transmission (Tab. 9).

Tab. 9: Analysis of imprinting effects for short-term memory.

Marker	all (n = 397)			SD ≥ 1.5 (n = 365)			SD ≥ 2.0 (n = 249)		
	$P_{mat.}$	$P_{pat.}$	P_{test}	$P_{mat.}$	$P_{pat.}$	P_{test}	$P_{mat.}$	$P_{pat.}$	P_{test}
rs1012586	**0.0001***	0.5518	**0.0006***	**0.00006***	0.5062	**0.0030**	**0.0001***	0.5586	**0.0004***
rs2268119	**0.0026**	0.7532	**0.0136**	**0.0060**	0.9747	**0.0417**	**0.0062**	0.6833	**0.0131**
rs2216128	**0.0089**	0.6438	**0.0289**	**0.0037**	0.7624	**0.0256**	**0.0246**	0.9828	0.1212
rs2192973	**0.0086**	0.3850	**0.0143**	**0.0039**	0.4879	**0.0144**	**0.0343**	0.5745	**0.0491**

Nominal significant SNPs in the QTDT were analyzed across different severity groups (all (SD ≥ 1.0), SD ≥ 1.5, SD ≥ 2.0). *P*-values are presented for maternal transmissions ($P_{mat.}$), paternal transmissions ($P_{pat.}$) and the test for significant differences between the parental transmissions (P_{test}). *P*-values are bold if ≤ 0.05. *P*-values that withstand correction for multiple testing (4 SNPs, 9 test statistics) are indicated by an asterix (*).

In the overall group, we observed significant maternal effects for all four SNPs ($P_{mat.}$-values ≤ 0.01, Tab. 9), but not for the alleles transmitted from the father. To show whether the transmission rate is significantly biased between the two parental alleles, P_{test}-values were calculated. They were statistically significant for all four SNPs. The most significant SNP, rs1012586, withstands correction for multiple testing ($P_{mat.corr.}$ = 0.0036). Stratification for severity revealed similar results for all four SNPs across different affection grades, with again rs1012586 remaining significant after correction for the number of comparisons (SD ≥ 1.5, $P_{mat.corr.}$ = 0.0022; SD ≥ 2.0, $P_{mat.corr.}$ = 0.0036).

We next investigated the phenotypic consequences of the significant maternal effects.

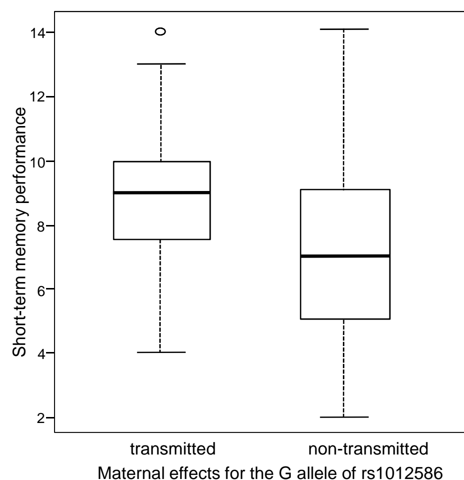


Fig. 11: Boxplots for short-term memory performance with respect to transmission of the G allele of rs1012586. The performance in short-term memory tasks is shown for individuals carrying a maternally transmitted G allele (left) versus those in whom the G allele has not been transmitted maternally (right). Middle black line – median, rectangle - the quantiles (75%, top line; 25%, bottom line), dashed vertical lines - distribution (minimum to maximum) of the performance scores, circle - outliers (more than 1.5fold inter-quartile distance).

The phenotypic measures for 'short-term memory' in individuals carrying maternally transmitted alleles were compared to individuals that carried non-maternally transmitted alleles. For the most significant SNP, rs1012586, 113 mother-child trios were genotypically informative. The G allele was observed to be maternally transmitted in 60 individuals and not transmitted in 53 individuals. Fig. 11 shows that individuals with a maternally inherited G allele performed significantly better in short-term memory tasks in comparison to individuals

where the G allele had not been maternally transmitted. The difference in the mean for short-term memory was 0.7 sd, resulting in a *P*-value of 0.0001. Similar results were obtained for rs2268119, rs2216128 and rs2192973 (see Attachment V).

4.1.3 Investigation of *LRRTM1* and human handedness

It has been suggested that variations within the *leucine-rich repeat transmembrane neuronal 1 gene* (*LRRTM1*) contribute to the lateralization of the human brain (Francks *et al.* 2007). More precisely, association was found with respect to the parental origin of a three-marker haplotype upstream of the gene.

14 SNPs within the genomic locus of *LRRTM1* were analyzed in 398 parent-child trios of the German DYS-sample, including available siblings of the probands. Selected SNPs comprised the three haplotype-forming markers (rs1446109, rs1007371, rs723524), as identified by Francks and colleagues, and 11 tSNPs (Tab. 10). Genotyping was successful for all SNPs.

Tab. 10: Quantitative trait analysis for handedness and SNPs located in *LRRTM1*.

Marker		TDT	QTDT for handedness			
			(1) QTDT	(2) Parent-of-origin effect		
SNP-ID	Position*	*P*-value	*P*-value	$P_{mat.}$	$P_{pat.}$	P_{test}
rs13019601	80,367,436	0.1794	**0.0462**	0.2040	0.4709	0.6776
rs1930	80,368,183	0.2700	**0.0136**	0.9792	0.1610	0.2480
rs1446110	80,371,264	0.3668	**0.0235**	0.0826	0.9082	0.2656
rs10170020	80,373,325	0.1400	0.0782	**0.0505**	0.7801	0.2011
rs6718055	80,373,709	0.6107	**0.0174**	0.8597	0.4592	0.4544
rs2862286	80,374,852	0.6205	**0.0055**	0.8312	0.3018	0.3259
rs6712681	80,377,689	0.6107	**0.0426**	0.3263	0.6266	0.8213
rs6733871	80,383,467	0.9475	**0.0376**	**0.0160**	0.6632	0.0974
rs1446109[a]	80,391,930	0.6014	0.2685	**0.0322**	0.8251	**0.0396**
rs11126755	80,396,062	0.2534	0.1714	**0.0476**	0.3829	0.2546
rs6755232	80,398,998	0.8005	0.5310	0.6138	0.7978	0.8198
rs767587	80,399,434	0.9068	0.8300	0.5556	0.1830	0.2853
rs1007371[a]	80,406,856	0.9287	0.2054	**0.0364**	0.9967	0.0627
rs723524[a]	80,435,312	0.8005	0.3139	0.1060	0.7492	0.3398

SNPs within *LRRTM1* are presented with their respective positions (in bp) according to dbSNP129 (*). *P*-values are given for the TDT with dyslexia as qualitative trait, and for the QTDT analysis with 'handedness'. Here, (1) refers to the nominal results of the QTDT analysis, and (2) to the imprinting effects ($P_{mat.}$ - maternal / $P_{pat.}$ - paternal transmissions). P_{test} refers to whether the difference beween the two parental transmissions is also significant. [a] – Markers forming the significantly associated haplotype in Francks *et al.* 2007. *P*-values in bold if ≤ 0.05.

As shown in Tab. 10, TDT analysis on dyslexia as qualitative trait revealed no significant *P*-values. The lowest *P*-value observed was $P = 0.14$ for rs10170020. The QTDT analysis using 'handedness' as continuous trait yielded seven SNPs with nominal significant *P*-values (Tab. 10, lowest *P*-value = 0.0055 for rs2862286). Although none of these results withstood correction for multiple testing, the number of significant SNPs was more than the expected number of false-positives (14 x 0.05 = 0.7). The significant SNPs included none of the upstream haplotype-forming markers, but the coding SNP rs6733871. Of the seven nominally significant SNPs, rs6733871 was the only one showing an imprinting effect in the subsequent analysis of parental

effects ($P = 0.0160$). This P-value was the lowest among the five SNPs showing significant maternal effects, and rs6733871 also provided with a trend of significance for the difference between the parental transmissions ($P_{test} = 0.0974$). No significant paternally-driven association was found for any of the 14 SNPs. A sufficiently powerful haplotype analysis was not possible due to the limited number of left-handers in our sample (n = 63) and the relatively low frequency of the previously associated three-marker haplotype.

We next assessed the phenotypic consequences of the maternal transmission effect, comparing the phenotypic measures for maternally transmitted and non-transmitted alleles for rs6733871.

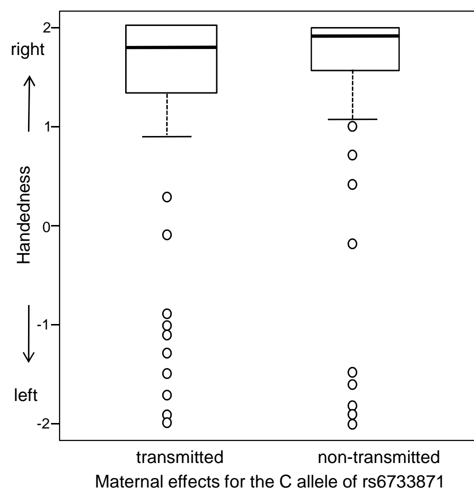


Fig. 12: Boxplots for the quantitative measure 'handedness' with respect to the parent-of-origin for the C allele of rs6733871. The group measures for handedness are shown for individuals carrying a maternally transmitted C allele (right) versus those in whom the C allele has not been transmitted maternally (left). For each group, the median (middle black line) and the quantiles (75%, top line; 25%, bottom line) are represented by the rectangle. The dashed vertical lines illustrate the distribution of the handedness measures. Outliers, showing more than 1.5fold inter-quartile distance, are represented by (°).

From Fig. 12 it can be deduced that children with a maternally transmitted C allele showed a stronger tendency towards left-handedness, with mean values of 1.18 for the maternally transmitted group, and 1.58 for the maternally non-transmitted group.

4.2 Genome-wide approach

To identify new susceptibility loci for dyslexia, a genome-wide analysis was performed. 200 dyslexia probands of the German DYS-sample were genotyped on HumanHap300K BeadChips. These data were subsequently analyzed with respect to two different questions:

i) In context of the NeuroDys consortium, a genome-wide association study (GWAS) was conducted. Herefore, data of the 200 German cases were compared to 926 population-based German controls and subsequently combined with genome-wide data from a UK sample (385 cases, 1406 controls) in form of a meta-analysis.

ii) Endophenotypes such as event-related potentials (MMN as neural correlate for speech perception) and arithmetic skills were not part of the joint European effort. The genome-wide data of the 200 German individuals were therefore independently correlated with these quantitative measures.

4.2.1 Association analysis of dyslexia as qualitative trait

All German individuals provided with call rates ≥ 98%, thus no sample had to be excluded. *P*-values generated by the Armitage-trend test were subsequently combined with the results from the independent UK GWAS. After quality control for SNPs available in both data sets, 290,809 SNPs remained and were included in the joint analysis. To exclude that population stratification between UK and German samples would affect statistical analysis, a multi-dimensional-scaling (MDS) analysis was performed. No individual outlier in either of the samples was detected. To account for small population differences observed, the main axes of variation from the MDS analysis as well as country of origin were included as covariates in the final logistic regression model. The top SNPs of the analysis are depicted in Tab. 11.

Tab. 11: Top SNPs of the joint German-UK dyslexia GWAS (NeuroDys).

SNP	Location (dbSNP129)			German data			UK data			Joint data		
	Gene	Chr.	Position	*P*-value	MAF ca	MAF co	*P*-value	MAF ca	MAF co	*P*-value	Risk allele	Rank
rs10513829	*LPP*	3	189,971,027	**0.0064**	0.268	0.339	**3.3×10^{-06}**	0.279	0.368	**6.00×10^{-08}**	T	1
rs7840675		8	34,075,414	0.0829	0.102	0.079	**4.8×10^{-07}**	0.157	0.096	**1.40×10^{-06}**	C	2
rs11117425	*IRF8*	16	84,529,772	**0.0034**	0.270	0.347	**0.0003**	0.265	0.333	**1.62×10^{-06}**	C	3
rs7202472	*IRF8*	16	84,535,003	**0.0005**	0.156	0.234	**0.0026**	0.180	0.228	**1.75×10^{-06}**	G	4
rs747783		11	15,670,132	**0.0029**	0.316	0.243	**0.0002**	0.310	0.242	**2.59×10^{-06}**	T	5
rs2836341	*ERG*	*21*	*38,656,627*	*0.0532*	*0.464*	*0.470*	***4.6×10^{-05}***	*0.486*	*0.461*	***4.10×10^{-06}***	*G*	*6*
rs9465637	*MBOAT1*	6	20,222,087	**0.0064**	0.316	0.354	**0.0023**	0.335	0.376	**5.01×10^{-06}**	G	7
rs10123957	*C9orf5*	9	110,900,239	**0.0069**	0.449	0.525	**0.0002**	0.451	0.527	**5.26×10^{-06}**	C	8
rs4887111	*LOC388135*	15	71,815,338	**0.0025**	0.344	0.427	**0.0006**	0.312	0.379	**7.52×10^{-06}**	A	9
rs366078	*IRF8*	16	84,522,064	**0.0036**	0.114	0.175	**0.0184**	0.141	0.169	**7.98×10^{-06}**	T	10
rs10816767	*C9orf5*	9	110,822,491	**0.0033**	0.434	0.517	**0.0006**	0.450	0.519	**8.00×10^{-06}**	A	11
rs4327894	*ARHGEF10*	8	1,740,904	**0.0007**	0.148	0.224	**0.0034**	0.203	0.254	**8.57×10^{-06}**	G	12

SNPs showing *P*-values ≤ 10^{-05} are depicted in order of significance in the joint data set. For each of the two subsamples (Germany, UK), *P*-values and minor allele frequencies (MAF) in cases (ca) and controls (co) are represented. One SNP, rs2836341, appeared among the top SNPs although allele frequencies showed opposite directions between both samples (*in italics*). Thus, this SNP (ranked #6) was excluded from further analysis. For the remaining SNPs, risk alleles were the same in both samples. Ranks were deduced from the combined *P*-values. *P*-values are given in bold if significant.

The joint analysis did not reveal any SNP that reached genome-wide significance. 12 SNPs yielded nominal *P*-values ≤ 10^{-05}, the lowest *P*-value observed was $P = 6.0 \times 10^{-08}$ for rs10513829. One SNP (rs2836341) showed opposite allelic effects in both samples and was excluded from further analysis. Among the top hits were two loci with multiple hits (*IRF8* (rs11117425, rs7202472, rs366078) and *C9orf5* (rs10123957, rs10816767)). Besides the 12 top hits, 51 further SNPs showed *P*-values ≤ 10^{-04} (data not shown). Thus, in total, 63 SNPs provided with *P*-values ≤ 10^{-04} and were considered for the next step of independent replication in one iPlex reaction.

Replication of initial GWAS results

For assay design, SNPs were prioritized with respect to their respective rank in the joint GWAS. Except rs2836341, each of the 12 top SNPs was included, and the assay was then filled up with additional SNPs showing *P*-values ≤ 10^{-04}. High-ranking SNPs (top 20) that could not be

included in the assay due to technical reasons were tried to be replaced by tSNPs based on HapMap data. In total, 34 SNPs were included in the assay (Tab. 12).

Tab. 12: Replication of the NeuroDys GWAS.

SNP[a]	Location (dbSNP129)			Results GWAS combined			Results replication combined		*P*-value combined**
	Gene	Chr.	Position	Rank	*P*-value	Risk allele	*P*-value	Risk allele*	
rs10513829	*LPP*	3	189,971,027	1	**6.00×10^{-08}**	T	0.636		
rs7840675		8	34,075,414	2	**1.40×10^{-06}**	C	0.5450		
rs11117425 (rs11648084)	*IRF8*	16	84,529,772	3	**1.62×10^{-06}**	C	0.5360		
rs7202472	*IRF8*	16	84,535,003	4	**1.75×10^{-06}**	G	0.0858		
rs747783		11	15,670,132	5	**2.59×10^{-06}**	T	0.2550		
rs9465637 (rs13191158)	*MBOAT1*	6	20,222,087	7	**5.01×10^{-06}**	G	0.0503		
rs10123957	*C9orf5*	9	110,900,239	8	**5.26×10^{-06}**	C	0.0952		
rs4887111	*LOC388135*	15	71,815,338	9	**7.52×10^{-06}**	A	0.6290		
rs366078	*IRF8*	16	84,522,064	10	**7.98×10^{-06}**	**T**	**0.0456**	**C**	0.2526
rs10816767 (rs7034615)	*C9orf5*	9	110,822,491	11	**8.00×10^{-06}**	A	0.1220		
rs4327894	*ARHGEF10*	8	1,740,904	12	**8.57×10^{-06}**	G	0.8300		
rs6136213		20	17,798,340	13	**1.09×10^{-05}**	A	0.0587		
rs1181841		5	128,580,605	14	**1.12×10^{-05}**	G	failed		
rs10518444 (rs2271081)		4	125,945,654	16	**1.34×10^{-05}**	G	0.3020		
rs12454776		18	56,706,020	17	**1.43×10^{-05}**	C	failed		
rs10512712		5	39,728,089	18	**1.56×10^{-05}**	**C**	**0.0147**	**C**	**1.19×10^{-07}**
rs1429411	*ANKRD44*	2	197,852,247	19	**1.60×10^{-05}**	T	failed		
rs6984900	*LOC727677*	8	128,373,451	20	**1.66×10^{-05}**	T	0.2810		
rs9662100 (rs9659751)	*FMN2*	1	238,645,882	21	**1.90×10^{-05}**	G	0.3930		
rs902025		15	61,019,454	22	**1.99×10^{-05}**	C	failed		
rs7623540	*LPP*	3	189,972,234	23	**2.09×10^{-05}**	C	failed		
rs4747165	*CDH23*	10	72,969,308	24	**2.13×10^{-05}**	G	0.5220		
rs2077268	*RYR3*	15	31,661,043	25	**2.28×10^{-05}**	C	failed		
rs1003346 (rs11792635)	*C9orf5*	9	110,855,161	30	**2.74×10^{-05}**	T	0.1970		
rs6498274	*LOC92017*	16	12,273,877	43	**4.50×10^{-05}**	A	0.3020		
rs1872285		11	15,621,628	44	**4.51×10^{-05}**	T	0.3470		
rs1892577		21	32,353,866	46	**4.95×10^{-05}**	T	0.1490		
rs7411544	*PLXNA2*	1	206,428,793	47	**5.54×10^{-05}**	T	0.8270		
rs460420	*ADAMTS1*	21	27,105,523	48	**5.68×10^{-05}**	C	0.4920		
rs3821173	*ADAM23*	2	207,186,405	51	**6.32×10^{-05}**	A	0.5680		
rs958877		2	356,410	52	**6.34×10^{-05}**	A	0.1530		
rs9529688		13	69,663,347	54	**6.89×10^{-05}**	G	0.0845		
rs7541094		1	68,536,862	56	**7.85×10^{-05}**	G	0.4820		
rs905950	*LOC92017*	16	12,265,707	64	**9.95×10^{-05}**	C	0.2810		

34 SNPs were included in the replication, based on their ranking in the combined GWAS data (as indicated in 'rank'). [a] - SNPs in parentheses are tSNPs used in the replication assay. Replication results are only given if the SNP was genotyped successfully in all subsamples (else 'failed'). *- Risk alleles are presented for SNPs that showed nominal significance in the replication sample. ** - Combined *P*-values are given for SNPs that were significant in the replication sample. *P*-values in bold if ≤ 0.05.

Replication was performed in independent DNAs of the European NeuroDys sample (1409 cases, 1959 controls). Genotyped along were the 585 UK and German dyslexia cases of the GWAS, to ensure genotype consistencies between platforms and tSNPs. Statistical analysis of the replication data was performed using the Cochran-Mantel-Haenszel (CMH) test, with the different samples (= country of origin) analyzed as a covariate.

As shown in Tab. 12, genotyping was successful for 28 of the 34 SNPs. Two SNPs (rs7623540, rs2077268) did not reach CR_{SNPs} of 90% in the overall sample and were excluded. Another four SNPs (rs1181841, rs124547796, rs1429411, rs902025) failed genotyping in one of the populations and were also excluded. Quality control for the 585 samples genotyped on both platforms (Illumina and Sequenom) revealed concordancy rates > 99% for identical SNPs, and high LD for GWAS SNPs and their respective tSNPs genotyped on Sequenom ($r^2 \geq 0.98$, data not shown).
Statistical analysis of the 28 SNPs remaining after QC revealed two SNPs (rs366078, rs10512712) that provided with nominal significant *P*-values in the NeuroDys replication sample. Comparing the risk alleles between GWAS and replication identified the opposite direction of association for rs366078 (chr. 16: 84,522,064 bp, *IRF8*). For rs10512712 (chr. 5: 39,728,089 bp, intergenic), the risk allele was the same in both studies. The combined *P*-value of 1.19×10^{-07}, however, did not reach genome-wide significance. Calculation of the odds ratio for rs10512712 revealed an OR of 1.266 (CI = 1.158 - 1.384) for the risk allele C.

Replication after integrating pooling data

The method of pooling allows for an estimation of genome-wide allele frequencies in large samples, without the immense effort of genotyping DNAs individually. This approach helps to prioritize SNPs from the individual GWAS that should be taken into replication. As, for many SNPs, allele frequencies are expected to be different between populations, we classified the NeuroDys sample into three main population-based groups, and constructed DNA pools for Central Europe (532 cases, 912 controls; see Tab. 5, section 3.7.3), UK (426 cases, 219 controls) and Finland (286 cases, 321 controls). The pools contained samples from the NeuroDys consortium with the exception of the 585 samples that were part of the initial joint GWAS. Pools were analyzed genome-wide for 1,199,187 markers using HumanHap1M-DUO Illumina BeadChips. An overview of the most significant hits is represented in Attachment VI. These data were used to prioritize SNPs from the initial GWAS for a second step of replication. SNPs were chosen according to the following criteria:
i) SNPs with lowest *P*-values in joint pooling analysis and significant *P*-values in at least two populations, ii) SNPs with significant *P*-values in the joined pooling analysis and in the initial GWAS, iii) functional SNPs and SNPs in gene pathways and iv) SNPs with multiple additional hits in the same gene. This strategy for assay design resulted in one iPlex assay comprising 40 of the top SNPs. They were next individually genotyped in the entire European NeuroDys sample (1526 cases, 2261 controls), and statistical analysis was performed using CMH. Results are presented in Tab. 13.

Tab. 13: Results of replication including GWAS and pooling data.

SNP	Location according to dbSNP129			Pooling	GWAS	Results replication combined	
	Gene	Chr.	Position [bp]	*P*-value combined	*P*-value	*P*-value	Risk allele*
i) SNPs with significant *P*-values in both, initial GWAS and pooling							
rs1569012	*STON2*	14	80,923,160	**1.36×10^{-08}**	**0.0246**	0.5493	
rs268598	*TRAM1*	8	71,677,425	**1.93×10^{-05}**	**0.0284**	0.7904	
rs1350317	*ODZ3*	4	183,649,628	**1.38×10^{-04}**	**0.0428**	0.3456	
rs1581413	*VEPH1*	3	158,532,859	**2.94×10^{-04}**	**0.0061**	failed[1]	
ii) Top-SNPs of pooling							
rs10932727	*DIRC3*	2	218,313,957	**8.67×10^{-10}**	0.280	0.0502	
rs12290752		11	115,939,639	**9.00×10^{-09}**	n.a.	failed[2]	
rs12344734	*TMC1*	9	74,383,363	**7.65×10^{-08}**	n.a.	**0.0016**	A
rs10509910	*MXI1*	10	111,991,750	**8.73×10^{-08}**	0.914	0.5514	
rs6812487	*ODZ3*	4	183,697,713	**2.07×10^{-07}**	n.a.	0.6237	
rs7934218	*FAM168A*	11	72,884,377	**2.75×10^{-07}**	n.a.	0.1971	
rs2311445		16	17,456,460	**4.69×10^{-07}**	0.999	0.2328	
rs705790		6	166,286,499	**5.07×10^{-07}**	n.a.	**0.0485**	A
rs12352208		9	14,563,137	**5.57×10^{-07}**	0.823	0.8739	
rs2817764	*CDC2L6*	6	111,087,345	**5.99×10^{-07}**	n.a.	0.6110	
rs16932422	*DNAJC5B*	8	67,103,552	**7.06×10^{-07}**	0.477	0.1636	
rs7904542	*CEP55*	10	95,245,571	**7.43×10^{-07}**	n.a.	0.1962	
rs17615558		6	12,463,160	**1.17×10^{-06}**	n.a.	0.1299	
rs16900429	*RIPK2*	8	90,845,430	**1.26×10^{-06}**	n.a.	**0.0143**	C
rs9397276		6	156,298,468	**1.70×10^{-06}**	n.a.	0.0837	
rs7381	*FBLN1*	22	44,375,446	**1.80×10^{-06}**	0.432	**0.0029**	A
rs3736403	*CCDC108*	2	219,613,491	**3.65×10^{-06}**	n.a.	0.2334	
rs7686728		4	184,541,581	**3.84×10^{-06}**	n.a.	**0.0140**	G
rs4330611	*SGCE*	7	94,109,934	**4.94×10^{-06}**	0.746	0.5620	
rs11232875		11	81,191,362	**5.98×10^{-06}**	n.a.	failed[1]	
rs10821663	*ANK3*	10	61,480,286	**7.86×10^{-06}**	0.086	0.5228	
rs1546929	*BCKDHB*	6	81,104,278	**7.86×10^{-06}**	n.a.	0.3900	
rs2189167		4	104,953,292	**8.00×10^{-06}**	n.a.	failed[1]	
rs9324005		14	98,565,680	**8.14×10^{-06}**	n.a.	failed[3]	
rs4436151		8	114,902,482	**8.47×10^{-06}**	n.a.	**0.0240**	C
rs9916926		18	12,908,318	**9.84×10^{-06}**	0.069	0.1342	
rs34871518		19	63,046,077	**9.87×10^{-06}**	n.a.	0.2389	
rs9535442		13	49,821,641	**1.01×10^{-05}**	0.985	0.1151	
rs4510693		6	156,270,620	**1.10×10^{-05}**	n.a.	failed[3]	
iii) SNPs with significant *P*-values in pooling approach and at least two further hits at same locus							
rs12743401	*PPP1R12B*	1	200,743,271	**1.38×10^{-09}**	n.a.	failed[2]	
rs6687859	*PTGER3*	1	71,135,175	**4.19×10^{-08}**	n.a.	0.0599	
rs2289191	*IARS2*	1	218,366,658	**2.27×10^{-05}**	0.641	0.6825	
rs4655653	*WDR78*	1	67,104,024	**3.80×10^{-05}**	n.a.	**0.0123**	A
iv) SNPs with significant *P*-values in pooling approach and functional evidence[#]							
rs461119[a]	*GRIK1*	11	40,740,032	**1.30×10^{-08}**	0.465	**0.0015**	C
rs945386	*KIAA1984*	9	138,813,417	**1.89×10^{-05}**	n.a.	0.2446	
rs5063	*NPPA*	1	11,830,235	**1.99×10^{-05}**	n.a.	0.1699	

The table is structured with respect to the reason why SNPs were included in the replication assay (i – iv). Functional evidence ([#]) was defined based on expression data (UCSC browser). 33 SNPs were included in statistical analysis (else 'failed', for reasons of [1] - deviation from HWE, [2] - minor allele frequency ≤ 0.05, or [3] - call rate ≤ 90%). Some SNPs were not available in the GWAS (n.a.). * - risk allele only given for SNPs that showed nominal significance in the replication, [a] - rs420121 (GWAS, pooling) was replaced in the replication by tSNP rs461119. *P*-values are given in bold if ≤ 0.05.

Genotyping was successful for 33 of 40 SNPs, which were statistically analyzed. As presented in Tab. 13, seven SNPs could be replicated. After statistical correction for 33 SNPs, one SNP remained with a significant *P*-value (rs461119, $P_{corr.} = 0.0485$). Two additional SNPs (rs7381, rs12344734) provided with nominal *P*-values ≤ 0.003 which, at corrected level, still showed suggestive evidence ($P_{corr.} \leq 0.1$).

As these three SNPs presented the strongest evidence for being true findings, they were subsequently analyzed in the single samples, in order to exclude that findings were only triggered by single populations (Tab. 14).

Tab. 14: Frequencies of candidate SNPs in different populations within the NeuroDys sample.

Samples	rs461119, *GRIK1*			rs12344734, *TMC1*			rs7381, *FBLN1*		
	Cases	Controls	*P*-value	Cases	Controls	*P*-value	Cases	Controls	*P*-value
i) Pools									
Central Europe	0.39	0.33		0.21	0.14		0.20	0.15	
UK	0.40	0.32	**1.30×10^{-08}**	0.28	0.22	**7.65×10^{-08}**	0.18	0.13	**1.80×10^{-06}**
Finland	0.37	0.33		0.21	0.17		0.13	0.10	
ii) Replication sample									
Austria	0.30	0.28		0.12	0.11		0.06	0.06	
Germany NeuroDys	0.33	0.30		0.12	0.11		0.09	0.06	
Germany DYS-sample	0.31	0.28		0.14	0.10		0.08	0.06	
Switzerland	0.29	0.34		0.19	0.16		failed		
Hungary	0.34	0.31	**0.0015**	0.15	0.11	**0.0016**	0.05	0.06	**0.0029**
Netherlands	0.28	0.28		0.10	0.08		0.10	0.06	
France	0.30	0.29		0.10	0.11		0.06	0.06	
UK Cardiff	0.28	0.20		0.10	0.06		failed		
UK Oxford	0.33	0.25		0.12	0.10		0.08	0.07	
OR [CI 95%] in replication	**1.18 [1.07 - 1.31]**			**1.27 [1.09 - 1.47]**			**1.32 [1.10 - 1.58]**		

Minor allele frequencies in the groups of cases and controls, respectively, are given for each of the three SNPs showing $P_{corr.} \leq 0.1$. The table is separated in (i) results based on pooling (allele frequency estimates) and (ii) the individual genotyping (true allele frequencies). 'Failed' SNPs refer to QC exclusion criteria (MAF ≤ 0.05). OR – odds ratio, CI – confidence interval.

The replication data did not present with exclusively consistent results over all samples. As indicated by the allele frequencies (Tab. 14), for each of the three SNPs, one population showed opposite effects between cases and controls compared to the other samples (Switzerland for rs461119, France for rs12344734, Hungary for rs7381). For all other countries, the three candidate SNPs provided the same allelic direction. Calculation of the odds ratios (OR) for each of the SNPs revealed effect sizes ranging from OR = 1.18 to 1.32 for each respective risk allele (Tab. 14).

4.2.2 Genome-wide analysis of dyslexia-related endophenotypes

We aimed at correlating quantitative traits within the dyslexia cognitive spectrum to the genome-wide data of the 200 cases of the German DYS-sample. As analysis of measures directly involved in reading and writing belonged to the general intellectual property of the NeuroDys consortium, we analyzed two more abstractly related endophenotypes, i.e. event-related potentials (ERPs) as measured by mismatch negativity (MMN) and arithmetical skills (see section 2.1.4).

Analysis of event-related potentials

After quality control in the German DYS-sample, 297,086 SNPs were tested for association with the two components of MMN, namely MMNa and MMNb, using analysis of variance

(ANOVA). As depicted in Tab. 15, two SNPs showed genome-wide significant associations with the early component of MMN, MMNa (rs1365152: $P_{nom.} = 4.27 \times 10^{-08}$, $P_{corr.} = 0.013$; rs2114167: $P_{nom.} = 1.17 \times 10^{-07}$, $P_{corr.} = 0.035$). 17 other SNPs provided with nominal P-values < 10^{-05} for at least one of the two MMN components. These 19 SNPs were taken for replication in 186 independent cases from the German DYS-sample.

Tab. 15: Results of GWAS and replication for association with mismatch negativity (MMN).

SNP	Location (dbSNP 129)			MMN component	Genetic Model[1]	*P*-values		
	Gene	Chr.	Position			GWAS	Replication	combined
rs2487742	*MGC16664*	1	177,994,375	MMNa	carrier-T	**2.77 x 10^{-06}**	0.3020	
rs11300	*MPP4*	2	202,273,815	MMNb	allelic	**8.47 x 10^{-06}**	0.0760	
rs1365152	*CLSTN2*	3	141,213,956	MMNa	allelic	**4.27 x 10^{-08}**	0.8300	
					carrier-A	**9.02 x 10^{-07}**	0.9860	
rs2114167	*CLSTN2*	3	141,262,049	MMNa	allelic	**1.17 x 10^{-07}**	0.0868	
					carrier-G	**9.02 x 10^{-07}**	0.9861	
rs7683638		4	154,227,903	MMNa	carrier-G	**4.60 x 10^{-06}**	failed[2]	
rs4234898		4	157,217,991	MMNb	allelic	**3.29 x 10^{-06}**	**0.0049**	**1.44 x 10^{-07}**
					carrier-T	**6.20 x 10^{-06}**	**0.0014**	**5.14 x 10^{-08}**
rs4704133		5	73,667,082	MMNb	carrier-C	**7.56 x 10^{-06}**	failed[2]	
rs9390586		6	148,962,045	MMNb	genotypic	**2.24 x 10^{-06}**	0.4890	
					carrier-T	**4.79 x 10^{-07}**	1	
rs7793973	*MEOX2*	7	15,697,062	MMNa	genotypic	**9.78 x 10^{-06}**	0.3540	
rs1607924	*SAMD12*	8	119,671,788	MMNb	carrier-A	**9.91 x 10^{-06}**	0.9100	
rs965670	*ENPP2*	8	120,666,727	MMNb	allelic	**1.53 x 10^{-06}**	0.8950	
rs10996111		10	66,391,943	MMNa	carrier-G	**9.54 x 10^{-06}**	failed[3]	
rs4751178	*TXNL2*	10	131,882,235	MMNb	genotypic	**7.25 x 10^{-06}**	0.0861	
					carrier-G	**1.30 x 10^{-06}**	0.1080	
rs1777697		14	82,206,283	MMNa	genotypic	**9.83 x 10^{-07}**	0.5710	
					allelic	**1.16 x 10^{-06}**	0.5410	
					carrier-T	**3.36 x 10^{-06}**	0.6950	
rs4238922	*HS3ST4*	16	25,694,112	MMNa	genotypic	**7.48 x 10^{-06}**	0.8440	
					carrier-A	**1.20 x 10^{-06}**	0.5670	
rs11871364		17	42,538,151	MMNa	carrier-C	**9.69 x 10^{-06}**	0.7840	
rs7217223	*raptor*	17	76,260,258	MMNb	carrier-C	**4.57 x 10^{-06}**	failed[3]	
rs2612570		18	41,228,148	MMNb	carrier-C	**4.42 x 10^{-06}**	0.7720	
rs1736148		21	15,735,083	MMNa	allelic	**1.93 x 10^{-06}**	0.0985	

P-values are depicted for 19 SNPs that were included in the replication analysis for association with mismatch negativity (MMN). Four SNPs were excluded from statistical analysis ('failed'). For SNPs providing nominal significant associations in the replication, a combined *P*-value was calculated. [1] - all genetic models yielding *P*-values ≤ 10^{-05} in the GWAS are shown, with their corresponding *P*-values in the replication sample. [2] – SNPs excluded due to CR ≤ 98%, [3] - SNPs failed genotyping. *P*-values bold if significant.

As shown in Tab. 15, only rs4234898 was significantly associated with one MMN component (MMNb, $P_{nom.} = 0.00146$, $P_{corr.} = 0.028$) in the replication sample. Subsequent combination of the data from both samples ($n_{total} = 386$) yielded $P = 5.14 \times 10^{-08}$ for the carrier-T model. This result is genome-wide significant ($P_{corr.} = 0.015$). In addition, the combined analysis using the allelic model is also significant after correction for multiple testing ($P_{corr.} = 0.043$). To address the phenotypic consequences of the given association, we compared the phenotypic measures for MMNb between the two groups (carrier-T vs. non-carrier-T).

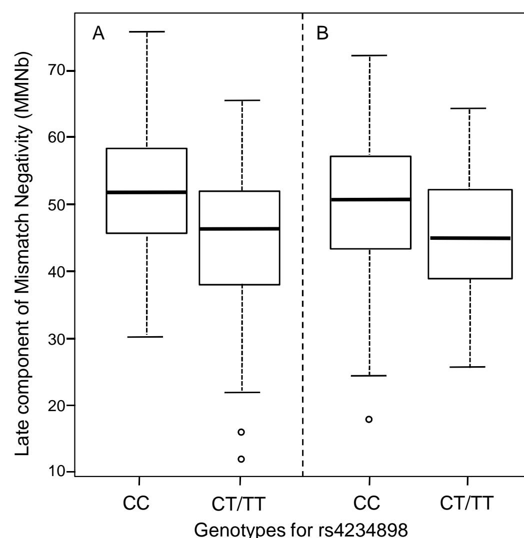


Fig. 13: Boxplot for the MMNb measures with respect to rs4234898 genotype in (A) initial and (B) replication sample. For each group, the median (middle black line) and the quantiles (75%, top line; 25%, bottom line) are represented by the rectangle. The dashed vertical lines illustrate the distribution of the performance scores. Outliers, showing more than 1.5fold interquartile distance, are represented by a circle.

As shown in the boxplot (Fig. 13), individuals that were homozygous for the C allele showed a larger MMNb than individuals who carried at least one copy of the T allele. This was true for both samples.

To further investigate the association findings with rs4234898, haplotypes at this locus were tested using a sliding window approach. In the initial sample, a significant two-marker haplotype consisting of rs4234898 and rs11100040 was identified ($P = 2.79 \times 10^{-06}$ for the allelic combination T-T). This finding was confirmed in the replication sample ($P = 0.0058$). In the combined sample, the association *P*-value for the two-marker haplotype was observed to be $P = 6.71 \times 10^{-08}$.

rs4234898 (chr. 4: 157,217,991 bp, NCBI build 36) is located in a gene desert. We therefore assessed potential regulatory effects using a publically available dataset, the 'mRNA-by-SNP-browser' (Dixon *et al.* 2007). Affymetrix-probe '*202497_x_at*' showed the most significant association ($P = 1.1 \times 10^{-05}$) with rs4234898 alleles, with the C allele being associated with a higher abundance of the probe in lymphoblastoid tissue. This probe represents the sequence of transcripts of *SLC2A3* (*solute carrier family 2 member 3*). Five other probes, representing partial transcripts of *SLC2A3*, were also significantly associated with rs4234898 (*P*-values from 4.7×10^{-05} to 1.5×10^{-04}).

We aimed at replicating the regulatory effects of rs4234898 on *SLC2A3* expression levels. 17 EBV-transformed cell lines, prepared from peripheral blood of our dyslexic children, were included in a functional assay (7 x CC, 10 x CT/TT for rs4234898). To investigate expression patterns in adults, mRNA from blood samples of 37 adult individuals were also analyzed (20 x CC, 17 x CT/TT). The abundance of *SLC2A3* transcripts was measured using quantitative RealTime-PCR (TaqMan).

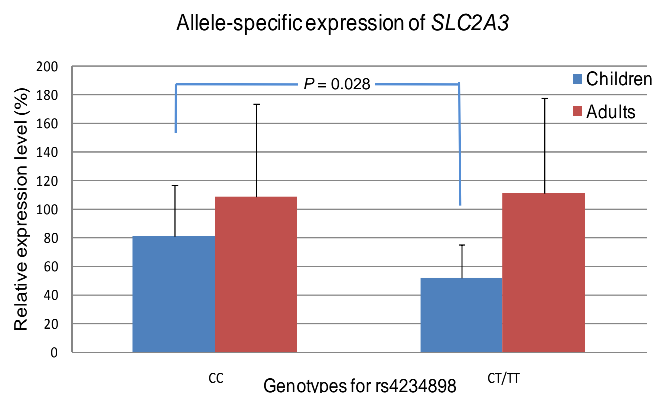


Fig. 14: Relative expression levels for *SLC2A3* with respect to rs4234898 genotypes. mRNA abundance was measured in EBV-transformed cell lines from children (n = 7 CC, n = 10 carrier-T) and peripheral blood from adults (n = 20 CC, n = 17 carrier-T). While expression levels do not vary significantly between the two adult groups, the children group of carrier-T shows significantly lower expression of *SLC2A3* than the CC-group ($P = 0.028$).

As shown in Fig. 14, we detected a significant lower expression of *SLC2A3* in cells from children carrying at least one T allele, in comparison to children that were non-T-carriers ($P = 0.028$). Overall, expression levels in children were lower compared to adults. In the adult samples, no significant associations of *SLC2A3* mRNA levels with rs4234898 genotypes were found.

We next analyzed whether the rs11100040 - rs4234898 two-marker haplotype also contributes to different expression levels of *SLC2A3* in our functional data. Herefore, we compared children with an unambiguous T-T haplotype for rs4234898 - rs11100040 (n = 4) to all other individuals included in the experiment (n = 13). It was observed that individuals carrying a T-T haplotype showed a significant lower expression level of *SLC2A3* ($P = 0.005$, one-sided) as compared to non-T-T-carriers. As the T-T group consisted of only 4 individuals, the haplotype T-T status was permuted and the *P*-value was recalculated. 55 of 10,000 permutations yielded better results than the original dataset, corresponding to an empirical *P*-value of 0.005 (95% CI = 0.0042 - 0.0073).

Analysis of arithmetical skills

The genome-wide data were also tested for association with performance in (i) exact calculation (EC), (ii) number comparison (NC) and (iii) the basic mathematical factor (BMF). As depicted in Tab. 16, analysis of variance identified three SNPs that fulfilled the criterion of genome-wide significance with at least one of these three mathematical phenotypes. Rs1399428 (chr. 9: 119,700,389 bp, lowest *P*-value = 3.19×10^{-09}) and rs4837521 (chr. 9: 119,694,282 bp, lowest *P*-value = 2.08×10^{-08}) showed significant associations with ‘number comparison’, while rs133885 (chr. 22: 24,489,289 bp) provided a genome-wide significant result for the basic mathematical factor in the carrier-A model ($P = 5.78 \times 10^{-09}$).

Tab. 16: Results for arithmetical abilities.

SNP (Phenotype)	Location according to dbSNP129			Genetic model*	*P*-values		
	Chr.	Position	Gene		GWAS	Replication	combined
rs1399428 (NC)	9	119,700,389	*LOC100129489*	genotypic	**2.40×10^{-08}**	0.4670	7.01×10^{-06}
				het. / hom.	**3.19×10^{-09}**	0.2540	1.09×10^{-06}
rs4837521 (NC)	9	119,694,282	*LOC100129489*	genoptyic	1.53×10^{-07}	0.0997	5.67×10^{-07}
				het. / hom.	**2.03×10^{-08}**	0.0402	8.46×10^{-08}
rs133885 (BMF)	22	24,489,289	*MYO18B*	genotypic	4.39×10^{-08}	**0.0089**	**2.75×10^{-09}**
				allelic	1.24×10^{-06}	**0.0024**	**9.83×10^{-09}**
				carrier-A	**5.78×10^{-09}**	**0.0098**	**8.81×10^{-10}**

In the GWAS, three SNPs provided with genome-wide significance in at least one genetic model (*). The associated phenotypes are given below SNP-IDs (NC – number comparison, BMF – Basic mathematical factor). *P*-values are highlighted in bold if they are genome-wide significant (GWAS) or withstand correction for multiple testing for 3 SNPs ($P \leq 0.016$). het. – heterozygous, hom. – homozygous.

The three SNPs were chosen for replication in 186 independent individuals from the German DYS-sample. As shown in Tab. 16, rs4837521 and rs133885 were also nominally significant associated with the respective quantitative measure in this second sample. However, only rs133885 remains significant after correction for multiple testing ($P_{corr.} = 0.0294$). Combining the data from initial GWAS and replication showed that, for the two SNPs on chromosome 9 (rs1399428, rs4837521), higher *P*-values than in the initial GWA study were obtained. For rs133885, the combined analysis provided with a *P*-value of 8.81×10^{-10} for the basic mathematical factor. This result withstands conservative correction for the number of SNPs and three mathematical phenotypes ($P_{corr.} = 0.00079$). In each sample (initial and replication), the single measures ‘exact calculation’ and ‘number comparison’ also showed significant associations with rs133885, however, *P*-values were less significant than for the combined mathematical factor (data not shown). Haplotype analysis did not yield any haplotype that improved *P*-values as compared to rs133885 alone.

Next, the effect size for rs133885 on mathematical performance was estimated in the combined data set, using r^2 goodness-of-fit measure. This analysis revealed an effect size of 9.2% in the overall group ($SD \geq 1.0$). Stratifying the subsample for severity of dyslexia yielded stronger effects with increasing dyslexia severity (9.8% for $SD \geq 1.5$, 10.4% for $SD \geq 2.0$). In the most severely affected group ($SD \geq 2.5$), effect size for rs133885 was observed to be 18.4%.

As depicted in Fig. 15, correlating genotypes for rs133885 and the measures of arithmetical skills revealed that A-carriers perform better in mathematical tasks compared to non-A-carriers. This was observed in both samples, even though the group difference was larger in the initial GWAS sample.

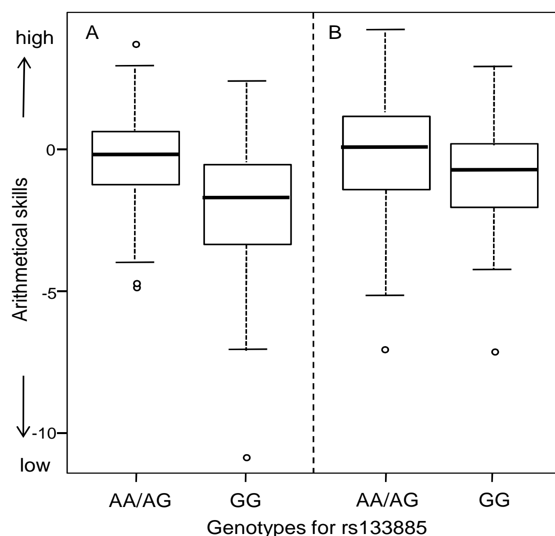


Fig. 15: Boxplot for the arithmetical skills with respect to rs133885 genotype in (A) initial and (B) replication sample. For each group, the median (middle black line) and the quantiles (75%, top line; 25%, bottom line) are represented by the rectangle. The dashed vertical lines illustrate the distribution of the performance scores. Outliers, showing more than 1.5fold interquartile distance, are represented by a circle.

In a next step, the findings of rs133885 were to be replicated in two independent samples that provided similar arithmetic measures, namely a dyslexia sample from Austria (n = 510) and the population-based cohort of the Twin Early Development Study (TEDS, n = 1,081). In the Austrian sample, data on 'dot counting' and 'exact calculation' were available. ANOVA on the combined mathematical measure yielded $P = 0.046$ for the carrier-A model (one-sided). In the TEDS sample, data on reading, spelling and mathematical skills were only available at ordinal scale (ranked in levels from 1 to 5), not as quantitative measure. Here, the distribution of rs133885 genotypes was compared for the mathematical performance between individuals that appeared in the worst-performing writing group (level 1, n = 78), to children in the four other levels. This analysis yielded a borderline significant, one-sided *P*-value of $P = 0.055$.

In both samples, correlating genotypic and phenotypic measures revealed the same direction of effect as in the German sample, with A-carriers performing better in mathematical tasks as compared to individuals homozygous for the G allele. Combining results from all three samples using Fisher's Product revealed a *P*-value of $P = 1.27 \times 10^{-09}$. This value withstands conservative correction for multiple testing (297,086 SNPs, three phenotypes), yielding $P_{corr.} = 0.011$.

Rs133885 is a non-synonymous variant within *myosin 18B (MYO18B)*, coding for a substitution of glutamic acid (E) to glycine (G) at protein level. Database search in SwissProt suggested the existence of two different isoforms, Q8IUG5-1 (MYO18B_long) and Q8IUG5-2 (MYO18B_short). These isoforms share the entire amino acid sequence of MYO18B_short (= MYO18B core), but MYO18B_long has an extended set of 483 aa at the N-terminus of the protein. The amino acid exchange E/G, coded for by rs133885, is located at position 44 of MYO18B_long (E44G).

Expression of *MYO18B_short* and *MYO18B_long* was analyzed in cDNAs derived from multiple human tissues from both, fetal and adult stages, respectively. PCR products for *MYO18B_short* were obtained in all tissues investigated, regardless of their developmental stage. To specifically address the expression pattern of *MYO18B_long*, we investigated the same tissues with a primer pair uniquely addressing the N-terminus. A PCR product of 418 bp length was expected.

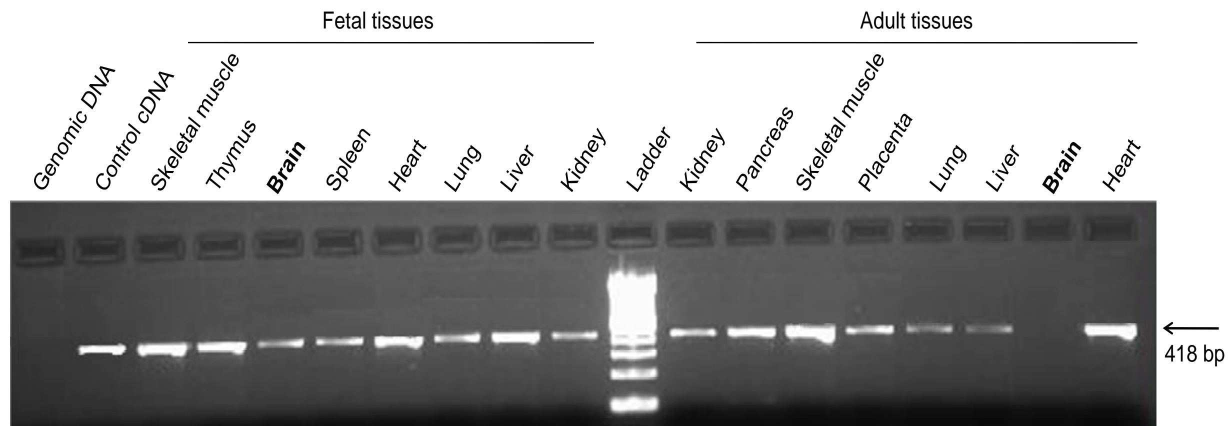


Fig. 16: Expression analysis of *MYO18B_long*. The abundance of the long *MYO18B* isoform was analyzed in diverse human tissues from two developmental stages (fetal, adult). Note the distinct expression change for brain tissues (marked in bold). Tissue order in both panels differs due to different tissue contents. Ladder = 100 bp ladder.

As shown in Fig. 16, *MYO18B_long* is expressed in all fetal tissues. In adults, *MYO18B_long* is expressed in most of the investigated tissues (kidney, pancreas, skeletal muscle, lung, liver, heart; Fig. 16). Notably, no expression signal is observed in adult brain tissue.

In an external cooperative study conducted at the Max-Planck Institute of Psychiatry in Munich, it was analyzed whether the associated variant contributes to anatomical brain changes in 79 healthy controls, using structural MRI. Analysis was restricted to the intraparietal sulcus (IPS), a brain area involved in numerical processing, and results are presented in Fig. 17.

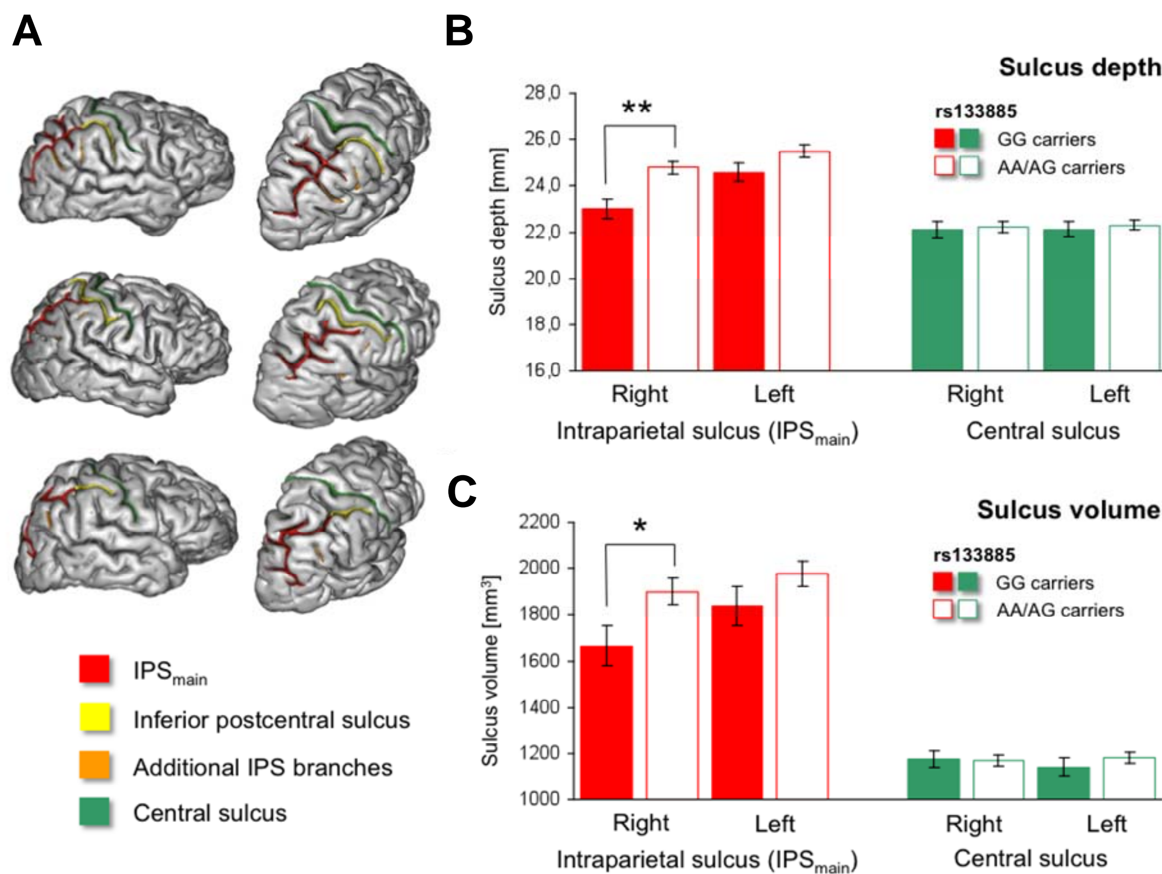


Fig. 17: Rs133885 genotype-dependent differences of intraparietal sulcus (IPS) morphology. (A) Three representative brain cases are shown in standard views. Right IPS segmentations and central sulcus have been automatically assigned. (B) The depths of the right and left IPS_{main} is compared between rs133885 non-A-carriers and A-carriers. The central sulcus is used as control. (C) Analysis of sulcus volumes, with respect to rs133885 A-carriers and non-A-carriers. $*P < 0.05$; $**P < 0.0063$ (Bonferroni correction). Bars show estimated means after correction for age, gender and hemisphere average sulcus depth or total sulcus volume, respectively, and one standard error of mean. Figure by P. Saemann (Ludwig *et al.* in preparation).

As depicted in Fig. 17, the group of rs133885 non-A-carriers showed significantly lower depth of the right intraparietal sulcus (IPS_{main} segmentation) as compared to A-carriers ($P = 0.0010$). This *P*-value was obtained after adjustment for age, gender and depth average. A trend effect was found for the depth of the left IPS_{main} ($P = 0.0587$). Similar, the right IPS_{main} also showed significantly lower volume in non-A-carriers as compared to A-carriers ($P = 0.0334$). After correction for multiple testing, only the depth of the right IPS_{main} remained significant ($P_{corr.} = 0.0080$). No genotype-dependent effects were seen for depth or volume of the central sulcus, which was used as control structure.

5. DISCUSSION

5.1 Genetic factors for dyslexia as qualitative trait

In the general human population, reading and writing skills are normally distributed, and some children perform at the very bottom-end of this Gaussian curve (Shaywitz *et al.* 1992; Shaywitz *et al.* 1995). If other causes such as hearing problems, neurological disorders or inadequate social environments can be excluded, it is probable that these children are affected with dyslexia, one of the most common neurodevelopmental disorders (Shaywitz *et al.* 1990). Dyslexia, also known as reading and spelling disability, is defined as an impairment of reading and writing skills which (i) are unexplained given a child's general intelligence, sociocultural opportunity and educational environment (Schulte-Körne *et al.* 2001b) and (ii) do not disappear with adolescence (Shaywitz *et al.* 1999). The disorder has been shown to cluster in families (Hinshelwood 1907; Stephenson 1907), which suggests the general presence of genetic factors. However, since it is also known that environmental factors play a role in its etiology, dyslexia is best described as a human complex disorder of multifactorial origin (Schumacher *et al.* 2007).

Linkage scans in dyslexia families have identified nine chromosomal loci (DYX1 – DYX9) that were expected to harbor candidate genes conferring susceptibility for dyslexia (Williams & O'Donovan 2006; Schumacher *et al.* 2007). As the results of the linkage studies were pretty consistent between different studies, it was assumed that a limited number of common genetic variants would be involved in dyslexia susceptibility. So far, genetic variations in six candidate genes have been reported to increase risk for dyslexia (Anthoni *et al.* 2007; Schumacher *et al.* 2007). Functional studies have suggested that at least three of them (*DCDC2*, *KIAA0319*, *DYX1C1*) are involved in the process of neuronal migration (Galaburda *et al.* 2006). The migration of neurons to their final location in the highly structured cerebral cortex is a key process during brain development. Disturbed neuronal migration has been proposed to be a neurological mechanism implicated in dyslexia development by both, postmortem brain studies (Galaburda *et al.* 1985) and functional analyses (Chang *et al.* 2005). A fourth gene, *ROBO1*, is known to be involved in axonal guidance and thus also exhibits a particular function that can be associated with dyslexia symptoms (Galaburda *et al.* 2006). The functions of the remaining two genes, *MRPL19* and *C2orf3* (Anthoni *et al.* 2007), still remain elusive, however, an action in energy metabolism has been suggested by the authors.

Although the current data suggest a dyslexia-related functional role for at least four of the six candidate genes, the functional variants causing the given genetic association findings have not yet been conclusively identified. For *DCDC2*, Meng and colleagues (2005) had reported a

deletion of 2,445 bp in intron 2 to be of putative causative function (Meng *et al.* 2005b). This deletion contains a sequence of different short tandem repeats (STRs) with variable numbers of the single repetitive elements. This compound STR marker harbors potential binding sites for a number of transcription factors. If the deletion or the compound STR marker would be truly causative for dyslexia, a significantly biased transmission for any of these alleles would be observed in our familial dyslexia (DYS-) sample. We therefore genotyped the intronic deletion region and sequenced the different STR alleles in our German parent-child trios of the DYS-sample.

To assess whether population differences between Germany and the initial US-sample are present, we first compared the frequency of the deletion and the STR alleles for the parental chromosomes. In the German DYS-sample, we could not observe three of the alleles that were present in the US-sample (alleles number 2, 7, 8; see section 4.1.1). However, we identified two new alleles (19, 20) that had not been found in the US-study by Meng and colleagues (2005). As each of these alleles provided with a very low allele frequency in its respective sample ($\leq 0.5\%$), the failure of detection in the other sample was in accordance with our expectations. For alleles observed in both samples, allele frequencies were similar between the two studies, indicating that no population-specific effects are present and statistical results are comparable.

Transmission disequilibrium test (TDT) did not yield significant *P*-values for the deletion or any of the STR alleles in our data set (lowest *P*-value observed for allele 9, $P = 0.2$, see section 4.1.1). Also, combining the minor alleles as suggested in the initial study (Meng *et al.* 2005b) did not reveal any obvious association. In their US-study, Meng and colleagues had suggested that the deletion and different alleles of the STR marker would alter the sequence of transcription factor binding sites such as those for two brain-specific factors, PEA3 (binding site AGGAAA) and NF-ATp (AGGAAG; (Meng *et al.* 2005b)). PEA3 has been found involved in peripheral motor neuron arborization (Laing *et al.* 2000), and NF-ATp is suggested to mediate embryonic axon extension that is necessary for the formation of neuronal connections (Graef *et al.* 2003). Given the failed replication in our sample, we cannot support the hypothesis raised by Meng and colleagues (2005b) and exclude the intronic deletion or any of the STR marker alleles to be causative for dyslexia in the German population.

We ourselves have previously reported a two-marker risk haplotype, located in intron 7 of *DCDC2*, to be associated with dyslexia ([rs793862-rs807701 (A-C)]; (Schumacher *et al.* 2006)). It was proposed that, given its intronic location, the risk haplotype might have a regulatory effect on *DCDC2* expression levels (Schumacher *et al.* 2006). In the same study, we also investigated whether SNPs within *KIAA0319*, the second candidate gene within the dyslexia susceptibility locus DYX2 on chr. 6p22, would confer genetic risk for dyslexia. We could not show any

association with dyslexia as qualitative trait, suggesting that variants in *KIAA0319* do not contribute to the dyslexia phenotype in the German population (Schumacher *et al.* 2006). However, our analysis at that time had not included all SNPs that were subsequently reported to be associated with dyslexia in a joint analysis of two UK samples (Harold *et al.* 2006). In that study, nine SNPs predominantly located in and around exon 1 of *KIAA0319* showed significant *P*-values for both, dyslexia as qualitative trait and various dyslexia-related endophenotypes. The strongest associated SNP was rs761100 ($P = 0.00004$). To avoid that the failed replication of *KIAA0319* in our first study (Schumacher *et al.* 2006) was only caused by low linkage disequilibrium (LD) with the true causal variant, we performed a second investigation of the *KIAA0319* locus and included all *KIAA0319* SNPs that were (i) significant in Harold *et al.* (2006) and (ii) had not been analyzed in our sample previously. For none of these six *KIAA0319* SNPs, significant *P*-values for an association with dyslexia were obtained in this new investigation of the German DYS-sample. We were thus not able to replicate the findings by Harold *et al.* (2006) and, even though we further extended the SNP set, could not find any evidence for a contribution of *KIAA0319* to dyslexia in the German population. However, given the consistent findings for SNPs in *KIAA0319* to be associated with dyslexia in English-speaking populations (Francks *et al.* 2004; Cope *et al.* 2005a; Harold *et al.* 2006), *KIAA0319* cannot be generally excluded as dyslexia candidate gene. It might be possible that *KIAA0319* contributes to a different subgroup of dyslexia patients, and that the specific ascertainment strategies or recruitment criteria biases the samples towards the one or the other genetic variant. The involvement of a language specific effect (German / English) is rather unlikely, as the findings for *KIAA0319* also failed replication in an US-American study (Meng *et al.* 2005b).

As mentioned before, *KIAA0319* and *DCDC2* show functional similarities, as both genes are involved in neuronal migration (Galaburda *et al.* 2006). Using RNA interference assays in murine embryos, it was demonstrated that a lack of *Kiaa0319* and *Dcdc2* results in an interruption of the normal direction of the elongating neurons in the developing brain (Meng *et al.* 2005b; Paracchini *et al.* 2006). Hereby it is not yet known whether both genes act independently from one another, or whether they show interaction effects in particular steps of the same pathway. To obtain a clearer picture of this, we performed a statistical interaction analysis on the available SNP data. Statistical interaction is observed in situations where the genetic variant at one locus modifies the genetic effect at a second locus. Such epistatic effects are likely to occur between genes involved in same pathways or regulatory networks (Carlson *et al.* 2004). Harold and colleagues (2006) had reported first evidence for interaction between two SNPs in *KIAA0319* (rs761100, rs4504469) and the *DCDC2* intronic risk haplotype (lowest

P-value = 0.007). We thus investigated the same marker combinations in our German DYS-sample.

No significant interaction *P*-values were found for dyslexia as qualitative trait. However, we observed a trend of interaction (*P* < 0.1) for the *DCDC2* two-marker risk haplotype and rs4504469, located in exon 4 of *KIAA0319* (*P* = 0.053). Although this *P*-value does not reach significance at statistical level, this is a remarkable finding given the limited sample size, and provides some further evidence for the presence of interaction between *KIAA0319* and *DCDC2*. Additionally, we found a significant interaction for the *DCDC2* two-marker risk haplotype and rs761100, located in intron 1 of *KIAA0319*, for the quantitative subdimension 'word-reading' (*P* = 0.0351). This nominally significant *P*-value, which does not require correction for multiple testing as we replicated a specific hypothesis, is of particular interest, as the quantitative measure 'word reading' reflects the main inclusion criteria for the UK sample. As the German DYS-sample is ascertained based on spelling ability, the results suggest that an analysis of the subdimension 'reading' is better comparable to the UK sample and thus provides with more power to detect a genetic effect. These results suggest that in presence of the intronic *DCDC2* risk haplotype, the genetic effect of rs761100 is modified and a contribution of *KIAA0319* can also be seen in the German DYS-sample. Our interaction findings thus replicate the results of the original study (Harold *et al.* 2006) and, therefore, they can be seen as further evidence for an interaction between *KIAA0319* and *DCDC2* for the particular endophenotype. Given that both genes share functional similarities in the process of neuronal migration, one possible explanation for this epistatic effect could be that a lower expression level of *KIAA0319* (as mediated by the intronic variant rs761100) alone does not cause dyslexia, but would do so in individuals where, additionally, *DCDC2* is also expressed at very low amounts due to the presence of the two-marker risk haplotype. This would be a synergistic episatic effect between the two loci which is difficult to detect with single marker analysis (Carlson *et al.* 2004). This provides an explanation why the analysis of *KIAA0319* SNPs alone has not yielded any significant results in our German DYS-sample. Notably, epistatic interactions have already been observed in other diseases such as Hirschsprung disease (Bolk *et al.* 2000) and myocardial infection (Tiret *et al.* 1994; Butt *et al.* 2003), implementing that interaction mechanisms might generally contribute to human disease susceptibility.

The effect size that each of the dyslexia candidate genes known to date has on dyslexia susceptibility is considered to be rather small (Fisher & Francks 2006), which suggests that other, so far unidentified genetic factors do exist. The identification of new genetic variants that confer risk for dyslexia as categorically defined trait was therefore attempted using a genome-wide association study (GWAS). This approach, in which frequencies of genetic variants are

compared between groups of affected and unaffected individuals, provides a systematic tool for the correlation of genetic variants with disease status over the entire human genome. One main advantage hereby is that no prior hypotheses on potential genomic locations or candidate genes are required. Thanks to recent advances in the field, with international consortia building extensive SNP and haplotype maps of the human genome (HapMap Consortium 2003; 2005), and companies such as Affymetrix and Illumina developing new high-throughput genotyping technologies, the performance of large-scale GWAS became feasible in the last years. In our study, 200 individuals of the German DYS-sample were analyzed genome-wide for about 300,000 SNPs. Allele frequencies in the group of cases were compared to those in 926 population-based controls, and results were combined with genome-wide data from the UK (385 cases, 1406 controls) in a meta-analysis, to increase power. In order to control for population stratification, statistical control measures such as multi-dimensional-scaling (MDS) were included in the analysis (Li & Yu 2008; McCarthy *et al.* 2008). Interestingly, MDS analysis between the UK and German data did not reveal any individual outliers at sample level. Some negligible differences between the populations were shown to exist, which is in accordance with previous findings (Heath *et al.* 2008; Neale *et al.* 2008), however, their extent indicated that population stratification between the two samples would not affect statistical results considerably. It was thus decided to include the information on 'country' as covariate in the statistical analysis, to control for these minor population stratifications.

The statistical analysis of the combined data from 585 cases and 2326 controls did not yield any SNP to show genome-wide significance ($P < 7.2 \times 10^{-08}$; (Dudbridge & Gusnanto 2008)), and only a limited number of SNPs showed consistent results between the two single samples from the UK and Germany. It is likely that ascertainment strategies and language-specific characteristics such as different orthographic transparencies (which presumably also partially involve some different cognitive performances) might account for part of the inconsistencies and for the limited power. However, 63 SNPs provided with P-values below 10^{-04} (see section 4.2.1). Using two different strategies, markers for replication in the European-wide NeuroDys sample were chosen. While the first approach was a top-down strategy based on the list of the aforementioned top SNPs, the second approach involved a large separate genome-wide pooling analysis on about one million SNPs. The two replication rounds included all different NeuroDys samples that were available at the respective points in time (section 3.7.3, Tab. 5).

For the first replication, 28 SNPs derived from the top list of 63 markers were successfully genotyped in 1409 cases and 1959 controls. Statistical analysis yielded one marker, rs10512712, to be significantly associated with dyslexia in the replication sample ($P = 0.0147$). Combining GWAS and replication data revealed a P-value of $P = 1.19 \times 10^{-07}$. This was lower than the

P-value provided by the GWAS data set alone (P = 1.56 x 10^{-05}), which is accepted as additional measure for independent replication and confirms that the association found in both samples referred to the same risk allele (Skol *et al.* 2006). Calculation of the odds ratio (OR) for the C allele of rs10512712 in the combined sample revealed an OR of 1.2 [CI = 1.16 – 1.38], indicating that rs10512712 provides with a small but distinct genetic effect on dyslexia susceptibility. This OR is in the range of many other risk variants which have been identified in complex disorders (Altshuler *et al.* 2008), suggesting a true contribution of rs10512712 to dyslexia risk.

Rs10512712 is located in an intergenic region on chr. 5p13.1. Flanking genes reported so far are the *homo sapiens disabled homolog 2, mitogen-responsive phosphoprotein* (*DAB2*) gene, which is located 300 kb in telomeric direction, and the *prostaglandin E receptor 4* (*PTGER4*) gene, almost 1 Mb distant centromerically. *DAB2* has been found expressed in normal ovarian epithelial cells and showed significant downregulation in ovarian cancer cell lines (Mok *et al.* 1994). The protein encoded by *PTGER4* is one of four receptors known to bind prostaglandin E2 (PGE2), thus participating in prostaglandin signaling (Regan *et al.* 1994). Knockout studies in mice suggest that PTGER4 is involved in the formation of the circulatory system (Segi *et al.* 1998) and in the initiation of skin immune responses (Kabashima *et al.* 2003). Variation within *PTGER4* have previously been found associated with Crohn disease (Libioulle *et al.* 2007), a disorder that causes inflammation of the gastrointestinal tract. Thus, neither *DAB2* nor *PTGER4* shows an explicit function that would directly imply an involvement in the neurocognitive aspects of dyslexia. However, it is possible that rs10512712 executes its function through yet unidentified genes within the region or through long-range *cis-* or *trans*-effects.

With rs10512712, only one out of 28 GWAS top SNPs was successfully replicated. This suggests that dyslexia as qualitative trait is far more heterogeneous than was initially thought based on the consistent linkage findings (Schumacher *et al.* 2007). The inclusion of roughly 600 dyslexia cases and almost four times as many controls seems not to be enough to detect moderate risk variants with reasonable power. This is in concordance with current estimates, which suggest that the detection of a risk allele of 20% frequency and an effect size of 1.2 with 90% power would require about 8,600 samples (Altshuler *et al.* 2008). The respective sample size would be further increased if variants with lower allele frequencies or smaller genetic effect sizes are addressed. We therefore assumed that in our sample, power issues prevent the true causative variants from appearing among the top SNPs of the GWAS. Thus, SNPs showing the lowest *P*-values in the GWAS are likely to be false-positives, explaining the low rate of replicated markers in our study.

To circumvent the need to individually genotype several thousands of top markers in the replication sample, we performed a second, independent GWAS using large numbers of pooled

cases and controls. This method has been shown to be an effective first genome-wide screening (Kirov *et al.* 2008) and was performed by genotyping pools of three different populations, namely Finland (286 cases / 321 controls), UK (426 cases / 219 controls) and Central Europe (532 cases / 912 controls). Splitting the NeuroDys sample in three groups was performed in order (i) to control for allele frequency biases due to population stratification, and (ii) to use them for replication purposes as independent sets of cases / controls. Allele frequency estimates (AFEs) were calculated based on the allele-specific fluorescent ratios (see section 3.8.2).

Statistical analysis of the pooling data yielded a list of SNPs showing allele frequency differences in the same direction in all three populations. In combination with results from the individual GWAS data, pooling results were used to prioritize SNPs for a second step of replication. Additionally, SNPs being highly significant in the pooling analysis alone or providing functional evidence / multiple hits at one locus were included. In the second replication, 33 SNPs were successfully genotyped in 1526 cases and 2261 controls from the NeuroDys sample. To assess the concordance between pooling and individual genotyping, we compared the estimated allele frequencies (pooling data) and the true frequencies (based on individual genotyping) for these SNPs. This analysis revealed obvious frequency differences between the two data sets. Reasons therefore are expected to be mainly due to technical errors, which include differential allelic amplification, pool formation errors and frequency measurements (Jawaid & Sham 2009). Furthermore, it points out that pooling contributes to an over- or underestimation of allele frequencies, thus increasing the risk for both, false-positive and false-negative findings. While false-positive findings would be identified by subsequent individual genotyping and can be removed, false-negative findings cannot be corrected for and, thus, will result in the miss of true association findings.

Of the 33 SNPs, seven SNPs were significantly associated with dyslexia in the replication sample. One SNP, rs461119, provided with a significant *P*-value even after correction for multiple testing ($P = 0.0015$, $P_{corr.} = 0.0485$). Two other promising SNPs, rs12344734 ($P = 0.0016$) and rs7381 ($P = 0.0029$), remained with a trend of significance after statistical correction ($P_{corr.} \leq 0.1$, see section 4.2.1). For each of these three candidate SNPs, the calculated ORs were in the expected range for complex disorders (rs461119, OR = 1.18 [1.07 – 1.31]; rs12344794, OR = 1.27 [1.09 – 1.47]; rs7381, OR = 1.32 [1.10 – 1.58]). Notably, neither of the three SNPs (or their respective tSNPs) appeared among the top SNPs of the individual GWAS ($P = 0.465$ (rs420121, as tSNP for rs461119), $P = 0.432$ (rs7381), $P = 0.86$ (rs9942926, as tSNP for rs12344734)). This further supports the aforementioned power issues and illustrates the limited size of our initial GWAS sample.

Rs461119 is located on chr. 21q21.3, in intron 1 of the *ionotropic glutamate receptor, kainate 1* (*GRIK1*) gene. *GRIK1* belongs to the gene family coding for ionotropic glutamate receptors which are considered to be the predominant excitatory neurotransmitters in the mammalian brain (Bowie 2008). They are activated in a variety of neurophysiologic processes, so e.g. during the induction of short- and long-term potentiation (Contractor *et al.* 2001). Notably, *GRIK1* is subject to alternative splicing and, particularly, massive RNA editing within its second transmembrane domain (Barbon & Barlati 2000). RNA editing is a posttranscriptional mechanism by which mRNA sequences are enzymatically modified by deamination (removal of an amino group) or transamination (addition of an amino group; (Chan 1993)). These modifications lead to a change in the mRNA sequence, as for instance adenosine is transformed into the guanosine-like inosine, or cytosine becomes uracil. As consequence, the diversity of mRNAs and proteins increases (Schaub & Keller 2002). RNA editing also contributes to the regulation of gene expression levels as has been suggested for non-coding RNAs (Kurokawa *et al.* 2009) and small RNAs (Carthew & Sontheimer 2009). The influence of RNA editing on mammal brain and cognition and its contribution to neurodevelopmental disorders has already been demonstrated (Sodhi *et al.* 2001; Rula *et al.* 2008; Iwamoto *et al.* 2009). Rula and colleagues (2008) showed that the ratio of edited / non-edited forms of the mRNA coding for the alpha3 subunit of the gamma-aminobutyric acid (GABA(A)) receptors in murine brain is regulated in a spatiotemporal manner. The authors also observed that the non-edited form is more rapidly activated and more slowly deactivated as compared to the edited form, which might reflect the excitatory robustness that is critical for normal synapse formation. In an earlier study, Sodhi and colleagues (2001) had already shown that the serotonin-2C receptor shows a reduced RNA editing level in frontal cortex samples obtained from schizophrenia patients, as compared with control samples. The RNA editing of *GRIK1* is suggested to alter the calcium flow regulation and has been shown to be of different extent in specific brain regions (Paschen & Djuricic 1994). Together, these results provide evidence for a possible functional impact that might be mediated by the observed association of rs461119 with dyslexia.

Rs12344734 (chr. 9q21.13) locates to an intronic region within the *transmembrane channel-like 1* (*TMC1*) gene. Although its particular function is not well established, *TMC1* is known to be required for normal function of the hearing system, as mutations in *TMC1* have been found associated with both, postlingual hearing loss (OMIM 606705) and profound prelingual deafness (OMIM 600974; (Kurima *et al.* 2002)). A normal functioning of the hearing system is required for an appropriate development of reading and spelling skills, as the segments of spoken language have to be matched to their appropriate visual representation, i.e. the written graphemes (Wallace 2009). This process is disturbed in some, even though not all dyslexic

children (Galaburda *et al.* 1994), and the present data suggest that subtle variations within *TMC1* might contribute to dyslexic subgroups that mainly suffer from impairments in some aspects of auditory processing.

The third candidate SNP, rs7381, is located within the *fibulin 1 isoform D* (*FBLN1*) gene on chr. 22q13.31. This gene codes for a secreted glycoprotein that is predominantly found in human connective tissues (Argraves *et al.* 1989). *In situ* hybridization revealed that the mouse homologue *Fbln1* is expressed in multiple tissues during embryonic organogenesis, including also mesenchymal parts of the central nervous system (Zhang *et al.* 1996). Mice with targeted homozygous inactivation of *Fbln1* died at birth due to ruptures of small blood vessels, malformations of kidney and a delayed development of lung alveoli (Kostka *et al.* 2001). In humans, the entire *FBLN1* gene is deleted in the 22q13.3 deletion syndrome, a complex malformation syndrome which is additionally characterized by marked speech and language delay (Cusmano-Ozog *et al.* 2007). Although a direct connection between the *FBLN1* gene and the speech-related impairments in this syndrome has not yet been established, the current data provide support for a potential involvement of the gene in cognitive aspects which might also be of importance for the development of dyslexia.

By combining pooling and individual genotyping data, we were able to identify three additional new candidate SNPs located in genes with putative dyslexia-related biological relevance. These findings provide promising new starting points for further functional investigation, however, some limitations have to be considered. Replication has not been performed in an independent sample, as a part of the individuals included in the replication study overlapped with those contained in the pooling analysis. Also, for each of the three associated SNPs, one NeuroDys sample failed to replicate the direction of effects. Notably, these samples provided with very small sample sizes and can therefore be considered as false-negative findings.

All together, the GWAS performed in this thesis suggests four new dyslexia susceptibility variants (rs10512712, rs461119, rs7381, rs12344734). Replication in independent samples will now be the next step to support these findings and to identify the underlying biological mechanisms.

5.2 Quantitative measures of dyslexia-related endophenotypes

The analysis of dyslexia as categorically defined trait provided evidence for four SNPs to be associated with this neurodevelopmental disorder in the European population. Although the effect sizes for each of these new candidate markers were in the range that is expected for

complex disorders (Altshuler *et al.* 2008), the SNPs only explain a small fraction of dyslexia cases. This suggests that dyslexia is among the more heterogeneous human traits, and that the analysis of the qualitative trait, so far, lacks the power to detect the underlying common genetic risk variants (Gottesman & Gould 2003). It has been postulated that the correlation of quantitative endophenotypes, of which a complex phenotype is made up, would increase the chance for true findings, as heterogeneity could be restricted and the measures would be more direct than their combination in compound disorders (Gottesman & Gould 2003). We thus correlated some of the quantitative measures that are associated with reading and spelling with genotypic data on both, candidate gene and genome-wide level.

Several studies have shown that individuals with dyslexia have marked weaknesses in phonological working memory and perform poorly in memory-related tasks (Baddeley & Wilson 1993; Palmer 2000; Swanson 2006). A typical task used to investigate verbal working memory is a digit span task, which requires a subject to temporarily store digits and then retrieve them from phonological memory in order, in both forward and backward direction. This quantitative measure is stable even in adults, and is thus considered to be one cognitive correlate within the complex dyslexia phenotype (Hulslander *et al.* 2004).

In 2006, it was shown that the *N-methyl-D-aspartate receptor subunit 2B* (*GRIN2B*) gene contributes to human memory performance in the general population (de Quervain & Papassotiropoulos 2006). *GRIN2B* is located within the chr. 12p12-p13 region that had provided suggestive, but replicated evidence for linkage to phonological memory in US-American families (Brkanac *et al.* 2008). The gene has also been found to be associated with attention-deficiency / hyperactivity disorder (ADHD) in a Canadian sample (Dorval *et al.* 2007). ADHD is a comorbid disorder of dyslexia, as approximately 25% to 40% of children with dyslexia also present with symptoms of ADHD (Pennington 2006). This comorbidity may, at least partially, be attributable to common genetic influences (Willcutt *et al.* 2007). Working memory is among the cognitive processes suggested to be shared between the two disorders (Tiffin-Richards *et al.* 2008). We therefore considered *GRIN2B* a candidate gene for memory-related aspects in dyslexia, and analyzed the genetic variation within this gene in the German DYS-sample.

We did not find any evidence for an association between markers in *GRIN2B* and dyslexia as categorical trait, as only one nominal significant *P*-value was obtained (rs933614, $P = 0.013$). However, we found four SNPs to be significantly associated with the quantitative subdimension 'verbal short-term memory' (rs1012586, rs2268119, rs2216128, rs2192973; lowest *P*-value = 0.0243 for rs2268119, see section 4.1.2). All four associated SNPs are located in intron 3 of *GRIN2B*, in close proximity to the markers suggested for general human memory performance (de Quervain & Papassotiropoulos 2006) and for ADHD (Dorval *et al.* 2007).

As there is increasing evidence that parent-of-origin effects play a role in human brain, behavior and cognition (Cattanach & Kirk 1985; Isles & Wilkinson 2000), we also considered parent-of-origin effects in our study. This strategy was supported by some controversial evidence for imprinting effects in ADHD (Hawi *et al.* 2005; Laurin *et al.* 2007). We thus followed-up the positive signals of the single marker analysis and found that for each of the four aforementioned SNPs, a statistical significant biased transmission for 'short-term memory' was observed when maternal transmissions only were considered (lowest P-value = 0.0001 for rs1012586). Probands who inherited the risk allele from their mother displayed a significantly better performance in the respective memory tasks. This suggests the presence of a maternal parent-of-origin effect for *GRIN2B* on memory performance in dyslexic children.

Parent-of-origin effects refer to the silencing of one parental allele during early development which, in its most extreme, result in monoallelic gene expression (Isles & Wilkinson 2000). The underlying imprinting mechanisms probably involve allele-specific DNA methylation and / or histone modification (Delaval & Feil 2004), however, the specific silencing procedure for *GRIN2B* remains to be identified. A contribution of imprinting mechanisms to human cognitive performance has already been demonstrated, for instance for Angelman syndrome and Prader-Willi syndrome. Both disorders present varying degrees of cognitive impairments and have been shown to be associated with an imprinted gene cluster located on chr. 15q11 – q13 (Isles & Humby 2006).

GRIN2B shows a distinct expression pattern in human brain, with high levels being found in the frontal cortex as well as in hippocampal pyramidal cells (Schito *et al.* 1997). The gene was also shown to influence synaptic plasticity (Kutsuwada *et al.* 1996). An overexpression of *Grin2b* in the forebrain of mice results in an increased activation of N-methyl-D-aspartate (NMDA) receptors, with mice displaying a superior performance in learning and memory tasks (Tang *et al.* 1999). Interestingly, according to the UCSC genome browser, the genomic region of *GRIN2B* contains several expressed sequence tags (ESTs) that do not belong to any known exons of *GRIN2B*. Two of them, CD514667 and AL133734, are located in the region between the four associated SNPs (rs1012586 and rs2192973), and have been detected in white matter and amygdala, respectively. It is therefore also possible that the causal variant of the given association findings for *GRIN2B* and verbal short-term memory might be attributable to other yet uncharacterized expressed sequences within the genomic region of *GRIN2B*.

Notably, in the study by Dorval and colleagues (2007) on ADHD, no significant P-values were observed for the endophenotype measure 'short-term memory', even though that study had applied exactly the same measures as were used in the present study (digit span forwards / backwards). This suggests that the effect of *GRIN2B* on memory performance is more

prominent in, if not exclusive to, children with dyslexia. Alternative hypotheses may be that the Dorval *et al.* (2007) study failed to find this association because they did not include any of the four markers found to be significantly associated in our study, or that different variants within the same gene are causative for ADHD and dyslexia. To further follow-up this question, it would be interesting to see whether the four significant markers from our study show association with memory performance in the Canadian ADHD sample.

Our findings of genetic imprinting effects within *GRIN2B* in the German DYS-sample adds to the increasing evidence that parent-of-origin effects play an important role in human cognition (Isles & Wilkinson 2000; Wilkinson *et al.* 2007). In another study included in this thesis, we analyzed parent-of-origin effects within a novel imprinted gene on chr. 2p12, the *leucine-rich repeat transmembrane neuronal 1 (LRRTM1*) gene. *LRRTM1* was considered a dyslexia-related candidate gene because variation within the gene has been suggested to contribute to human brain asymmetry and schizophrenia (Francks *et al.* 2007). Asymmetrical brain function is pronounced in humans and has been implicated in human cognition, behavior and emotions (Hughdal & Davidson 2003), and several studies have reported that a disturbed lateralization can be found in dyslexic brains (Eckert & Leonard 2003; Wijers *et al.* 2005; Penolazzi *et al.* 2006). As brain asymmetry cannot be easily assessed in probands, the measure of 'relative hand skill' has been suggested as respective intermediate phenotype: Left-handedness in humans is associated with reductions or reversals of normal brain asymmetries, and has been shown in particular for cerebral cortex areas implicated in the language system (Geschwind *et al.* 2002; Mevorach *et al.* 2005).

Our German DYS-sample also provides with quantitative data on the extent of left- and right-handedness in the dyslexic probands and their siblings. We used these data to analyze whether imprinted genetic variants at the *LRRTM1* locus are involved in asymmetrical brain function in dyslexic patients.

Fourteen SNPs were genotyped. Three of them constituted the associated three-marker haplotype (rs1446109, rs1007371, rs723524) as suggested by Francks and colleagues (2007), and 11 tSNPs captured the genetic variation at the entire locus. After statistical analysis using TDT, none of the markers revealed evidence for an involvement of *LRRTM1* in dyslexia as qualitative trait, which is in accordance with previous reports that also did not reveal any association with dyslexia *per se* (Francks *et al.* 2007). However, seven out of 14 markers showed significantly biased transmission when 'relative hand skill' was analyzed (lowest *P*-value = 0.0055, rs2862286). This is more than would be expected by chance alone, indicating that *LRRTM1* might generally contribute to handedness. The fact that none of the *P*-values withstands correction for multiple testing could be explained by the limited power of our sample, due to

the low number of left-handed individuals (n = 63). Alternatively, the seven nominally significant SNPs could also provide a haplotypic background for the true causative variant which had not been included in our study.

The subsequent analysis of potential imprinting effects provided evidence for a maternal effect for five out of 14 SNPs. Among these SNPs were two of the three haplotype-forming markers (rs1446109, $P = 0.0322$; rs1007371, $P = 0.0364$) and a non-synonymous variant, rs6733871 ($P = 0.016$). For the haplotype-forming SNPs, the analysis of the phenotypic direction revealed that the minor alleles were significantly overtransmitted from the mothers to left-handed individuals. This is in the opposite direction compared to the initial study, so the findings reported by Francks and colleagues (2007) could not be replicated in our approach. Our findings for rs6733871, which codes for an amino acid substitution (Asp330Ser) of the LRRTM1 protein, provide new evidence for the contribution of a structural variant to human brain lateralization. However, rs6733871 was also analyzed in the initial study by Francks and colleagues (2007) but did not show a significant effect. This might be due to different ascertainment criteria of the samples or could be explained by the fact that neither of the two studies has identified the true causative variant. *LRRTM1* is predominantly expressed in human forebrain regions such as thalamus and cerebral cortex (Francks *et al.* 2007). The encoded peptide LRRTM1 is a transmembrane protein that induces presynaptic differentiation in contacting axons (Linhoff *et al.* 2009). *Lrrtm1*(-/-) mice show an altered distribution of the vesicular glutamate transporter (VGLUT1; (Linhoff *et al.* 2009)), which provides evidence for LRRTM1 to be involved in synaptic signaling mechanisms. Although the biological functions that are mediated by rs6733871 still have to be identified, our findings suggest a new imprinted structural variant within *LRRTM1* to contribute to human brain asymmetry.

Apart from the candidate gene approach, we also analyzed quantitative measures in genome-wide data of 200 German dyslexia individuals. As traits closely related to reading and writing were protected from independent analysis by the NeuroDys consortium agreement, we focused our approaches on more distantly related endophenotypes. These analyses do not require any control samples, as genetic variants are correlated with quantitative measures that are available within the dyslexia probands.

Some cognitive endophenotypes of dyslexia can be assessed by electroencephalography (EEG) measurements such as mismatch negativity (MMN), in which the automatic response of the brain to any change in auditory stimulation is represented (Schulte-Körne *et al.* 1998a). MMN is considered to be the neural correlate of speech perception and has been repeatedly found attenuated in dyslexic children (Schulte-Körne *et al.* 1998a; 2001a; Lachmann *et al.* 2005; Alonso-Bua *et al.* 2006; Corbera *et al.* 2006). Analyzing the genome-wide data with respect to

performance in both components of MMN, namely MMNa and MMNb (see section 2.1.4, Fig. 2), yielded 19 SNPs showing *P*-values ≤ 10^{-05}. These SNPs were included in a subsequent replication study in an independent set of 186 children.

In the initial GWAS comprising 200 individuals of the German DYS-sample, two SNPs (rs1365152, $P = 4.27 \times 10^{-08}$; rs2114167, $P = 1.17 \times 10^{-07}$) were strongly associated with the earlier component of the mismatch negativity, MMNa. Both SNPs are located on chr. 3q23 within the genomic region of *calsyntenin 2* (*CLSTN2*), a gene coding for a post-synaptic membrane protein (Hintsch *et al.* 2002). Genetic variation within *CLSTN2* has previously been found associated with verbal working memory in the general population (Papassotiropoulos *et al.* 2006). *CLSTN2* shows a cell-type specific expression pattern and was suggested to play a role in excitatory synaptic transmission (Hintsch *et al.* 2002). The two SNPs are highly correlated (D′ = 1, $r^2 = 0.91$), indicating that they are not transmitted independently from one another and thus depict the same genetic effect. The initial findings on both SNPs could not be replicated in the second independent sample of 186 German dyslexics (rs1365152, $P = 0.83$; rs2114167, $P = 0.09$). This was unexpected, as the initial results provided with genome-wide significance and were thus considered to be true findings. The failure of replication might be explained by the different dyslexia severity grades of the individuals used in both samples: The children included in the replication step were less severely affected with dyslexia compared to the probands of the initial sample. It is thus possible that the genetic effects of rs1365152 and rs2114167 on MMNa are restricted to children who are severely affected with dyslexia.

Rs1365152 and rs2114167 are located in intronic regions of *CLSTN2*. It can thus be suggested that they provide with regulatory effects on *CLSTN2* expression levels. If this regulation occurs in a spatially-restricted manner, e.g. in synapses, it is envisageable that *CLSTN2* contributes to different transmission rates, and the gene could be involved in the automatic brain response to auditory stimulation. Evidence for a region-specific expression difference in brain has indeed already been suggested by Hintsch and colleagues (2002), who showed that *CLSTN2* mRNA levels vary between different types of neurons (Hintsch *et al.* 2002). The genetic data, together with the functional evidences described above, make *CLSTN2* remaining a promising candidate gene for an involvement in some aspects of MMN measurements, at least in severely affected dyslexic children.

Another SNP, rs4234898, was found to be significantly associated with MMNb in both, the GWAS ($P = 6.20 \times 10^{-06}$) and the replication sample ($P = 0.00146$). The association result in the combined sample (n = 386) withstood genome-wide correction for multiple testing ($P_{corr.} = 0.015$), making rs4234898 a highly probable genetic variant for a contribution to MMN performance. A second SNP at the same locus (rs11100040, 8.5 kb distant from rs4234898) was

also significantly associated with MMNb in the GWAS sample ($P = 0.0006$). However, this *P*-value did not make rs11100040 appear among the top SNPs that were initially chosen for replication. We thus subsequently genotyped rs11100040 in the replication sample and confirmed the initial results, with $P = 0.03$. Both SNPs, which show neither interaction nor high LD, form a two-marker haplotype which is also significantly associated with MMNb ($P = 6.71 \times 10^{-08}$). As this *P*-value is not as low as for rs4234898 alone, this indicates that rs11100040 does not contribute to the given association independently from rs4234898.

Rs4234898 lies within a gene desert region on chr. 4q32.1, with the nearest genes (*cathepsin O preproprotein (CTSO), trypthophan 2,3-dioxygenase (TDO2)* and *platelet-derived growth factor C (PDGFC)*) being more than 150kb away. Given its intergenic location, we hypothesized that rs4234898 might mediate its biological function via regulatory effects on either neighbouring (*cis*) or more distant (*trans*) genes. To address this question, an expression data set based on Dixon *et al.* (2007) was analyzed, the 'mRNA-by-SNP browser'. This database contains genome-wide genotypes of 400 children which are correlated with Affymetrix-based expression data obtained from lymphoblastoid cell lines of the same individuals. These data thus provide valuable information on allele-specific expression.

This *in silico* analysis yielded evidence for a *trans*-regulational effect on the *solute carrier family 2 member 3 (SLC2A3)*. We subsequently confirmed this finding experimentally in EBV-transformed blood cell lines from children of our German DYS-sample. In both data sets, individuals carrying at least one T allele at rs4234898 showed lower expression levels of *SLC2A3*. For rs11100040, the second associated SNP at the locus, similar results were found in both, the 'mRNA-by-SNP browser' and our own expression data. To further strengthen our observation, we also analyzed a potential effect of the rs4234898-rs11100040 haplotype in the EBV-transformed cell lines. As expected, individuals carrying the T-T haplotype showed lower *SLC2A3* expression levels than individuals without the T-T haplotype. Given the small sample numbers included in this analysis (4 vs. 13), statistical simulation using permutation was done, and the results indicated that the findings were indeed true.

In our expression data, the two-marker haplotype was more significantly associated with the expression levels of *SLC2A3* as compared to rs4234898 alone. Given the small sample numbers, this effect could either be random, or it could indicate that the two markers provide a haplotypic background for the true causal variant. However, the latter is not reflected in the genetic data where the haplotype was less significantly associated than rs4234898 alone. This discrepancy will have to be assessed in further studies, and the respective results will provide insight into whether rs4234898 is the causal variant, or whether the latter resides on the haplotypic background formed by rs4234898 and rs11100040.

SLC2A3 belongs to the family of facilitative glucose transporters (SLC2A) and, among these, it is the predominant member abundant in neurons, where it has been suggested to provide energy for synaptic transmission (Maher & Simpson 1994; McCall *et al.* 1994). Mouse studies have shown that the rodent homologue of *SLC2A3*, *Glut3*, plays a role in axonal and / or dendritic transport (Mantych *et al.* 1992). Members of the SLC2A-family show high sequence similarities with one another (Joost & Thorens 2001). We thus compared the sequences of the probes used within Dixon *et al.* (2007) to the different SLC2A-sequences and identified that they did not exclusively bind to *SLC2A3*. Instead, based on their sequence, they were also able to detect transcripts of other genes of the SCL2A group, in particular *SLC2A14*. Therefore, in our functional assay, we designed the probe in a 5′ region that is specific for *SLC2A3* and distinguishes this specific variant from other members of the family. As we observe similar effects in our data as compared to Dixon and colleagues (2007), the *trans*-regulational effect of rs4234898 seems to be specific to *SLC2A3*.

One important consideration in interpretation of our expression results is that the functional analyses refer to expression in cells or cell lines that were extracted from blood. As samples from human brain, the target tissue for dyslexia-related findings, are not easily available, analysis of blood offers the possibility to provide first hints for putative regulational effects. However, the question remains whether the given findings on *SLC2A3* expression can be transferred to human brain. Currently, only one genome-wide expression data set of human postmortem cortical brain samples does exist (Myers *et al.* 2007). Neither rs4234898 nor rs11100040 were available in this data set, due to the different SNP-contents of Illumina and Affymetrix SNP-chips. Six proximal SNPs were available on the Affymetrix chip, but they provided with too low LD in order to be considered as tSNPs for rs4234898 ($0.001 < r^2 < 0.5$). An analysis of the *trans*-regulational effect in brain tissue was therefore not possible.

Recent genome-wide data revealed that *trans*-regulated genes can be found widespread in the human genome, however, their abundance compared to *cis*-elements is estimated to be rather low (Cheung *et al.* 2005; Dixon *et al.* 2007; Myers *et al.* 2007). Several possible mechanisms for *trans*-effects can be imagined. For instance, loci or genes on different chromosomes could locate close to one another in the three-dimensional space of the nucleus, a mechanism refered to as ‘interchromosomal interaction’ (Spilianakis *et al.* 2005). A second way of regulation in *trans* are indirect effects mediated by intermediate proteins: The causal variant at the rs4234898 locus could regulate transcription factors which, subsequently, would bind to promoter regions of *SLC2A3*. As we did not detect any regulatory effects apart from *SLC2A3*, it might be that such intermediate transcription factor is not expressed in blood and, thus, could not be targeted. This hypothesis is supported by findings in murine brain which showed that the *brain-derived*

neurotrophic factor (*Bdnf)*, located on mouse chr. 2, increases *Glut3* (mouse chr. 6) expression in developing cortical neurons (Burkhalter *et al.* 2003). As *BDNF* is not expressed in blood, we were not able to identify this effect. Again, this illustrates the limitations that arise if it is not the correct target tissue that is analyzed. Apart from *Bdnf*, also other *trans*-activators of *Glut3* expression such as Sp1 and Sp3 have been described, which regulate the transcription of *Glut3* in cultured murine neuroblast cells (Rajakumar *et al.* 1998; Rajakumar *et al.* 2004).

Both, the 'mRNA-by-SNP-browser' and the experiments in EBV-transformed lymphoblastoid cell lines, provided expression results only for children. In a next step, we thus investigated whether the *trans*-regulational findings could also be replicated in adults. However, neither the publically available data set GENEVAR (Stranger *et al.* 2007), nor the blood samples of 37 adults yielded evidence for a *trans*-regulational effect of rs4234898 on *SLC2A3* expression in grown-up human individuals. This suggests that the *trans*-regulational effect mediated by the 4q32.1 locus is developmentally restricted to human childhood. Although this finding is astonishing at first glance, there are indeed several arguments supporting this observation. In mice, it was shown that the expression levels of *Glut3* in brain vary between different developmental stages (Khan *et al.* 1999). In that study, low amounts of *Glut3* mRNA were detected during prenatal phases, and a maximum was reached in the first two to three weeks after birth. Thereafter, mRNA levels of *Glut3* declined and then remained constant during adulthood. Notably, the time point of maximal *Glut3* expression was directly correlated with the developmental stage in which synaptogenesis takes place in the murine brain (Aghajanian & Bloom 1967; Sachs *et al.* 1986). In human brain, synaptogenesis starts during embryonic development and continues at high levels until approximately two years of age (Huttenlocher 1979). Thereafter, until adolescence, unused synapses are degraded and the general synaptic density decreases (Huttenlocher 1979; Huttenlocher & de Courten 1987). The plasticity of synapses is reorganized during that time, with new branches that form and new synaptic connections that are established (Johnston 2003). This reorganization of the synaptic structures until adolescence well correlates with the developmental time point at which our analyses were undertaken.

Additionally, studies using positron emission tomography in humans have shown that the cerebral utilization of glucose rises from birth to an age of four years, to a level which is then maintained until approximately 10 years of age. Thereafter, the metabolic rates of glucose utilization gradually decline (Chugani 1998). Thus, cell migration processes such as elongating of axons, branching of synapses or formation of synaptic connections rather take place in 4 to 10 year old children than in adults. These processes trigger an increased demand of glucose in the human brain (Deza & Eidelberg 1967; Maher & Simpson 1994). A lower expression level of *SLC2A3* mediated by rs4234898, as suggested by our findings, might thus result in reduced

cerebral amount of glucose. Such an interrelation has indeed already been demonstrated by Liu and colleagues (2008), who showed that a decreased mRNA level of *SLC2A3* is responsible for the impaired glucose metabolism observed in patients suffering from Alzheimer's disease (Liu *et al.* 2008). In rats, an induced glucose deficit was reported to immediately cause slower reaction times and modified amplitudes in EEG measurements (Suh *et al.* 2003). These findings suggest that children with reduced glucose levels, caused by lower abundance of the glucose transporter SLC2A3, might show slower reaction times upon representation of a stimulus. This could explain the attenuated speech MMN and provides a pathological mechanism for the observed association between rs4234898, *SLC2A3* expression levels and MMN.

Apart from MMN as cognitive endophenotype, we analyzed a second dyslexia-related trait in terms of underlying genetic factors. Epidemiological studies have indicated that between 11% and 56% of dyslexic children also present with serious impairments in the acquisition of arithmetical abilities (Dirks *et al.* 2008), and it has been suggested that arithmetic and literacy skills depend on similar cognitive factors (Hecht *et al.* 2001; Geary & Hoard 2002). As formal genetic studies suggest a genetic contribution for both, arithmetic abilities alone and their comorbidity with dyslexia, respectively (Gillis *et al.* 1992; Alarcon *et al.* 1997), we correlated our genome-wide data with different mathematical measures established in the German DYS-sample. Three SNPs (rs1399428, $P = 3.19 \times 10^{-09}$, rs4837521, $P = 2.03 \times 10^{-08}$, both chr. 9q33.1; and rs133885, $P = 5.78 \times 10^{-09}$, chr. 22q12.1) provided with genome-wide significant associations with at least one of the three investigated measures, namely 'exact calculation', 'number comparison' and the principal component extracted from both ('basic mathematical abilities factor', see section 3.6.1). None of the three SNPs showed association with dyslexia *per se* in the German sample (rs1399428, $P = 0.456$; rs4837521, $P = 0.594$; rs133885, $P = 0.115$), indicating that the findings are restricted to mathematical skills only or represent the comorbidity between the two learning impairments.

The two associated SNPs of the 9q33.1 locus (rs1399428, rs4837521), which were genome-wide significant in the GWAS, are located in a region devoid of any known genes. They show an intermarker distance of about 6.1 kb and considerable LD ($r^2 = 0.71$, $D' = 0.88$). However, both SNPs failed replication in the independent German sample set of 186 cases, which was unexpected given the genome-wide significant *P*-values in the GWAS. As for our MMN-study, children included in the initial sample were more severely affected with dyslexia compared to children in the replication sample, again suggesting that the effect of the 9q33.1 locus on mathematical abilities is restricted to more severely affected dyslexic individuals. A second

possible explanation would be that the two markers appeared among the top hits simply by chance, without a true contribution to the quantitative trait analyzed. Given these controversial findings, further experiments will have to be conducted in order to dissect the contribution of the 9q33.1 locus on arithmetic abilities.

The third SNP that showed a genome-wide significant association in the GWAS was rs133885 ($P = 5.78 \times 10^{-09}$), located on chr. 22q12.1. This finding could be fully confirmed in the replication study ($P = 0.0098$). In both samples, we found the effect to be in the same direction, with children homozygous for the G allele performing worse in mathematical tasks than carriers of at least one A allele ($P_{combined} = 8.81 \times 10^{-10}$). Our primary association finding was with the basic mathematical abilities factor, however, also the underlying single measures 'exact calculation' and 'number comparison' showed significant associations with rs133885 throughout different genetic models (data not shown). Calculating the effect size for rs133885 on mathematical skills revealed an effect range of 9% to 18% for different severity groups, suggesting that the identified variant provides with a large genetic contribution to mathematical abilities. The effect size was observed to increase with higher dyslexia affection status, indicating that the extent of comorbidity has an influence on risk. This could be explained by the fact that although dyscalculia and dyslexia have separate cognitive deficits, these are overlapping in children with both disorders (Landerl *et al.* 2009). Thus, with a more severe degree of dyslexia, less compensatory mechanisms for arithmetical skills are available, and it can be suggested that rs133885 contributes to this specific cognitive aspect.

Two other children samples, the Austrian dyslexia sample and the English Twin Early Development Study (TEDS), also provide qualitative information on their proband's skills in reading, spelling and mathematical abilities. These two samples were used for further independent replication of our association finding. Although sample ascertainment and applied test batteries differed between the samples, they provided further support: We detected an association with rs133885 in the same direction, although the *P*-values were only borderline significant ($P_{Austria} = 0.046$, $P_{TEDS} = 0.055$). In these samples, the effect of rs133885 on mathematical skills in children with general reading / writing inabilities was not as strong as in the German discovery sample.

As the probands included in the Austrian and the TEDS sample were not exclusively dyslexic individuals, analysis of their arithmetic skills provided the possibility to dissect the given association in terms of whether the underlying trait is (i) general arithmetic ability or (ii) the comorbid aspect. When we analyzed the arithmetic skills without considering the proband's

reading and writing skills, we did not find any evidence for an association. As already suggested by the increasing effect size with higher dyslexia severity grade in the German DYS-sample, this observation adds to the evidence that the given association contributes to the comorbidity between dyslexia and impairments in mathematical skills, and not to isolated arithmetics.

Rs133885 is located within the genomic region of *myosin 18B (MYO18B)*. The cellular function of MYO18B, which belongs to the family of unconventional myosins, has been predominantly related to myocardial structures (Salamon *et al.* 2003). Myosins are generally known as actin-based molecular motors which are ubiquitously expressed in all eukaryotic cells (Hasson & Mooseker 1995; Sellers 2000). Differences between members of the myosin family are related to altered amino acid sequences, which cause modified protein structures and interaction properties (Sellers 2000). For MYO18B, database research revealed the existence of two distinct isoforms, MYO18B_short (Q8IUG5-2) and MYO18B_long (Q8IUG5-1), which vary in the abundance of the 483 N-terminal amino acids (aa). Rs133885 is located in exon 3 of *MYO18B*, where it codes for a non-synonymous change of glutamic acid (Glu, A allele) to glycine (Gly, G allele) on position 44 of the MYO18B protein. The SNP thus exhibits its effect only in the long *MYO18B* isoform. The exchange from Glu to Gly results in the loss of a COOH group within the N-terminus of MYO18B_long. *In silico* analysis of MYO18B revealed that the middle area of the MYO18B N-terminus contains the conserved amino acids Val153Leu154, which have been identified to be essential for binding of actin filaments in co-sedimentation experiments (Ajima *et al.* 2008). It has further been hypothesized that many ligands for F-actin contain proline-rich domains (Witke 2004), which can also be found in the N-terminus of MYO18B. The Glu44Gly substitution is located close to this proline-rich region, and a loss of the carboxy-group might thus influence the binding properties of MYO18B.

Although microarray data (UCSC browser) suggested an ubiquitous expression pattern of *MYO18B*, we analyzed the expression of *MYO18B_long* in various tissues, at different developmental stages. We were able to show that *MYO18B_long* is ubiquitously expressed in all human fetal tissues (see section 4.2.2, Fig. 16). The highest rates were observed in myocardial structures such as heart and skeletal muscle, which is in accordance with previous findings (Salamon *et al.* 2003). The abundance of *MYO18B_long* in fetal brain seems to be less strong than in most of the other tissues, however, one has to keep in mind that the PCR-based expression analysis provides only semi-quantitative results. Given that *MYO18B_long* would truly be lower expressed in whole brain tissue, this could point towards a regional specificity of *MYO18B_long*

expression in particular regions of the fetal brain. In adult tissues, expression was observed in all tissues but brain. Given this expression difference between the two different developmental stages, it can be hypothesized that the long isoform of MYO18B plays an important role in the structural development of the brain.

We next assessed potential impacts of the given rs133885 association using structural brain data. In previous studies, the right intraparietal sulcus (IPS) was identified as one brain structure involved in mathematical processing: Fias and colleagues (2007) had shown that the IPS plays a crucial role in the cognitive representation of numerical quantity (Fias *et al.* 2007). Additionally, morphological analysis of the right IPS had revealed an abnormal length, depth and sulcal geometry in patients with Turner Syndrome, a complex X-linked disorder that has served as model for genetically influenced dyscalculia (Molko *et al.* 2003). In our study, structural magnetic resonance imaging (MRI) in 80 healthy German controls provided evidence that individuals homozygous for the G allele of rs133885 show a lower depth of the right IPS when compared to carriers of the A allele ($P = 0.0010$, see section 4.2.2). No such effect was found for any of the control sulci, suggesting that this observation was specific for the IPS. As an involvement of the IPS in numerical processing was shown in young children and remained continuous into adulthood (Cantlon *et al.* 2006), it can be suggested that the underlying structural networks develop prior to birth or during childhood. This correlates well with our expression findings, indicating that the rs133885 genotype-dependent effect manifests during brain development. Given the present data, a functional mechanism for the observed association between rs133885 and the impaired arithmetical skills in dyslexic children can conclusively be suggested. Together with the results presented for the cognitive MMN phenotype, we were thus able to successfully apply the GWAS approach for the identification of new genetic variants for two dyslexia-related quantitative endophenotypes.

6. SUMMARY

Reading and writing are abilities that are unique to the human species. Acquiring these skills involves a complex network of cognitive processes (e.g. visual processing, phonological decoding or orthographic processing) which have to interact efficiently and in a time-constraint manner. Deficits in at least some of these cognitive functions contribute to the development of dyslexia, one of the most prominent neurodevelopmental disorders. It is widely accepted that genetic factors herefore play an essential role. In this thesis, genetic analyses have been conducted in order to further investigate known genetic variants which have been reported in the context of dyslexia, and to identify new genetic variants that would explain the heritability observed for both, dyslexia as qualitative trait and dyslexia-related quantitative endophenotypes.

We and others have suggested two dyslexia candidate genes on chromosome 6p22, namely *DCDC2* and *KIAA0319*. Causal susceptibility variants have been reported for both of these genes, however, these findings required replication in an independent sample. In our German dyslexia (DYS-) sample, we were not able to replicate prior reports on an intronic deletion or a compound STR marker in intron 2 of *DCDC2* to be causative for dyslexia. Also, the analysis of further common variants in and around the first exon of *KIAA0319* did not provide evidence for a genetic effect in the German DYS-sample. However, we were able to confirm previous findings on epistatic gene-gene interactions between an intronic *DCDC2* two-marker risk haplotype and SNPs within *KIAA0319*. This suggests that genetic variation within *KIAA0319* might provide with small modifying effects on dyslexia susceptibility in the presence of the *DCDC2* risk haplotype.

We also analyzed quantitative measures in the dyslexia sample, using candidate-gene approaches. Dyslexic children often perform poorly in verbal short-term memory tasks, and *GRIN2B* has been reported to play an important role in human memory and cognition. Our genetic results provide evidence that variation within intron 3 of *GRIN2B* contribute to the weak performance of dyslexic children in verbal short-term memory. Notably, the effect was shown to involve genomic imprinting, with children inheriting the associated allele from their mothers performing significantly better in memory-related tasks. Although the exact mechanisms behind this observation still remain to be identified, the results contribute to our understanding of memory-related cognitive processes and might partly explain the comorbidity observed between dyslexia and ADHD.

A non-synonymous variant within a second gene, *LRRTM1*, was found to be associated with relative hand skill. We were able to show that carriers of a maternally inherited C allele of rs6733871 show a stronger tendency towards left-handedness. This measure represents a correlate for asymmetrical brain function, which has been suggested to partly underlie the neural signature of dyslexia.

We also attempted to identify new variants or genes contributing to the development of dyslexia using a genome-wide association approach. As the number of individuals included in this study provided only limited power, we combined individual genotyping with large pooling efforts to choose SNPs for subsequent replication. Our study revealed four new susceptibility loci on chr. 5p13.1, 9q21.3, 21q21.3 and 22q13.31. Promising candidate genes within these regions include *GRIK1*, *TMC1* and *FBLN1*. As each of the identified variants only provides with a small effect size, replication in large, independent samples will be required to confirm our findings.

The application of the genome-wide approach on quantitative dimensions of dyslexia has led to the identification of new susceptibility variants for two dyslexia-related endophenotypes. We present strong evidence for an association of rs4234898, an intergenic marker on chromosome 4q32.1, with mismatch negativity, a neural correlate of speech perception. Our data indicate that the locus mediates its function via *trans*-regulational effects on the expression level of *SLC2A3*, a member of the facilitative glucose transporters expressed in brain. We were able to show that the associated effect is functionally relevant during human childhood, when an increased amount of glucose substrate is required for the formation of synaptic connections and branching in the human brain.

In a second quantitative analysis using the genome-wide data, we found that arithmetical skills are associated with rs133885, a non-synonymous marker in the long isoform of *MYO18B*. This isoform was found to be expressed in human fetal brain but did not appear in adult brain tissue, which suggests a predominant function in structural brain development. This was subsequently confirmed by structural MRI data, which provided evidence that in contrast to the non-risk group, carriers of the risk allele show a lower depth and volume of the right intraparietal sulcus (IPS), a structural entity involved in numerical processing.

To our knowledge, the latter two studies are the first ones which investigated the genetic basis of dyslexia and some of the related endophenotypes on a genome-wide level. The new findings presented within this thesis might contribute to a better understanding of dyslexia susceptibility, genetic effects on related endophenotypes and the functional mechanisms that underlie human reading and writing skills.

7. OUTLOOK

During the last years, research in the field of dyslexia has led to the identification of a number of dyslexia susceptibility loci and candidate genes. However, many of them failed replication in independent samples, and none has yet provided with a true causative variant. The genome-wide association study conducted in the context of this thesis is the first dyslexia GWAS known to date. We revealed suggestive evidence for four new common variants that confer susceptibility for the development of dyslexia. Replication in independent samples is now required to confirm the present findings. Potential further samples include a large population-based sample from Iceland (deCODE) or samples from the Twin Early Development Study (TEDS), which both provide ordinal measures for reading and spelling ability. Replicated variants would then provide a starting point for the identification of the responsible causative variants, by using finemapping and targeted resequencing approaches. They would also be excellent candidates for functional analyses such as allele-specific expression and / or MRI studies, which will be conducted to further understand the pathological mechanisms of dyslexia. Interdisciplinary collaborations that combine expertise from a variety of fields are strongly required to identify the biological consequences of risk variants and to understand the cognitive mechanisms underlying the found associations.

As the large genetic heterogeneity of dyslexia and phenotypic differences between the samples seem to contribute to the limited power of our analyses, a further identification of other susceptibility variants will require larger sample numbers and an extension of analysis strategies towards structural variants (CNVs) and pathway information. Additionally, rare variants with alleles occurring at frequencies below 1% should also be addressed. New technologies such as Next-Generation Sequencing will contribute to accurately map these low-frequency variants at both, structural and single-base level, and will offer new possibilities in the search for disease candidate genes.

Apart from investigating dyslexia as qualitative trait, it will become more and more important to analyze the underlying endophenotypes with respect to genetic contribution. Quantitative measures, if they are ascertained in a clinically appropriate and comparable manner, provide with higher power to detect risk variants in complex disorders, as genotype-phenotype correlations are more direct than in compound diagnoses. Based there upon, it was now decided to analyze the quantitative measures 'reading' and 'spelling' within the entire GWAS data of the NeuroDys consortium. This collaborative study is currently undergoing and will hopefully provide new susceptibility variants for dyslexia.

Another aim of future research will be to establish a better understanding of both, gene-gene and gene-environment interactions. It is currently assumed that environmental factors such as pre-school trainings or the literacy environment at home might contribute to the development of dyslexia, but only some of these factors have been conclusively identified so far. If such factors can be modulated, future dyslexia prevention and individual genetic risk profiling could be envisaged. For affected dyslexic individuals, a precise knowledge of exact genotype-phenotype correlations allows for an appropriate classification of individuals into subgroups based on their specific cognitive impairments. This, in turn, would enable the development and application of personalized therapeutic approaches that are individually adapted for each patient and will increase the chance for successful remediation of the young affected dyslexia patients.

8. REFERENCES

Aghajanian GK, Bloom FE. 1967. The formation of synaptic junctions in developing rat brain: a quantitative electron microscopic study. Brain Res 6(4):716-27.

Ajima R, Akazawa H, Kodama M, Takeshita F, Otsuka A, Kohno T, Komuro I, Ochiya T, Yokota J. 2008. Deficiency of Myo18B in mice results in embryonic lethality with cardiac myofibrillar aberrations. Genes Cells 13(10):987-99.

Alarcon M, DeFries JC, Light JG, Pennington BF. 1997. A twin study of mathematics disability. J Learn Disabil 30(6):617-23.

Alexander AW, Slinger-Constant AM. 2004. Current status of treatments for dyslexia: critical review. J Child Neurol 19(10):744-58.

Alonso-Bua B, Diaz F, Ferraces MJ. 2006. The contribution of AERPs (MMN and LDN) to studying temporal vs. linguistic processing deficits in children with reading difficulties. Int J Psychophysiol 59(2):159-67.

Altmüller 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(5):936-50.

Altshuler D, Daly MJ, Lander ES. 2008. Genetic mapping in human disease. Science 322(5903):881-8.

Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C and others. 2000. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet 26(1):76-80.

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(11):2243-52.

Anthoni H, Zucchelli M, Matsson H, Müller-Myhsok B, Fransson I, Schumacher J, Massinen S, Onkamo P, Warnke A, Griesemann H and others. 2007. A locus on 2p12 containing the co-regulated MRPL19 and C2ORF3 genes is associated to dyslexia. Hum Mol Genet 16(6):667-77.

APA. 1994. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Washington DC: American Psychiatric Association.

Argraves WS, Dickerson K, Burgess WH, Ruoslahti E. 1989. Fibulin, a novel protein that interacts with the fibronectin receptor beta subunit cytoplasmic domain. Cell 58(4):623-9.

August GJ, Garfinkel BD. 1990. Comorbidity of ADHD and reading disability among clinic-referred children. J Abnorm Child Psychol 18(1):29-45.

Baddeley A, Wilson BA. 1993. A developmental deficit in short-term phonological memory: implications for language and reading. Memory 1(1):65-78.

Balding DJ. 2006. A tutorial on statistical methods for population association studies. Nat Rev Genet 7(10):781-91.

Balding DJ, Bishop M, Cannings C. 2007. Handbook of Statistical Genetics. Chichester, England: John Wiley & Sons.

Bansal A, van den Boom D, Kammerer S, Honisch C, Adam G, Cantor CR, Kleyn P, Braun A. 2002. Association testing by DNA pooling: an effective initial screen. Proc Natl Acad Sci U S A 99(26):16871-4.

Barbon A, Barlati S. 2000. Genomic organization, proposed alternative splicing mechanisms, and RNA editing structure of GRIK1. Cytogenet Cell Genet 88(3-4):236-9.

Bates TC, Lind PA, Luciano M, Montgomery GW, Martin NG, Wright MJ. 2009. Dyslexia and DYX1C1: deficits in reading and spelling associated with a missense mutation. Mol Psychiatry.

Becker T. 2007. Genidentifizierung bei multifaktoriellen Krankheiten. Medizinische Genetik 19:300-303.

Bellini G, Bravaccio C, Calamoneri F, Donatella Cocuzza M, Fiorillo P, Gagliano A, Mazzone D, del Giudice EM, Scuccimarra G, Militerni R and others. 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(3):311-4.

Birnbaum S, Ludwig KU, Reutter H, Herms S, Steffens M, Rubini M, Baluardo C, Ferrian M, Almeida de Assis N, Alblas MA and others. 2009. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nat Genet 41(4):473-7.

Blondal T, Waage BG, Smarason SV, Jonsson F, Fjalldal SB, Stefansson K, Gulcher J, Smith AV. 2003. A novel MALDI-TOF based methodology for genotyping single nucleotide polymorphisms. Nucleic Acids Res 31(24):e155.

Bolk S, Pelet A, Hofstra RM, Angrist M, Salomon R, Croaker D, Buys CH, Lyonnet S, Chakravarti A. 2000. A human model for multigenic inheritance: phenotypic expression in Hirschsprung disease requires both the RET gene and a new 9q31 locus. Proc Natl Acad Sci U S A 97(1):268-73.

Bonferroni CE. 1937. Teoria statistica delle classi e calcolo delle probabilita. Volume in Onore di Ricarrdo dalla Volta: Universita di Firenza. p 1-62.

Bowie D. 2008. Ionotropic glutamate receptors & CNS disorders. CNS Neurol Disord Drug Targets 7(2):129-43.

Bradley L, Bryant PE. 1978. Difficulties in auditory organisation as a possible cause of reading backwardness. Nature 271(5647):746-7.

Brähler E, Holling V, Leutner D, Pettermann F. 2002. Brickenkamp Handbuch psychologischer und pädagogischer Tests. Göttingen: Hogrefe.

Brkanac Z, Chapman NH, Igo RP, Jr., Matsushita MM, Nielsen K, Berninger VW, Wijsman EM, Raskind WH. 2008. Genome scan of a nonword repetition phenotype in families with dyslexia: evidence for multiple loci. Behav Genet 38(5):462-75.

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(6):781-3.

Bruck M. 1992. Persistence of dyslexics' phonological awareness deficits. Dev Psychol 28:874-886.

Bruck M. 1998. Outcomes of adults with childhood histories of dyslexia. In: Hulme C, Joshi RM, editors. Reading and spelling: development and disorders. Mahwah, NJ: L Erlbaum. p 179-200.

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):1901-17.

Burkhalter J, Fiumelli H, Allaman I, Chatton JY, Martin JL. 2003. Brain-derived neurotrophic factor stimulates energy metabolism in developing cortical neurons. J Neurosci 23(23):8212-20.

Bustin SA. 2002. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. J Mol Endocrinol 29(1):23-39.

Butt C, Zheng H, Randell E, Robb D, Parfrey P, Xie YG. 2003. Combined carrier status of prothrombin 20210A and factor XIII-A Leu34 alleles as a strong risk factor for myocardial infarction: evidence of a gene-gene interaction. Blood 101(8):3037-41.

Cantlon JF, Brannon EM, Carter EJ, Pelphrey KA. 2006. Functional imaging of numerical processing in adults and 4-y-old children. PLoS Biol 4(5):e125.

Cardon LR, Bell JI. 2001. Association study designs for complex diseases. Nat Rev Genet 2(2):91-9.

Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC. 1994. Quantitative trait locus for reading disability on chromosome 6. Science 266(5183):276-9.

Cardullo RA, Agrawal S, Flores C, Zamecnik PC, Wolf DE. 1988. Detection of nucleic acid hybridization by nonradiative fluorescence resonance energy transfer. Proc Natl Acad Sci U S A 85(23):8790-4.

Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N and others. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet 22(3):231-8.

Carlson CS, Eberle MA, Kruglyak L, Nickerson DA. 2004. Mapping complex disease loci in whole-genome association studies. Nature 429(6990):446-52.

Carthew RW, Sontheimer EJ. 2009. Origins and Mechanisms of miRNAs and siRNAs. Cell 136(4):642-55.

Castles A, Coltheart M. 1996. Cognitive correlates of developmental surface dyslexia : a single case study. Cognitive neuropsychology 13:25-50.

Castles A, Coltheart M. 2004. Is there a causal link from phonological awareness to success in learning to read? Cognition 91(1):77-111.

Cattanach BM, Kirk M. 1985. Differential activity of maternally and paternally derived chromosome regions in mice. Nature 315(6019):496-8.

Chakravarti A. 1999. Population genetics--making sense out of sequence. Nat Genet 21(1 Suppl):56-60.

Chan L. 1993. RNA editing: exploring one mode with apolipoprotein B mRNA. Bioessays 15(1):33-41.

Chang BS, Ly J, Appignani B, Bodell A, Apse KA, Ravenscroft RS, Sheen VL, Doherty MJ, Hackney DB, O'Connor M and others. 2005. Reading impairment in the neuronal migration disorder of periventricular nodular heterotopia. Neurology 64(5):799-803.

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 131B(1):67-75.

Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT. 2005. Mapping determinants of human gene expression by regional and genome-wide association. Nature 437(7063):1365-9.

Chugani HT. 1998. A critical period of brain development: studies of cerebral glucose utilization with PET. Prev Med 27(2):184-8.

Clay M. 1985. The early detection of reading difficulties. Auckland, New Zealand: Heinemann.

Contractor A, Swanson G, Heinemann SF. 2001. Kainate receptors are involved in short- and long-term plasticity at mossy fiber synapses in the hippocampus. Neuron 29(1):209-16.

Cooper DN, Krawczak M. 1993. Human Gene Mutation. Oxford, UK: BIOS Scientific Publishers Limited.

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(4):581-91.

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(3):237-8.

Coquelle FM, Levy T, Bergmann S, Wolf SG, Bar-El D, Sapir T, Brody Y, Orr I, Barkai N, Eichele G and others. 2006. Common and divergent roles for members of the mouse DCX superfamily. Cell Cycle 5(9):976-83.

Corbera S, Escera C, Artigas J. 2006. Impaired duration mismatch negativity in developmental dyslexia. Neuroreport 17(10):1051-5.

Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA. 1993. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 261(5123):921-3.

Cotter RJ. 1992. Time-of-flight mass spectrometry for the structural analysis of biological molecules. Anal Chem 64(21):1027A-1039A.

Cusmano-Ozog K, Manning MA, Hoyme HE. 2007. 22q13.3 deletion syndrome: a recognizable malformation syndrome associated with marked speech and language delay. Am J Med Genet C Semin Med Genet 145C(4):393-8.

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(9):652-7.

de Quervain DJ, Papassotiropoulos A. 2006. Identification of a genetic cluster influencing memory performance and hippocampal activity in humans. Proc Natl Acad Sci U S A 103(11):4270-4.

DeAngelis MM, Wang DG, Hawkins TL. 1995. Solid-phase reversible immobilization for the isolation of PCR products. Nucleic Acids Res 23(22):4742-3.

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(2):128-38.

DeFries JC, Fulker DW, LaBuda MC. 1987. Evidence for a genetic aetiology in reading disability of twins. Nature 329(6139):537-9.

Delaval K, Feil R. 2004. Epigenetic regulation of mammalian genomic imprinting. Curr Opin Genet Dev 14(2):188-95.

Demb JB, Boynton GM, Heeger DJ. 1998. Functional magnetic resonance imaging of early visual pathways in dyslexia. J Neurosci 18(17):6939-51.

Denckla MB, Rudel R. 1974. Rapid "automatized" naming of pictured objects, colors, letters and numbers by normal children. Cortex 10(2):186-202.

Dennis MY, Paracchini S, Scerri TS, Prokunina-Olsson L, Knight JC, Wade-Martins R, Coggill P, Beck S, Green ED, Monaco AP. 2009. A common variant associated with dyslexia reduces expression of the KIAA0319 gene. PLoS Genet 5(3):e1000436.

Denton C, Mathes P. 2003. Intervention for struggling readers. In: Foorman BR, editor. Preventing and remediating reading difficulties: Bringing science to scale. Timonium, MD: York Press. p 229-252.

Deuel TA, Liu JS, Corbo JC, Yoo SY, Rorke-Adams LB, Walsh CA. 2006. Genetic interactions between doublecortin and doublecortin-like kinase in neuronal migration and axon outgrowth. Neuron 49(1):41-53.

Deza L, Eidelberg E. 1967. Development of cortical electrical activity in the rat. Exp Neurol 17(4):425-38.

Ding C, Cantor CR. 2003. A high-throughput gene expression analysis technique using competitive PCR and matrix-assisted laser desorption ionization time-of-flight MS. Proc Natl Acad Sci U S A 100(6):3059-64.

Dirks E, Spyer G, van Lieshout EC, de Sonneville L. 2008. Prevalence of combined reading and arithmetic disabilities. J Learn Disabil 41(5):460-73.

Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, Taylor J, Burnett E, Gut I, Farrall M and others. 2007. A genome-wide association study of global gene expression. Nat Genet 39(10):1202-7.

Dorval KM, Wigg KG, Crosbie J, Tannock R, Kennedy JL, Ickowicz A, Pathare T, Malone M, Schachar R, Barr CL. 2007. Association of the glutamate receptor subunit gene GRIN2B with attention-deficit/hyperactivity disorder. Genes Brain Behav 6(5):444-52.

Dudbridge F, Gusnanto A. 2008. Estimation of significance thresholds for genomewide association scans. Genet Epidemiol 32(3):227-34.

Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A and others. 2006. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science 314(5804):1461-3.

Dufor O, Serniclaes W, Sprenger-Charolles L, Demonet JF. 2007. Top-down processes during auditory phoneme categorization in dyslexia: a PET study. Neuroimage 34(4):1692-707.

Eckert MA, Leonard CM. 2003. Developmental disorder: dyslexia. In: Hughdal K, Davidson RJ, editors. The Asymmetrical Brain. Cambridge, MA: MIT press. p 651-679.

Eden GF, Jones KM, Cappell K, Gareau L, Wood FB, Zeffiro TA, Dietz NA, Agnew JA, Flowers DL. 2004. Neural changes following remediation in adult developmental dyslexia. Neuron 44(3):411-22.

Eden GF, Moats L. 2002. The role of neuroscience in the remediation of students with dyslexia. Nat Neurosci 5 Suppl:1080-4.

Eden GF, VanMeter JW, Rumsey JM, Maisog JM, Woods RP, Zeffiro TA. 1996. Abnormal processing of visual motion in dyslexia revealed by functional brain imaging. Nature 382(6586):66-9.

Eliez S, Rumsey JM, Giedd JN, Schmitt JE, Patwardhan AJ, Reiss AL. 2000. Morphological alteration of temporal lobe gray matter in dyslexia: an MRI study. J Child Psychol Psychiatry 41(5):637-44.

Erskine L, Williams SE, Brose K, Kidd T, Rachel RA, Goodman CS, Tessier-Lavigne M, Mason CA. 2000. Retinal ganglion cell axon guidance in the mouse optic chiasm: expression and function of robos and slits. J Neurosci 20(13):4975-82.

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(9):664-9.

Fedorcsak I, Ehrenberg L. 1966. Effects of diethyl pyrocarbonate and methyl methanesulfonate on nucleic acids and nucleases. Acta Chem Scand 20(1):107-12.

Feuk L, Carson AR, Scherer SW. 2006. Structural variation in the human genome. Nat Rev Genet 7(2):85-97.

Fias W, Lammertyn J, Caessens B, Orban GA. 2007. Processing of abstract ordinal knowledge in the horizontal segment of the intraparietal sulcus. J Neurosci 27(33):8952-6.

Fisher SE, Francks C. 2006. Genes, cognition and dyslexia: learning to read the genome. Trends Cogn Sci 10(6):250-7.

Fisher SE, Francks C, Marlow AJ, MacPhie IL, Newbury DF, Cardon LR, Ishikawa-Brush Y, Richardson AJ, Talcott JB, Gayan J and others. 2002. Independent genome-wide scans

identify a chromosome 18 quantitative-trait locus influencing dyslexia. Nat Genet 30(1):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(1):146-56.

Francks C, Maegawa S, Lauren J, Abrahams BS, Velayos-Baeza A, Medland SE, Colella S, Groszer M, McAuley EZ, Caffrey TM and others. 2007. LRRTM1 on chromosome 2p12 is a maternally suppressed gene that is associated paternally with handedness and schizophrenia. Mol Psychiatry 12(12):1129-39, 1057.

Francks C, Paracchini S, Smith SD, Richardson AJ, Scerri TS, Cardon LR, Marlow AJ, MacPhie IL, Walter J, Pennington BF and others. 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(6):1046-58.

Frauenheim JG, Heckerl JR. 1983. A longitudinal study of psychological and achievement test performance in severe dyslexic adults. J Learn Disabil 16(6):339-47.

Galaburda AM, Kemper TL. 1979. Cytoarchitectonic abnormalities in developmental dyslexia: a case study. Ann Neurol 6(2):94-100.

Galaburda AM, LoTurco J, Ramus F, Fitch RH, Rosen GD. 2006. From genes to behavior in developmental dyslexia. Nat Neurosci 9(10):1213-7.

Galaburda AM, Menard MT, Rosen GD. 1994. Evidence for aberrant auditory anatomy in developmental dyslexia. Proc Natl Acad Sci U S A 91(17):8010-3.

Galaburda AM, Sherman GF, Rosen GD, Aboitiz F, Geschwind N. 1985. Developmental dyslexia: four consecutive patients with cortical anomalies. Ann Neurol 18(2):222-33.

Gayán J, Olson RK. 2001. Genetic and environmental influences on orthographic and phonological skills in children with reading disabilities. Dev Neuropsychol 20(2):483-507.

Gayán J, Smith SD, Cherny SS, Cardon LR, Fulker DW, Brower AM, Olson RK, Pennington BF, DeFries JC. 1999. Quantitative-trait locus for specific language and reading deficits on chromosome 6p. Am J Hum Genet 64(1):157-64.

Gayán J, Willcutt EG, Fisher SE, Francks C, Cardon LR, Olson RK, Pennington BF, Smith SD, Monaco AP, DeFries JC. 2005. Bivariate linkage scan for reading disability and attention-deficit/hyperactivity disorder localizes pleiotropic loci. J Child Psychol Psychiatry 46(10):1045-56.

Geary DC, Hoard MK. 2002. Learning disabilities in basic mathematics: Deficits in memory and cognition. In: Royer JM, editor. Mathematical cognition. Greenwich, CT: Information Age Publishing. p 93-115.

Geschwind DH, Miller BL, DeCarli C, Carmelli D. 2002. Heritability of lobar brain volumes in twins supports genetic models of cerebral laterality and handedness. Proc Natl Acad Sci U S A 99(5):3176-81.

Gillis JJ, DeFries JC, Fulker DW. 1992. Confirmatory factor analysis of reading and mathematics performance: a twin study. Acta Genet Med Gemellol (Roma) 41(4):287-300.

Gottesman, II, Gould TD. 2003. The endophenotype concept in psychiatry: etymology and strategic intentions. Am J Psychiatry 160(4):636-45.

Graef IA, Wang F, Charron F, Chen L, Neilson J, Tessier-Lavigne M, Crabtree GR. 2003. Neurotrophins and netrins require calcineurin/NFAT signaling to stimulate outgrowth of embryonic axons. Cell 113(5):657-70.

Grigorenko EL, Chang JT. 1997. An extension of affected-pedigree-member analyses to triads of relatives. Genet Epidemiol 14(6):1005-10.

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 118B(1):89-98.

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(1):120-9.

Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS. 2005. A genome-wide scalable SNP genotyping assay using microarray technology. Nat Genet 37(5):549-54.

Haff LA, Smirnov IP. 1997. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. Genome Res 7(4):378-88.

Hall MH, Schulze K, Rijsdijk F, Picchioni M, Ettinger U, Bramon E, Freedman R, Murray RM, Sham P. 2006. Heritability and reliability of P300, P50 and duration mismatch negativity. Behav Genet 36(6):845-57.

Hallgren B. 1950. Specific dyslexia (congenital word-blindness); a clinical and genetic study. Acta Psychiatr Neurol Suppl 65:1-287.

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(4):e50.

HapMap Consortium. 2003. The International HapMap Project. Nature 426(6968):789-96.

HapMap Consortium. 2005. A haplotype map of the human genome. Nature 437(7063):1299-320.

Hardy GH. 1908. Mendelian Proportions in a Mixed Population. Science 28(706):49-50.

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(4):373-84.

Harold D, Paracchini S, Scerri T, Dennis M, Cope N, Hill G, Moskvina V, Walter J, Richardson AJ, Owen MJ and others. 2006. Further evidence that the KIAA0319 gene confers susceptibility to developmental dyslexia. Mol Psychiatry 11(12):1085-91, 1061.

Hasson T, Mooseker MS. 1995. Molecular motors, membrane movements and physiology: emerging roles for myosins. Curr Opin Cell Biol 7(4):587-94.

Hawi Z, Segurado R, Conroy J, Sheehan K, Lowe N, Kirley A, Shields D, Fitzgerald M, Gallagher L, Gill M. 2005. Preferential transmission of paternal alleles at risk genes in attention-deficit/hyperactivity disorder. Am J Hum Genet 77(6):958-65.

Hawke JL, Wadsworth SJ, DeFries JC. 2006. Genetic influences on reading difficulties in boys and girls: the Colorado twin study. Dyslexia 12(1):21-9.

Heath SC, Gut IG, Brennan P, McKay JD, Bencko V, Fabianova E, Foretova L, Georges M, Janout V, Kabesch M and others. 2008. Investigation of the fine structure of European populations with applications to disease association studies. Eur J Hum Genet 16(12):1413-29.

Hecht SA, Torgesen JK, Wagner RK, Rashotte CA. 2001. The relations between phonological processing abilities and emerging individual differences in mathematical computation skills: a longitudinal study from second to fifth grades. J Exp Child Psychol 79(2):192-227.

Hillmer AM, Brockschmidt FF, Hanneken S, Eigelshoven S, Steffens M, Flaquer A, Herms S, Becker T, Kortüm AK, Nyholt DR and others. 2008. Susceptibility variants for male-pattern baldness on chromosome 20p11. Nat Genet 40(11):1279-81.

Hinshelwood J. 1895. Word-blindness and visual memory. Lancet(146):1564-1570.

Hinshelwood J. 1907. Four cases of congenital word-blindness occuring in the same family. British Medical Journal(1):608-609.

Hintsch G, Zurlinden A, Meskenaite V, Steuble M, Fink-Widmer K, Kinter J, Sonderegger P. 2002. The calsyntenins--a family of postsynaptic membrane proteins with distinct neuronal expression patterns. Mol Cell Neurosci 21(3):393-409.

Hirschhorn JN, Daly MJ. 2005. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6(2):95-108.

Horwitz B, Rumsey JM, Donohue BC. 1998. Functional connectivity of the angular gyrus in normal reading and dyslexia. Proc Natl Acad Sci U S A 95(15):8939-44.

Hsiung GY, Kaplan BJ, Petryshen TL, Lu S, Field LL. 2004. A dyslexia susceptibility locus (DYX7) linked to dopamine D4 receptor (DRD4) region on chromosome 11p15.5. Am J Med Genet B Neuropsychiatr Genet 125B(1):112-9.

Hughdal K, Davidson RJ, editors. 2003. The Asymmetrical Brain. Cambridge, MA: MIT Press.

Hulslander J, Talcott J, Witton C, DeFries J, Pennington B, Wadsworth S, Willcutt E, Olson R. 2004. Sensory processing, reading, IQ, and attention. J Exp Child Psychol 88(3):274-95.

Huttenlocher PR. 1979. Synaptic density in human frontal cortex - developmental changes and effects of aging. Brain Res 163(2):195-205.

Huttenlocher PR, de Courten C. 1987. The development of synapses in striate cortex of man. Hum Neurobiol 6(1):1-9.

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 141B(1):15-27.

Isles AR, Humby T. 2006. Modes of imprinted gene action in learning disability. J Intellect Disabil Res 50(Pt 5):318-25.

Isles AR, Wilkinson LS. 2000. Imprinted genes, cognition and behaviour. Trends Cogn Sci 4(8):309-318.

Iwamoto K, Bundo M, Kato T. 2009. Serotonin receptor 2C and mental disorders: Genetic, expression and RNA editing studies. RNA Biol 6(3).

Jansen H, Mannhaupt G, Marx H, Skowronek H. 2002. Bielefelder Screening zur Früherkennung von Lese- und Rechtschreibschwierigkeiten (BISC). Second edition. Göttingen: Hogrefe.

Jawaid A, Sham P. 2009. Impact and quantification of the sources of error in DNA pooling designs. Ann Hum Genet 73(1):118-24.

Jeffreys AJ, Kauppi L, Neumann R. 2001. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat Genet 29(2):217-22.

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F and others. 2001. Haplotype tagging for the identification of common disease genes. Nat Genet 29(2):233-7.

Johnston MV. 2003. Brain plasticity in paediatric neurology. Eur J Paediatr Neurol 7(3):105-13.

Joost HG, Thorens B. 2001. The extended GLUT-family of sugar/polyol transport facilitators: nomenclature, sequence characteristics, and potential function of its novel members (review). Mol Membr Biol 18(4):247-56.

Jurcak V, Tsuzuki D, Dan I. 2007. 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. Neuroimage 34(4):1600-11.

Kabashima K, Sakata D, Nagamachi M, Miyachi Y, Inaba K, Narumiya S. 2003. Prostaglandin E2-EP4 signaling initiates skin immune responses by promoting migration and maturation of Langerhans cells. Nat Med 9(6):744-9.

Kaminen N, Hannula-Jouppi K, Kestilä M, Lahermo P, Muller K, Kaaranen M, Myllyluoma B, Voutilainen A, Lyytinen H, Nopola-Hemmi J and others. 2003. A genome scan for developmental dyslexia confirms linkage to chromosome 2p11 and suggests a new locus on 7q32. J Med Genet 40(5):340-5.

Kaplan BJ, Dewey DM, Crawford SG, Wilson BN. 2001. The term comorbidity is of questionable value in reference to developmental disorders: data and theory. J Learn Disabil 34(6):555-65.

Kaplan DE, Gayán J, Ahn J, Won TW, Pauls D, Olson RK, DeFries JC, Wood F, Pennington BF, Page GP and others. 2002. Evidence for linkage and association with reading disability on 6p21.3-22. Am J Hum Genet 70(5):1287-98.

Katusic SK, Colligan RC, Barbaresi WJ, Schaid DJ, Jacobsen SJ. 2001. Incidence of reading disability in a population-based birth cohort, 1976-1982, Rochester, Minn. Mayo Clin Proc 76(11):1081-92.

Kenmochi N, Suzuki T, Uechi T, Magoori M, Kuniba M, Higa S, Watanabe K, Tanaka T. 2001. The human mitochondrial ribosomal protein genes: mapping of 54 genes to the chromosomes and implications for human disorders. Genomics 77(1-2):65-70.

Khan JY, Rajakumar RA, McKnight RA, Devaskar UP, Devaskar SU. 1999. Developmental regulation of genes mediating murine brain glucose uptake. Am J Physiol 276(3 Pt 2):R892-900.

Kidd T, Bland KS, Goodman CS. 1999. Slit is the midline repellent for the robo receptor in Drosophila. Cell 96(6):785-94.

Kim KK, Won HH, Cho S, Park JH, Kim MJ, Kim S, Kim JW. 2009. Comparison of identical single nucleotide polymorphisms genotyped by the GeneChip Targeted Genotyping 25K, Affymetrix 500K and Illumina 550K platforms. Genomics.

Kirov G, Nikolov I, Georgieva L, Moskvina V, Owen MJ, O'Donovan MC. 2006. Pooled DNA genotyping on Affymetrix SNP genotyping arrays. BMC Genomics 7:27.

Kirov G, Zaharieva I, Georgieva L, Moskvina V, Nikolov I, Cichon S, Hillmer A, Toncheva D, Owen MJ, O'Donovan MC. 2008. A genome-wide association study in 574 schizophrenia trios using DNA pooling. Mol Psychiatry.

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(2):493-500.

Koizumi H, Tanaka T, Gleeson JG. 2006. Doublecortin-like kinase functions with doublecortin to mediate fiber tract decussation and neuronal migration. Neuron 49(1):55-66.

Kostka G, Giltay R, Bloch W, Addicks K, Timpl R, Fassler R, Chu ML. 2001. Perinatal lethality and endothelial cell abnormalities in several vessel compartments of fibulin-1-deficient mice. Mol Cell Biol 21(20):7025-34.

Kurima K, Peters LM, Yang Y, Riazuddin S, Ahmed ZM, Naz S, Arnaud D, Drury S, Mo J, Makishima T and others. 2002. Dominant and recessive deafness caused by mutations of a novel gene, TMC1, required for cochlear hair-cell function. Nat Genet 30(3):277-84.

Kurokawa R, Rosenfeld MG, Glass CK. 2009. Transcriptional regulation through noncoding RNAs and epigenetic modifications. RNA Biol 6(3).

Kutsuwada T, Sakimura K, Manabe T, Takayama C, Katakura N, Kushiya E, Natsume R, Watanabe M, Inoue Y, Yagi T and others. 1996. Impairment of suckling response, trigeminal neuronal pattern formation, and hippocampal LTD in NMDA receptor epsilon 2 subunit mutant mice. Neuron 16(2):333-44.

Lachmann T, Berti S, Kujala T, Schroger E. 2005. Diagnostic subgroups of developmental dyslexia have different deficits in neural processing of tones and phonemes. Int J Psychophysiol 56(2):105-20.

Laing MA, Coonrod S, Hinton BT, Downie JW, Tozer R, Rudnicki MA, Hassell JA. 2000. Male sexual dysfunction in mice bearing targeted mutant alleles of the PEA3 ets gene. Mol Cell Biol 20(24):9337-45.

Lander ES. 1996. The new genomics: global views of biology. Science 274(5287):536-9.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W and others. 2001. Initial sequencing and analysis of the human genome. Nature 409(6822):860-921.

Landerl K, Bevan A, Butterworth B. 2004. Developmental dyscalculia and basic numerical capacities: a study of 8-9-year-old students. Cognition 93(2):99-125.

Landerl K, Fussenegger B, Moll K, Willburger E. 2009. Dyslexia and dyscalculia: two learning disorders with different cognitive profiles. J Exp Child Psychol 103(3):309-24.

Landerl K, Wimmer H, Moser E. 1997. Salzburger Lese- und Rechtschreibtest. Bern: Huber.

Laurin N, Feng Y, Ickowicz A, Pathare T, Malone M, Tannock R, Schachar R, Kennedy JL, Barr CL. 2007. No preferential transmission of paternal alleles at risk genes in attention-deficit hyperactivity disorder. Mol Psychiatry 12(3):226-9.

Levine M, Tjian R. 2003. Transcription regulation and animal diversity. Nature 424(6945):147-51.

Li Q, Yu K. 2008. Improved correction for population stratification in genome-wide association studies by identifying hidden population structures. Genet Epidemiol 32(3):215-26.

Li WH, Sadler LA. 1991. Low nucleotide diversity in man. Genetics 129(2):513-23.

Liberman IY, Schankweiler D, Liberman AM. 1989. Phonology and reading disability: Solving the reading puzzle. In: Shankweiler D, Liberman IY, editors. International academy for research in learning disabilities monograph series. Ann Arbor: University of Michigan press.

Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, Franchimont D, Vermeire S, Dewit O, de Vos M, Dixon A and others. 2007. Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. PLoS Genet 3(4):e58.

Linhoff MW, Lauren J, Cassidy RM, Dobie FA, Takahashi H, Nygaard HB, Airaksinen MS, Strittmatter SM, Craig AM. 2009. An unbiased expression screen for synaptogenic proteins identifies the LRRTM protein family as synaptic organizers. Neuron 61(5):734-49.

Liu Y, Liu F, Iqbal K, Grundke-Iqbal I, Gong CX. 2008. Decreased glucose transporters correlate to abnormal hyperphosphorylation of tau in Alzheimer disease. FEBS Lett 582(2):359-64.

Livak KJ, Flood SJ, Marmaro J, Giusti W, Deetz K. 1995. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. PCR Methods Appl 4(6):357-62.

Loo SK, Fisher SE, Francks C, Ogdie MN, MacPhie IL, Yang M, McCracken JT, McGough JJ, Nelson SF, Monaco AP and others. 2004. Genome-wide scan of reading ability in affected sibling pairs with attention-deficit/hyperactivity disorder: unique and shared genetic effects. Mol Psychiatry 9(5):485-93.

LoTurco JJ, Wang Y, Paramasivam M. 2006. Neuronal migration and dyslexia susceptibility. In: Rosen GD, editor. The dyslexic brain: New pathways in neuroscience discovery-. Mahwah, New Jersey: Lawrence Erlbaum Associates. p 119-128.

Ludwig KU, Mattheisen M, Mühleisen TW, Roeske D, Schmäl C, Breuer R, Schulte-Körne G, Müller-Myhsok B, Nöthen MM, Hoffmann P and others. 2009a. Supporting evidence for LRRTM1 imprinting effects in schizophrenia. Mol Psychiatry 14(8):743-5.

Ludwig KU, Roeske D, Herms S, Schumacher J, Warnke A, Plume E, Neuhoff N, Bruder J, Remschmidt H, Schulte-Körne G and others. 2009b. Variation in GRIN2B contributes to weak performance in verbal short-term memory in children with dyslexia. Am J Med Genet B Neuropsychiatr Genet.

Ludwig KU, Roeske D, Sämann P, Alexander M, Landerl K, Becker J, Spieler D, Czich M, Plomin R, Nöthen MM and others. in preparation. A non-synonymous change in Myo18b is associated with low math performance in dyslexic children and correlates with structural abnormalities of the intraparietal sulcus.

Ludwig KU, Roeske D, Schumacher J, Schulte-Körne G, König IR, Warnke A, Plume E, Ziegler A, Remschmidt H, Müller-Myhsok B and others. 2008a. Investigation of interaction between DCDC2 and KIAA0319 in a large German dyslexia sample. J Neural Transm 115(11):1587-9.

Ludwig KU, Schumacher J, Schulte-Körne G, König IR, Warnke A, Plume E, Anthoni H, Peyrard-Janvid M, Meng H, Ziegler A and others. 2008b. Investigation of the DCDC2 intron 2 deletion/compound short tandem repeat polymorphism in a large German dyslexia sample. Psychiatr Genet 18(6):310-2.

Maher F, Simpson IA. 1994. The GLUT3 glucose transporter is the predominant isoform in primary cultured neurons: assessment by biosynthetic and photoaffinity labelling. Biochem J 301 (Pt 2):379-84.

Mantych GJ, James DE, Chung HD, Devaskar SU. 1992. Cellular localization and characterization of Glut 3 glucose transporter isoform in human brain. Endocrinology 131(3):1270-8.

Marino C, Giorda R, Luisa Lorusso M, Vanzin L, Salandi N, Nobile M, Citterio A, Beri S, Crespi V, Battaglia M and others. 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(4):491-9.

Massinen S, Tammimies K, Tapia-Paez I, Matsson H, Hokkanen ME, Soderberg O, Landegren U, Castren E, Gustafsson JA, Treuter E and others. 2009. Functional interaction of DYX1C1 with estrogen receptors suggests involvement of hormonal pathways in dyslexia. Hum Mol Genet.

Matsuzaki H, Dong S, Loi H, Di X, Liu G, Hubbell E, Law J, Berntsen T, Chadha M, Hui H and others. 2004. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. Nat Methods 1(2):109-11.

Maughan B. 1995. Annotation: long-term outcomes of developmental reading problems. J Child Psychol Psychiatry 36(3):357-71.

Maurer U, Bucher K, Brem S, Benz R, Kranz F, Schulz E, van der Mark S, Steinhausen HC, Brandeis D. 2009. Neurophysiology in Preschool Improves Behavioral Prediction of Reading Ability Throughout Primary School. Biol Psychiatry.

McCall AL, Van Bueren AM, Moholt-Siebert M, Cherry NJ, Woodward WR. 1994. Immunohistochemical localization of the neuron-specific glucose transporter (GLUT3) to neuropil in adult rat brain. Brain Res 659(1-2):292-7.

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9(5):356-69.

McCrory EJ, Mechelli A, Frith U, Price CJ. 2005. More than words: a common neural basis for reading and naming deficits in developmental dyslexia? Brain 128(Pt 2):261-7.

McGrath LM, Smith SD, Pennington BF. 2006. Breakthroughs in the search for dyslexia candidate genes. Trends Mol Med 12(7):333-41.

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(1):87-90.

Meng H, Smith SD, Hager K, Held M, Liu J, Olson RK, Pennington BF, DeFries JC, Gelernter J, O'Reilly-Pol T and others. 2005b. DCDC2 is associated with reading disability and modulates neuronal development in the brain. Proc Natl Acad Sci U S A 102(47):17053-8.

Mevorach C, Humphreys GW, Shalev L. 2005. Attending to local form while ignoring global aspects depends on handedness: evidence from TMS. Nat Neurosci 8(3):276-7.

Meyer MS, Wood FB, Hart LA, Felton RH. 1998. Longitudinal course of rapid naming in disabled and nondisabled readers. Ann of Dyslexia 48:89-114.

Miller SA, Dykes DD, Polesky HF. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16(3):1215.

Mok SC, Wong KK, Chan RK, Lau CC, Tsao SW, Knapp RC, Berkowitz RS. 1994. Molecular cloning of differentially expressed genes in human epithelial ovarian cancer. Gynecol Oncol 52(2):247-52.

Molko N, Cachia A, Riviere D, Mangin JF, Bruandet M, Le Bihan D, Cohen L, Dehaene S. 2003. Functional and structural alterations of the intraparietal sulcus in a developmental dyscalculia of genetic origin. Neuron 40(4):847-58.

Mullis KB, Faloona FA. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol 155:335-50.

Myers AJ, Gibbs JR, Webster JA, Rohrer K, Zhao A, Marlowe L, Kaleem M, Leung D, Bryden L, Nath P and others. 2007. A survey of genetic human cortical gene expression. Nat Genet 39(12):1494-9.

Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A fine-scale map of recombination rates and hotspots across the human genome. Science 310(5746):321-4.

Naylor CF, Felton RH, Wood FB. 1990. Adult outcome in developmental dyslexia. In: Pavlidis G, editor. Perspectives on dyslexia: cognition, language and treatment. Chichester, England: John Wiley & Sons. p 29.

Neale BM, Sham PC, Purcell S, Banaschewski T, Buitelaar J, Franke B, Sonuga-Barke E, Ebstein R, Eisenberg J, Mulligan A and others. 2008. Population differences in the International Multi-Centre ADHD Gene Project. Genet Epidemiol 32(2):98-107.

Nolan T, Hands RE, Bustin SA. 2006. Quantification of mRNA using real-time RT-PCR. Nat Protoc 1(3):1559-82.

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(10):658-64.

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(10):771-5.

Nyholt DR. 2004. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet 74(4):765-9.

Oeth P, Beaulieu M, Park C, Kosman D, del Mistro G, van den Boom D, Jurinke C. 2005. iPLEX Assay: Increased Plexing Efficiency and Flexibility for Mass ARRAY System Through Single Base Primer Extension with Mass-Modified Terminators. Sequenom Application Note:1-12.

Olson RK. 2002. Dyslexia: nature and nurture. Dyslexia 8(3):143-59.

Olson RK, Forsberg H, Wise B. 1994. Genes, environment, and development of orthographic skills. In: Berninger VW, editor. The varieties of orthographic knowledge I: theoretical and developmental issues. Dordrecht, Netherlands: Kluwer Academic Publishers. p 27-71.

Owen FW, Adams PA, Forrest T, Stolz LM, Fisher S. 1971. Learning disorders in children: sibling studies. Monogr Soc Res Child Dev 36(4):1-77.

Palmer S. 2000. Phonological recoding deficit in working memory of dyslexic teenagers Journal of Research in Reading 23(1):28-40.

Papassotiropoulos A, Stephan DA, Huentelman MJ, Hoerndli FJ, Craig DW, Pearson JV, Huynh KD, Brunner F, Corneveaux J, Osborne D and others. 2006. Common Kibra alleles are associated with human memory performance. Science 314(5798):475-8.

Paracchini S, Steer CD, Buckingham LL, Morris AP, Ring S, Scerri T, Stein J, Pembrey ME, Ragoussis J, Golding J and others. 2008. Association of the KIAA0319 dyslexia susceptibility gene with reading skills in the general population. Am J Psychiatry 165(12):1576-84.

Paracchini S, Thomas A, Castro S, Lai C, Paramasivam M, Wang Y, Keating BJ, Taylor JM, Hacking DF, Scerri T and others. 2006. The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. Hum Mol Genet 15(10):1659-66.

Paschen W, Djuricic B. 1994. Extent of RNA editing of glutamate receptor subunit GluR5 in different brain regions of the rat. Cell Mol Neurobiol 14(3):259-70.

Paulesu E, Demonet JF, Fazio F, McCrory E, Chanoine V, Brunswick N, Cappa SF, Cossu G, Habib M, Frith CD and others. 2001. Dyslexia: cultural diversity and biological unity. Science 291(5511):2165-7.

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):143-57.

Pennington BF. 2006. From single to multiple deficit models of developmental disorders. Cognition 101(2):385-413.

Penolazzi B, Spironelli C, Vio C, Angrilli A. 2006. Altered hemispheric asymmetry during word processing in dyslexic children: an event-related potential study. Neuroreport 17(4):429-33.

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(6):507-17.

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(2):125-6.

Plomin R, DeFries JC, McClearn GE, McGuffin P. 1997. Behavioral Genetics: W.H. Freeman & Co.

Plomin R, Kovas Y. 2005. Generalist genes and learning disabilities. Psychol Bull 131(4):592-617.

Pritchard JK. 2001. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 69(1):124-37.

Purvis KL, Tannock R. 1997. Language abilities in children with attention deficit hyperactivity disorder, reading disabilities, and normal controls. J Abnorm Child Psychol 25(2):133-44.

Rabin M, Wen XL, Hepburn M, Lubs HA, Feldman E, Duara R. 1993. Suggestive linkage of developmental dyslexia to chromosome 1p34-p36. Lancet 342(8864):178.

Ragoussis J, Elvidge GP, Kaur K, Colella S. 2006. Matrix-assisted laser desorption/ionisation, time-of-flight mass spectrometry in genomics research. PLoS Genet 2(7):e100.

Rajakumar A, Thamotharan S, Raychaudhuri N, Menon RK, Devaskar SU. 2004. Trans-activators regulating neuronal glucose transporter isoform-3 gene expression in mammalian neurons. J Biol Chem 279(25):26768-79.

Rajakumar RA, Thamotharan S, Menon RK, Devaskar SU. 1998. Sp1 and Sp3 regulate transcriptional activity of the facilitative glucose transporter isoform-3 gene in mammalian neuroblasts and trophoblasts. J Biol Chem 273(42):27474-83.

Ramus F, Rosen S, Dakin SC, Day BL, Castellote JM, White S, Frith U. 2003. Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. Brain 126(Pt 4):841-65.

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(7):699-711.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W and others. 2006. Global variation in copy number in the human genome. Nature 444(7118):444-54.

Regan JW, Bailey TJ, Pepperl DJ, Pierce KL, Bogardus AM, Donello JE, Fairbairn CE, Kedzie KM, Woodward DF, Gil DW. 1994. Cloning of a novel human prostaglandin receptor with characteristics of the pharmacologically defined EP2 subtype. Mol Pharmacol 46(2):213-20.

Reich DE, Lander ES. 2001. On the allelic spectrum of human disease. Trends Genet 17(9):502-10.

Reik W, Walter J. 2001. Genomic imprinting: parental influence on the genome. Nat Rev Genet 2(1):21-32.

Reiner O, Coquelle FM, Peter B, Levy T, Kaplan A, Sapir T, Orr I, Barkai N, Eichele G, Bergmann S. 2006. The evolving doublecortin (DCX) superfamily. BMC Genomics 7:188.

Rice TK, Schork NJ, Rao DC. 2008. Methods for handling multiple testing. Adv Genet 60:293-308.

Roeske D, Ludwig KU, Neuhoff N, Becker J, Bartling J, Bruder J, Brockschmidt FF, Warnke A, Remschmidt H, Hoffmann P and others. 2009. First genome-wide association scan on neurophysiological endophenotypes points to trans-regulation effects on SLC2A3 in dyslexic children. Mol Psychiatry.

Rogers LJ, Andrew R, editors. 2002. Comparative Vertebrate Lateralization. Cambridge, UK: Cambridge University Press.

Ross P, Hall L, Smirnov I, Haff L. 1998. High level multiplex genotyping by MALDI-TOF mass spectrometry. Nat Biotechnol 16(13):1347-51.

Rula EY, Lagrange AH, Jacobs MM, Hu N, Macdonald RL, Emeson RB. 2008. Developmental modulation of GABA(A) receptor function by RNA editing. J Neurosci 28(24):6196-201.

Rüsseler J. 2006. Neurobiologische Grundlagen der Lese-Rechtschreib-Schwäche. Zeitschrift für Neuropsychologie 17(2):101-111.

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(16):2007-12.

Rutter M, Yule W. 1975. The concept of specific reading retardation. J Child Psychol Psychiatry 16(3):181-97.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL and others. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409(6822):928-33.

Sachs GM, Jacobson M, Caviness VS, Jr. 1986. Postnatal changes in arborization patterns of murine retinocollicular axons. J Comp Neurol 246(3):395-408.

Salamon M, Millino C, Raffaello A, Mongillo M, Sandri C, Bean C, Negrisolo E, Pallavicini A, Valle G, Zaccolo M and others. 2003. Human MYO18B, a novel unconventional myosin heavy chain expressed in striated muscles moves into the myonuclei upon differentiation. J Mol Biol 326(1):137-49.

Salmelin R, Helenius P, Service E. 2000. Neurophysiology of fluent and impaired reading: a magnetoencephalographic approach. J Clin Neurophysiol 17(2):163-74.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74(12):5463-7.

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(11):853-7.

Schaub M, Keller W. 2002. RNA editing by adenosine deaminases generates RNA and protein diversity. Biochimie 84(8):791-803.

Schito AM, Pizzuti A, Di Maria E, Schenone A, Ratti A, Defferrari R, Bellone E, Mancardi GL, Ajmar F, Mandich P. 1997. mRNA distribution in adult human brain of GRIN2B, a N-methyl-D-aspartate (NMDA) receptor subunit. Neurosci Lett 239(1):49-53.

Schulte-Körne G. 2001. Lese-Rechtschreibstörung und Sprachwahrnehmung - Psychometrische und neurophysiologische Untersuchungen zur Legasthenie. (Dyslexia and Speech perception). Münster: Waxman Verlag.

Schulte-Körne G, Deimel W, Bartling J, Remschmidt H. 1998a. Auditory processing and dyslexia: evidence for a specific speech processing deficit. NeuroReport 9(2):337-40.

Schulte-Körne G, Deimel W, Bartling J, Remschmidt H. 2001a. Speech perception deficit in dyslexic adults as measured by mismatch negativity (MMN). Int J Psychophysiol 40(1):77-87.

Schulte-Körne G, Deimel W, Müller K, Gutenbrunner C, Remschmidt H. 1996. Familial aggregation of spelling disability. J Child Psychol Psychiatry 37(7):817-22.

Schulte-Körne G, Deimel W, Remschmidt H. 2001b. Diagnosis of reading and spelling disorder. Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie 29:113-116.

Schulte-Körne G, Grimm T, Nöthen MM, Müller-Myhsok B, Cichon S, Vogt IR, Propping P, Remschmidt H. 1998b. Evidence for linkage of spelling disability to chromosome 15. Am J Hum Genet 63(1):279-82.

Schulte-Körne G, Ludwig KU, el Sharkawy J, Nöthen MM, Müller-Myhsok B, Hoffmann P. 2007a. Genetics and Neuroscience in Dyslexia: Perspectives for Education and Remediation. Mind, Brain and Education 1(4):162-72.

Schulte-Körne G, Remschmidt H. 2003. Legasthenie - Symptomatik, Diagnostik, Ursachen, Verlauf und Behandlung. Deutsches Ärzteblatt 100(7):396-406.

Schulte-Körne G, Ziegler A, Deimel W, Schumacher J, Plume E, Bachmann C, Kleensang A, Propping P, Nöthen MM, Warnke A and others. 2007b. Interrelationship and familiality of dyslexia related quantitative measures. Ann Hum Genet 71(Pt 2):160-75.

Schumacher J, Anthoni H, Dahdouh F, König IR, Hillmer AM, Kluck N, Manthey M, Plume E, Warnke A, Remschmidt H and others. 2006. Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. Am J Hum Genet 78(1):52-62.

Schumacher J, Hoffmann P, Schmäl C, Schulte-Körne G, Nöthen MM. 2007. Genetics of dyslexia: the evolving landscape. J Med Genet 44(5):289-97.

Segi E, Sugimoto Y, Yamasaki A, Aze Y, Oida H, Nishimura T, Murata T, Matsuoka T, Ushikubi F, Hirose M and others. 1998. Patent ductus arteriosus and neonatal death in prostaglandin receptor EP4-deficient mice. Biochem Biophys Res Commun 246(1):7-12.

Sela M, Anfinsen CB, Harrington WF. 1957. The correlation of ribonuclease activity with specific aspects of tertiary structure. Biochim Biophys Acta 26(3):502-12.

Sellers JR. 2000. Myosins: a diverse superfamily. Biochim Biophys Acta 1496(1):3-22.

Sham P, Bader JS, Craig I, O'Donovan M, Owen M. 2002. DNA Pooling: a tool for large-scale association studies. Nat Rev Genet 3(11):862-71.

Shaywitz BA, Fletcher JM, Shaywitz SE. 1995. Defining and classifying learning disabilities and attention-deficit/hyperactivity disorder. J Child Neurol 10 Suppl 1:S50-7.

Shaywitz BA, Shaywitz SE, Blachman BA, Pugh KR, Fulbright RK, Skudlarski P, Mencl WE, Constable RT, Holahan JM, Marchione KE and others. 2004. Development of left occipitotemporal systems for skilled reading in children after a phonologically- based intervention. Biol Psychiatry 55(9):926-33.

Shaywitz BA, Shaywitz SE, Pugh KR, Mencl WE, Fulbright RK, Skudlarski P, Constable RT, Marchione KE, Fletcher JM, Lyon GR and others. 2002. Disruption of posterior brain systems for reading in children with developmental dyslexia. Biol Psychiatry 52(2):101-10.

Shaywitz SE. 1998. Dyslexia. N Engl J Med 338(5):307-12.

Shaywitz SE. 2003. Overcoming dyslexia: A new and complete science-based program for reading problems at any level. New York: Alfred A. Knopf.

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(3):145-50.

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(6):1351-9.

Shaywitz SE, Shaywitz BA. 2005. Dyslexia (specific reading disability). Biol Psychiatry 57(11):1301-9.

Shaywitz SE, Shaywitz BA. 2008. Paying attention to reading: the neurobiology of reading and dyslexia. Dev Psychopathol 20(4):1329-49.

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(8):998-1002.

Shianna KV, Willard HF. 2006. Human genomics: in search of normality. Nature 444(7118):428-9.

Shifman S, Bhomra A, Smiley S, Wray NR, James MR, Martin NG, Hettema JM, An SS, Neale MC, van den Oord EJ and others. 2008. A whole genome association study of neuroticism using DNA pooling. Mol Psychiatry 13(3):302-12.

Skipper L, Li Y, Bonnard C, Pavanni R, Yih Y, Chua E, Sung WK, Tan L, Wong MC, Tan EK and others. 2005. Comprehensive evaluation of common genetic variation within LRRK2 reveals evidence for association with sporadic Parkinson's disease. Hum Mol Genet 14(23):3549-56.

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(2):209-13.

Smith SD, Kimberling WJ, Pennington BF, Lubs HA. 1983. Specific reading disability: identification of an inherited form through linkage analysis. Science 219(4590):1345-7.

Sodhi MS, Burnet PW, Makoff AJ, Kerwin RW, Harrison PJ. 2001. RNA editing of the 5-HT(2C) receptor is reduced in schizophrenia. Mol Psychiatry 6(4):373-9.

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(3):506-16.

Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. 2005. Interchromosomal associations between alternatively expressed loci. Nature 435(7042):637-45.

Steemers FJ, Chang W, Lee G, Barker DL, Shen R, Gunderson KL. 2006. Whole-genome genotyping with the single-base extension assay. Nat Methods 3(1):31-3.

Steemers FJ, Gunderson KL. 2007. Whole genome genotyping technologies on the BeadArray platform. Biotechnol J 2(1):41-9.

Stephenson S. 1907. Six cases of congenital word-blindness affecting three generations of one family. Ophthalmoscope(5):482-484.

Stevenson J. 1991. Which aspects of processing text mediate genetic effects. Read Writ 3:249-269.

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(2):229-47.

Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D and others. 2007. Population genomics of human gene expression. Nat Genet 39(10):1217-24.

Strehlow U, Kluge R, Moller H, Haffner J. 1992. [Long-term course of dyslexia beyond the school years: catamnesis from pediatric psychiatric ambulatory care]. Z Kinder Jugendpsychiatr 20(4):254-65.

Suarez BK, Taylor C, Bertelsen S, Bierut LJ, Dunn G, Jin CH, Kauwe JS, Paterson AD, Hinrichs AL. 2005. An analysis of identical single-nucleotide polymorphisms genotyped by two different platforms. BMC Genet 6 Suppl 1:S152.

Suh SW, Aoyama K, Chen Y, Garnier P, Matsumori Y, Gum E, Liu J, Swanson RA. 2003. Hypoglycemic neuronal death and cognitive impairment are prevented by poly(ADP-ribose) polymerase inhibitors administered after hypoglycemia. J Neurosci 23(33):10681-90.

Swanson HL. 2006. Working memory and reading disabilities: Both phonological and executive processing deficits are important. In: Alloway TP, Gathercole SE, editors. Working memory and neurodevelopmental disorders Hove, UK: Psychology Press. p 59-88.

Taipale M, Kaminen N, Nopola-Hemmi J, Haltia T, Myllyluoma B, Lyytinen H, Muller K, Kaaranen M, Lindsberg PJ, Hannula-Jouppi K and others. 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(20):11553-8.

Tang K, Fu D, Kotter S, Cotter RJ, Cantor CR, Koster H. 1995. Matrix-assisted laser desorption/ionization mass spectrometry of immobilized duplex DNA probes. Nucleic Acids Res 23(16):3126-31.

Tang YP, Shimizu E, Dube GR, Rampon C, Kerchner GA, Zhuo M, Liu G, Tsien JZ. 1999. Genetic enhancement of learning and memory in mice. Nature 401(6748):63-9.

Temple E, Deutsch GK, Poldrack RA, Miller SL, Tallal P, Merzenich MM, Gabrieli JD. 2003. Neural deficits in children with dyslexia ameliorated by behavioral remediation: evidence from functional MRI. Proc Natl Acad Sci U S A 100(5):2860-5.

Temple E, Poldrack RA, Protopapas A, Nagarajan S, Salz T, Tallal P, Merzenich MM, Gabrieli JD. 2000. Disruption of the neural response to rapid acoustic stimuli in dyslexia: evidence from functional MRI. Proc Natl Acad Sci U S A 97(25):13907-12.

Tewes U. 1983. HAWIK-R. Hamburg-Wechsler Intelligenztest für Kinder. Bern: Hans Huber.

Thomas DC. 2004. Statistical methods in genetic epidemiology. New York: Oxford University Press.

Tiffin-Richards MC, Hasselhorn M, Woerner W, Rothenberger A, Banaschewski T. 2008. Phonological short-term memory and central executive processing in attention-deficit/hyperactivity disorder with/without dyslexia--evidence of cognitive overlap. J Neural Transm 115(2):227-34.

Tiret L, Bonnardeaux A, Poirier O, Ricard S, Marques-Vidal P, Evans A, Arveiler D, Luc G, Kee F, Ducimetiere P and others. 1994. Synergistic effects of angiotensin-converting enzyme and angiotensin-II type 1 receptor gene polymorphisms on risk of myocardial infarction. Lancet 344(8927):910-3.

Torgesen JK. 1995. Phonological awareness: A critical factor in dyslexia. Orton: Dyslexia Society.

Torgesen JK, Wagner RK, Rashotte CA, Rose E, Lindamood P, Conway T, Garvin C. 1999. Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. Journal of Educational Psychology 91:1-15.

Tunmer W, Chapman J. 2003. The reading recovery approach to preventive early intervention: As good as it gets? Read Psychol 24:405-428.

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 127B(1):117-24.

Unnewehr S, Schneider S, Markgraf J. 1998. Kinder-DIPS: diagnostisches Interview bei psychischen Störungen von Kindern und Jugendlichen. Berlin: Springer-Verlag.

Valdois S, Bosse M-L, Ans B, Carbonnel S, Zorman M, David D, Pellat J. 2003. Phonological and visual processing deficits can dissociatie in developmental dyslexia: Evidence from two case studies. Reading and Writing: An interdisciplinary journal 16:541-572.

Vellutino F, Scanlon D, Jaccard J. 2003. Toward distinguishing between cognitive and experiential deficits as primary sources of difficulty in learning to read: A two-year follow-up of difficult to remediate and readily remediated poor readers. In: Foorman BR, editor. Preventing and remediating reading difficulties: Bringing science to scale. Timonium, MD: York Press. p 73-120.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA and others. 2001. The sequence of the human genome. Science 291(5507):1304-51.

Wadsworth SJ, Knopik VS, DeFries JC. 2000. Reading disability in boys and girls: no evidence for a differential genetic etiology. Read Writ 13:133-145.

Wagner R, Torgesen JK. 1987. The nature of phonological processes and its causal role in the acquisition of reading skills. Psychol Bull 101:192-212.

Walker L, Cole E. 1965. Familiar patterns of expression of specific reading disability in a population sample. Bull Orton Soc 15:12-24.

Wallace MT. 2009. Dyslexia: bridging the gap between hearing and reading. Curr Biol 19(6):R260-2.

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J and others. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280(5366):1077-82.

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(2):515-22.

Wehner DT, Ahlfors SP, Mody M. 2007. Effects of phonological contrast on auditory word discrimination in children with and without reading disability: a magnetoencephalography (MEG) study. Neuropsychologia 45(14):3251-62.

Weinberg W. 1908. Über den Nachweis der Vererbung beim Menschen. Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg 64:368-382.

Weiß RH. 1998. Grundintelligenz Skala 2 (CFT-20). Göttingen: Hogrefe.

Weiß RH, Osterland J. 1997. Grundintelligenz Skala 1 (CFT-1). Göttingen: Hogrefe.

Westfall PH, Young SS. 1993. Resampling-based Multiple Testing. New York: John Wiley & Sons.

WHO. 1993. The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research. Geneva: World Health Organization.

Wijers AA, Been PH, Romkes KS. 2005. Dyslexics show a deviant lateralization of attentional control: a brain potential study. Neurosci Lett 374(2):87-91.

Wilkinson LS, Davies W, Isles AR. 2007. Genomic imprinting effects on brain development and function. Nat Rev Neurosci 8(11):832-43.

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(3):293-301.

Willcutt EG, Pennington BF, Olson RK, DeFries JC. 2007. Understanding comorbidity: a twin study of reading disability and attention-deficit/hyperactivity disorder. Am J Med Genet B Neuropsychiatr Genet 144(6):709-14.

Williams J, O'Donovan MC. 2006. The genetics of developmental dyslexia. Eur J Hum Genet 14(6):681-9.

Winkelmann J, Schormair B, Lichtner P, Ripke S, Xiong L, Jalilzadeh S, Fulda S, Pütz B, Eckstein G, Hauk S and others. 2007. Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. Nat Genet 39(8):1000-6.

Witke W. 2004. The role of profilin complexes in cell motility and other cellular processes. Trends Cell Biol 14(8):461-9.

WTCCC. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447(7145):661-78.

Yuan W, Zhou L, Chen JH, Wu JY, Rao Y, Ornitz DM. 1999. The mouse SLIT family: secreted ligands for ROBO expressed in patterns that suggest a role in morphogenesis and axon guidance. Dev Biol 212(2):290-306.

Zhang HY, Timpl R, Sasaki T, Chu ML, Ekblom P. 1996. Fibulin-1 and fibulin-2 expression during organogenesis in the developing mouse embryo. Dev Dyn 205(3):348-64.

Ziegler A, König IR. 2006. A statistical approach to genetic epidemiology. Weinheim: Wiley VCH Verlag.

Ziegler A, König IR, Deimel W, Plume E, Nöthen MM, Propping P, Kleensang A, Müller-Myhsok B, Warnke A, Remschmidt H and others. 2005. Developmental dyslexia--recurrence risk estimates from a german bi-center study using the single proband sib pair design. Hum Hered 59(3):136-43.

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(3):169-93.

Zielenski J, Tsui LC. 1995. Cystic fibrosis: genotypic and phenotypic variations. Annu Rev Genet 29:777-807.

9. LIST OF PUBLICATIONS

Mangold E*, **Ludwig KU***, Birnbaum S*, Baluardo C, Ferrian M, Herms S, Reutter H, Almeida de Assis N, Al Chawa T, Mattheisen M, Steffens M, Barth S, Kluck N, Paul A, Becker J, Lauster C, Schmidt G, Braumann B, Scheer M, Reich RH, Hemprich A, Pötzsch S, Blaumeiser B, Moebus S, Krawczak M, Schreiber S, Meitinger T, Wichmann E, Steegers-Theunissen RP, Kramer F-J, Cichon S, Propping P, Wienker TF, Knapp M, Rubini M, Mossey PA, Hoffmann P and Nöthen MM: **Genome-wide association study identifies new susceptibility loci for non-syndromic cleft lip with or without cleft palate.** ***In press, Nature Genetics.***

Roeske D*, **Ludwig KU***, Neuhoff N, Becker J, Bartling J, Bruder J, Brockschmidt FF, Warnke A, Remschmidt H, Hoffmann P, Müller-Myhsok B, Nöthen MM and Schulte-Körne G: **First genome-wide association scan on neurophysiological endophenotypes points to transregulation-effects on *SLC2A3* in dyslexic children.** ***Molecular Psychiatry*. Advance online publication, doi 10.1038/mp.2009.102.**

Birnbaum S, **Ludwig KU**, Reutter H, Herms S, Almeida de Assis N, Diaz-Lacava A, Barth S, Lauster C, Schmidt G, Scheer M, Saffar M, Martini M, Reich RH, Schiefke F, Hemprich A, Pötzsch S, Pötzsch B, Wienker TF, Hoffmann P, Knapp M, Kramer F-J, Nöthen MM and Mangold E: ***IRF6* gene variants in Central European patients with non-syndromic cleft lip with or without cleft palate.** ***In press, European Journal of Oral Sciences.***

Nikopensius T, Ambrozaitytė L, **Ludwig KU**, Birnbaum S, Jagomägi T, Saag M, Matulevičienė A, Linkevičienė L, Herms S, Knapp M, Hoffmann P, Nöthen MM, Kučinskas V, Metspalu A and Mangold E: **Replication of novel susceptibility locus for non-syndromic cleft lip with or without cleft palate on chromosome 8q24 in Estonian and Lithuanian patients.** ***American Journal of Medical Genetics, Part A* 2009: 149A(11):2551-3.**

Ludwig KU, Mattheisen M, Mühleisen TW, Roeske D, Schmäl C, Breuer R, Schulte-Körne G, Müller-Myhsok B, Nöthen MM, Hoffmann P, Rietschel M and Cichon S: **Supporting evidence for *LRRTM1* imprinting effects in schizophrenia.** ***Molecular Psychiatry* 2009: 14(8):743-5.**

Ludwig KU*, Roeske D*, Herms S, Schumacher J, Warnke A, Plume E, Neuhoff N, Bruder J, Remschmidt H, Schulte-Körne G, Müller-Myhsok B, Nöthen MM and Hoffmann P: **Variation in *GRIN2B* contributes to weak performance in verbal short-term memory in children with dyslexia.** ***American Journal of Medical Genetics, Part B.* Advance online publication, doi 10.1002/ajmg.b.31007.**

Treutlein J, Cichon S, Ridinger M, Wodarz N, Soyka M, Zill P, Maier W, Moessner R, Gaebel W, Dahmen N, Fehr C, Scherbaum N, Steffens M, **Ludwig KU**, Frank J, Wichmann H-E, Schreiber S, Dragano N, Sommer W, Leonardi-Essmann F, Lourdusamy A, Gebicke-Haerter P, Wienker TF, Sullivan PF, Nöthen MM, Kiefer F, Spanagel R, Mann K and Rietschel M: **Genome-wide association study of alcohol dependence.** *Archives of General Psychiatry* **2009: 66(7):773-84.**

Engels H, Wohlleber E, Zink A, Hoyer J, **Ludwig KU**, Brockschmidt FF, Wieczorek D, Moog U, Hellmann-Mersch B, Weber RG, Willatt L, Kreiß-Nachtsheim M, Firth HV and Rauch A: **A novel microdeletion syndrome involving 5q14.3-q15: Clinical and molecular cytogenetic characterisation of three patients.** *European Journal of Human Genetics.* **2009: 17(12):1592-9.**

Treutlein J, Mühleisen TW, Frank J, Mattheisen M, Herms S, **Ludwig KU**, Treutlein T, Schmäl C, Strohmaier J, Böβhenz KV, Breuer R, Paul T, Witt SH, Schulze TG, Schlösser RGM, Nenadic I, Sauer H, Becker T, Maier W, Cichon S, Nöthen MM and Rietschel M: **Dissection of phenotype reveals possible association between schizophrenia and G*lutamate Receptor Delta 1* (*GRID1*) gene promoter.** *Schizophrenia Research* **2009: 111(1-3):123–130.**

Birnbaum S*, **Ludwig KU***, Reutter H, Herms S, Steffens M, Rubini M, Baluardo C, Ferrian M, Almeida de Assis N, Alblas MA, Barth S, Freudenberg J, Lauster C, Schmidt G, Scheer M, Braumann B, Bergé SJ, Reich RH, Schiefke F, Hemprich A, Pötzsch S, Steegers-Theunissen RP, Pötzsch B, Moebus S, Horsthemke B, Kramer F-J, Wienker TF, Mossey PA, Propping P, Cichon S, Hoffmann P, Knapp M, Nöthen MM* and Mangold E*: **Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24.** *Nature Genetics* **2009: 41(4):473–477.**

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 and Hoffmann P: **Investigation of the *DCDC2* intron 2 deletion/compound short tandem repeat polymorphism in a large German dyslexia sample.** *Psychiatric Genetics* **2008: 18:310–312.**

Ludwig KU, Roeske D, Schumacher J, Schulte-Körne G, König IR, Warnke A, Plume E, Ziegler A, Remschmidt H, Müller-Myhsok B, Nöthen MM and Hoffmann P: **Investigation of interaction between *DCDC2* and *KIAA0319* in a large German dyslexia sample.** *Journal of Neural Transmission* **2008: 115(11):1587-1589.**

Schulte-Körne G, **Ludwig KU**, el Sharkawy J, Nöthen MM, Müller-Myhsok B and Hoffmann P: **Genetics and Neuroscience in Dyslexia: Perspectives for Education and Remediation.** ***Mind, Brain and Education*** **2007: 1(4):162-172.**

* Authors contributed equally to this work.

10. ATTACHMENTS

Attachment I: Overview of primers used in the present thesis.
Primers are depicted with their respective sequences in 5′ - 3′ direction. For SNPs included in Sequenom assays, the two standard primers (1st, 2nd) and the UEP primers are given for each SNP. The column 'UEP conc.' refers to the concentration of the UEP primers in µM, based on previous primer adjustment. Different mass groups per assay are separated by grey – white shadings. cont. – Table continued, FAC3 – Working title for basic mathematical factor.

A) Primer sequences for PCR and TaqMan assays

Locus	Name	Sequence 5' - 3'	Product size (bp)	Remarks / PCR conditions differing from standard protocol
MYO18B	Myo18B_Fn	GCCATCTCATCACGCCTCGC	418 bp	-
	Myo18B_Rn	TGGTTGGTGTCGCACTGGAGC		
Cyclo-philin	Cyc_F	unknown	unknown	ABI TaqMan assay 4326316E
	Cyc_R	unknown		
	Cyc probe	unknown	probe	
SLC2A3	SLC2A3_UTR/ex2F	ACCCCTAGATCTTTCTTGAAGACTTGA	98 bp	Custom ABI TaqMan assay SLC2A3-1ex2
	SLC2A3_UTR/ex2R	GATTGTAGCAACTGTGATGGCAAAT		
	SLC2A3_ex1/2	CAGAAGGTCACCCCAGCTC	probe	
DCDC2	Del_F_original	TGTAAAACGACGGCCAGTAGCCTGCCTACCACAGAGAA	215 bp (deleted) 525 bp (non-deleted)	Protocol as in Meng *et al.* 2005b
	Del_R_original	TCACACAGGAAACAGCTATGACTGAAACCCCGTCTCTACTGAA		
	Del_C_original	TCACACAGGAAACAGCTATGACGGAACAACCTCACAGAAATGG		

B) Primer used in Sequenom assays

Assay	SNP-ID	1st primer (sequence 5' - 3')	2nd primer (sequence 5' - 3')	UEP primer (sequence 5' - 3')	UEP conc. (µM)
KIAA0319	**rs3212236**	ACGTTGGATGCAAATGTCCCAACCAAAAGC	ACGTTGGATGATATGCCAGACACTGCAAG	ACACTGCAAGAGGTAG	7
	rs1555090	ACGTTGGATGGAGTACCTGTGGAATGCCAT	ACGTTGGATGTCTGATTCCAAGACTCTCCC	AAAAGACACTTCCTCCTTTCT	9.3
	rs761100	ACGTTGGATGAGACCAGACTCATTTTCATC	ACGTTGGATGTGTGGCTCACCATTCAGAAG	cTTAAGTTTCCAGTTTCTCC	9.3
	rs7766230	ACGTTGGATGCTGGTGAAGGGTTAGCAAAG	ACGTTGGATGATCTCCCTGATTAGTGCACC	TGCATATGCAGAGATACTGTTG	11.6
	rs17491230	ACGTTGGATGTTCCCAACCTCACGAATTGC	ACGTTGGATGGCAAATAATAGGAAAGATCAC	gCACATAAATAGTTCTGAATGAA	11.6
	rs2179515	ACGTTGGATGTTACTCAGTTCATTTTGCCC	ACGTTGGATGCATGTCCTGAATGTAGGAGC	ATGTAGGAGCCTCCTTCCAAATACTGC	14
LRRTM1	**rs6718055**	ACGTTGGATGCTCTGCTCCATAGTGATTGC	ACGTTGGATGAAAATAAAAGAAACACAGTC	cAACACAGTCTCAGTT	7
	rs10170020	ACGTTGGATGTATAAATGATTGCACCTGCG	ACGTTGGATGCTCTCATGGTGAGTCTGATA	TAGTCCTCTTGAGAATG	9.3
	rs2862286	ACGTTGGATGGGCAGACATAAACTTGTCCC	ACGTTGGATGGCAAGCTACTGAACCTTTTC	CAGATTTGAGTCTTCTTG	9.3
	rs6755232	ACGTTGGATGCAACCTGGACTTTGCTCTTG	ACGTTGGATGTTTGTTTCTGTGTGCCCCTC	CCCTGTTTTCCTTTATGCC	9.3
	rs11126755	ACGTTGGATGCTCAAGTGTCCATGATCAAC	ACGTTGGATGAATTGCCATCCCTGTGAGTG	CATCCTGGATTGTCCAGTCTTC	11.6
	rs13019601	ACGTTGGATGGGGTACAACATCCACATACA	ACGTTGGATGGGAAGACTTTGCTACACCTG	CTACACCTGTGTTATAGTAT	11.6
	rs1446109	ACGTTGGATGTACATAATGCCAAATTGTC	ACGTTGGATGGCAGCAAAGCAACTATTGAG	GGATGAAAAGTGTGAAATCTT	11.6
	rs767587	ACGTTGGATGTTTGTGAAGCTGGGCCTTTC	ACGTTGGATGCTTGTCTGCATTGCTCTACG	GGGCTTGGAAAGAATCCTA	11.6
	rs1007371	ACGTTGGATGAGCATGGCAGTTTTTGGATA	ACGTTGGATGCAAATTAGGGAAAGTGGATGC	AGGGAAAGTGGATGCTATTGTTT	14
	rs1446110	ACGTTGGATGTACAACATATCTGTGGTAT	ACGTTGGATGCCCCTTGAAAATACTATTTAG	gTTTTCCTAATTACTCCCCCTCCCTG	14
	rs1930	ACGTTGGATGGGATGACACTGATGGTGACG	ACGTTGGATGATTTTGCACATGCTGGGCAC	gTCAGTAAGTGTTATTACTATC	14
	rs6712681	ACGTTGGATGATGCAGGACTCTGCCAATAG	ACGTTGGATGGGGTCGGACTGAATGTACTG	GGACTGAATGTACTGGCCATTTCT	14
	rs6733871	ACGTTGGATGATTGCGGGCGCAACGTGTGT	ACGTTGGATGACTGCAAGTTGCCATCGTAG	aATCGTAGCGCCCCTGGAAGTTG	14
	rs723524	ACGTTGGATGTACATGACTTGATCATGTG	ACGTTGGATGATCTCCCACATGCTGTGTTG	CTGTGTTGGAAATATATTTCCCTAT	14
GRIN2B - Plex1	**rs7974275**	ACGTTGGATGGAAACTCTTGCTGAATTTTG	ACGTTGGATGGACTAATGTAACAACCAACTG	CCAACTGGTCTCCAT	7
	rs220557	ACGTTGGATGTTCAGGTTTTGTCCTCCCAG	ACGTTGGATGCTTCGATGCCCAGGACAATG	TGAGGTAGCTTGGTG	7
	rs2284418	ACGTTGGATGAGTCTGAAGTTCCTACTGTG	ACGTTGGATGTGGTTTGAAGGGGCATGGTC	GCATGGTCACTGTTCA	7
	rs2192973	ACGTTGGATGCTAAACAAAAGTGCCTGAAAG	ACGTTGGATGCTCAGAGGTATTTCCAGACA	cgTTCCAGACACTCCCA	7
	rs220549	ACGTTGGATGAGGAGTGACTAAAATACGGC	ACGTTGGATGGGGACTCCTCAAATTGAGC	acTGAGCTATTGTGCCC	7
	rs17220663	ACGTTGGATGAGCCTGAACATGGAATTCCC	ACGTTGGATGCTGAGCTTGGGCTTTCAAGG	GCTTTCAAGGTGGTTAA	7
	rs2284424	ACGTTGGATGGTTCAACTATGAAGAAGAAG	ACGTTGGATGCTCTTCCTCCTCTCTCAAAC	gCTCTCAAACCTCCCTTA	7
	rs1158541	ACGTTGGATGTTGGCAAGGCAGAGAGTAGG	ACGTTGGATGACAGGCATGGAAACCATCAG	CATCAGAGGAAGCATACA	9.3
	rs2284425	ACGTTGGATGGTCCTGAACTGGATCTTTTC	ACGTTGGATGCTGAAGATACAGGTTAACTAC	gaccAATGCAGTCCCCAAC	9.3
	rs2268120	ACGTTGGATGTACTGTTTCAGATGAAGGGG	ACGTTGGATGTGGTGTCTAGGACATAGCAG	tgGGACATAGCAGGTCCCC	9.3
	rs220597	ACGTTGGATGAGATAGGCAGCAGACATTTG	ACGTTGGATGGGCTGAATCTCAGTACCAAT	TGAGGAAAAGTACCCTGAG	9.3
	rs2193150	ACGTTGGATGTTCAGATATGGTGCCATCGC	ACGTTGGATGCTTCTGCTCCCAATTAGCTG	CTCCCAATTAGCTGTGTAAC	9.3

Assay	SNP-ID	1st primer (sequence 5' - 3')	2nd primer (sequence 5' - 3')	UEP primer (sequence 5' - 3')	UEP conc. (µM)
GRIN2B - Plex1 (cont.)	**rs220563**	ACGTTGGATGTACCTGTCATGCCTGCCTTG	ACGTTGGATGTAGATAGAGGTAGAGGAATC	GAGGTAGAGGAATCATATTG	9.3
	rs1805555	ACGTTGGATGGCAATCCAGGCTAACAAAAC	ACGTTGGATGTTTGAAGCTCTCATTGGAGG	gcTTGGAGGATAAGAAGGAA	9.3
	rs2300252	ACGTTGGATGCTGAACAGTACTGCAAGAAG	ACGTTGGATGCATACACAGAGAGGTTAATC	gttcTCCAAGGTGACAGGTAT	11.6
	rs2300242	ACGTTGGATGGGGAGTTTCAAGTCAAAGTG	ACGTTGGATGGGATAATATCTTGGTTCTCC	agATCTTGGTTCTCCTTATACA	11.6
	rs2110984	ACGTTGGATGGCAGACTTAGCAGCAAAGAG	ACGTTGGATGACTTCGCACTCTCTCTTGTC	ctaACTCTCTCTTGTCACTCTGA	11.6
	rs2216127	ACGTTGGATGGATCAGTTTATCAGTGTATC	ACGTTGGATGCATACCAAACAAGGGAGATG	ATTCTATTGTACTTGACTGTAAG	11.6
	rs765688	ACGTTGGATGAGGCTCCAGAGTAACATGTC	ACGTTGGATGTTTACCATAATGGCCCAAGG	aCCAAGGAATCAATAATGATTAG	11.6
	rs7966866	ACGTTGGATGAGCCTAACCATATCAGGTTG	ACGTTGGATGGTCTTCCAATCTCAATTTC	TCCAATCTCAATTTCATAAGTTTC	11.6
	rs11055582	ACGTTGGATGGGGAGAGCATCAGGAAAAAC	ACGTTGGATGCCTATCAACCCATCACCTAAA	caacCCCATCACCTAAATATTAAGC	11.6
	rs12809496	ACGTTGGATGGTACCTGGGAATTGGAGAAG	ACGTTGGATGAATGCCCCATGACTGATTTC	ccctTGATTTCAGTGGAATATCTTC	14
	rs11612284	ACGTTGGATGAAGGAATGTAGCTGGAGTGC	ACGTTGGATGACCATGATGATGAAGCCACG	gattATGAAGCCACGGATCTCAAAT	14
	rs1861452	ACGTTGGATGACAAAGCCCATTTACTATG	ACGTTGGATGATGAAACCACAGCTGGCTGA	ggccTAGCAAGAGGTACAAAAGCCA	14
	rs2268119	ACGTTGGATGTTCCAAAGGGACAAGTTAGC	ACGTTGGATGCACATAGCTACTTTTTGCATC	cctCTACTTTTTGCATCTTTACATTG	14
	rs11055593	ACGTTGGATGCTGAAGATTGGGGAGAAATG	ACGTTGGATGCAGCCTCATGACCACATTAC	gACTATTTTCTTATCTCTCTGATAAAT	14
	rs17833967	ACGTTGGATGTTTTCCCCGTGTCTATCTTC	ACGTTGGATGGAAGTCTAGCCAAATAAACA	GAACAAAGAATAATACTTTACTAAATG	14
	rs2300251	ACGTTGGATGTGAGCACTTATTTATGCTA	ACGTTGGATGCTTGTATGTAATAGGAAAATG	ATGTAATAGGAAAATGATATTTTATAG	14
GRIN2B - Plex2	**rs10845837**	ACGTTGGATGAGACTTGCATGGCAGTCCAC	ACGTTGGATGAAGGAGAGTAAGACACAAAG	CATATGCCACGACAAG	7
	rs220583	ACGTTGGATGCTGTGAGACGGACTCAGAAG	ACGTTGGATGGGCTTCTTAAAAAAGGGCAC	CACCAAGTCAAGGACTG	7
	rs2268122	ACGTTGGATGCTTTATTGGGTGTTTTGTTC	ACGTTGGATGATGAAAGCTGGCTAAAGAAG	GGCTAAAGAAGAGAGGG	7
	rs2216128	ACGTTGGATGACACTTCAGACTGAGAGACT	ACGTTGGATGTGAGGAACAGTGAGGAAGTG	TGGCCTGGGGATGCTTCA	9.3
	rs1012586	ACGTTGGATGGGAAAGCCACTCTTTAACTC	ACGTTGGATGCCAGGGATTCAACTCACTTC	CCAAATACCCCATAAAGAC	9.3
	rs7970407	ACGTTGGATGGAGGAATTCCAAGGTTATGC	ACGTTGGATGAAATCAGATCGATCCAGAGG	ACTAATAGTGGCAATGCTA	9.3
	rs17833639	ACGTTGGATGCTGTGAAATCCCTTTCCCAC	ACGTTGGATGGGAAATCAGGGCTATTACAC	AGGGCTATTACACTATGTAT	11.6
	rs2300273	ACGTTGGATGCCATTTGCAGGCTGGCAATC	ACGTTGGATGTGCCTTACCTCACTAAACAC	CCTCACTAAACACTCTGTCCA	11.6
	rs11055616	ACGTTGGATGGCCTGTATTTATTGTCCCTC	ACGTTGGATGTAAAAGGGCCCGTTCAAGAG	TTCAAGAGTGCCTAGTCCTCC	11.6
	rs11055608	ACGTTGGATGAGAGGGAAACCAGATTGATT	ACGTTGGATGGTACTGTATAATATCATAGGG	ATAGGGAATATAATCCTGTTAG	14
	rs2300245	ACGTTGGATGAGGGCTGTAAAAGCCTGTG	ACGTTGGATGATTGTCCCTGATGCCATCAC	CCCTGATGCCATCACCATAAAGA	14
	rs220599	ACGTTGGATGTAACCCTCACAGCTACATAC	ACGTTGGATGAGGTCCACCCTCTCAGTTAG	GTAAAAGATCTAAACATAATCAATC	14
	rs2160734	ACGTTGGATGTGAGCTGGCATCTTCTAGGG	ACGTTGGATGAGCTGGGAGAATTCTCAGTG	TGGGAGAATTCTCAGTGTTTATTAA	14
	rs1005549	ACGTTGGATGTCGTTAGCACAAGTCAATAC	ACGTTGGATGAATAACAAATATTCATTTGG	ATAACAAATATTCATTTGGTATAAATA	14
GRIN2B - Plex3	**rs1805247**	ACGTTGGATGTGGGCTGCCTGAAGAAGTAG	ACGTTGGATGCAAAACCCTTTCATCCCCAC	CCAGTGCTTGCTCCA	7
	rs7301328	ACGTTGGATGTCGCCCAGATCCTCGATTTC	ACGTTGGATGATTATCATAGAGGAGCCCCC	GTGGATGCCCAGGAT	7
	rs12830358	ACGTTGGATGGTCTTGGTCAGTTGGAGGTA	ACGTTGGATGCAAATAGGGCCTGAGGTAAC	cAACCTCCAGCTTCTGC	7
	rs2192977	ACGTTGGATGCATCTAGATGAGAAAAACCAC	ACGTTGGATGATAAGATGTGGGTTTGACTG	ACTTTTGACCAGCACTT	7
	rs10845838	ACGTTGGATGGACTTTTGCTTACTTTGTG	ACGTTGGATGAAAGATGGATAACCTGGGGC	gcCCTGGGGCAAAGTAC	7
	rs11055711	ACGTTGGATGGGTTAAAAATACATAAAACTC	ACGTTGGATGTTATTTCTTCTGCCAAAGGG	gaCAAAGGGGGACATCG	7
	rs219876	ACGTTGGATGCCAGGAAGTGATAGAAACAG	ACGTTGGATGAGTTGAATACCTCTGCCTTC	TCTGCCTTCTCTTTGTCAT	9.3
	rs2284416	ACGTTGGATGCTCTTCTCACCATTCTTAGG	ACGTTGGATGGAAATTATTGACGATGTTCC	CGATGTTCCTTAAATGCCA	9.3
	rs2284411	ACGTTGGATGACATGACTTTTTTCCCCTAC	ACGTTGGATGGTCTGCTAGGAGCATAAAAG	CATAAAAGGAGAGCATCAA	9.3
	rs17820659	ACGTTGGATGACAGTTTGCAGAATAGCTTC	ACGTTGGATGCTCTTATCACTCAAACCACC	CTCAAACCACCTCACAGATC	9.3
	rs12818068	ACGTTGGATGGGTTACGTGATGTAGATCC	ACGTTGGATGCTCTGGGCTTCATCTTCAAC	aaTTCATCTTCAACTCGTCGA	9.3
	rs11055557	ACGTTGGATGTCTTTCGTTTATTCAGCAGG	ACGTTGGATGATCATCACTGTCCTCAGCAC	ATTCCTGACTCCTGATTGTTG	9.3
	rs3026160	ACGTTGGATGATGGCTCTCAGCCTCATCAC	ACGTTGGATGCACCCATAAAGCAATGTCGG	aTGCCAATAGAAAAGGTGTTC	11.6
	rs2268115	ACGTTGGATGGAAACTAACACTACTATATG	ACGTTGGATGCAAAATAGAGGGCATTTCGC	ccCCTATCATTCCATTCCTGTC	11.6
	rs1805482	ACGTTGGATGTGTTCTAACCCTGAGCACTG	ACGTTGGATGCATCACAAACATCATCACCC	cACATCATCACCCATACGTCAGC	11.6
	rs2300256	ACGTTGGATGGGAATAGTCTTGCTCTGAAG	ACGTTGGATGAGTGTTTACTGCATGTTAC	ttCATGTTACTTTTCCTAAGAGT	11.6
	rs2041986	ACGTTGGATGTCTATTATCTTGTCTGCAGC	ACGTTGGATGACCTCCCAAAGAGGTTATAG	ggCCAAAGAGGTTATAGCTATTC	11.6
	rs9971835	ACGTTGGATGCACTGCTCCAGAAGGGAAGA	ACGTTGGATGACTCCTTCTTCGGGTCACTC	ACCAAACATTTCTAGATTCCATGA	11.6
	rs1421109	ACGTTGGATGTTTTTTCCTCACCTGGGGTA	ACGTTGGATGCAAAAAGATCCACTTGAGGG	cGGGACTCGAAGAGATAGATAAAA	14
	rs12828473	ACGTTGGATGCTCACCCACTTTTAATTTCTG	ACGTTGGATGCTGTCTAATATGTTGTGTTC	catcTGTTGTGTTCATTATAAAGCA	14
	rs1806201	ACGTTGGATGGGATGTTGGAGTGTGTGTTG	ACGTTGGATGAGCGCCAGTCTGTAATGAAC	CGCCAGTCTGTAATGAACTCCCCCAC	14
	rs933614	ACGTTGGATGATGAGGTGGAGTGTGCTTAC	ACGTTGGATGCAGAATGGTAGGCACTGTTA	cgtCACTGTTAAAATAATGAGGTAAA	14
	rs2284407	ACGTTGGATGGTGCCTGAAATTGATTGGCT	ACGTTGGATGACTCATTGAGAGAGAGGAAG	gaaggGAAAGTCAGCTATTTTGAGAG	14
	rs7964322	ACGTTGGATGTCTGCTGCCCCATAATGTAG	ACGTTGGATGCACAGAGCTTCTTGAGCAAT	ccTTATCAGGTTTGTTCAATTCACAGC	14
NeuroDys - Replication	**rs958877**	ACGTTGGATGGGCATCAATTTACTGATGCG	ACGTTGGATGGCACCCAAATGTTCTCTGGC	CCTCAGGCTCTGTGG	7
	rs9465637	ACGTTGGATGAAAAAATAAAATTGAAATTC	ACGTTGGATGTTACTGTTGACCACCCATAC	TGCTGGGTTGGTGTT	7
	rs1003346	ACGTTGGATGTGTAGAATGGTGGTAAGGGT	ACGTTGGATGCATAGAGCTCACTGTGTACC	CTGTGTACCAGGCAAT	7
	rs6498274	ACGTTGGATGAGGTATCTGTGTCCTGGTTG	ACGTTGGATGCACTTAGCCATGGGACTTAG	AGAGCAGTGGAGCCCG	7
	rs366078	ACGTTGGATGACACTTTCCTTAGTGAGGGC	ACGTTGGATGCAGAGGAGGCAGCGTGTGG	GAGCGTGTGGTAAGTG	7
	rs10123957	ACGTTGGATGAGCCATCTATCTTCCCTTAG	ACGTTGGATGGCCTACTCTACAGATACAAG	GTTGCTTAGCTCTGTGG	7
	rs4327894	ACGTTGGATGCATTTCAACATCATCCCCTC	ACGTTGGATGCTTCCTTGTACAGGTGAGTG	TAGGTGAGTGAGAGACC	7
	rs4887111	ACGTTGGATGAATGCCCAGTGCCTAGCACG	ACGTTGGATGTGGAACCTGGTGATGGAGAG	GTGATGGAGAGAAGTCA	7

Assay	SNP-ID	1st primer (sequence 5' - 3')	2nd primer (sequence 5' - 3')	UEP primer (sequence 5' - 3')	UEP conc. (µM)
NeuroDys - Replication (cont.)	**rs11117425**	ACGTTGGATGTCTAACTGATTCAGGGCTGG	ACGTTGGATGGCGGGCCATTTGCAGAACAT	TGCAGAACATTCCTTTGA	9.3
	rs2077268	ACGTTGGATGAGGGAATGTTGGCCCTTGTC	ACGTTGGATGTGCAATCCCTGCAAAGTGTG	GTGCTACGCTATTGTAGA	9.3
	rs10512712	ACGTTGGATGTCTCCCTCCAGTGGTCAATC	ACGTTGGATGAGGTATTTTAGGCAGGTAGC	GAGCAGCAGGTGAAAAGC	9.3
	rs1872285	ACGTTGGATGAGTGGATGGTTTTCAGGCAG	ACGTTGGATGGTTGGAAGTTGAAAGTAATGC	TTGAAAGTAATGCATGCAC	9.3
	rs6984900	ACGTTGGATGTCTAGAGGCACACAGTAGAG	ACGTTGGATGGATGGATCTAGAATTCATTCC	TTTTCCTTTTCACCTTAACT	9.3
	rs9529688	ACGTTGGATGACTGACCTGGTGATCTTATC	ACGTTGGATGTCTTGCTCAAGAGAAAGGAC	CAAGATCCTTTGCTAACCTA	9.3
	rs7840675	ACGTTGGATGAGCATCACCACAAATGTGAG	ACGTTGGATGATTCCTGTTGCCTAGTGCTG	GCCAACATAATGTGGTACTG	9.3
	rs1429411	ACGTTGGATGTACAATACCAGGGAGCCTAC	ACGTTGGATGATGCTGTTTCTCAGGCCATC	AAAGGACAAATGACGTTTTA	9.3
	rs10816767	ACGTTGGATGGGGAGTTGGTTCATCTTTTC	ACGTTGGATGAAACTAAGGACGAAAAGAGG	GACGAAAAGAGGAAAAAAAG	9.3
	rs7623540	ACGTTGGATGGTAAATAACTGTACCAGGGC	ACGTTGGATGCAAGTTAGGCAAAATTCTCTC	CAAATTCTCTCTTGCAGATTT	11.6
	rs10513829	ACGTTGGATGAATTACCCAGCTGTAGCGTG	ACGTTGGATGCTCCTGAGAATGAAGAGTAG	TCGTTTAGGAGACCGTACTAA	11.6
	rs4747165	ACGTTGGATGTTAGACGCATGTCAAGCCTG	ACGTTGGATGGAAATGGAGATAAAAGGAG	GGGAGATAAAAGGAGCTGTTA	11.6
	rs10518444	ACGTTGGATGAGATGAGAGAACTCCATCTG	ACGTTGGATGATCATAAGCCTAAGATTCC	GCCTAAGATTCCTAATGATATT	11.6
	rs12454776	ACGTTGGATGCCCCATTTAGAATAGTCTAC	ACGTTGGATGGTGCTTACTGTTGAACTAAC	CAATTCAAAAGGAGACAAACTC	11.6
	rs6136213	ACGTTGGATGGTGCTGCAATGGGACAGACT	ACGTTGGATGGGTCTCCACTCCTCGTTATG	CCTCCTCGTTATGGTAACTCTCG	11.6
	rs1181841	ACGTTGGATGCACTAAGAGTAGAAGAATAC	ACGTTGGATGGTTTGCTAAAATTGTATGAC	ACAAAATTGTATGACAAGCCTTC	11.6
	rs747783	ACGTTGGATGTGGTTCTTCTACAACCGGTG	ACGTTGGATGTCTTAGTAGTCTTGCTCTGG	GGAATGCTGACATTATTGATTCA	11.6
	rs902025	ACGTTGGATGATGATGTTGCCTACACCCAG	ACGTTGGATGCTCATTTAGAGTCAGACAAGG	TTAGAGTCAGACAAGGAAGTTCA	11.6
	rs7541094	ACGTTGGATGCTCCCTGGATGAAAATCAGC	ACGTTGGATGCTGGAGAGCAGACTGTATGG	TGGGTGGGCAAGAGCAGCAGGGA	14
	rs9662100	ACGTTGGATGGGGCCTAGACATTTCTGGAG	ACGTTGGATGACAGAAGACATGGCTTGAGG	AGGTTAATAGACTGCTTCTAATAC	14
	rs905950	ACGTTGGATGAGCCTTCCTCGGGAAATTAC	ACGTTGGATGATCCTCTGCATGGGCTTTTG	GAAGGTTCTCTTCTTAAGGGCAGG	14
	rs7202472	ACGTTGGATGGCTCCATGACTCAGAGTCAC	ACGTTGGATGGACACACTAACTGTTCTGCC	GTTCTGCCTAAAGAAGAGCCATGGA	14
	rs460420	ACGTTGGATGTCCAACAGATTAACTCACAG	ACGTTGGATGTCTCCTGTTTGTTTCTGGGC	AGGGTTGGAGATAAGATCATCTATA	14
	rs1892577	ACGTTGGATGGGACAGTTAGAGTGTACAGC	ACGTTGGATGCAGCTGGTAATTGGCAAAAT	ATTATACAACTCAAGTCTTTCTGGGA	14
	rs3821173	ACGTTGGATGATTAGTCGTAGAGTCCCAGG	ACGTTGGATGCCACATGCCTATGTTCGTTT	GGGATTTCATAGAGAAATTCCTAAGA	14
	rs7411544	ACGTTGGATGGTGACTTCTGAAACTGCAGG	ACGTTGGATGGAGAGGAGAAATGTGAGAGT	AGTATACTGCATACTGCAATGTGTAAA	14
NeuroDys - Replication after integrating pooling data	**rs7686728**	ACGTTGGATGGGAAGTATGAGAAAATGACC	ACGTTGGATGGGAAAGGGAAAACCATCAGT	CCATAGCACCTACGG	7
	rs10932727	ACGTTGGATGGACGCACAGCAAGTGAAAAC	ACGTTGGATGACGAGACAGCTGTGCGGAAG	GGAAGCTGTGATGCC	7
	rs3736403	ACGTTGGATGTTGATTTCATCTGAACGTCC	ACGTTGGATGTACTGTGCTAGGCACCAGAC	AGGCTCCTACCTCCAA	7
	rs461119	ACGTTGGATGAACAGCAAAATAGCACACAG	ACGTTGGATGAAAAGGGCAACAGGAAAGGC	GCTGCTCCATGGTGAG	7
	rs16900429	ACGTTGGATGGGGATTGCTAACTCCTGGAA	ACGTTGGATGTGTGATCATGGAGACAAGGG	AGGGAGAGAAGTGCTC	7
	rs1581413	ACGTTGGATGTTGGTTGCCAAGAGCTAGGA	ACGTTGGATGCCATCACCCTAAATGGAACC	ACTGGAACCCCCACATC	7
	rs9397276	ACGTTGGATGATCCTAAAACAAGAAACGGG	ACGTTGGATGGGAGACAAGGAAGATCTTGC	AGCCAAGTTTTTGCCTC	7
	rs4655653	ACGTTGGATGCATGTTTTCTGCAAGTGGAC	ACGTTGGATGAGGAGAAGGAATGAAGCTCG	TTGAAGCTCGGTCATTT	7
	rs9535442	ACGTTGGATGTGCATGACCAAGAAAGGTG	ACGTTGGATGTATCAACTGGGAAGGATGAG	GACCTTGGTCAAGATCAT	7
	rs2289191	ACGTTGGATGCAAAAGCATGTCAGGTACAC	ACGTTGGATGCATTTTCTGAAGACTTTTGGG	ACTTTTGGGGTACAGTAT	7
	rs10509910	ACGTTGGATGGTTACACAGCCTGTGTGAGT	ACGTTGGATGGATCATAAAGACCCATGGTG	CCCATGGTGTTAGAGAAA	9.3
	rs12290752	ACGTTGGATGAATCAGGGCTGCACAAACAC	ACGTTGGATGCCTCCTGAAACGATGTCAAG	CCCCACTCTAAAGCCAACA	9.3
	rs9916926	ACGTTGGATGTCCATTCTCCTGAAGTTGCC	ACGTTGGATGAAACCAGAAATGGCACCACC	CCTCATAAAACTGTCACAA	9.3
	rs6812487	ACGTTGGATGTGAGTGCTCAAGTGCTATTC	ACGTTGGATGCAAACCTGGTTTACACATCC	CACATCCAATATCTTCCTTC	9.3
	rs1350317	ACGTTGGATGCACTAGCCATAAATGTTTTTG	ACGTTGGATGCTACATACCAAAACTACCAC	CCAAAACTACCACAGTATCA	9.3
	rs5063	ACGTTGGATGGCCCTACCTTGAAATCCATC	ACGTTGGATGAGGTCAGACCAGAGCTAATC	CCCCCTCCCATGTACAATGCC	9.3
	rs945386	ACGTTGGATGGATGTTGCTGCGCAGGAGG	ACGTTGGATGTCGGGCACTCCAGATCCAAG	CCAAGGGGTGAAAGAGAATA	9.3
	rs705790	ACGTTGGATGAATTCTGAAGTGAAGCAGGG	ACGTTGGATGACAAAAAGGGAGCTCAGCAG	CAAACACGGATTTTCATTTTA	9.3
	rs2817764	ACGTTGGATGTCCAAAGCTGGTATTTCTGC	ACGTTGGATGCATGTTTCTGCATAGTGTCC	GGGATTGTCCTTTCAATGCCT	9.3
	rs34871518	ACGTTGGATGAGGATTCAATGACAGAAGGC	ACGTTGGATGACAGGAGAGCTGTCACTACG	GAAATGCCTTTTGAATGTAAT	9.3
	rs7904542	ACGTTGGATGGTCCGCTAGCATTCATAATC	ACGTTGGATGCACAGCAGGGAGTGTTAACG	AAGTGAGTGTTAACGGGAGAG	11.6
	rs7934218	ACGTTGGATGGAGGTGGTATCTTGGTGTAG	ACGTTGGATGCCAACAAGCGTATGAAGGTA	AGGAAGGTATGTTCAACATCAT	11.6
	rs12743401	ACGTTGGATGGGAAATTCACTAGAATGTTA	ACGTTGGATGAGCTAAGAAAAAATTGTCC	CATTGTCCATAACCATCCATAAC	11.6
	rs4510693	ACGTTGGATGCAACACTAGGATCAGCTGAG	ACGTTGGATGAAGTCACTGGTCACTCATCC	CACCGGTCACTCATCCAGACTTA	11.6
	rs1569012	ACGTTGGATGCTAGAGATAGTCAGGAACAC	ACGTTGGATGACCTAAACCCATCTCATGTC	ACCGCATGTCTTATTCACGTTGA	11.6
	rs4436151	ACGTTGGATGATGTAAGTCAGTTCAGCCAC	ACGTTGGATGGTGGGTCAAATGGTAGTTATG	TAGTTATGTTTTAACTCCTTTGA	11.6
	rs12344734	ACGTTGGATGTACAGACTTTGCCGTAGGAC	ACGTTGGATGTTTGTAACCTCCCTCCTGTC	TCCCATCCCTCCTGTCACTAGCTC	11.6
	rs9324005	ACGTTGGATGTACCATGCCACTGGACTTTG	ACGTTGGATGTGATGGGCACTTCTGGATAC	ATGGATACATGGTCACTTATTAGC	11.6
	rs4330611	ACGTTGGATGGGGATATTCCACTATCCAGC	ACGTTGGATGCCATTCCAACCTCCAACTGA	CCTACAACCTCCAACTGATTTACAG	11.6
	rs1546929	ACGTTGGATGTAAGTCAAAACAGGCATACG	ACGTTGGATGTTCTCATTGGTCCAAGGAGC	AATCCTGATTTATCAGATAATCCTA	11.6
	rs11232875	ACGTTGGATGAGGAGTCATGAGGCTTCAAC	ACGTTGGATGTGCTTTGAGATGTCTCTGTC	GGGATGAAACTAATGTACTTCTTAC	14
	rs6687859	ACGTTGGATGAGGGAAAGAAATCCAGCGTC	ACGTTGGATGGTGTTTCAGACCCAAACCTC	CACATCAGAAAATATCCCACAATATG	14
	rs268598	ACGTTGGATGGCAATGTCACATATTAAGC	ACGTTGGATGTCTCCTTTCCCTATGGGAAC	GCCTATGGGAACTATATAAATAGTCA	14
	rs17615558	ACGTTGGATGGCTATGCACTTGGACAATTA	ACGTTGGATGGAGCTGATCACCGGCTAGAA	GGGTTGATCACCGGCTAGAATTTATA	14
	rs10821663	ACGTTGGATGTCTTGACACTTTGGGATTGG	ACGTTGGATGACTGCCCTAGAGAGAAAGAG	AGAGAAAGAGGAAGAGAATATCTATT	14

Assay	SNP-ID	1st primer (sequence 5' - 3')	2nd primer (sequence 5' - 3')	UEP primer (sequence 5' - 3')	UEP conc. (µM)
NeuroDys - Replication after integrating pooling data (cont.)	rs2311445	ACGTTGGATGAGGCACAAAAGAGTACAGAC	ACGTTGGATGCTATCGTGACTTCTAACCCC	CCCAACAGGTACTTTTGCCTCTTTTTA	14
	rs16932422	ACGTTGGATGTGAGGTTGGAGAGATGAGTG	ACGTTGGATGCAGTTCTTCCTGCTGCCAC	CTCTTCTGGGCCCTCATCTCCTCATCTT	14
	rs2189167	ACGTTGGATGCAACTTCAATCTCTAGCACC	ACGTTGGATGGTGTTTACTATATTTCCATTC	GGTTTACTATATTTCCATTCTTGTTTAC	14
	rs12352208	ACGTTGGATGCAAACATGGAGTGTCTGCTG	ACGTTGGATGATGGAAGGAAGTACTCCAGC	ACTAGTTCATATGACCAGCATCCAATAC	14
	rs7381	ACGTTGGATGTAAGGACATGGCTGCTGTAG	ACGTTGGATGGCATTCTGTCATGCTGCTTC	CGGATTCTGGGGATAATTTATCATCCAC	14
MMN - FAC3 - Plex1	rs4234898	ACGTTGGATGCTCTGGGTGGTTTTCACTGT	ACGTTGGATGGCTGGCATCTTCAAGTTTTC	TCTGCCATGTCCCCCTC	7
	rs2114167	ACGTTGGATGAGGACACTCAGACAGCCAG	ACGTTGGATGTGTAGATTCCAGGAAATCCC	GAAATCCCTGGTCCTC	7
	rs133885	ACGTTGGATGGGGCTTCATTAAGCAACTGG	ACGTTGGATGAGGCGACAGCTAACTGCTTC	TGGCCTCTTTTTCAGTC	7
	rs1365152	ACGTTGGATGACACAGGCTAAGAGTCGAAG	ACGTTGGATGGACAGGCTGTCCAGAGCTT	GGAGCTTGGTGGCGATGC	9.3
	rs4837521	ACGTTGGATGACAGGCTTTTAGTAGCTGAC	ACGTTGGATGCTCCTGCCACTGGTTTAATG	GGTTGCTATAAGCAAAAGG	9.3
	rs1399428	ACGTTGGATGGCTAGGAACTCTTGGAGTAG	ACGTTGGATGCTACTTTAGATGATGATGAC	ATGATAACTATGATTCGATGA	11.6
MMN - Plex2	rs9307938	ACGTTGGATGGAATGGCATAAAAGTGAAGTG	ACGTTGGATGCCAGTGGTTTATACACAGGC	CAGGCAACGGTTAGG	7
	rs12500004	ACGTTGGATGGGGATTATGTAAGGAGTCCA	ACGTTGGATGAGTTATTTGCTTTCTCCCTC	TTTCTCCCTCTCCCTTTC	9.3
	rs4555581	ACGTTGGATGCCTCAGGAAGCTTCTAATCG	ACGTTGGATGCTGTGACTATCTCTCTTGGC	CATGAGACACATTTGCTC	9.3
	rs4690853	ACGTTGGATGAAAGAACACTGACCCACAGG	ACGTTGGATGTGCTCAAAACAGCCTGTGCC	CAGACAATGTCATCACACA	9.3
	rs6845616	ACGTTGGATGTATTTGGGGAGTACAATAC	ACGTTGGATGGGAGTCCTGAATTCAAAGCC	GAATTCAAAGCCTAGTACAT	11.6
	rs11100040	ACGTTGGATGAACACAGTTGTTTACAGGAC	ACGTTGGATGGTAAGGAAAATGACAGTAGC	AAAATGACAGTAGCACTATC	11.6
	rs4600965	ACGTTGGATGCACCTAGATTTGTATAAAG	ACGTTGGATGCCATGTGTGTTTCTTGGGTC	TGTTTCTTGGGTCTTTTAAGTTA	11.6
	rs4256192	ACGTTGGATGAAATGCCAGTCCAGTTGTCC	ACGTTGGATGATGGCAATCAGTTAATCTC	AGAGATGTTGGCTTTTCAATAAA	14
FAC3 - Plex2	rs13058434	ACGTTGGATGTCCTTTCCAGTATCTGTGGC	ACGTTGGATGCAAGCCAGAGAAGACTCATC	CCCATGACGCCCCCC	7
	rs542162	ACGTTGGATGCCCACTTTCTGGAATCTAGG	ACGTTGGATGTGTGGCTGGAGGAACACAAG	AGGACCTCATGGCTC	7
	rs4822644	ACGTTGGATGGGTCACAGGGATGGATTTTC	ACGTTGGATGAACTGGAGAACAGGCCTATG	AGGCCTATGCGTGTT	7
	rs9624894	ACGTTGGATGTGGTGGCCATTTGGGGAATC	ACGTTGGATGTCCCACACTCAAACCCACAG	CCTCCTGTCCTTAGGC	7
	rs4822661	ACGTTGGATGTCTCCACCCCTAGAACTCAC	ACGTTGGATGCAAGTGAAATTGTCAAGGGC	TGAGTCGAGCAGAGGT	9.3
	rs133866	ACGTTGGATGGGACATCAGAAATCTCTGCC	ACGTTGGATGTGAACAGTTAGAAATGAGCC	TAGAAATGAGCCTGAGAA	9.3
	rs133903	ACGTTGGATGTTTAGATGGCTCCAGAGCAC	ACGTTGGATGTAAACCAGCAAATGGCACCG	GCCCATCTTCTACCTTACAC	11.6
	rs9620553	ACGTTGGATGTGCCTCGTCAGCTGCTGTTT	ACGTTGGATGAGACTCGCGATTTTATACCG	TTTATACCGTGAATTCCAAA	11.6
	rs4822649	ACGTTGGATGATCTTGGACAATGCCAGCAC	ACGTTGGATGGTTTGAATGCAAGAGGAGGG	AGGGGGAGAATCCCAGATTCC	11.6
	rs3859865	ACGTTGGATGCTTATCAAGAACCTCTGTGC	ACGTTGGATGATCTCCTTTGAATTCCTCCC	TTGAATTCCTCCCTTATCAGTT	14
	rs133847	ACGTTGGATGACAAACATCCACTGAGCACG	ACGTTGGATGCAGAGAAAGTCCCACTGAAG	TACAGTGACTGTTACAGTCCAC	14
	rs133871	ACGTTGGATGTGACTGTAGCATTGAAAGAG	ACGTTGGATGCTGCGAATTGGAGGTACTCT	TGACATTCTCTAGTTATCGGACTTC	14
	rs3848858	ACGTTGGATGGAAGGATGCTGCTTGCAGAC	ACGTTGGATGGGGATCCCTCTGCATCTGA	GGGATCCCTCTGCATCTGAGGTGAC	14

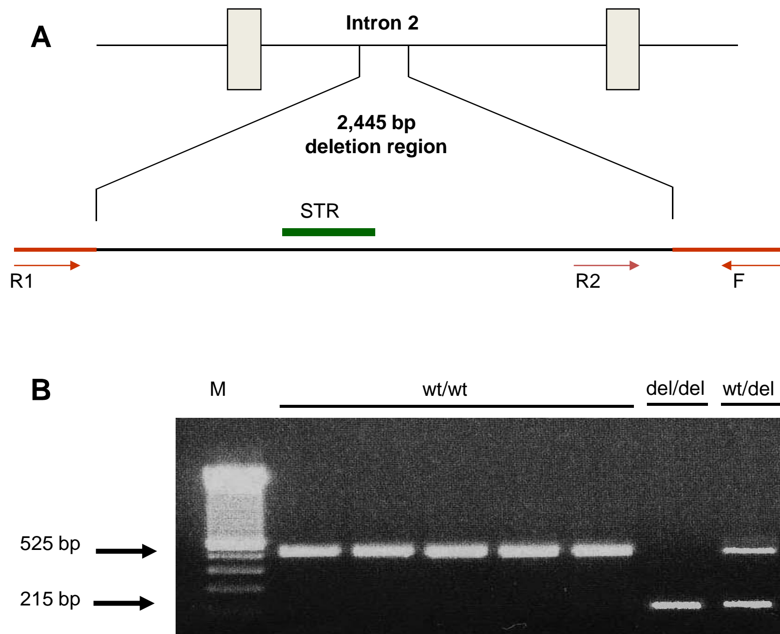
Attachment II: Custom TaqMan assay for *SLC2A3*.

The cDNA sequence for *SLC2A3* is represented based on RefSeq annotation (chr. 12: 7,963,092…7,980,159 bp, NCBI build 36), from its 5'UTR until beginning of exon 5. Exons are given in alternating blue and black-colored letters, with codons separated by blue-white shadings. 5' and 3' UTRs are given in red-colored letters and are highlighted by blue shadings. The probe is shown in yellow, while the flanking primers are marked in green. The TaqMan assay was designed in a region that distinguishes *SLC2A3* from other closely related members of the SLC2-protein family.

GTGGGGTGGGGTGGGGCTGGGGGCTTGTCGCCCTTTCAGGCTCCACCCTTTGCGGAGA TTATAAATAGTCATGATCCCAGCGAGACCCAGAGATGCTGTAATGGTAAGACTTTGGA TCCTTCCTGAGGACGTGGAGAAAACTTGCTGCTGAGAAGGACATTTTGAAGGTTTTGT TGGCTGAAAAAGCTGTTTCTGGAATCACCCCTAGATCTTTCTTGAAGACTTGAATTAG ATTACAGCGATGGGGACACAGAAGGTCACCCCAGCTCTGATATTTGCCATCACAGTTG CTACAATCGGCTCTTTCCAATTTGGCTACAACACTGGGGTCATCAATGCTCCTGAGAA GATCATAAAGGAATTTATCAATAAAACTTTGACGGACAAGGGAAATGCCCCACCCTCT GAGGTGCTGCTCACGTCTCTCTGGTCCTTGTCTGTGGCCATATTTTCCGTCGGGGGTA TGATCGGCTCCTTTTCCGTCGGACTCTTCGTCAACCGCTTTGGCAGGCGCAATTCAAT GCTGATTGTCAACCTGTTGGCTGTCACTGGTGGCTGCTTTATGGGACTGTGTAAAGTA GCTAAGTCGGTTGAAATGCTGATCCTGGGTCGCTTGGTTATTGGCCTCTTCTGCGGAC TCTGCACAGGTTTTGTGCCCATGTACATTGGAGAGATCTCGCCTACTGCCCTGCGGGG TGCCTTTGGCACTCTCAACCAGCTGGGCATCGTTGTTGGAATTCTGGTGGCCCAGATC TTTGGTCTGGAATTCATCCTTGGGTCTGAAGAGCTATGGCCGCTGCTACTGGGTTTTA	SLC2A3_UTR/ex2F SLC2A3_ex1/2 SLC2A3_UTR/ex2R

Attachment III: Setup for *DCDC2* deletion screening.

(A) The deletion region covers 2,445 bp in intron 2 of *DCDC2*. Within the deletion, a STR marker consisting of different numbers of repeats is located. The positions of the primers are indicated. In presence of the deletion, the R2 primer cannot bind and a product is formed between R1 and F (product size 215 bp). For non-deleted alleles, R2 and F primer form the product, as R1 is too far away given the used PCR conditions (525 bp). (B) PCR results for non-deleted (wt) and deleted (del) alleles. Non-deleted alleles were sequenced to assess STR alleles.

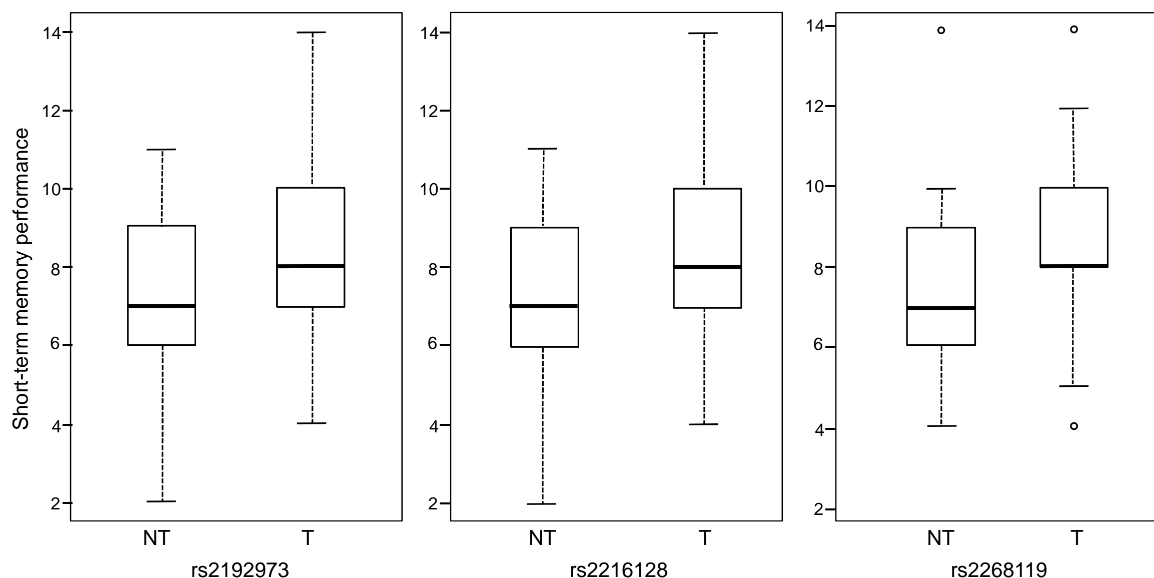


Attachment IV: Summary of the statistical analysis for SNPs included in the *GRIN2B* study.
66 SNPs were included in the assay, 5 SNPs failed genotyping ('failed'). SNPs are presented in consecutive order, with the results of the TDT (dyslexia as qualitative trait) and the QTDT ('short term memory').* – Position according to dbSNP129. Minor allele frequencies (MAFs) are given for the parental chromosomes (n.a. – not analyzed for failed SNPs). [1] – SNPs with $P < 0.1$ in Dorval *et al.*, [2] – SNPs contained in multi-SNP cluster in de Quervain & Papassotiropoulos, [3] – SNP corresponds to hCV2682119 in de Quervain & Papassotiropoulos, [4] – SNP corresponds to hCV3289602 in de Quervain & Papassotiropoulos. *P*-values are bold if ≤ 0.05. SNPs shaded in grey showed nominal significant *P*-values for short-term memory performance and were taken forward into subsequent analysis.

Marker		MAF	*P*-values	
SNP-ID	Position*		TDT	QTDT 'short-term memory'
rs17820659	13,580,175	0.101	0.3673	0.5825
rs1805247	13,607,242	0.082	0.5806	0.7065
rs1806201	13,608,775	0.271	0.2733	0.1650
rs3026160	13,611,310	0.121	0.5834	0.4174
rs1805482	13,656,041	0.319	0.1275	0.1597
rs11055557	13,680,094	0.327	0.0901	0.2417
rs11055582	13,720,442	0.093	0.8648	0.8175
rs17833639	13,721,044	0.076	0.2564	0.5915
rs765688	13,721,348	0.218	0.756	0.6990
rs2193150	13,722,606	0.183	0.2772	0.8518
rs2192977[2]	13,730,269	0.414	0.7935	0.3924
rs2300242	13,731,564	n.a.	failed	failed
rs2300245	13,732,562	0.234	0.8527	0.3717
rs2284407[1]	13,733,473	0.399	0.8742	0.2365
rs1158541	13,737,940	0.35	0.4593	0.6070
rs7966866	13,745,523	n.a.	failed	failed
rs1012586	13,746,899	0.286	0.6506	**0.0401**
rs11055593	13,748,274	0.066	0.4235	0.2185
rs12809496	13,749,176	0.377	0.8378	0.5344
rs17220663	13,750,851	0.055	0.5784	0.6272
rs2300251	13,752,862	0.437	0.9606	0.3589
rs2300252	13,752,951	0.306	0.7872	0.3523
rs2284411[1]	13,757,439	0.356	0.7564	0.3575
rs2300256[1]	13,759,677	0.495	0.8043	0.7036
rs2268115[1]	13,760,992	n.a.	failed	failed
rs2268119	13,763,901	0.204	0.7054	**0.0243**
rs2110984	13,766,847	0.225	0.6695	0.0727
rs11612284	13,766,943	0.18	0.6985	0.3603
rs7970407	13,768,117	n.a.	failed	failed
rs2268120	13,769,155	0.099	0.4957	0.6637
rs10845837	13,771,435	0.3	0.6229	0.5799
rs2216128	13,774,281	0.225	1	**0.0406**
rs10845838[2,3]	13,785,663	0.243	0.8559	0.4900
rs2192973	13,787,822	0.203	0.8510	**0.0381**
rs2268122	13,788,728	0.139	0.3096	0.9736
rs1005549	13,793,082	0.224	0.3338	0.5373
rs1805555	13,799,334	0.031	0.7728	0.9037
rs1861452	13,800,997	0.144	0.2825	0.2009
rs11055608	13,804,693	0.45	0.8821	0.1981
rs2284416	13,810,481	0.449	0.9205	0.1911
rs2216127	13,825,676	0.156	0.5355	0.5854
rs220549	13,828,587	0.431	0.8797	0.1528
rs11055616	13,831,108	0.24	0.7638	0.1900
rs2284418	13,834,895	0.201	0.183	0.0674
rs220557	13,839,047	0.343	0.4985	0.8289
rs220563	13,841,135	0.203	0.368	0.7412
rs7974275	13,841,844	0.284	0.2858	0.0510
rs17833967	13,846,345	0.084	0.1721	0.0584
rs220583	13,852,010	0.257	0.2433	0.2123
rs220597	13,859,453	0.359	0.9569	0.4367
rs220599[2]	13,866,565	0.328	0.34	0.5548
rs2160734	13,875,616	0.455	0.5076	0.9323
rs2284424	13,880,137	0.294	0.2482	0.6380
rs2284425	13,880,286	0.292	0.1876	0.7553

Marker		MAF	*P*-values	
SNP-ID	Position*		TDT	QTDT 'short-term memory'
Table continued				
rs2300273	13,881,701	0.443	0.7213	0.8750
rs7301328	13,910,044	0.398	0.4691	0.9025
rs12818068	13,910,424	0.121	0.1744	0.2282
rs2041986	13,929,237	n.a.	failed	failed
rs9971835	13,953,134	0.357	0.5604	0.1063
rs219876	13,972,890	0.078	0.5672	0.1176
rs12828473[2,4]	13,996,995	0.474	0.3708	0.7249
rs1421109[2]	14,024,078	0.029	0.1119	0.4576
rs11055711	14,044,935	0.059	0.6275	0.6229
rs7964322	14,070,657	0.064	0.833	0.5338
rs12830358	14,091,910	0.181	0.3494	0.4625
rs933614	14,111,149	0.174	**0.0130**	0.4285

Attachment V: Short-term memory performance in relation to maternal transmissions for rs2192973, rs2216128 and rs2268119. As for rs1012586, individuals with maternally transmitted alleles for three significant SNPs rs2192973, rs2216128 and rs2268119 performed better in short-term memory tasks. The figure illustrates the phenotypic differences for each of the three SNPs, always representing performance of individuals with maternally non-transmitted alleles (NT) and maternally transmitted alleles (T). For all three SNPs, the difference between the groups is significant, with $P < 0.01$. Numbers of transmitted / non-transmitted alleles are as follows: 62 T vs. 52 NT (rs2192973), 68 T vs. 63 NT (rs2216128), 55 T vs. 48 NT (rs 2268119).



Attachment VI: Results of the pooling approach using allele frequency estimates.
Pooling results are depicted for SNPs providing *P*-values ≤ 10^{-04} in the combined analysis of all three pools (Fisher's Product). SNPs showing nominal significant *P*-values in at least two of the three pools were considered for analysis. For each of these SNPs, functional evidence based on UCSC expression data was assessed (x).

SNP	Location (dbSNP129)			*P*-values of the different populations			Combined Fisher's *P*-value	Functional evidence
	Gene	Chr	Position	Finnish pools	German pools	UK pools		
rs36019094		5	40,273,131	4.15×10^{-01}	**5.06×10^{-10}**	**3.28×10^{-04}**	**3.38×10^{-11}**	
rs3817222	*PPP1R12B*	1	200,731,383	**9.21×10^{-02}**	**3.59×10^{-09}**	**8.97×10^{-04}**	**1.32×10^{-10}**	
rs10932727	*DIRC3*	2	218,313,957	**4.26×10^{-02}**	**5.48×10^{-07}**	**9.58×10^{-05}**	**8.67×10^{-10}**	
rs11661017		18	9,865,106	**1.41×10^{-09}**	**4.89×10^{-03}**	5.55×10^{-01}	**1.43×10^{-09}**	
rs12290752		11	115,939,639	**4.04×10^{-02}**	1.74×10^{-01}	**4.00×10^{-09}**	**9.00×10^{-09}**	
rs13297028		9	76,729,283	2.13×10^{-01}	**9.25×10^{-09}**	**1.53×10^{-02}**	**9.62×10^{-09}**	
rs420121	*GRIK1*	21	30,068,479	7.29×10^{-02}	**7.23×10^{-07}**	**7.96×10^{-04}**	**1.30×10^{-08}**	x
rs1569012	*STON2*	14	80,923,160	**7.48×10^{-05}**	**1.47×10^{-06}**	4.01×10^{-01}	**1.36×10^{-08}**	x
rs6687859	*PTGER3*	1	71,135,175	**7.69×10^{-06}**	**8.80×10^{-04}**	**2.22×10^{-02}**	**4.19×10^{-08}**	
rs12344734	*TMC1*	9	74,383,363	**3.53×10^{-02}**	**9.31×10^{-05}**	**8.81×10^{-05}**	**7.65×10^{-08}**	x
rs10509910	*MXI1*	10	111,991,750	**2.25×10^{-04}**	**3.17×10^{-04}**	**4.70×10^{-03}**	**8.73×10^{-08}**	
rs6812487	*ODZ3*	4	183,697,713	**2.09×10^{-07}**	**1.92×10^{-02}**	2.15×10^{-01}	**2.07×10^{-07}**	x
rs13261597	*ADRA1A*	8	26,684,235	3.99×10^{-01}	**1.02×10^{-05}**	**2.16×10^{-04}**	**2.10×10^{-07}**	x
rs7934218	*FAM168A*	11	72,884,377	**6.53×10^{-08}**	**2.93×10^{-02}**	6.18×10^{-01}	**2.75×10^{-07}**	
rs6992898		8	126,629,069	**8.90×10^{-03}**	**1.72×10^{-05}**	**9.52×10^{-03}**	**3.33×10^{-07}**	
rs16918472		11	92,441,542	1.47×10^{-01}	**1.40×10^{-03}**	**1.03×10^{-05}**	**4.67×10^{-07}**	
rs2311445		16	17,456,460	**2.24×10^{-06}**	**8.26×10^{-03}**	1.15×10^{-01}	**4.69×10^{-07}**	
rs705790		6	166,286,499	**9.36×10^{-03}**	**4.33×10^{-06}**	5.72×10^{-02}	**5.07×10^{-07}**	
rs12352208		9	14,563,137	**1.18×10^{-05}**	**5.77×10^{-04}**	3.79×10^{-01}	**5.57×10^{-07}**	
rs2817764	*CDC2L6*	6	111,087,345	**6.43×10^{-04}**	**4.37×10^{-06}**	9.92×10^{-01}	**5.99×10^{-07}**	
rs16932422	*DNAJC5B*	8	67,103,552	**1.28×10^{-06}**	**1.01×10^{-02}**	2.59×10^{-01}	**7.06×10^{-07}**	
rs7904542	*CEP55*	10	95,245,571	**1.04×10^{-04}**	1.48×10^{-01}	**2.29×10^{-04}**	**7.43×10^{-07}**	
rs12919062		16	83,122,879	**1.79×10^{-06}**	**1.66×10^{-02}**	1.32×10^{-01}	**8.16×10^{-07}**	
rs17022928		2	83,081,762	**8.65×10^{-07}**	5.44×10^{-01}	**1.03×10^{-02}**	**9.88×10^{-07}**	
rs2001363	*NLRP1*	17	5,403,378	2.80×10^{-01}	**2.63×10^{-06}**	**7.79×10^{-03}**	**1.15×10^{-06}**	
rs17615558		6	12,463,160	**1.13×10^{-06}**	1.25×10^{-01}	**4.16×10^{-02}**	**1.17×10^{-06}**	
rs16900429	*RIPK2*	8	90,845,430	**1.64×10^{-05}**	**4.07×10^{-04}**	9.56×10^{-01}	**1.26×10^{-06}**	
rs9397276		6	156,298,468	**7.03×10^{-06}**	**1.69×10^{-03}**	7.51×10^{-01}	**1.70×10^{-06}**	
rs7381	*FBLN1*	22	44,375,446	**2.01×10^{-02}**	**5.30×10^{-05}**	**8.90×10^{-03}**	**1.80×10^{-06}**	
rs17203209	*OSBPL1A*	18	20,137,360	**2.30×10^{-02}**	**7.30×10^{-05}**	**5.70×10^{-03}**	**1.82×10^{-06}**	
rs7121541		11	12,247,388	**3.61×10^{-02}**	**3.00×10^{-07}**	9.22×10^{-01}	**1.89×10^{-06}**	
rs752527	*SNX29*	16	12,491,562	**1.87×10^{-04}**	**2.38×10^{-02}**	**3.46×10^{-03}**	**2.77×10^{-06}**	
rs3736403	*CCDC108*	2	219,613,491	**1.65×10^{-05}**	**2.82×10^{-02}**	**4.47×10^{-02}**	**3.65×10^{-06}**	
rs7686728		4	184,541,581	**2.33×10^{-02}**	6.67×10^{-01}	**1.42×10^{-06}**	**3.84×10^{-06}**	
rs9361115		6	77,462,787	1.87×10^{-01}	**1.04×10^{-05}**	**1.20×10^{-02}**	**4.02×10^{-06}**	
rs4330611	*SGCE*	7	94,109,934	1.86×10^{-01}	**1.66×10^{-02}**	**9.51×10^{-06}**	**4.94×10^{-06}**	
rs6805519	*GLB1*	3	33,057,192	**9.64×10^{-03}**	9.97×10^{-01}	**3.27×10^{-06}**	**5.26×10^{-06}**	
rs3744749		22	44,832,872	**1.96×10^{-05}**	8.41×10^{-01}	**2.19×10^{-03}**	**5.95×10^{-06}**	
rs11232875		11	81,191,362	**1.10×10^{-03}**	**3.85×10^{-05}**	8.59×10^{-01}	**5.98×10^{-06}**	
rs2280711	*TAOK3*	12	117,134,337	**2.18×10^{-02}**	8.05×10^{-02}	**2.28×10^{-05}**	**6.53×10^{-06}**	
rs7627525	*UBE2E2*	3	23,440,163	**2.53×10^{-02}**	**4.64×10^{-06}**	3.91×10^{-01}	**7.38×10^{-06}**	
rs11575489	*DDC*	7	50,511,385	**7.80×10^{-04}**	**2.81×10^{-04}**	2.10×10^{-01}	**7.39×10^{-06}**	x
rs1546929	*BCKDHB*	6	81,104,278	**1.58×10^{-06}**	**3.18×10^{-02}**	9.51×10^{-01}	**7.64×10^{-06}**	
rs10821663	*ANK3*	10	61,480,286	**7.84×10^{-04}**	6.39×10^{-01}	**9.84×10^{-05}**	**7.86×10^{-06}**	x
rs9324005		14	98,565,680	1.80×10^{-01}	**2.80×10^{-02}**	**1.02×10^{-05}**	**8.14×10^{-06}**	
rs1959967	*LOC728755*	14	27,175,888	**7.36×10^{-03}**	**7.43×10^{-06}**	9.39×10^{-01}	**8.15×10^{-06}**	
rs4436151		8	114,902,482	**4.66×10^{-03}**	**2.31×10^{-05}**	4.99×10^{-01}	**8.47×10^{-06}**	
rs1036196		5	150,992,450	**3.94×10^{-03}**	**1.64×10^{-05}**	9.34×10^{-01}	**9.42×10^{-06}**	
rs9916926		18	12,908,318	**8.22×10^{-03}**	1.57×10^{-01}	**4.92×10^{-05}**	**9.84×10^{-06}**	
rs34871518		19	63,046,077	**2.80×10^{-05}**	**2.65×10^{-03}**	8.59×10^{-01}	**9.87×10^{-06}**	
rs6751361	*TRAF3IP1*	2	238,920,448	**3.61×10^{-02}**	**3.49×10^{-05}**	5.14×10^{-02}	**1.00×10^{-05}**	
rs9535442		13	49,821,641	1.58×10^{-01}	**4.92×10^{-05}**	**8.41×10^{-03}**	**1.01×10^{-05}**	
rs7291798	*SGSM1*	22	23,568,729	**2.37×10^{-02}**	**7.21×10^{-05}**	**4.01×10^{-02}**	**1.05×10^{-05}**	x
rs4510693		6	156,270,620	**1.22×10^{-05}**	**9.03×10^{-03}**	6.51×10^{-01}	**1.10×10^{-05}**	
rs17542525	*MS4A5*	11	59,958,103	**4.20×10^{-03}**	**1.63×10^{-03}**	**1.06×10^{-02}**	**1.10×10^{-05}**	
rs10499129	*NKAIN2*	6	124,975,906	2.39×10^{-01}	**4.94×10^{-02}**	**8.14×10^{-06}**	**1.42×10^{-05}**	
rs945386	*KIAA1984*	9	138,813,417	**1.34×10^{-04}**	**4.80×10^{-02}**	**2.06×10^{-02}**	**1.89×10^{-05}**	
rs268598	*TRAM1*	8	71,677,425	7.00×10^{-02}	**1.59×10^{-04}**	**1.23×10^{-02}**	**1.93×10^{-05}**	
rs5063	*NPPA*	1	11,830,235	**1.02×10^{-04}**	**1.02×10^{-02}**	1.35×10^{-01}	**1.99×10^{-05}**	

SNP	Location (dbSNP129)			*P*-values of the different populations			Combined Fisher's *P*-value	Functional evidence
	Gene	Chr	Position	Finnish pools	German pools	UK pools		
Table continued								
rs2289191	*IARS2*	1	218,366,658	**9.80×10^{-06}**	4.58×10^{-01}	**3.64×10^{-02}**	**2.27×10^{-05}**	
rs4655653	*WDR78*	1	67,104,024	**3.14×10^{-05}**	6.85×10^{-01}	**1.37×10^{-02}**	**3.80×10^{-05}**	
rs2254143	*KIAA1984*	9	138,820,929	**1.99×10^{-04}**	5.12×10^{-01}	**8.25×10^{-03}**	**9.48×10^{-05}**	

ERKLÄRUNG

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet habe.

Bonn, den 30. November 2009

Kerstin U. Ludwig