Max-Planck-Institute of Psychiatry
Statistical Genetics

Identification of genetic variants involved in dyslexia pathogenesis by joint analysis of QTLs and epistasis

Nazanin Karbalai Mirza Agha

Vollständiger Abdruck der von der Fakultät für Medizin der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (**Dr. rer. nat**)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Markus Ploner

Prüfer der Dissertation: 1. apl. Prof. Dr. Bertram Müller-Myhsok

2. Univ.-Prof. Dr. Hans-Werner Mewes

Die Dissertation wurde am 12.08.2014 bei der Technischen Universität München eingereicht und durch die Fakultät für Medizin am 17.12.2014 angenommen.

Identification of genetic variants involved in dyslexia pathogenesis by joint analysis of QTLs and epistasis

Nazanin Karbalai Mirza Agha

For my parents

Contents

Acronyms

A Adenine. 23

ADHD attention deficit hyperactivity disorder. 33, 36, 104

ALSPAC Avon Longitudinal Study of Parents and Children. 36

ALU Arithmetic Logic Unit. 43

API application programming interface. 43

ASD autism spectrum disorder. 104, 107

BP biological process. 94, 96

bp base-pair. 62, 68, 78, 88

BWA Burrows-Wheeler Alignment. 60

C Cytosine. 23, 24

ChIP chromatin immunoprecipitation. 52, 53, 63, 87–89, 106

CI confidence interval. 69

CNS central nervous system. 108, 109

CNV copy number variation. 107

CPU central processing unit. 42, 43

CUDA Compute Unified Device Architecture. 20, 43

d.f. degree of freedom. 65

DNA deoxyribonucleic acid. 23, 27, 52, 63, 105

DZ dizygotic. 34

ENCODE Encyclopedia of DNA Elements. 52, 62, 63, 66, 87–89, 105–107

eQTL expression quantitative trait locus. 20, 107

eSNP expression single nucleotide polymorphism. 23

FAIRE Formaldehyde Assisted Isolation of Regulatory Elements. 52

FDR false discovery rate. 40, 41, 51, 60, 63

fMRI functional magnetic resonance imaging. 38

FWER familywise error rate. 40, 41

G Guanine. 23

GO gene ontology. 51, 95

GPU graphics processing unit. 20, 42–47, 49–51

GSEA gene set enrichment analysis. 51, 63, 93, 95, 96, 109, 110

GWA genome-wide association. 28, 40

GWAS genome-wide association study. 21, 28, 30, 31, 51, 65, 69, 102, 112

GWIA genome-wide interaction analysis. 30, 31, 40, 61, 65, 67, 104, 112

GWIS genome-wide interaction study. 65

HGP Human Genome Project. 27

HS hypersensitive site. 52, 63, 87, 105–107

HSIC Hilbert-Schmidt Independence Criterion. 47–49

HWE Hardy-Weinberg equilibrium. 24, 25, 59, 60

IBD identity-by-distance. 59

IBS identity-by-state. 26, 59

IC inbreeding coefficient. 59

Kb kilo base pairs. 88, 92, 108

LD linkage disequilibrium. 25, 26, 38, 40, 59, 67, 68, 73, 82, 89, 92, 105, 107

LINE long interspersed nuclear element. 67, 105

LTR long terminal repeat. 77, 106

MAF minor allele frequency. 24, 59, 60, 78, 101

Mb mega base pairs. 62, 91, 92

MDR multifactor dimension reduction. 111

MDS multidimensional scaling. 26, 56, 59

MMN mismatch negativity. 38

MRI Magnetic Resonance Imaging. 33

mRNA messenger ribonucleic acid. 38

mSNP methylation single nucleotide polymorphism. 23

MZ monozygotic. 34

NMJ neuromuscular junction. 109

NRSF neuron-restrictive silencer factor. 63, 87

NS nervous system. 96, 107

NWR non-word reading. 60, 61, 65, 77, 78, 86, 87, 92, 93, 101, 108

PA phonological-awareness. 58–61, 65, 73, 74, 87, 92, 93, 106, 109

PCA Principal Component Analysis. 26

PCER per-comparison error rate. 41

Q-Q Quantile-Quantile. 62, 86

QC quality control. 23–25, 55, 57, 59–62, 88, 91

QTL quantitative trait locus. 17, 24, 37, 47, 73, 106

RD reading disability. 19, 55, 57, 61, 107

RNA ribonucleic acid. 27, 52, 63, 87–89, 106

SD standard deviation. 55–57

SE standard error. 61

SIMD Single Instruction Multiple Data. 42

SLD speech and language disorder. 104, 105

SLI specific language impairment. 36–38, 104

SNP single nucleotide polymorphism. 20, 23–26, 28, 31, 32, 38–40, 42, 43, 45–52, 55–57, 59–63, 65–68, 73, 77, 78, 86–89, 91–93, 99–102, 104–108

SP spelling. 55, 58, 60, 61, 65, 82, 87, 89, 95, 107, 110

SSD speech sound disorder. 33, 36

SVM support vector machine. 111

SWR single-word reading. 37, 56–58, 60, 61, 65, 67, 69, 70, 78, 86–88, 91–93, 100, 104–106, 109

T Thymine. 23, 24

TF transcription factor. 53, 63, 87–89, 105–107

UCSC University of California, Santa Cruz. 52, 53, 60, 62, 63, 78, 88

UK United Kingdom. 36, 37

US United States. 36, 37

List of Figures

List of Tables

Abstract

The aim of this thesis was the identification of genetic susceptibility factors contributing to developmental dyslexia. Dyslexia is a highly prevalent disorder in children, characterized by deficits in reading and spelling [Schulte-Körne et al., 2001].The clinical picture of dyslexia can be categorized into various endophenotypes, describing different cognitive abilities. The underlying hierarchical processes of these cognitive competences suggest a highly heterogeneous genetic architecture composing the disorder, which obstructs single locus scans [Carrion-Castillo et al., 2013]. From this perspective, the question arose whether disturbances of cognitive processes in dyslectics could be a result of genetic interactions. Here, we therefore went beyond established single-locus association methods in order to analyze higher order genetic interactions.

So far, genetic studies of dyslexia and other complex traits have focused on single-locus effects, utilizing the present gold standard of genome-wide scans. Like all approaches, also this method has its limitations, such as the inability to capture higher order connections. However, single-locus effects explain just a low proportion of heritability. In fact, the interplay of individual factors is an essential component of the underlying genetical processes. Due to computational challenges, this central aspect was nevertheless often neglected.

By exploiting the technical and methodological progress, we were able to utilize a genome-wide two-locus interaction scan using the tool GLIDE [Kam-Thong et al., 2012]. We identified genetic interactions affecting susceptibility for altered cognitive skills in dyslexic individuals. Indeed, dyslexia proved to be an ideal phenotype for the genome-wide search of epistatic effects.

The genotype profiles of 862 dyslexic children from different ethnic origins (Germany, USA, Great Britain, and France) were determined. Measurements for the quantitative dyslexia endophenotypes single-word reading, phonological awareness, non-word reading, and spelling were conducted for each individual in a battery of psychometric tests. Within these samples, we performed exhaustive two-locus interaction searches for epistatic effects.

We were able to detect strong evidence for genetic interactions affecting various dyslexia endophenotypes, such as single-word reading and phonological awareness. The results included highly relevant genomic loci, comprising both novel and previously known dyslexia-sensitive loci. An example is the interaction between 18q11.2, a genomic region linked to various quantitative trait loci (QTLs) with intronic variants of *NCAM1*, a gene involved in the development of the nervous system. These interactions were found for

the dyslexia endophenotypes phonological awareness and single-word reading. An additional interesting interaction that was detected took place between intergenic variants on chromosome 9 and intronic variants of *FOXP2*, a gene associated with linguistic deficits. *FOXP2* was recently shown to play a role in the etiology of developmental dyslexia [Wilcke et al., 2011].

Our findings suggest that the multifacetedness and the etiology of dyslexia is likely to be explained by multi-genetic mechanisms like allelic interactions, with both single dependent and independent factors contributing to the disorder.

Taking into account the complexity of genetics and quantitative traits, the ubiquity of epistasis, and the success of our study, we suggests that epistasis should be shifted further into the focus of investigations. This would likely lead to a better understanding of complex traits and genetic heredity.

1 Introduction

1.1 Research question and motivation

Dyslexia is one of the most common neurodevelopmental disorders, with a prevalence of approximately 5–12% in school-aged children [Ludwig et al., 2008]. Affected children show pronounced difficulties in learning to read and spell, with a possible social impact on the child reaching until adulthood, and affecting their career prospects [Shaywitz et al., 1990].
Several genetic factors for dyslexia have been identified. Among those are genes involved in memory-related aspects of the brain [Ludwig et al., 2010], neuronal migration [Ludwig et al., 2008], as well as a major neuronal glucose transporter [Roeske et al., 2011]. Some of these genes are known to be strongly expressed and regulated in the cortex, hypothalamus, amygdala, and hippocampus [Meng et al., 2005].
The disorder comprises a variety of different endophenotypes. Affected children can exhibit difficulties in reading, spelling, phonological abilities or even show various disabilities simultaneously. Several endophenotypes have been shown to be associated with certain genomic regions (e.g. 6p22) or genes (*DCDC2*, *KIAA0319*, *GRIN2B*, *SLC2A3*) [Carrion-Castillo et al., 2013, Grigorenko, 2001]. Both the cluster of different symptoms and the involvement of various genetic factors indicate that the predisposition for dyslexia is likely to be explained by complex coherences and networks, conclusively the interaction of genetic factors. The research group of Statistical Genetics at the Max-Planck-Institute of Psychiatry [Czamara et al., 2013, Ludwig et al., 2008, Roeske et al., 2011] and other research groups has already conducted several studies in the field of dyslexia. The extensive findings regarding the pathogenesis of dyslexia pointed to the assumption that a highly heterogeneous pattern underlying the disorder [Carrion-Castillo et al., 2013].
Given the complexity of the phenotype, the knowledge about the pathogenesis of dyslexia, and the availability of eligible datasets as well as a tool for rapid detection of interactions between two genetic loci, we formed the hypothesis that a genome-wide analysis beyond single-locus effects may help to unveil unknown genetic contributions to the development of reading disability (RD) among individuals. The assumption that interaction effects are strong on the level of genetic markers seems reasonable, since such effects are consequences of evolution. Epistatic effects are more likely to withstand purifying selection, as they are protected by epistatic shielding [Moore and Williams, 2009].

The strategy of this thesis comprise of two major sections: a genome-wide single nucleotide polymorphism (SNP) interaction search and a subsequent candidate-gene analysis. The term epistasis may have different meanings: on one hand for a biologist it might be an aspect of biological mechanisms, on the other hand for a statistician the interaction between two factors represents deviation from a mathematical model for joint effects of several factors [Clayton, 2009]. The challenge is to satisfy the demands of both disciplines. The statistical evidence has to support the biological understanding.
Keeping this idea in mind, the first stage of the thesis was a hypothesis-free approach, in which a genome-wide two-locus scan was employed to unmask possible marker interactions within genotypic information of dyslexic subjects. The second stage was conceived as the validation phase, in which genes detected to be significantly associated to the phenotypic pattern in the initial stage, was used as candidates for an expression quantitative trait locus (eQTL) analysis in hippocampal expression profiles.
The hippocampus plays an important role in associative behaviors, not only in the transduction of information from short to long-term memory, but also in spatial memory and attention functions that likely participate in the development of dyslexia [Carrion-Castillo et al., 2013, Ludwig et al., 2008, Moser and Moser, 1998, Smith, 2007].
The aim of this thesis was to identify novel genetic factors involved in dyslexia pathogenesis through a joint analysis of epistasis and eQTLs. The analyses comprised novel combinations of stablished bioinformatic and statistical methods and biological comprehension. In order to provide a more in-depth understanding of complex traits and predispositions the focus was set on methods beyond single SNP associations.
Whole-genome exhaustive epistasis search was enabled via GLIDE [Kam-Thong et al., 2012], a tool for calculation of SNP-SNP interactions, implemented in Compute Unified Device Architecture (CUDA) for fast computation on graphics processing units (GPUs) of graphics cards. Utilizing GPUs for parallel computation reduced calculation time by a factor of 100 compared to conventional methods, which opens a wide dimension for novel analyses.
Provided datasets for the project included: (1) four dyslexia case samples comprising 200 German, 92 French, 377 British, and 194 American dyslexic individuals. For a wide spectrum of dyslexia related phenotypes, such as single-word reading, spelling, non-word reading and phonological-awareness measurements were recorded. (2) For candidate gene analyses, hippocampal gene expression data, as well as genotypic data of 138 epilepsy-patients were available. In this dataset, we already identified 360 cis- and 75 trans-acting eQTL associations on a Bonferroni-corrected genome-wide significance level (publication in preparation).
The most straightforward approach for an analysis was to examine whether detected dyslexia-associated interaction marker pairs represent hippocampal eQTLs. Gene expression data constitutes an excellent biological phenotype for the identification of regulatory genomic regions of candidate genes. The motivation here was to evaluate statistical epistasis findings with biologically appropriate expression profiling data with

respect to their potential as novel markers that could help to infer new hypotheses for dyslexia pathogenesis.

Epistasis analysis may help to improve and extend classical genome-wide association studies (GWASs) effectively, opening a new area of investigations, and moreover, enlighten genetic causes of dyslexia.

2 Background

2.1 Genetic background

2.1.1 Genetic factors

This section addresses basic genetic factors that are crucial for effect analysis in genome association mapping. Each subsection will cover the biological basics and address the quality control (QC) steps to avoid confounding impacts on the statistical analysis.

2.1.1.1 Single Nucleotide Polymorphism

A single nucleotide polymorphism (SNP) describes genetic variations involving only a single nucleotide (Adenine (A), Cytosine (C), Guanine (G) or Thymine (T)) exchange on a certain genomic position on the deoxyribonucleic acid (DNA) that is observed between individuals in a population. It is assumed that there are up to 30 million variations covering the DNA, where each variant is located at a specific site (locus) and occurs with a particular frequency within populations [Kwok, 2003].
These polymorphisms are point mutations that withstood natural selection and are manifested in the genome. SNPs are genetic determinants of the individual development and measurable phenotypes, such as complex diseases and disorders, which make them biological markers of interest. With the progress of modern technologies, especially the development of microarray platforms and sequencing strategies, the whole genome of hundreds of individuals can be scanned and single variations can be subjected to surveys. On the basis of these variations genomic studies are conducted, where the correlation of an explicit locus to a trait of interest is ascertained in a cohort.
SNPs are unequally distributed over the whole genome, they are more frequent in non-coding regions than in more conserved coding-regions of the genome. Occurrences of SNPs in coding regions of genes can be classified in synonymous and non-synonymous mutations, where the latter lead to a change in the amino-acid sequence of the encoded protein while the former does not [Kwok, 2003].
Biologically functional SNPs are further categorized depending on the traits they are affecting. A SNP regulating respectively affecting the expression of a certain transcript is annotated as expression single nucleotide polymorphism (eSNP). If any impact on methylation processes can be estimated it is referred to methylation single nucleotide polymorphism (mSNP). In genetic mapping studies of continuous traits, a specific locus

that is detected to significantly affect the trait outcome it is annotated/declared as a QTL.
In the genotyping process, the pair of alleles on the homologous chromosome (diploid cells consist of pairs of chromosomes, where one chromosome is transmitted maternally and one paternally) are taken into account. An individuals genotype can be homozygous by harboring two copies of the same allele (e. g., TT or CC), or heterozygous with one of each alleles (TC). An allele is either less (minor allele) or more frequent (major allele) in various populations and can differ between those. The linkage/connection between the prevalence of a risk-allele and the affected disease are often controversy discussed, and will be briefly subjected in the following subsection 2.1.1.2.

2.1.1.2 Genotype and allele frequencies

Genotype frequency describes the prevalence of a given SNP genotype in a population by the simple expression $\frac{\text{\# individual genotypes}}{\text{total \# of individuals}}$. Allele frequency in turn defines the rate of each single allele in the population by $\frac{\text{\# of each allele}}{2 \text{ x \# of individuals}}$ [Kam-Thong, 2012]. It is established that a minor allele frequency (MAF) less than 1%-5% in a given population is considered as a rare allele, and alleles with a frequency above that threshold are common alleles.
It is highly discussed whether causative mutations for complex traits are more likely to be common or rare [Salyakina, 2007, Manolio et al., 2009]. One standpoint is that common variants, which occur often in a population are less likely to be malicious as the statistically associated disease would be highly prevalent in the population. On the other hand, if genetic variants account only for 10-15% of a trait [Manolio et al., 2009] and environmental factors have an important role in complex traits, then it is possible that the interplay of both factors causes the phenotype. Therefore common complex diseases may be caused by common variants.
A notable example is a study published by Thorleifsson et al., where 25% of the general population are homozygous for the highest-risk haplotype associated with exfoliation glaucoma [Thorleifsson et al., 2007].
In order to avoid misclassification, bias, high false positive or negative rates a QC criterion is the exclusion of SNPs with a MAF $< 5\%$. The intention is that biomarkers with a MAF less than 0.05 do not show much variation across the population and detection of effects becomes unlikely unless the effect sizes are very large, such as in monogenic conditions [Manolio et al., 2009]. Facing the complexity of genetics, both mentioned mechanisms may exist in various combinations [Salyakina, 2007].

2.1.1.3 Hardy-Weinberg equilibrium

The principle of the Hardy-Weinberg equilibrium (HWE) is an ideal scenario where the allele frequency in a population remains constant (Eq. 2.1 on the facing page) over generations, without being influenced by mutation, selection, genetic drift, or

non-random mating.

$$P(A)^2 + 2 \times P(A)P(B) + P(B)^2 = 1 \tag{2.1}$$

P(A): allele frequency of allele A
P(B): allele frequency of allele B

Deviation from HWE within genotyped SNPs can indicate inbreeding, population stratification, genotyping errors [Wigginton et al., 2005] or imputation inconsistencies. In genetic studies deviation from HWE can provide wrong evidence for association due to incorrect genotyping. The usual QC actions include a HWE test to remove biomarkers violating the expectation of HWE. HWE tests are commonly performed using a simple χ^2 goodness-of-fit test [Wigginton et al., 2005]. A commonly defined QC threshold for a genome-wide genotype chip (550K) is $p \leq 5 \times 10^{-5}$, meaning that each SNP violating the HWE assumption with a significance level less then the defined p-value will be excluded from further analysis. On the other hand there is the QC dilemma that susceptibility variants often violate HWE expectation and are excluded via HWE testing.

2.1.1.4 Linkage disequilibrium

Linkage disequilibrium (LD) is the genetic phenomenon of non-random allelic relations at different loci usually on one chromosome, whereby the physical distance does not always explain the level of linkage and can differ from case to case. The association of two or multiple markers are influenced by recombination and is more or less frequent than expected from haplotypes. A haplotype can be one locus or a set of loci on one chromosome inherited together [Li et al., 2003]. The degree (D) of LD between loci can be measured and is expressed as the deviation of the observed frequency $f(A_iB_i)$ of a haplotype from that expected for independent alleles $f(A_i) \times f(B_i)$, via D' (Eq. 2.2) or the correlation coefficient r^2 (Eq. 2.4 on the following page).

$$\begin{aligned} D' &= \frac{D}{D_{max}} \\ &= \frac{f(A_iB_i) - f(A_i)f(B_i)}{D_{max}} \end{aligned} \tag{2.2}$$

$$D_{min/max} = \begin{cases} max(-f(A_i)f(B_i), -f(A_j)f(B_j)), & \text{D } < 0 \\ min(f(A_i)f(B_j), f(A_j)f(B_i)), & \text{D } > 0 \end{cases} \tag{2.3}$$

D_{max} is the maximum disequilibrium at an allelic frequency of 0.5.

$$r^2 = \frac{D^2}{f(A_i)f(B_i)f(A_j)f(B_j)} \tag{2.4}$$

The values of D' and r^2 of two markers can range between 0 and 1, where 0 implies a total independence and 1 a complete LD for D' and a perfect LD for r^2.*Ther*^{2} is usually the measure of choice in population genetics, as a value of 1 between two loci can just be reached when identical allele frequencies are given and the occurrence of an allele at a locus perfectly predicts the allele at the other locus [Salyakina, 2007].
A genomic region on a chromosome with high LD, or conserved LD, which is structured into a small number of haplotypes is referred to as an LD-block [Li et al., 2003] (figure 2.1). In genetic studies it is expected that variants of an LD-block have demonstrate to a certain extent, and in dependence to the r^2 value associations to the same phenotypic trait. In other words, if a tag SNP is associated to a phenotypic trait, then marker in high LD ($r^2 \leq 0.7$) should likewise exhibit the effect. A tag SNP is a representative variant of a LD-block. Well chosen tag SNPs can provide enough information to predict information about other variants in the corresponding LD-block. The extent of LD structures, haplotypes and the prevalence are population specific [Amaral et al., 2008].

2.1.1.5 Population stratification

A confounding factor in genetic studies may be population stratification, specially in studies with a huge sample size and more over in those with admixture cohorts. Population stratification describes ancestral differences, resulting in variation of marker allele frequencies among subpopulations. Samples comprising multiple populations can corrupt LD structures and lead to incorrect and false positive associations driven for instance by unexpected relatedness of individuals.
To intercept inaccuracy and avoid nonexistent associations, machine learning algorithms can be applied to account for population stratification. A conceivable approach, besides the general Principal Component Analysis (PCA) approach, is the complete linkage agglomerative clustering provided by the open-source tool PLINK v1.07 (http://pngu.mgh.harvard.edu/purcell/plink/) [Purcell et al., 2007]. Using PLINK a multidimensional scaling (MDS) analysis on the pairwise identity-by-state (IBS) distance matrix of each individual, measured by the respective whole genome SNP data, can be performed. Coexisting subpopulations, the relatedness degree between individuals and single outliers can be detected and either excluded or corrected for, by utilizing the MDS-components as covariates in statistical analyses. Applying genomic

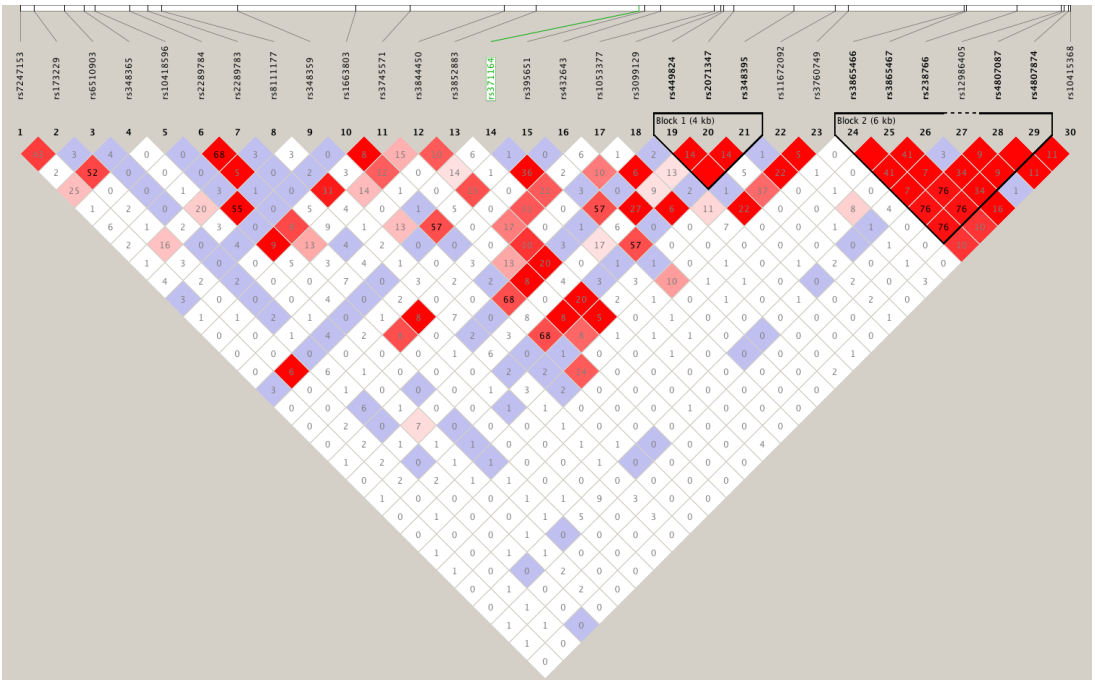


Figure 2.1: An Haploview example of a haplotype structure on chromosome 19. The figure depicts a chromosomal region with sparse LD-structure. Values in the squares annotate the r^2 of the SNP to its neighboring SNPs, supported by the color gradient. Increasing r^2 is indicated by a darker color, while white implies no or very little LD, red would mark higher correlation between the SNPs. In black framed triangles indicate LD-blocks at the sequence position.

scans heritability estimates can be generated free of potential confounding, due to unmeasured environment factors [Manolio et al., 2009].

A confounding factor describes a known or hidden factor disturbing the association between two other variables of consideration, due to relationship with both of them, the dependent and independent variables [Mosby, Inc, 2009]. In statistical genetics analyses confounding factors can originate from diverse sources. As mentioned above, confounders can arise by mingled populations or cohorts being genotyped in diverse laboratories. Especially in microarray (e. g., expression or methylation) analyses researchers are faced with confounders, like processing batch or measured ribonucleic acid (RNA) quality, unknown or specific environmental factors, medication and other contamination factors. Such factors can reduce the power of association detection and even facilitate false-positive associations [Stegle et al., 2012].

2.1.2 Genetic variation and complex traits

Since the establishment of genetic science by Gregor Mendel (19th century), the discovery of the molecular DNA structure by James D. Watson and Francis Crick up to the sequencing of the human genome in 2001 – 2003 by Craig Venter and the Human

Genome Project (HGP), researchers are on the hunt for genetic factors affecting phenotypic outcomes, with the aim to reveal disease susceptibilities and understanding basic genetic concepts for a better diagnosis, treatment, and prevention.
Apart from monogenic diseases, common human diseases and heritable traits are suggested to be shaped by several genetic and environmental factors following complex inheritance patterns. Individuals within the same species show heritable attributes for many traits of biological and medical interest, but until recently the identification of genetic factors contributing to these *complex traits* has been slow and arduous [Manolio et al., 2009, Bloom et al., 2013].

A complex trait is a phenotype that is determined by multiple factors, either genetic, or environmental. Generally, several genetic factors contribute to the development of a continuous outcome, or increase susceptibility for a disease. Susceptibility factors, or more specific in genetics a **susceptibility allele** describes a *risk allele* that is associated to a specific disease and increases the probability of the carrier developing the deleterious phenotype. Investigation of such complex traits are often complicated, as each contributing factor has usually a small effect.

GWASs have been a landmark in the investigation of genetic susceptibility factors in complex traits. With the advancement of high throughput genotyping technologies, up to one million common SNPs can be assayed in thousands of individuals, representing a powerful opportunity for investigating the genetic architecture of complex species. Compared to whole-genome sequencing, genotyping is much more time- and cost-efficient. Consequently a lot more individuals can be scanned for specific genetic variants of interest.
Large sample sizes are beneficial to gain statistical power for the detection of genetic effects and increase reliability of the results. In the past years these studies have identified more than a thousand loci harboring genetic variants affecting over 165 common human diseases and phenotypes [Zuk et al., 2012], providing precious insights into the complexities of human genetics.
The genome-wide association (GWA) method represents a hypothesis-free approach and is an important advance beyond candidate gene studies and family-based linkage studies that are both limited in their sample sizes and the assayed variants. Despite many convincing successes in monogenic Mendelian traits the moderate success of linkage studies has been attributed to their low power and resolution for variants of low effect. However even markers identified by genome-wide mapping meant to account for only 5–15% of the given phenotypic variance in human beings. A famous example is human height with an estimated heritability of about 80%. Meanwhile 40 loci have been identified to be associated to height, nevertheless they explain only about 5% of

the trait variance despite studies with reasonable sample size [Manolio et al., 2009]. If 15% of heritability is explained by common genetic variants and another proportion by environmental conditions or rare and structural variants, what explains the remaining proportion? This leads to the question what explains *missing heritability* in genetic epidemiology. Heritability of continuous traits, formally defined as the *„proportion of phenotypic variance in a population attributable to additive genetic factors (narrow-sense heritability), which can be inflated by non-additive genetic effects“* (dominance and epistasis, or gene-gene interaction) [Manolio et al., 2009], which are captured under the term *broad-sense heritability*. The next section 2.1.3 will cover different aspects of epistasis and its role in nowadays mapping studies.

2.1.3 Diversity of the epistasis term

As discussed in the previous section the identification of single genetic marker mapping to an attribute does not fully explain underlying variance. A possible explanation apart from environmental or undiscovered factors, rare and structural variants, or other heritable epigenetic factors and unforeseen sources could be the interplay between unlinked loci [Bloom et al., 2013], the so-called *epistasis*.
Epistasis nomenclature has a variety of different notations and meanings. Starting from the very beginning, the first definition of epistasis was given by Bateson and Punnett [Bateson, 1909] to define the masking action of one gene by another. They observed in a chicken and a follow-up pea flower experiment that alleles at one locus could mask the effects of the alleles at another locus forcing a totally new phenotypic outcome [Miko, 2008]. Another famous example of epistasis is the mouse coat color, where the homozygotes of the recessive alleles (e. g., aa) at one locus (A) alter the phenotypic effect of alleles at the other locus (B) regardless of the alleles B being recessive or dominant [Miko, 2008]. In this sense the term of epistasis would be limited to a kind of inhibiting or interfering effect, nevertheless, meanwhile the broad sense of epistasis covers a variety of interactions between genes.
The first specification of this kind of interplay beyond masking effects was made by Fisher, by defining the term of *epistacy* to describe deviation from the addition of superimposed effects between Mendelian factors [Fisher, 1918], in other words, epistacy is the deviation from additive linear effects of two loci on the phenotype. Under this definition interaction can be modeled mathematically (statistical epistasis), it is said to explain the combined effect of multiple genes on an outcome, which cannot be predicted by the sum of their separate effects [Frankel and Schork, 1996].
In the past years, the term epistasis was adopted by Fisher's more general definition of genetical interactions [Phillips, 2008]. One has to be aware about the biological interpretation of statistical interaction as a statistically detected significant interaction does not necessarily mean an interaction in biological senses and vice versa, biologically discovered epistasis does not have to be detectable statistically, for example due to

non-linear interaction patterns, making the whole survey in this field even more complicated. In the statistical sense, interaction is just the non-additive effect of multiple factors on an independent variable.
Very little is known about the role of epistasis in human biology, as initially the focus was placed on single locus effects and investigations on the interaction level were neglected. Phillips [Phillips, 2008] defined three major categories of epistasis: *functional epistasis*, *compositional epistasis* and *statistical epistasis*.
Functional epistasis describes any kind of molecular interactions of genetic elements, such as proteins that operate within the same pathway or directly build complexes with one another.
Compositional epistasis addresses the traditional usage of epistasis denoted by Bateson as the blocking effect of one allele by another allele at a different locus.
Finally, and most important for our demands, is the **statistical epistasis**, in which the deviation of allele combinations is estimated over all other genotypes present within a population [Phillips, 2008].
Considering the complexity of genetics and the ubiquity of epistasis it is natural that accounting for interactive effects is one step towards unveiling missing heritability, at least partially.
Examples of phenotypes for which synergistic effects, this is the variation addressed by the interaction of multiple factor that is greater as the sum of their single effects, between loci have indeed demonstrated to be reliable predictor variables of the phenotypic variance include diseases such as type 1 and type 2 diabetes, hypertension [Kam-Thong et al., 2012] and increased risk for schizophrenia [Nicodemus et al., 2010b].
Ashworth et al. observed interactions in the context of cancer cell proliferation and listed examples detailing the different nature of genetic interactions enhancing or suppressing cancer mutations with new therapeutic treatments proposed to target these interactions [Ashworth et al., 2011, Kam-Thong et al., 2012].
Rohlfs et al. noticed unusual allelic association, not attributable to population structure, between the coevolving interacting genes ZP3 and ZP3R. Coevolving interacting genes undergo complementary mutations to maintain their interaction. Alleles of such coevolving genes interact differently and can create several varying degrees of fitness. They mentioned that if the created fitness differential is adequately large the resulting selection for allele matching could maintain allelic association, even between physically distant loci [Rohlfs et al., 2010].
Statistical modeling of interactions may be helpful in identifying genes influencing disease susceptibility that otherwise would remain unidentified. It is important to note that the presence of statistical epistasis can help generate new hypotheses, but requires an in-depth investigation of the elementary molecular mechanisms involved to substantiate the findings.
Keeping all that in mind and with the aim to understand functional and pathophysiological properties of epistasis we tried to take a step beyond GWASs towards genome-wide

interaction analyses (GWIAs) (section 2.1.4).

2.1.4 From GWA to GWIA

GWASs are the gold-standard in single-locus analysis strategies, wherewith genotyped variants are assessed individually for association with a specific phenotype. GWASs harbor the lack of discounting the existence of interactions either environmental or genetical. The complex relationship between genotype and phenotype, thus, may be inadequately described by simply summing the modest effects from several contributing loci. Instead, the relationship may depend in a fundamental way on epistasis between multiple loci and/or genotype and environment [Culverhouse et al., 2002].
It is not deniable that the hypothesis-free approach of single-locus mapping, besides linkage studies and targeted candidate gene studies, has been incredibly successful in the last decade and has offered a great insight in the genetical complexity of disease susceptibility, but each technique has, beyond its benefits, also its individual drawbacks and limitations. Naturally, genetic factors work through nested mechanisms that involve multiple genes and environmental factors. The detection of such convoluted effects would be totally missed if the gene is examined in isolation without allowing for its potential intrinsic interactions with other unknown genetical factors [Cordell, 2009].
Targeting this deficiency of GWASs and utilizing the recent progress in adequate hard- and software new methods (examples are addressed in chapter 2.3) were developed addressing potential multiple loci interactions by scanning the effectiveness of two or more loci on a specific phenotype. Schematically GWIA does not differ substantially from GWASs. In either method the association relies on a linear (in dichotomous traits: logistic) link between a given genotype and phenotype. In GWIA this link involves multiple, respectively two loci. In fact the simplest and most direct way is the exhaustive search between all loci. In a two-locus interaction search all possible SNP-pairs are considered for their effects while allowing for interactions, meaning that the single effect of a locus will be a component of all tests that involve that locus [Cordell, 2009]. Such an exhaustive search, that can comprise up to 10^{14} tests in the two-dimensional space, is incredibly time consuming and computationally intense but feasible with new technologies (see section 2.3.3). Meanwhile a variety of tools have been developed to compute interactions; and controversial debates about the reasonability and adequately of approaches have been argued and examined. Discussions about the validity of the results split the opinions of researchers. A very common theory expects a marginal penetrance model, in which marginal single locus effects of both markers have to be present to explain the validity of the association and biological interpretation. This approach is often taken into consideration, apart from the assumption of biological causality, to shrink data dimensionality and thus the search space concerning time and computation feasibility.
However, the so-called *purely epistatic model* does not expect additive or dominance

variation at any of the susceptibility or involved loci. The idea is that the effect on a phenotype is only perceivable through the interaction. Studies proposed complex theoretical penetrance models by accounting for multiple loci interactions without displaying any main effects. Culverhouse et al. [Culverhouse et al., 2002] illustrated that two-locus models can exist without marginal effect at either locus involved, which nonetheless accounts for a large portion of the population variance. He summarized that a purely epistatic model includes both incomplete penetrance and phenocopies and that by now it is very ambitious to judge the extent to which purely epistatic interactions are manifested in human disease. He also mentioned that in situations in which single locus analyses do not account for the predicted genetic variance it is worth to comprise interaction effects.
Since 2002, few publications were able to validate statistical epistasis with biological functionality e. g., Nicodemus et al. [Nicodemus et al., 2007, Nicodemus et al., 2010a] which is abstracted in the discussions (*section* 5). Still we are far to tell what is literally right or wrong, but it is reasonable to assume that scenarios displaying small marginal effects can account for more variation and seems more *natural* [Culverhouse et al., 2002]. Nevertheless, considering only SNP-sets on the basis of marginal effects would lead to a loss of information as effects between non-significantly associated variants, which could well be a part of complex genetics, would be completely uncovered.

2.1.5 Dyslexia, a common complex disorder

Dyslexia (specific reading disability) is a common neurodevelopmental disorder mainly characterized by difficulties in reading and spelling in children. Developmental dyslexia was first verified in English-speaking populations; a language where the relation between graphemes and phonemes is inconsistent [Grigorenko, 2001]. As meanwhile known dyslexia occurs in all languages with a prevalence of approximately 5–12% in children all over the world [Carrion-Castillo et al., 2013, Schulte-Körne et al., 2007].
Thus, cross-linguistic approaches, as in the case of this thesis, would be more promising to uncover more universal aspects of the disorder development. Longitudinal studies have proven that the disorder involves an extremely stable developmental disturbance, where some individuals are able to compensate their learning deficits with adolescence while others remain functionally reading-impaired their whole life [Pammer, 2014].
According to the International Classification of Diseases 10th Revision (ICD-10 Version:2010, `http://apps.who.int/classifications/icd10/browse/2010/en`) the pathogenesis is characterized by *„pronounced difficulty in learning to read and spell despite conventional instruction, adequate intelligence and sociocultural opportunity“*. It can have substantial impact on affected individuals and may impair their whole conduct of life. Documentations indicate an elevated appearance of depression [Maag and Behrens, 1989] and anxiety disorders [Smith, 1991] among dyslexic individuals, due to difficulties at school and work.

The first detectable symptoms are deficits in word reading and spelling, but several other cognitive components impact the core phenotype [Galaburda, 1999]. Studies implicate phonological deficits as central to dyslexia, which is not deceptive at all considering the complex hierarchical process of reading embracing cognitive systems being mutually in an influential relationship like visual recognition, symbol mapping, phonological, semantic, and syntactic processing and memory [Grigorenko, 2001].
Hence facing all these factors a limitation to a categorical definition (affected vs. not affected) may be too simple. Over the past years it has been established to work directly with psychometric measures for assessing relationships between molecular elements and the disorder. For such estimations a battery of cognitive tests (details of assessment can be found in section 3.1) are assessed for the study subjects, where the individual's performance on single word reading, spelling, orthographic processing, phonological-awareness, non-word reading, rapid automatized naming and phonological short-term memory are measured [Carrion-Castillo et al., 2013]. These so called *endophenotypes* are defined as quantitative indices close to the underlying biological phenomena and that are conceivably easier to link with the genetic factors [Gottesman and Gould, 2003].
The etiology of reading is not fully described not even in its non-disturbed normal processes. A handicap of cognitive and especially dyslexia research is that many findings have been poorly replicated, suggesting that dyslexia may have several manifestations at different stages of development and that cognitive systems are highly exposed to environmental systems affecting measurable outcomes [Grigorenko, 2001]. Nonetheless, it is plausible and known that learning deficits underly biological dysfunctions. Studies characterized cognitive deficits attributable to neurological abnormalities, which may affect other disorders as well. Comorbidity studies of dyslexia, attention deficit hyperactivity disorder (ADHD) and speech sound disorder (SSD) demonstrate genetical and cognitive commonalities between these conditions [Czamara et al., 2013, Smith, 2007].
Regarding the variety of the underlying systems such disorders should be considered in an interdisciplinary manner. Beginning by studying the brain structure many different observations via autopsy or structural Magnetic Resonance Imaging (MRI) techniques were published identifying varying sizes or unusual symmetry in different brain areas between dyslexic and non-dyslexic individuals. Furthermore the thalamus, insula, hippocampus, and other regions were mapped to the dyslexic brain [Grigorenko, 2001]. An important component is the neurophysiology of dyslexia. Humphreys et al. [Humphreys et al., 1990] described unusually organized nerve cells and suggested fetal developmental disturbances hindering neuronal migration. Meanwhile, the role of cell migration and axon guidance during the development of the nervous system and disturbances inducing dyslexic endophenotypes are well studied and widely accepted [Carrion-Castillo et al., 2013, Smith, 2007, Grigorenko, 2001]. Wood et al. [Wood et al., 1991] observed irregulary distributed metabolic activities through the brain in dyslexic persons compared to more equally distribution in unaffected persons. The observation of differentially regulated metabolites in cases and controls was also recognized in a

study by Garret et al. [Garret et al., 1997].
Overall the basal patterns of dyslexia are neither simple nor straightforward and it seems that no single mechanism or region can be allocated to the various disabilities of the phenomenon of dyslexia. The suggestion that the disorder reflects selective disturbances of a cognitive system seems more reasonable [Grigorenko, 2001]. Going deeper in the pathogenesis, the elementary question revolves around the genetic fundamentals and the heredity of the phenotype.

2.1.5.1 Genetics of dyslexia (Genetic epidemiology)

Family and twin studies provide evidence that developmental dyslexia is highly hereditary with an estimated proportion of variance that is explained by genetic factors ranging from 0.4 to 0.8 [Schumacher et al., 2007, Carrion-Castillo et al., 2013]. Wolff et al. detected higher risk for sibs showing dyslexic endophenotypes with both parents being affected, than those with one affected parent, indicating patterns of additive effects [Wolff and Melngailis, 1994]. Twin studies estimated concordance rates of 20% to 55% for dizygotic (DZ) twins and 68% up to 100% for monozygotic (MZ) twins [Hermann, 1959, Zerbin-Rüdin, 1967, Bakwin, 2008].
Since heritability studies provide strong evidence for solid genetic impact on the phenotype many surveys searching for genetic background and abnormalities were conducted. Considering the mentioned nebulous and multifaceted biological processes involved in the mastering of reading or spelling it stands to reason that the genetic architecture is not less complex or multifactorial. For example Olson et al. [Olson et al., 1999] demonstrated that genetic effects on phonological decoding and orthographic coding are due to shared and independent components. Moreover, it is important to keep in mind that in reading and spelling involved processes underly huge environmental influences, but nevertheless there is a high broad-sense heritability suggesting the contribution of genetic factors [Grigorenko, 2001]. More and more genetic risk factors are identified causing susceptibility to single endophenotypes of the disorder. In this section an overview of detected loci and genes being published in the field of dyslexia will be provided.
Despite the fact that currently the genetic architecture of dyslexia appears very complicated, there are some genes identified to be involved in the pathogenesis, however not in a Mendelian way. The trait pattern rather indicates to be a result of the interplay of genetic factors involving combinations of polygenicity, heterogeneity [Carrion-Castillo et al., 2013] and pleitropic genes [Grigorenko, 2001] with small effects manifesting the clinical picture [Paracchini, 2011].
The introduction of reported susceptibility regions should be initiated with the nine most popular candidates that indeed own their nomenclature from the disorder, namely DYX1 (*Dyslexia susceptibility 1*) to DYX9 [Schumacher et al., 2007] (figure 2.2), enumerated by the order of their detection (Nomenclature is assigned by the HUGO Gene Nomenclature Committee (`http://www.gene.ucl.ac.uk/nomenclature/`).

DYX1 on *15q21* was one of the first detected and replicated regions for reading and spelling disability [Smith et al., 1983, Bates et al., 2007]. Nopola-Hemmi et al. [Nopola-Hemmi et al., 2000] discovered *DYX1C1* as a candidate gene in a Finnish family study in which balanced translocations involve the region *15q* and different chromosome arms of chromosome 2, co-segregating with reading and writing difficulties. Markers in *DYX1C1* have been reported to be associated with developmental dyslexia and with short-term memory performance in affected females [Dahdouh et al., 2009]. Furthermore Tammimies et al. [Tammimies et al., 2013] detected interactions between *DYX1C1*, *DCDC2* and *LIS1* (a protein implicated in lissencephaly, a rare brain disorder caused by severely disrupted neuronal migration) [Schumacher et al., 2007].

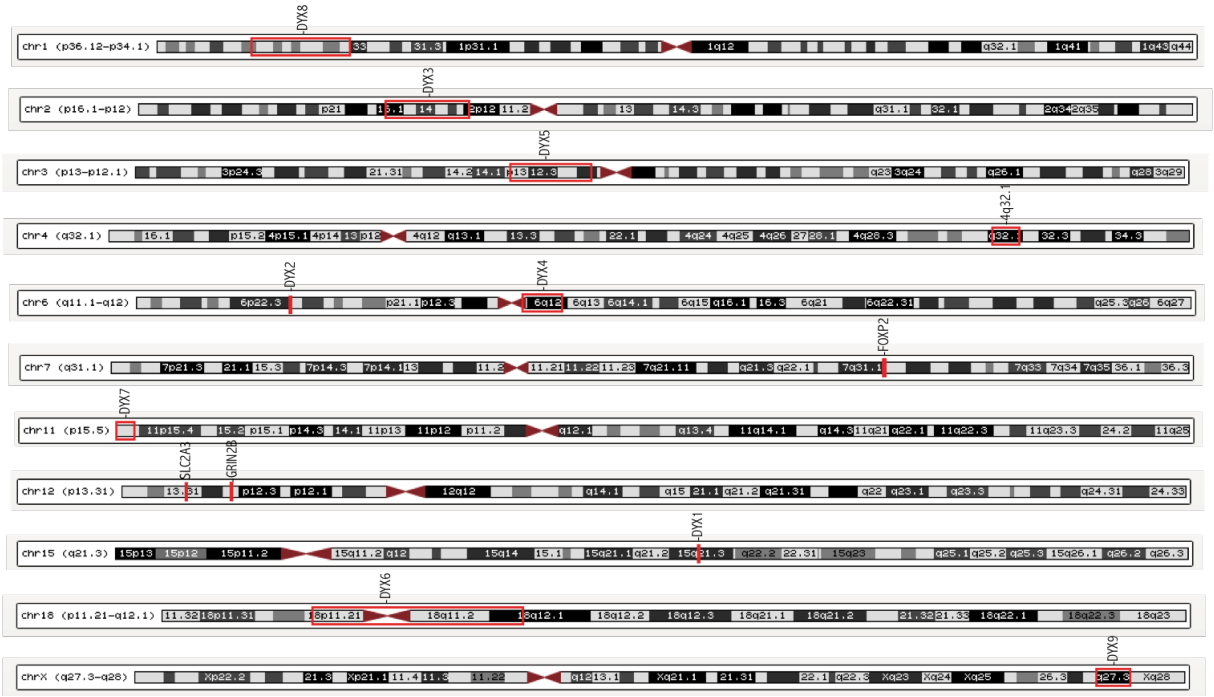


Figure 2.2: Ideogram of published dyslexia susceptibility loci. Chromosomes (1, 2, 3, 4, 6, 7, 11,12, 15, 18, X) of interest are shown with the cytogenetic bands. Red bars or boxes respectively indicate approximate cited susceptibility regions for dyslexia, with the corresponding locus or gene names. *Reference: UCSCs Genome Browser (hg19)*

DYX2 *DCDC2* (*6p22.2*) a gene that clusters among others (*VMP*, *KIAA0319*, *RREAP*, *THEM2*) in the DYX2 locus, one of the most consistent findings in dyslexia genetics [Smith, 2007], was reported originally by Cordon et al. [Cordell, 2009]. DYX2 is one of the most replicated DYX loci [Gibson and Gruen, 2008], especially the intra-cluster genes *DCDC2* and KIA00319 [Harold et al., 2006, König et al., 2011, Ludwig et al., 2008, Pinel et al., 2012]. Variants in *DCDC2* are described for association to reading by Meng et al. [Meng et al., 2005]), as well as to the quantitative phenotype spelling by Schumacher et al. [Schumacher et al., 2005] and recently an allele has been identified by Marino et al.

being associated to memory [Marino et al., 2012]. Precise function of *DCDC2* is not fully enlightened, but it is suggested to be involved in cell and neuronal migration [Smith, 2007]. Beside all findings there were also documented surveys that could not replicate any associations of this gene with dyslexia [Schumacher et al., 2007].
KIAA0319 seems to be more robust in replication studies than *DCDC2*, it is found to be significantly associated with risk to dyslexia by Francks et al. [Francks et al., 2004] in populations from the United Kingdom (UK) and the United States (US), thus was later confirmed by Paracchini et al. with further involvement of the gene in neuronal migration [Paracchini et al., 2006] and associations to reading skills in the general population [Paracchini et al., 2008, Paracchini, 2011]. Studies have also investigated the question whether *KIAA0319* variants might have impact across different neurodevelopmental disorders. Scerri et al. mentioned the association of *KIAA0319* with reading and spelling scores and relations to comorbid disorders such as ADHD and specific language impairment (SLI) [Scerri et al., 2011]. Subsequently in 2012 Scerri et al. found along with markers in *KIAA0319* also variants in the *MRPL19/C2ORF3* gene on the DYX3 locus (*2p12–2p16*) to be significantly associated with verbal and performance IQ in an investigation in the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort of 5, 000 individuals [Scerri et al., 2012].

DYX3 locus (*2p12–2p16*) was identified by Fagerheim et al. [Fagerheim et al., 1999] for linkage to dyslexia in a multigenerational Norwegian family where dyslexia is inherited as an autosomal dominant trait. Kamine et al. [Kaminen et al., 2003] confirmed linkage to DYX3 in a sample ascertained from Finland. Another study realized by Anthoni et al. supported DYX3 as susceptibility locus, by encountering risk haplotypes located in an intergenic region between *FLJ13391* and *MRPL19/C2ORF3* [Anthoni et al., 2007].

DYX4 In comparison to the previously mentioned loci DYX4 is poorly reported. Petryshen et al. [Petryshen et al., 2001] suggested linkage of the qualitative phenotype phonological coding and the continuos phenotypes: phonemic awareness, phonological decoding, rapid naming, and spelling to the region *6q11.2–q12* and designated this locus as DYX4 [Carrion-Castillo et al., 2013, Grigorenko, 2005].

DYX5 The location *3p12–q13* referred to as DYX5 is replicated more often. Initially Nopola-Hemmi et al. [Nopola-Hemmi et al., 2001] described in an analysis of dyslexic subjects a shared identical copy of a haplotype on chromosome 3, later this region led to the identification of the *ROBO1* gene with alleles associated to dyslexia showing an altered (attenuated) expression level [Smith, 2007]. Stein et al. [Stein et al., 2004] detected also linkage of the locus to SSD. Bates et al. [Bates et al., 2007] replicated this region in a genome-wide linkage analysis for reading and spelling. Homologous genes of *ROBO1* in mouse and Drosophila melanogaster seems to be involved in axon guidance cross the brain. Hannula-Jouppi et al. [Hannula-Jouppi et al., 2005] suggested

that a slight disturbance in neuronal axon navigation, dendrite guidance, or another function of *ROBO1* may manifest as a specific reading disability in humans.

DYX6 In a QTL based genome-wide scanning in UK and US samples [Fisher et al., 2002] DYX6 (*18p11–q12*) showed increased evidence for linkage to single-word reading (SWR) and measures related to phonological and orthographic processing. In a subsequent replication with a second UK sample they could detect strongest evidence from phoneme awareness measures to that loci. With a combined analysis of both UK families Fisher et al. substantially validated that the *18p* QTL is probably a general risk factor for dyslexia, influencing several reading-connected mechanisms. Furthermore Bates et al. [Bates et al., 2007] could detect linkage of DYX6 to phonological decoding and awareness as well as SWR and orthographic awareness.

DYX7 Apart from those finding Fisher et al. investigated the region *11p15* being linked to phonological-awareness in their UK sample. Results of a study of Hsiung et al.[Hsiung et al., 2004] have provided significant evidence for linkage of dyslexia to *11p15.5* which is assigned to DYX7.

DYX8 (*1p34-p36*) could repeatedly be mapped to qualitative and quantitative dyslexia phenotypes [Tzenova et al., 2004]. Grigorenko et al reported not only chromosome 1 being associated to dyslexia they also assumed interaction between a *1p36* locus and a *6p22.2* locus (DYX2) [Grigorenko, 2001, Grigorenko et al., 2001]. Again Bates et al. [Bates et al., 2007] were able to replicate this region in their replication study of reported linkages for dyslexia and spelling.

DYX9 Last but not least DYX9 which is assigned to the X chromosome (*Xq27.3*), was identified by de Kovel et al. as a genome-wide significant peak of linkage in a Dutch family. Their observations led to the assumption that the supposed risk allele has to be dominant with reduced penetrance and more variable effects in females [de Kovel et al., 2004, Carrion-Castillo et al., 2013].

Apart from the DYX regions there are few other genes and loci that are not considered as dyslexia candidate loci, as solid evidence for linkage or association is still not proven, they are rather correlated to other neurological abnormalities or disorders demonstrating comorbidity with dyslexia.
One of the most popular genes in neurodevelopmental speech and language disturbances is *FOXP2*. *FOXP2* (*7q31*) was the first candidate gene studied in SLI, a point mutation in exon 14 of the gene was found to be involved in the developmental process of speech and language [Lai et al., 2001]. Later MacDermot et al. [MacDermot et al., 2005] investigated the entire coding region of *FOXP2* in subjects affected with verbal

dyspraxia and detected protein sequence altering variants in 3 of 49 probands cosegregating with speech and language difficulties. Vernes et al. [Vernes et al., 2008] reported down-regulating actions of *FOXP2* on *CNTNAP2*, a gene expressed in the developing human cortex where polymorphisms in that gene are detected for association to SLI. A second paper of Vernes et al. [Vernes et al., 2011] indicated that *FOXP2* modulates neuronal network formation by directly and indirectly regulating messenger ribonucleic acids (mRNAs) involved in the development and plasticity of neuronal connections. Besides the monogenic syndrome of *FOXP2*, causing impaired speech development and linguistic deficits, reduced dosage of the gene result in abnormal synaptic plasticity and impaired motor-skill learning in mice, and disrupts vocal learning in songbirds [Fisher and Scharff, 2009, Kurt et al., 2012]. Effects of *FOXP2* were often investigated in the field of dyslexia with moderate success. Recently Wilcke et al. were able to find a variant of the gene significantly associated with dyslexia in a case-control and functional magnetic resonance imaging (fMRI) study [Wilcke et al., 2011]. A fMRI survey of Pine et al. could detect three intronic SNPs of *FOXP2* significantly associated with reading activation in two brain regions (frontal regions of the left hemisphere: the inferior frontal gyrus and the dorsal part of the precentral gyrus), and intriguingly they also observed a SNP within the *KIAA0319/TTRAP/THEM2* locus (DYX2) associated with temporal functional asymmetry [Pinel et al., 2012].

Further discoveries in the field of genetic dyslexia were done by Roeske et al. [Roeske et al., 2011] and Ludwig et al. [Ludwig et al., 2010], where two genes on chromosome 12 were detected for association to neurophysiological endophenotypes of dyslexia. Roeske et al. identified a marker (*4q32.1*) to be correlated with late mismatch negativity (MMN) component, which reflects automatic speech deviance processing that is altered in dyslectics; and a second SNP in LD, both in turn being significantly association with expression levels of *SLC2A3* (*12p13.31*), suggesting trans-regulation of the gene that is involved in glucose transport in neurons, which might lead to glucose deficits in dyslexic children and could explain their attenuated MMN in passive listening tasks [Roeske et al., 2011].

The next gene on chromosome 12 supposed to be related to dyslexia is *GRIN2B*, Ludwig et al. describe three intronic SNPs of the gene associated with short-term memory in dyslexia. They observed even stronger effects when only maternal transmission were considered [Ludwig et al., 2010].

Even though all the promising findings in the complex and multifaceted nature of dyslexia it appear complicated to discover highly effective genetic factors, probably due to genetic heterogeneity and relatively small single-locus effect sizes. It is conceivable that the syndrome is rather explainable by considering the interplay of genetic factors on the different neurophysiologic endophenotypes of dyslexia. From this perspective it seems reasonable to take an epistatic look on the genetics of dyslexia.

2.2 Statistical background

2.2.1 Basics of the epistasis model

Statistical interaction can be well described in relation to a regression model that depicts the relationship between an outcome variable and predictor variables [Cordell, 2009]. Depending on the outcome variable i.e., qualitative or quantitative, one needs to distinguished between the logistic- or linear regression model. Assuming a dichotomous trait e. g., the disease affectedness of the observation sample, the logistic approach is the method to choose (subsection 2.3.4.1) and the log-odds ratio between the binary values is mapped as the outcome variable. The genetic effect on quantitative traits, obviously not classifiable in to two distinct groups, is estimated by applying the linear model (subsection 2.2.1.1).

2.2.1.1 Multiplicative interaction - a deviation from the additive model

By definition, epistasis is the statistical deviation from additive linear effects of two involved loci on the phenotype [Fisher, 1918]. The basic mathematical approach to model statistical epistasis of two independent loci and their impact on a quantitative phenotype is the linear regression model, in simplified terms posed by Eq. 2.5:

$$Phenotype = Intercept + \alpha SNP_A + \beta SNP_B + \gamma SNP_A SNP_B \tag{2.5}$$

αSNP_A and βSNP_B present the main effects of each of the two SNPs (SNP_A and SNP_B), and $\gamma SNP_A SNP_B$ the interactive effect of both SNPs on the $Phenotype$. Genetic variants are represented by numerically coded alleles (detailed explanation available under section 2.3.4). The term $Intercept$ absorbs any bias that is not accounted for by the given terms, in genetic models it could be variations arising from the environmental or other confounding factors. In the case of the **additive model** [Wade et al., 2001] the interaction coefficient $\gamma SNP_A SNP_B$ is equal to zero, as no multiplicative effect is present, meaning that the single SNPs affect the relation to the trait independently only in an additive way; in other words, there is no interaction present. The realization of the linear model in statistical epistasis tools is provided in section 2.3.4.

2.2.2 Problem of multiple testing

In genetic association testing, where simultaneously hundreds of thousands SNPs are tested for their effect on a specific outcome the risk of discovering false positive effects, in other words incorrectly rejecting the null hypothesis, is very high.
The null hypothesis is rejected if a rare event is detected, but the larger the number of tests, the easier it is to find such events and therefore the easier to interpret an effect as rare when there is none [Abdi, 2006].

To prevent wrong interpretations based on multiple testing many statistical methods were developed by which the number of independently performed tests are taken into account for correction of the multiple testing hypothesis. The usual 5% level, meaning that there is a 5% chance that a result is incorrectly true, has to be corrected for the number of tests for the determination of the significance level. In a GWA analysis that would be the number of the considered SNPs.
One of the most commonly applied methods to control familywise error rate (FWER) (the probability of one or more false positive discoveries (type I errors) among multiple tests) is the Bonferroni-correction, which is also the most conservative technique with a very stringent threshold. The Bonferroni-correction is the simple approximation method of the Sidàk equation in which a test reaches significance with a smaller probability as $\frac{0.05}{\#tests}$ [Abdi, 2006].
To ensure that significance is not due to randomness and within the expectation of independence this method is surely the appropriate choice, on the other hand taking all SNPs (which indeed are not all independent due to LD structures) into account the significance level will converge to highly small probabilities and would possibly lead to false negative results. Hence two approaches are often applied, first selecting a limited number of LD-pruned loci for consideration, which shrinks the search space and, thus, the significance level correction, or the adoption of a permutation based method, which might be computationally infeasible [Cordell, 2009, Culverhouse et al., 2002].
Facing interaction analyses one can imagine that the problem of multiple testing gets even more server and stringent considering the number of tests, n^2. Tim Becker et al. [Becker et al., 2010] published a promising modification approach of the Bonferroni-correction for GWIA, which is also applied in this thesis. The ***Bonferroni-Becker approach*** implies that a correction with roughly $0.44 \times m$ (m being the number of tests, i. e., the number of SNP-pairs: $m = \frac{n(n-1)}{2}$ and 0.44 the correction factor) is appropriate for GWIA. The correction factor can be calculated for each dataset individually, the methodology relies on the combination of permutation and Sidàk strategy, which given computational constraints is not always feasible.
Basically, the permutation procedure is repeated x times for a given dataset, the obtained minimum p-value of each run is kept, from which the specific correction factor for GWIA is inferred via the reverse Sidàk equation (in R the function *ks.test*).Despite its benefit for accuracy this is a highly time consuming and computational expensive approach that is not realizable in every case; Becker et al. demonstrated that the correction factor of 0.44 is appropriate for GWIAs with the Illumina® 550 HumanHap chip, or either any genotyping data with up to 500,000 SNPs.
Another less stringent method is the false discovery rate (FDR) approach. FDR was developed by Benjamini and Hochberg [Benjamini and Hochberg, 1995] to control error in the multiple-testing situation. It is used to estimate false positive results. The formula is based on the FWER with the traditional situations described in table 2.1.

	True H_0	True H_1	Total
Declared significant	V	S	R
Declared non-significant	U	T	m - R
Total	m_0	m - m_0	m

Table 2.1: Hypotheses definition for familywise error rate testing m null hypotheses

m is the number of (null) hypotheses H_0 to be tested, of which m_0 are the true null hypotheses. R is the number of rejected hypotheses (non-null hypotheses (H_1)). R is an observable random variable; U, V, C and T are unobservable random variables. If each individual null hypothesis is tested separately at level α (usually 0.05 or 0.1), then $R = R(\alpha)$ is increasing in α. In terms of these random variables, the per-comparison error rate (PCER) (the probability of a result in the absence of any formal multiple hypothesis testing correction) is $E(V/m)$ and the FWER is the probability $P(V \geq 1)$.
In association studies the FDR is more commonly applied. It is defined by the random unobserved variable $Q = V/(V+S)$ the proportion of rejected null hypotheses which are rejected erroneously. Obviously, the unobserved random variable Q would be null if the number of false positives (V) and true positives (S) is zero. Therefore the FDR expectation corresponds to Q (Eq. 2.6). If all null hypotheses are true, the FDR is equivalent to FWER, otherwise FDR is smaller than or equal to the FWER.

$$FDR = E(Q) = E\left\{\frac{V}{(V+S)}\right\} = E\left(\frac{V}{R}\right) \tag{2.6}$$

2.3 Technical background

2.3.1 R

R (`http://www.r-project.org`) is an open source software environment for statistical computing and graphics. It runs on a wide variety of operating systems. It is a GNU project based on the S language and environment that was developed at Bell Laboratories (formerly ATT, now Lucent Technologies) by John Chambers and his colleagues. R provides a huge variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, etc.) and graphic techniques and is readily extensible [Core, 2008]. It offers a huge range of packages especially for biological questions, for example the packages from `Bioconductor` [Gentleman et al., 2004]. All statistical analyses for this thesis that were not performed with specific tools were realized utilizing R.

2.3.2 QUANTO

The software QUANTO (`http://biostats.usc.edu/Quanto.html`) [Gauderman, 2002] is a 32-bit Windows application for the computation of detection power for various association study models like single-locus, gene-environment-, or gene-gene interactions. It includes different study designs (for example, case-control or case-only) and can be used over a graphical user interface [Gauderman, 2002]. Study power for statistical interaction analyses can be estimated using the allele frequencies, population size, single SNP effect size, interaction effect size, and genome-wide significance level (in our case: $p = 1.61 \times 10^{-12}$).

2.3.3 Principles of Graphics Processing Unit computing

The massive data production, due to the impressively fast progress in biotechnology confronts computer experts with the challenges of data handling, memory usage, and computational power. While single locus association studies have become manageable with standard computer resources, such as high-performing single desktops, multi-core processor servers or cloud-computing, higher order (interaction) calculations remain a huge performance problem; another issue is the outsourcing of clinical data, due to security standards the use of cloud computing is often prohibited.
In quest of new strategies addressing these limitations the idea arose to leverage the power of the multiple cores available on GPUs on graphics cards to enhance computation speed arose. The high-performance ability of graphic cards are established in game consoles, but were barely taken into account as computing processors as a result of insufficient implementation abilities and expertise of GPUs as arithmetic units
Nowadays high performance computing solutions are ubiquitous. A new approach is heading towards heterogeneous systems where the central processing unit (CPU) of the system handles the serial part of a task and helps to coordinate the parallel environment on GPU for massive parallel computation. In this model CPU and GPU work in a heterogeneous co-processing computing model, where the CPU carry over the sequential part of the task and the GPU the computationally intensive part in a Single Instruction Multiple Data (SIMD) manner [Kam-Thong et al., 2012].
In the field of bioinformatics and biostatistics the demand for higher performance solutions increases, amongst others when it comes to high dimensional data analysis, such as the systematically brute force search for epistatic interactions between genetic variants. Meanwhile GPU programming is constantly growing in biosciences and several open-source software tools are designed to perform epistasis searches on GPUs , such as SHEsisEpi [Hu et al., 2010], GBOOST [Yung et al., 2011], EpiGPU [Hemani et al., 2011], and those developed in our group EPIBLASTER [Kam-Thong et al., 2010], EPIGPUHSIC [Kam-Thong et al., 2011] and GLIDE [Kam-Thong et al., 2012].
To enable GPUs as processors there is the need of special interfaces, like NVIDIAs Compute Unified Device Architecture (CUDA) (see subsection 2.3.3.2) which is a parallel

computing architecture application facilitating GPUs to operate as co-processors within a host computer. Each GPU has its own memory and processing elements that are separate from the host computer. The following subsections (2.3.3.1, 2.3.3.2) will give an overview of the implemented soft- and hardware of the epistasis tools developed in the *research group Statistical Genetics* at the *Max-Planck-Institute of Psychiatry*.

2.3.3.1 Graphics Processing Units

GPUs on graphics cards are composed of several hundred Arithmetic Logic Units (ALUs) providing the massive parallel environment in which an exhaustive search can be computed extremely time-efficient. Apart from the advantages concerning computational time, the economical aspects, as the lower price, lower power consumption and at least the less rack space requirement as for multi-core CPU machines, made GPU programming more popular and operative.
For our research question, we chose NVIDIAs CUDA computation architecture, available for application development with the graphics cards series GeForce, ION Quadro and Tesla. The consumer level GeForce GTX series fulfilled all our needs and allowed us to build a queue system. We opted for custom-built compute nodes based on a high-performance PC that accepts three GTX graphics cards (GTX580, offering 512 GPU cores) for GPU computations [Pütz et al., 2013]. Overall we could built a GPU-cluster by assembling four high-performance PCs with three graphics cards each.
The working mechanism of GPUs can be visualized well by figure 2.3, in general processing elements in GPUs are called threads that are grouped into blocks. Threads within the same block share a small amount of memory that can be used to share intermediate results among the threads. Hence a thread can be seen as one computation task for its own; for example in our instance one SNP-pair interaction calculation is assigned to one thread. As GPUs have their own memory, the CPU just need to calls GPU cores and immediately release control so the GPU works on its own. This allows for computations on the CPU and GPU to take place separately. As mentioned above GPU device has to be enabled for programing through an application programming interface (API). The most direct access to hardware utilization is offered by modules or libraries as the R package *gputools* [Buckner et al., 2010], which were deployed for EPIBLASTER and EPIGPUHSIC. In contrast, GLIDE the method applied in this thesis, is based on C programming language, explicitly CUDA the an extension to C (see 2.3.3.2).

2.3.3.2 Compute Unified Device Architecture

NVIDIAs CUDA C programming language is an extension to the C language specifically designed to facilitate not only CPU but also GPU computing. It can be seen as a C dialect with extensions for GPU organization and data exchange between host computer and integrated GPU. Code targeted for the GPU is developed in so-called kernels to perform the computations. In summary the duty of CUDA is the distribution of data

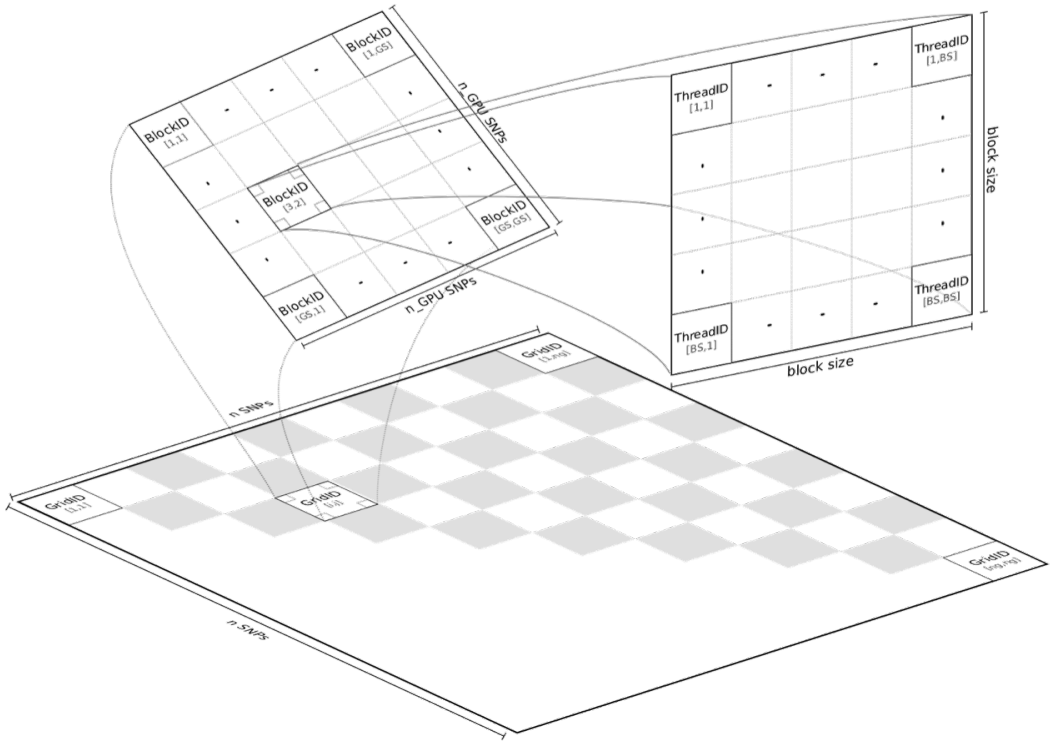


Figure 2.3: GPU threads cooperation. The epistatic interaction matrix of size $n \times n$ is divided into chunks; due to matrix symmetry only the greyed triangle needs to be computed. Chunks are divided into blocks of size $BS \times BS$. Each of those is computed in parallel by $BS \times BS$ threads [Kam-Thong et al., 2012].

from global (host memory) to shared memory on GPU, to allocate a task into sub-tasks, to process computations on multiple threads in parallel and to transfer results back to host memory [Kam-Thong et al., 2012].

2.3.4 Statistical Epistasis Tools

To overcome the computationally infeasible exhaustive search for the statistical effect of loci pairs on a phenotype we took benefit of the parallelizability of GPUs . By utilizing the computational power of GPUs, the detection of genetical interactions via regression models on a single desktop machine is enabled. Using the gain of efficiency our group has successfully developed three tools **EPIBLASTER** (2.3.4.1), **EPIGPUHSIC** (2.3.4.3) and **GLIDE** (2.3.4.5) to solve the computational demands for statistical epistasis detection.
As described in the previous section 2.3.3 the essential operating codes helps to unlock the parallel computational power of the GPU and process the statistical association of all possible bivariate interactions [Kam-Thong et al., 2012]. All three applications complete a hypothesis-free whole-genome brute force exhaustive search in less then a day on a single GPU, without imposing significant marginal effects of the single loci

involved.
In genome-wide studies the number of genetic variants can be in the order of millions, resulting in $10^{12} - 10^{14}$ possible SNP-pair combinations, which is the major cause why most epistasis tools are limited to a selection of single effect loci or loci being validated through experimental knowledge or biological pathways [Cordell, 2009].
The schematic for data preprocessing, organization and management behind all three applications are identical.

The preprocessing phase includes the appropriate **genotypic representation** as a numerical value ranging from 0 to 2 based on the count of reference allele of the selected SNP for an allele dosage model. In the context of a diallelic locus in diploid cells only three genotypes can be present *AA*, *Aa* or *aa*, with *a* being the minor, or rather the reference allele. In that circumstance the categorial values *0* would be assigned to the homozygous variant with the two major alleles *AA*, *1* to the heterozygous variant *Aa* and the carrier of the homozygous variant with two risk alleles would have the value *2*. This assumes that the reference allele affects the phenotype in an overall linearly increasing or decreasing manner, where the parameter will simply take on a negative value to express a negative linear correlation [Kam-Thong et al., 2012]. Furthermore recessive, dominant and heterozygous encodings can be applied with the assumption that the genotype effect is attributed to this conditions. Alternatively the genotype can yield any continuous numerical value of these estimated probabilities generated by an imputation algorithm.

Organization and management is the first instance of the main applications. This stage comprises the generation of a matrix to store the information of all n SNPs as column vectors and the measured quantitative values (phenotype), respectively collected disease status of the subjects along the rows. The dataset with n SNPs is then partitioned into chunks of SNPs, from where each of the chunks can be distributed across all GPU threads to be processed in parallel in an auto- and cross-correlation manner. Auto-correlation describes the analysis of a partition with itself, and cross-correlation the correlation analysis between two distinct SNP-sets. Departing from this point the subsequent mapping computation differ in their theoretical and practical implementation.

2.3.4.1 EPIBLASTER – A correlation-based interaction method for binary phenotypes

EPIBLASTER [Kam-Thong et al., 2010] was the first of three open-source multidimensional interaction tools developed on GPUs for binary phenotypes, like the case-control status of a cohort. The development strategy was to approximate the logistic regression model by applying Pearson's correlation in R (http://www.R-project.org).

More precisely, the tool relies on a two-stage approach whereby the first stage is the filtering phase calculating the difference of Pearson's correlation coefficients by performing an exhaustive two-locus interaction multiplicative search [Marchini et al., 2005] across all possible pairwise SNP combinations. The second stage performs a likelihood ratio test to compute the p-values on the estimated interaction coefficients of a subset of pairs indicating significance in the previous stage [Kam-Thong et al., 2010]. The reasoning for that methodology is that SNP pairs with the largest difference in correlation between affected and unaffected probands are most likely to show epistatic interaction, which is appealing from a biological standpoint; it ties up the concepts of epistasis with the evolutionary concept of co-selection of unlinked loci [Kam-Thong et al., 2010].

2.3.4.2 Algorithm and Implementation of EPIBLASTER

The algorithm computes the genotype vectors in pairs and the correlation coefficients are calculated for affected and unaffected subjects separately. The points of the bivariate model for diallelic loci lie on a 3×3 genotype matrix where the correlation coefficient is recorded. The difference between the correlation coefficient (Δp) in cases and controls is then used as an indication of the SNP pair contributing to the variation between affected and unaffected Eq. 2.7.

$$\begin{aligned} \Delta p = & \sum_{i \in \text{cases}} \frac{(SNP_{A_i} - \bar{SNP}_{A_{cases}})(SNP_{B_i} - \bar{SNP}_{B_{cases}})}{(n_1 - 1)\sigma_{SNPA_{cases}}\sigma_{SNPB_{cases}}} \\ & - \sum_{j \in \text{controls}} \frac{(SNP_{A_j} - \bar{SNP}_{A_{controls}})(SNP_{B_j} - \bar{SNP}_{B_{controls}})}{(n_0 - 1)\sigma_{SNPA_{controls}}\sigma_{SNPB_{controls}}} \\ = & \frac{1}{n_1 - 1} \sum_{i \in \text{cases}} \tilde{SNP}_{A_i}\tilde{SNP}_{B_i} - \frac{1}{n_0 - 1} \sum_{j \in \text{controls}} \tilde{SNP}_{A_j}\tilde{SNP}_{B_j} \end{aligned} \tag{2.7}$$

$\bar{SNP}_{Acases,controls}$ and $\bar{SNP}_{Bcases,controls}$ are the means of SNP_A and SNP_B at each loci for affected and unaffected, respectively; $\tilde{SNP}_A$ and $\tilde{SNP}_B$ represent the centered and scaled SNPs A and B for each subject group; and σ denotes the variance with n being the respective count of cases (n_1) and controls (n_0).
To utilize the full power of GPU resources a task requires to be partitioned into many smaller tasks that can be distributed in parallel. With our conditions and with respect to current GPU memory disposability, SNP-sets have to be partitioned into blocks containing 1000 or 2000 SNPs each. A two-level nested loop process runs, for cases and controls separately, through the entire data set and calculates the correlation coefficients in blocks of 2000 SNPs in an auto- and cross-correlation manner. Subsequently the difference of correlation coefficients between cases and controls and the p-values of each SNP-pair given that the distribution of the differences follows a normal distribution

[Gretton et al., 2007] are calculated. Result files store all SNP-pairs below the determined p-value threshold. The procedure is repeated across all partitions until all SNP-pairs are computed. As a function of the predetermined significance level, results of the first algorithm phase indicate consisting significant correlations for interaction terms. The second and final stage computes for those resulting SNP-pairs the fit using a full rank logistic model (Eq. 2.8).

$$Phenotype = \mu_i + \alpha x_{A_i} + \beta x_{B_i} + \gamma x_{A_i} x_{B_i} \tag{2.8}$$

The equation includes the intercept μ_i, the additive marginal effects of locus A (αx_{A_i}) and locus B (βx_{B_i}) and the interaction term of AB ($\gamma x_{A_i} x_{B_i}$).
Implementation of the algorithm on GPUs were realized via the R package *gputools* [Buckner et al., 2010] providing drop-in replacements for standard functions that make use of available GPU resource. The function 'gpuCor' allows the calculation of correlation coefficients for all possible pairwise interactions with dichotomous phenotype across vectors using the CUDA-enabled NVIDIA graphic cards [Kam-Thong et al., 2010].

2.3.4.3 EPIGPUHSIC - Correlation based interaction method for quantitative phenotypes

Nature has a disposition to continuos and complex traits, such as the popular example of human height. Genetic predispositions are results of combinations of genetic variation with small effects, the effect of a loci on a quantitative trait a so called QTL is currently the main subject of genetic research. To be able to capture the phenomena of loci interactions with the epistasis approach we had to overcome the limitation of EPIBLASTER on qualitative phenotypes.
The previous method, where association relied on the difference between the classes of subjects, is not applicable to continuos phenotypes. A time-efficient alternative approach to be applied in R on GPUs was the method developed by Gretton et al. [Gretton et al., 2005] derived from the Hilbert-Schmidt Independence Criterion (HSIC). The HSIC is a kernel-based method and an approximation to the linear regression model. The advantages of the HSIC approach, aside from the time-efficiency, is the applicability for binary and quantitative phenotypes in one tool. Subsection 2.3.4.4 will provide a broad outline of the complex statistics of the HSIC implementation, detailed mathematics of the model can be obtained from Gretton et al. [Gretton et al., 2005].

2.3.4.4 Algorithm and Implementation of EPIGPUHSIC

The development of **EPIGPUHSIC** [Kam-Thong et al., 2011] is highly related to EPIBLASTER. The implementation is likewise realized in R with the *gputools* package

[Buckner et al., 2010]. A variation is the compression of the genotyping data, which requires the popular R package, *GenABEL* (`http://cran.r-project.org/web/packages/GenABEL/`) [Aulchenko et al., 2007]. Each element is represented by the dosage model encoding $(0, 1, 2)$. When elements are recorded in floating-point numbers as in the case genotypic probabilities, for example imputed SNPs), this step is bypassed by partitioning data into smaller size files where the local memory limitation would not be of a constraint.
The principle of the HSIC is the measurement of independence of two random variables, where the squared correlation coefficient between two variables x and and y from the domains X and Y is computed in feature spaces $\mathcal{F}$ and $\mathcal{G}$ [Gretton et al., 2005].
In case of **qualitative phenotypes** the approach relies again on differences of correlation coefficients between two classes as an instance of HSIC (Eq. 2.9).

$$\begin{aligned} HSIC_{empirical}((X,Y),\mathcal{F},\mathcal{G}) &\propto \sum_{i,j} k(x_i,x_j)l(y_i,y_j) \\ &= \sum_{i,j} \tilde{x}_{Ai}\tilde{x}_{Bi}\tilde{x}_{Aj}\tilde{x}_{Bj}\psi(y_i)\psi(y_j) \\ &= \left(\sum_i \tilde{x}_{Ai}\tilde{x}_{Bi}\psi(y_i)\right)^2 \\ &= \left(\frac{1}{n_1}\sum_{i:y_i=1}\tilde{x}_{Ai}\tilde{x}_{Bi} - \frac{1}{n_0}\sum_{i:y_i=0}\tilde{x}_{Ai}\tilde{x}_{Bi}\right)^2 \\ &= \Delta\rho\left(X^{(A,B)},Y\right) \end{aligned} \tag{2.9}$$

Kernel k of space $\mathcal{F}$ and kernel l of space $\mathcal{G}$ are defined as:

$$\begin{aligned} k(x_i,x_j) = \phi(x_i)\phi(x_j) = \tilde{x}_{Ai}\tilde{x}_{Bi}\tilde{x}_{Aj}\tilde{x}_{Bj} \\ and \\ l(y_i,y_j) = \psi(y_i)\psi(y_j) \end{aligned} \tag{2.10}$$

$\psi(y_i)$ describes the state of the phenotype:

$$\psi(y_i) = \begin{cases} 1/n_1, & \text{if } y_i = 1 \\ -1/n_0, & \text{if } y_i = 0 \end{cases} \tag{2.11}$$

In the event of **quantitative phenotypes**, where a phenotype is an element of the real numbers $y_i \in \mathbb{R}$ the centered linear kernel $\psi(y_i) := \tilde{y}_i$ is chosen as kernel on the centered phenotype $\tilde{y}_i$. l is defined as kernel on continuos phenotypes with $l(y_i, y_j) = \tilde{y}_i\tilde{y}_j$ and k as a kernel on SNP pairs *(A,B)*:

$$HSIC_{empirical}((X,Y),\mathcal{F},\mathcal{G}) \propto \sum_{i,j} \tilde{x}_{Ai}\tilde{x}_{Bi}\tilde{x}_{Aj}\tilde{x}_{Bj}\psi(y_i)\psi(y_j) \tag{2.12}$$

For the implementation and a runtime of $O(m^2n)$, with n being the number of subjects and m the number of SNPs, the equation is rewritten as:

$$\begin{aligned} HSIC_{empirical}((X,Y),\mathcal{F},\mathcal{G}) &\propto \left(\sum_i \tilde{x}_{Ai}\tilde{x}_{Bi}\psi(y_i)\right)^2 \\ &= \left(\sum_i \tilde{x}_{Ai}\tilde{x}_{Bi}\tilde{y}_i)\right)^2 \end{aligned} \tag{2.13}$$

The HSIC algorithm is an approximation to the linear regression coefficients that are often used in statistical genetics to quantify the impact of variables on the phenotype [Kam-Thong et al., 2011]. Detailed information of the algorithm and the relationship between HSIC and linear regression is described in Kam-Thong et al. [Kam-Thong et al., 2011].

2.3.4.5 GLIDE - Linear regression based method for epistasis detection

GLIDE (GPU-based Linear regression for Detection of Epistasis) [Kam-Thong et al., 2012] is our most straightforward GPU -based implementation for interaction analysis. Interaction coefficients are systematically computed via linear regression in an exhaustive genome-wide brute force search in hundreds to thousands times faster than state-of-the-art implementations on CPUs.
Addressing all the limitations of previous GPU-based methods with the expectation to reveal part of the famous missing heritability, GLIDE is developed to perform genome-wide linear regression analysis without the need of data-pruning, or any limitations to discrete pheno- or genotypes. Hemani et al. [Hemani et al., 2011] released a GPU-based exhaustive search method applicable to continuos measures by calculating a F-test for the pairwise combinations. Despite its speed performance, the application is limited to SNP-pair combinations with a 3×3 contingency table of possible genotype combinations. In this context information of the SNP-pair combination can only fall into 9 possible classes and is not applicable to real-number input values, such as imputed genotype probabilities.
Therefore, our method aims to be general enough to be applicable to pairwise epistasis studies of various real or continuous predictor inputs (genetic and environmental factors) related to the phenotypic output [Kam-Thong et al., 2012]. This extension offers a huge spectrum of new opportunities and facilitates room for new innovative approaches in the field of biosciences. A very noteworthy example, due to the progress in the field

of methylation chip arrays, is the interaction detection between genetic marker and the methylation degree on the individual expression intensity as the phenotypic outcome. As generally known phenotypic variation is not purely driven by the individual genetic disposition. Environmental factors and biological processes are significantly involved in the measurable phenotypic variation. With tools like GLIDE any kind of biologically relevant data can be considered for possible interactions.

2.3.4.6 Algorithm and Implementation of GLIDE

GLIDE follows the same input and output criterion as the previously described epistasis tools, regarding SNP partition, GPU task distribution events and the assembly into result files. The implementation is the first linear regression method ported on the GPU to perform statistical measures on the bias (a bias term is required to remove any constant offset of the quantitative measure that does not embed any information which can be modeled by the independent/genetic variable(s)), univariate and interaction parameters of the basic epistasis model. The program allows for data in the continuous space for both, the phenotype (quantitative traits) and genotype (imputed SNPs) [Kam-Thong et al., 2012].
GLIDE is encoded in the C programming language using NVIDIA's CUDA extension. The epistasis computation relies on the standard linear regression model

$$\alpha^{ij} = (X^{ijT} X^{ij})^{-1} X^{ijT} y$$

where for each SNP pair a length-four coefficient vector α^{ij} (in total $\frac{SNP(SNP-1)}{2}$) such that:

$$X^{ij}\alpha^{ij} \approx y$$

has to be estimated. X is a $m \times n$ matrix, with m being the number of subjects and n the number of SNPs; let y be the $m \times 1$ phenotype vector. Given all this we want to discover the correlations between SNP pairs and the phenotype.
For each SNP pair $(i, j) \in \{1, ...n\}$ define the $m \times 4$ matrix:

$$X^{ij} = \begin{bmatrix} | & | & | & | \\ 1 & x_i & x_j & x_i \circ x_j \\ | & | & | & | \end{bmatrix} \tag{2.14}$$

where x_i is the i-th column of X (i. e., the i-th SNP over all subjects) and $x_i \circ x_j$ is the element-wise product of the i-th and j-th SNP columns (SNP_A and SNP_B).
The estimated output phenotype vector based on α^{ij} is:

$$\hat{y}_{ij} = X^{ij}\alpha^{ij}$$

with a residual sum of square error (Eq. 2.15).

$$Residual_{SSE}^{ij} = \sum_{k=1}^{m} \left(y_k - \hat{y}_k^{ij} \right)^2 \tag{2.15}$$

A subsequent t-test (Eq. 2.16) with $m - 4$ degrees of freedom should determine whether the estimated interaction term is significantly different from zero.

$$\frac{\alpha_4^{ij}}{\sqrt{\frac{\text{Residual}_{\text{SSE}}^{ij}}{m-4} \times \left[\left(X^{ijT} X^{ij} \right)^{-1} \right]_{4,4}}} \tag{2.16}$$

Retained SNP-pair p-values below the predefined threshold are append to a result-file. Compared to other tools, as for example the PLINK v1.07 epistasis function *FastEpistasis*, iterating with GLIDE over all genome-wide SNP-pairs recorded a speed-up factor of approximately 250. One would need to use a cluster of 250 CPUs to compute genetic interactions with FastEpistasis in the same amount of time as required by GLIDE on a single desktop GPU [Kam-Thong et al., 2012].
All epistasis interaction analyses and outcomes of this thesis are computed with GLIDE.

2.3.5 MAGENTA

MAGENTA is a pathway tool for enrichment analysis of genetic associations in predefined biological processes or sets of functionally related genes, using genome-wide genetic data as input. It provides pathways from several public databases as like PANTHER, Gene ontology (GO), KEGG or REACTOME and can be used to perform gene set enrichment analysis (GSEA). A table with variants associated p-values and their chromosome positions is required as input taken from a GWAS. Main result output is a nominal GSEA p-value and a FDR for each tested gene-set.SNPs not located directly in gene sequences are assigned to the closest gene estimated by a specific algorithm validating the assignment, details can be obtained in Segre et al. [Segrè et al., 2010].

2.3.6 Ariadne Pathway Studio

Pathway Studio® v9 is another pathway and ontology mapping software application developed for analysis of biological pathways, gene regulation networks and protein interaction. The included database of molecular networks is automatically assembled from scientific abstracts. It contains more than a million entries for regulation, interaction and modification between proteins, cell processes and small molecules. The database has been compiled by the application of the text-mining tool MedScan to the whole PubMed [Nikitin et al., 2003].

2.3.7 ENCODE

The Encyclopedia of DNA Elements (ENCODE) consortium is an international collaboration of research groups with the effort to build a comprehensive list of functional elements in the human genome, including elements at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active [Birney et al., 2007].
ENCODE data are available for the entire human and mouse genome on the University of California, Santa Cruz (UCSC) platform for download, search and visualization by specific tracks (`https://genome.ucsc.edu/ENCODE/`).
Assays and methods to identify functional elements were accomplished by sequencing RNA from a diverse range of sources, comparative genomics and integrative bioinformatic methods. Regulatory elements in various cell types were investigated through DNAse hypersensitive site (HS), DNA methylation and chromatin immunoprecipitation (ChIP) followed by high-throughput sequencing (ChIP-seq) (figure 2.4) [Rosenbloom et al., 2013, Birney et al., 2007]. Of interest for our SNP lookup were the tracks displaying

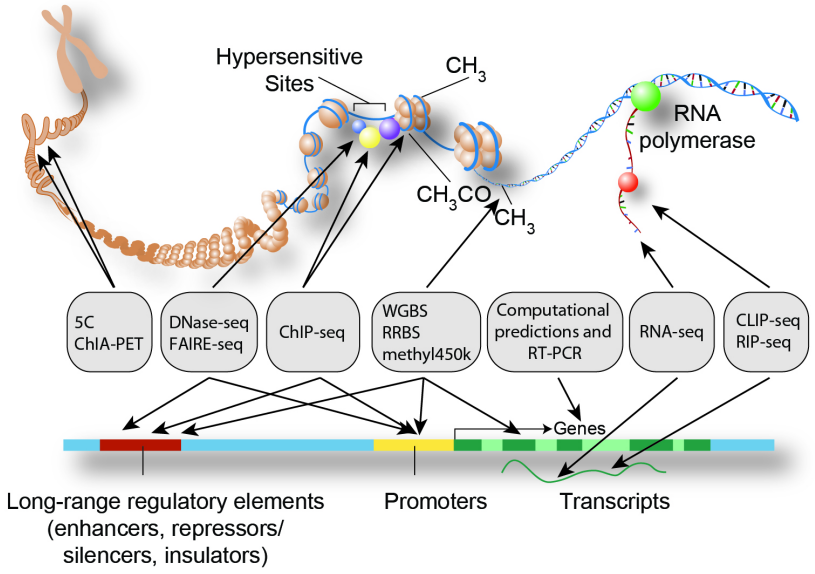


Figure 2.4: Functional elements identified by ENCODE Project Consortium with appropriate methods annotated in the grey rectangles. credits: http://genome.ucsc.edu/ENCODE/aboutScaleup.html

evidence of open chromatin segments, which are a sequences of accessible DNA to active regulatory elements like enzymes and molecules. This segments were identified using different techniques, such as DNaseI HS, Formaldehyde Assisted Isolation of Regulatory Elements (FAIRE), and ChIP for select regulatory factors. DNaseI HS and FAIRE provide assay cross-validation with commonly identified regions delineating the highest confidence areas of open chromatin. ChIP assays provide functional validation and preliminary annotation of a subset of open chromatin sites in different cell types

[Ho and Crabtree, 2010, Rosenbloom et al., 2013]. Further transcription factor (TF) binding tracks display specific TF binding sites identified by by ChIP-seq in chosen cells. Open chromatin locations and TF binding sites are indicated on the UCSC genome browser by either signals or peaks.
Raw Signals are a continuous signals that indicate density of aligned reads. The sequence reads were extended to the size-selected length (225 bp), and the read density computed as reads per million [Rosenbloom et al., 2013].
Peaks are sites with the greatest evidence of transcription factor binding, calculated using the MACS peak caller [Zhang et al., 2008], as enriched regions of high read density in the ChIP experiment relative to total input chromatin control reads.

3 Materials and Methods

3.1 Materials

3.1.1 Proband ascertainment

Four independent case-collections comprising individuals of German, British, French and American descent built the admixture sample for the study of genetic interactions. All participating individuals, school aged children with potential difficulties in reading or writing and who had been diagnosed with dyslexia, were referred to the investigators by parents, teachers, special educators or practitioners. Ethical approval was obtained from each Ethics Committees in the participating countries [Fisher et al., 2002, Francks et al., 2004, Ludwig et al., 2010, Schulte-Körne et al., 2007, Schumacher et al., 2005]. Collectively 862 (566 males, 296 females) probands passed all appropriate QCs and the subsequent genomic scan (figure 3.1). The sample composition can be broken down into the following cohorts:

Ascertainment of German sample German families with at least one affected child with RD were recruited at the Departments of Child and Adolescent Psychiatry and Psychotherapy at the Philipps University in Marburg and at the Julius-Maximilian University in Würzburg. To avoid any bias or misclassification subjects indicating any symptoms of ADHD, inattention and hyperactivity, bilingual education, IQ < 85, an uncorrected disorder of peripheral hearing or vision, a psychiatric or neurological disorder were excluded [Schulte-Körne et al., 2007]. For the diagnosis of dyslexia and as inclusion criterion a grade appropriate spelling (SP) test was conducted, resulting in a score using the T distribution of the general population. It is assumed that there is a correlation between IQ and SP of 0.4 [Schulte-Körne et al., 2001], based on this assumption an expected SP score can be calculated for each proband. A child was classified as affected, if the score showed at least a discrepancy of one standard deviation (SD) ($\geq$ 1 SD) between expected and observed score. Those who were classified as affected underwent a battery of psychometric tests [Schulte-Körne et al., 2007] (details in section 3.1.2). Overall 200 classified cases with a SD ≥ 1.25 were subjected to our study. Whole-genome SNP genotyping was performed at the LifeBrain Centre in Bonn (Germany) using the Illumina HumanHap 300K BeadChip (Illumina, Inc., San Diego, CA, USA) according to the manufacturer's standard protocols. After exclusion of probands violating the standard criteria for genomic analyses 199 (149 males, 50

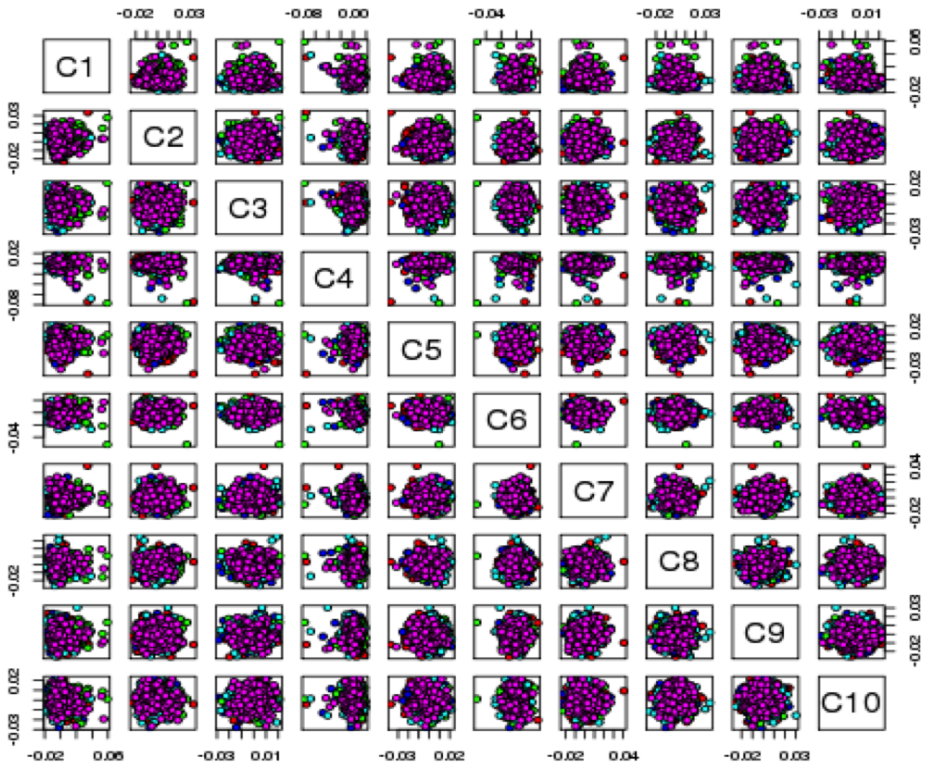


Figure 3.1: First ten computed MDS components of the combined sample. German individuals are indicated in red, British in blue and cyan, French in green and American samples are illustrated in magenta. No distinct clusters were detectable.

females) subjects of German origin with an age ranging between 8 to 19 (mean $\approx$ 12.5) were factored into the study collection.

Ascertainment of the French sample French samples were ascertained at the Laboratoire de Science Cognitives et Psycholinguistique in Paris, France. French probands underwent an age-appropriate reading test and were included as cases if their reading score were more than 1.25*SDs* below grade level on a standardized test of SWR, and if none of the exclusion criterions mentioned above (German sample) were applicable. Selected cases underwent then further psychometric tests.
Overall 92 (64 males, 28 females) participants classified as affected were assigned in our study. Children's age ranged at time of assessment between 8 and 12 (mean 10). All 92 withstood further filtering procedures regarding population stratification, individual call rate, a minimum IQ of 85.Whole-genome SNP genotyping was likewise conducted at the LifeBrain Centre in Bonn (Germany) using the Illumina HumanHap 660K BeadChip.

Ascertainment of UK sample The individuals of British origin were collected at the Royal Berkshire Hospital in Reading, UK, from unrelated nuclear families. Over 90% of

the original subjects were recruited on the basis of having SWR deficiency more than $2SD$ below that predicted by tests of verbal or non-verbal reasoning and evidence of RD in one or more siblings with dyslexia. Another part were required to have SWR ability with $\geq$ 1 SD below that predicted for their age and IQ $\geq$ 90. Psychometric tests were administered to all UK probands and siblings, regardless of recruitment [Francks et al., 2004]. With regard to the criteria of $\geq$ 1.25 SD below that predicted for individual age and a IQ $\geq$ 85, and all appropriate QCs we included overall 377 (266 male, 111 female) children with UK origin in our analyses. Sample age were at the time of assembly between 8 and 13 (mean 11). Whole-genome SNP genotyping was performed at the Wellcome Trust Centre for Human Genetics in Oxford (UK) using the Illumina HumanHap 550K BeadChip.

Ascertainment of US sample We comprised 194 (87 males, 107 females) unrelated probands from the twin study of RD from the Colorado Learning Disabilities Research Center (CLDRC) in Boulder, US [Compton et al., 2001, Francks et al., 2004]. Probands age at assembly ranged between 8 and 18 with a mean of 11.3. Classification of the participants were done by measuring word recognition performance on the Peabody Individual Achievement Test and a time-limited (two seconds) word recognition test designed by Olson et al. [Olson et al., 1999]. Additional sampling criteria took place similar to the remaining cohorts, like an IQ $\geq$ 85 and the exclusion of individuals with any neurological- or sensory-deficits.

3.1.2 Endophenotype ascertainment

All probands underwent a battery of age appropriate psychometric tests with continuous outcomes. The *t*- or *z*-scores were age-adjusted and standardized against a normative control data set. Related tasks were conducted to measure cognitive abilities that have been found to be correlated with the core symptoms of dyslexia and to be components of the reading process [Compton et al., 2001, Fisher and DeFries, 2002, Francks et al., 2004, Schulte-Körne et al., 2007]. None of the tests were administered to parents or teachers. Next subsections give a brief overview of the assorted endophenotypes that were available through all samples.

3.1.2.1 Single-word reading

The single-word reading test is a straightforward test with slight country based differences. Experiment participants have to read a list of language- and age-appropriate real words aloud as accurately and quickly as possible in a given time-limit. Performance is measured age adjusted and standardized against a given control group. Detailed population based task procedures can be gathered in the corresponding publications [Compton et al., 2001, Fisher et al., 2002, Francks et al., 2004, Olson et al., 1999, Schulte-Körne et al., 2007].

3.1.2.2 Spelling

Spelling is measured using an age-appropriate writing to dictation test, items were presented orally and had to be written down correctly. As mentioned above the SP score is calculated on the basis of the correlation between individual IQ and expected SP skills [Compton et al., 2001, Francks et al., 2004, Schulte-Körne et al., 2007].

3.1.2.3 Phonological awareness

Phonological-awareness (PA) or Phoneme awareness is a specific cognitive competence that is defined as *the capacity to reflect explicitly on the individual speech sounds that make up a phoneme* [Gayán and Olson, 2003]. It requires the ability to extract and manipulate the single words of full sentences. Different examples of word manipulation tasks can be conducted during a test, for example: probands have to add a sound to a given word, or they have to repeat a given word as a non-word by deleting a given character (phoneme deletion). Another task is the phoneme segmentation where a pseudo-word is partitioned into its phonemes. The last exercise is the phoneme reversal test in which children have to switch phonemes of a word [Compton et al., 2001, Gayán and Olson, 2003, Olson et al., 1999, Schulte-Körne et al., 2007]. Regularly all executed tasks are age specific averaged to obtain one measure for PA. It is assumed that PA is correlated with SWR skills, difficulties are recognized usually just in dyslexic children [Rack et al., 1992].

3.1.2.4 Non-word reading

Non-words are non-existing words without any meaning but can be pronounced by rules for mapping letters to speech sounds. Two different tasks can be administered to the participants: an oral or silent test, where in the oral test subjects have to read a non-word aloud and in the silent test they have to choose a word out of few that would sound like a real word if read aloud [Compton et al., 2001]. Affected individuals show problems in reading non-words compared to not reading disabled individuals, which is clearly shown in the speed of reading such non-words [Wimmer, 1996]. Either one test or both tests can be administered to the probands, whereby in the second case a composite score has to be calculated. The discrepancy between affected and not affected individuals indicate a correlation of the task to dyslexia.

3.1.3 Epilepsia sample

Biopsy samples of 148 patients with chronic pharmacoresistant temporal lobe epilepsy were collected in the Epilepsy Surgery Program at Bonn University. Epilepsy-surgery was conducted to achieve seizure control in all patient, whereby premortal human hippocampal segments were abstracted. Informed and written consent was obtained from all patients for additional studies. Procedures were carried out in accordance with

the Declaration of Helsinki and were approved by the local ethics committee.
Genotyping and profiling of hippocampal gene expression were processed on Illumina's Human660W-Quad and HumanHT-12v3, respectively.

3.2 Methods

Application of all described statistics (2.2) and tools (2.3) on the given study samples (3.1) were realized as following.

3.2.1 Data preprocessing

3.2.1.1 Dyslexia sample

Autosomal genotype data have been subjected to different QC procedures utilizing PLINK v1.07. SNPs with a MAF $< 5\%$, a deviation from HWE of $p <= 10^{-5}$ in the exact test and a SNP call rate $< 95\%$ were removed. Further an individual call rate likewise of $< 95\%$ excluded those probands whom genotyping rate violated the threshold.
Missing genotype data for German, apart of the UK and French samples were imputed via IMPUTE2 (`https://mathgen.stats.ox.ac.uk/impute/impute_v2.html`) [Howie et al., 2009] based on the reference sample 1000 Genomes Project (NCBI build 37 (hg19)). The remaining part of the UK samples and all US samples were imputed with MiniMach [Howie et al., 2012] likewise based on the 1000 Genomes reference sample (in Max-Planck-Institute for Psycholinguistics in Nijmegen, Netherlands). Accuracy threshold for the genotype prediction were set to $\geq 80\%$ for IMPUTE2 corresponding to a $R^2 >= 0.3$ for the MiniMac predictios.
Again the above described QC conditions (MAF, HWE, SNP and individual call rate) were applied to the imputed data. 3.520.975 variants withstood all QCs. To reduce time and memory costs LD-pruning with a $r^2 > 0.7$ took place. The threshold of $r^2 > 0.7$ was chosen on the expectation that SNPs with an LD greater or equal 0.7 should exhibit similar effects or a trend to an effect on the phenotype. Overall 393.802 genetic marker entered further statistical analyses. To account for population stratification, a complete linkage clustering of individuals (867) on the consistent set of 131.389 genotyped markers was performed. Outliers based on the MDS analysis on the IBS matrix and homogeneous samples with a PI_HAT (proportion identity-by-distance (IBD)) ≥ 0.0625 (corresponding to the inbreeding coefficient (IC)) were excluded. Five probands exceeding the IC were rejected and the remaining 862 subjects were referred to further statistics. For the endophenotype PA just a subset of 645 probands (414 males, 231 females) were included due to missing phenotype. The first ten MDS components served as correction factors in a regression analysis on the quantitative phenotypes. Obtained residuals conducted as phenotype equivalents for the analyses.

3.2.1.2 Epilepsia sample

Genotype QC were applied with the following parameters: a maximum deviation from HWE of $p = 10^{-5}$, a MAF of 5%, a SNP call rate $\geq$ 98%, a per sample call rate $\geq$ 98% and a FDR $\geq$ 1% for autosomal heterozygosity. Heterozygous mtDNA-linked, female Y-linked as well as male heterozygous X- and Y-linked genotypes were excluded from the analyses. In total 530.794 genetic variants and 138 individuals passed all QC criterias.
The expression probe sequences were realigned to the UCSC [Kent et al., 2002] version 18 of the human genome (hg18). Realignment of the Illumina probe sequences were conducted using Burrows-Wheeler Alignment (BWA) [Li and Durbin, 2009] allowing only perfect matches. Probes that could not be aligned or probes containing either intrinsic polymorphisms as listed in HapMap-CEU or matched to multiple positions in the human genome were excluded from the downstream analysis.
The remaining probes were pre-analyzed using the GenomeStudio v2010.1 gene expression module software (`http://www.illumina.com/informatics/sequencing-microarray-data-analysis/genomestudio.ilmn`) from Illumina to merge data of replicated samples of the same individuals as well as to identify probes with a detection p-value < 0.01 in at least 5% of all samples. Data not matching these criteria were considered as non-expressed transcripts and dropped from further analysis. The remaining 15.426 probes (accounting for 13.842 distinct transcripts) were then transformed and normalized to consider background noise using `vsn2` implemented in the R `Bioconductor` (`http://www.bioconductor.org` [Gentleman et al., 2004]) package `vsn` [Huber et al., 2002].
Furthermore a hidden confounder analysis were applied to correct for expression heterogeneity. Unknown confounding influences, such as samples history or subtle technical variation, commonly lead to expression heterogeneity, a pronounced correlation pattern within the expression profiles. A Bayesian factor analysis were used to infer the hidden determinants of this confounding expression variability with the software tool PEER (`https://github.com/PMBio/peer/wiki`) [Stegle et al., 2012]. We set the maximum number of factors to 30 and used the Gamma prior settings $p_a = 1 \times 10^{-2}$, $p_b = 1 \times 10^{-2}$ to regularize the effective number of factors used. After convergence 15 factors remained active. As proposed in Stegle et al., we calculated expression residuals, subtracting the effect of the confounding factors. The residual expression dataset was then used for all downstream analysis.

3.2.2 Study design

3.2.2.1 Association of univariate genetic variants in dyslexia

Genome-wide single locus effects were computed for each endophenotype (SWR, non-word reading (NWR), SP, PA) in the merged sample of German, French, British and

American descendants. Primarily the existence of significant univariate effects in this special cohort admixture should be explored, but particularly we were interested if any marginal effects are present in the resulting interaction SNP-pairs.
All markers were tested for univariate association to each corrected measured cognitive score in a linear regression model via PLINK v1.07. The p-value significance level after Bonferroni for multiple testing was determined to be $\leq 1.27 \times 10^{-7}$.

3.2.2.2 Association of bivariate genetic variants in dyslexia

The multi-language caucasian sample composed of German, French, British and American individuals were subjected to a two-dimensional statistical interaction analysis with the aim to reveal epistatic associations explaining possible predisposition for RD beyond single-loci effects.
Although several dyslexia susceptible loci were detected in the past, it has been shown to be very difficult to detect single genome-wide significant correlations explaining heritability of this disorder. From this perspective we anticipated to uncover a potentially more complex underlying nature of the disorder symptoms.
The procedure for bivariate analyses were applied on the imputed and QCed samples described above. The different imputed cohorts were merged using GTOOL (`http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html`). The R package `DatABEL` provides the function `impute2mach` which enable the conversion of the genotype probabilities into a single dosage value. Genotype dosage value can be any continuous decimal, ranging from 0 to 2. Homozygosity of major allele is indicated by values ~ 0, heterozygosity by values ~ 1 and finally homozygous state of minor allele with values ~ 2.
The customized data-set according to the required input standards for GLIDE [Kam-Thong et al., 2012] were referred to a systematically exhaustive search for epistatic interactions by mapping each endophenotype to all possible pairs of genetic loci. Overall 393.802 SNPs were analyzed for interaction effects independently on the endophenotypes SWR, NWR, PA and SP with $frac393.802 \times 393.8012$ tests each.
To avoid any influences on the statistical scores, due to four different populations and to capture possible heterogeneity, meta-analyses via the R package `rmeta` were subsequently applied to all interactions with $p < 1 \times 10^{-8}$, resulting in a subset of 3.274 SNP combinations over all endophenotypes.
Meta-analyses were performed with the function `meta.summaries` that computes a summary estimate from a collection of effect estimates (beta, β) and standard errors (SEs) [Lumley, 2013]. The essential values were obtained from the individual cohort-specific linear regression analyses for bivariate interactions, likewise realized in R. `meta.summaries` further allows to account for random effects, capturing heterogeneity in admixture samples.
Significance were determined with respect to the Bonferroni method with the modification for GWIAs according to Becker et al. [Becker et al., 2010] resulting in a genome-wide

significance level of $p = 1.61 \times 10^{-12}$.
To eliminate conjectures regarding genotyping or imputation inaccuracies and for validation of the interactions we repeated the whole procedure for each significant SNP-pair with all ahead excluded proxy-SNPs with a $r^2 \geq 0.7$.
Resulting candidate SNPs were subjected to a variety of functional analyses to define their biological relevance, precisely:

1. a regional based expression analysis (3.2.2.3)

2. sequence mapping on functional/regulatory ENCODE annotations (3.2.2.5)

3. and pathway analyses (3.2.2.4)

Further SNPs located in a repetitive region were tested for correct assignment on the genome via the fast online alignment tool BLAT (`http://genome.ucsc.edu/cgi-bin/hgBlat?command=start`) provided by UCSC. Affected marker plus 25 base-pair (bp) down- (3′) and upstream (5′) were mapped utilizing the tool to the human genome (hg19).
Finally to validate the methodological approach of interaction computation a Quantile-Quantile (Q-Q)-plot (figure 4.17) across 100.000 randomly sampled interactions against a standard normal distribution was constructed in R with the function `qqnorm`.

3.2.2.3 Regional based interaction in hippocampal expression

Regional based defines the restriction of genomic regions of interest, explicitly the restriction to a sequence length of 2mega base pairs (Mb) up- and downstream of each SNP retrieved as candidate locus from the perviously conducted dyslexia epistasis study.
To guarantee an equivalent SNP set and quality of the hippocampal data, SNP coordinates were initially updated using LiftOver (`http://genome.ucsc.edu/cgi-bin/hgLiftOver?hgsid=369473265_Fb9WaNxa4A5uRbG5nHTTMuIAh0Rb`) provided by UCSC, which converts genome coordinates and genome annotation files between assemblies, in our case from hg18 to hg19. Genotypes were then imputed using IMPUTE2 with identical QC conditions, regarding imputation accuracy and genomic quality, as applied to the dyslexia data (3.2.1.1).
The appropriate final dataset underwent statistical epistasis calculation with GLIDE. Four independent candidate analyses were performed by running each SNP set just against the defined interaction partner. Particularly SNPs on chromosome 9 were just tested against SNPs on chromosome 7 if there were an interaction detected in the dyslexia sample and so forth.
Bonferroni-Becker significance thresholds were estimated for each run separately due to varying SNP amount and are provided individually in the according result sections.

3.2.2.4 Pathway analyses

Pathway analyses were realized in two steps. The first step was the analysis performed via the software MAGENTA, to assign intergenic SNPs to a proper gene by the software algorithm. We took all marker that were involved in an interaction with $p \leq 1 \times 10^{-10}$ and performed an endophenotype specific GSEA. Correction for multiple testing were realized automatically by the software with FDR of 5%. Pathways or groups were declared as significant if they withstood FDR correction.
The second step was the overall pathway analysis by Pathway Studio® v9. In this stage all by MAGENTA retrieved genes were pooled together for a pathway/group enrichment analyses. Overall 16548 Pathways and groups are annotated in Pathway Studio® v9 and were tested for enrichment with the MAGENTA estimated gene-set. Significance level were determined after Bonferroni for multiple testing with $p \leq 3.1 \times 10^{-}6$.

3.2.2.5 Functional sequence mapping

An approach to validate the biological relevance of the findings was to map candidate SNPs on the annotated functional DNaseI HS and ChIP-seq estimated regulatory binding sites. For signal visualization on the UCSC Genome Browser we selected the regulation tracks for ENCODE open chromatin by DNaseI HS and FAIRE from the Duke group for the lymphoblastoid cell line (GM12878) produced from the blood of female donor with northern and western European ancestry, and male H1 human embryonic stem cells (H1-hESC 1). Further we enabled the ChIP-seq estimated binding sites for TF. Peaks for 161 transcription factors in 91 cell types are combined here into clusters to produce a summary display showing occupancy regions for each factor and motif sites within the regions when identified [Gerstein et al., 2012, Wang et al., 2012a, Wang et al., 2013]. Additionally tracks displaying binding sites for neuron-restrictive silencer factor (NRSF) and RNA Polymerase II POL2-4H8 were enabled [Fields, 2007, Rosenbloom et al., 2013]. The NRSF protein is a transcriptional repressor for neuronal genes in non-neuronal tissues and is suggested to function as a negative regulator of neurogenesis [Schoenherr and Anderson, 1995]. RNA Polymerase II estimated from neuronal cell lines derived from H1 embryonic stem cells (H1-neurons). RNA Polymerase II is an enzyme that catalyses the transcription of DNA, by binding to the promoter of a gene for initiation [Fields, 2007, Johnson et al., 2007].
All top hit SNPs that resulted as significant or with a strong trend towards significance in the epistasis studies were considered in UCSC Genome Browser (hg19/GRCh37).

4 Results

4.1 Association of univariate genetic variants in dyslexia

GWASs were conducted via PLINK v1.7 for each of the endophenotypes SWR, NWR, PA and SP with the combined sample of German, French, UK and US descendants. With respect to the Bonferroni adjusted significance level $p \leq 1.27 \times 10^{-7}$ none of the GWASs identified any significant associations. Testing each cohort individually likewise revealed no genome-wide significant hits.

4.2 Association of bivariate genetic variants in dyslexia

Four independent endophenotype (SWR, NWR, PA, SP) specific genome-wide interaction studies (GWISs) were realized utilizing the tool GLIDE [Kam-Thong et al., 2012]. With the defined GLIDE output threshold of $|t| > 5$ corresponding to $p \leq 6.95 \times 10^{-7}$ for the SWR, NWR, SP studies, and $p \leq 7.41 \times 10^{-7}$ for the PA study, according to one degree of freedom (d.f.) and sample size (862 and 645 (PA)) more than 2.36×10^5 statistical interactions exceeding the respective thresholds were obtained. Observed p-values ranged from the specified output threshold down to 6.45×10^{-13}.
To eradicate remaining uncertainties based on possible genomic heterogeneity, due to an admixed sample, meta-analyses were conducted to all SNP-pairs with $p < 1 \times 10^{-8}$. Interactions that showed at least a trend towards genome-wide significance (Bonferroni-Becker estimated significance level of $p \leq 1.61 \times 10^{-12}$) in the meta-analyses were subjected to the next analysis procedure.
In this phase we reiterated all applied steps by concentrating on the chosen candidate SNP-pairs including their previously excluded proxy SNPs with a $r^2 \geq 0.7$. This step should ensure that discovered interactions were not based on genotyping or imputation artifacts.
The GLIDE computations and the corresponding subsequent meta-analyses resulted in strong evidence for SNP-SNP interactions in each single survey (Fig. 4.1). Three (SWR, SP, PA) of the four analyses detected epistatic interactions achieving genome-wide significance ($p \leq 1.61 \times 10^{-12}$). GWIA for NWR barely failed the significance level.

Functional analyses with the aim to validate biological relevance of the statistical epistasis were realized via functional sequence mapping using ENCODE data (4.2.7), pathway (4.2.10) and expression (4.2.8) analyses.
Details of the resulted epistatic interactions are described in the according subsections for each of the associated endophenotype. Exposed p-values refer to the outcome of the conducted meta-analyses.

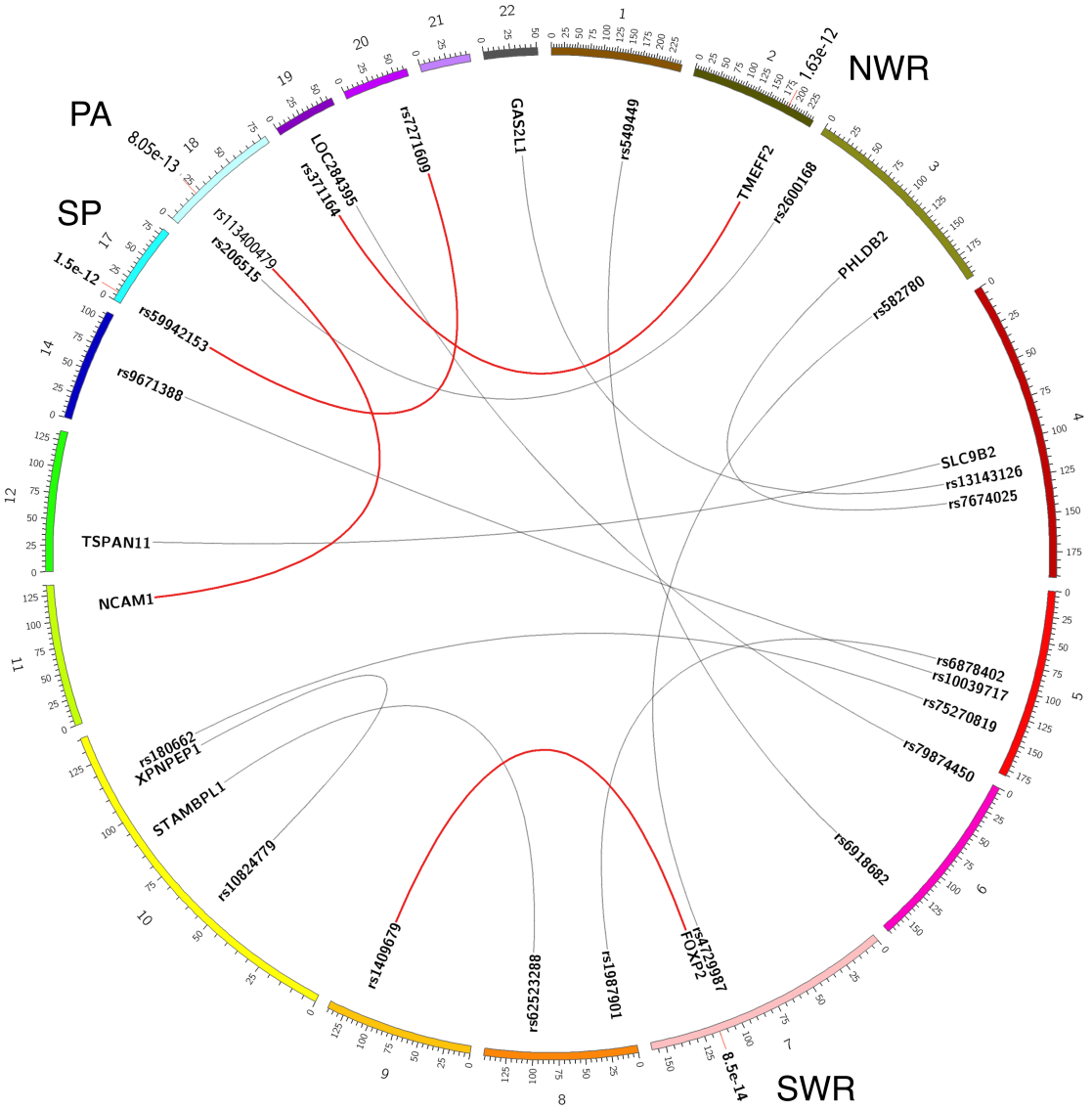


Figure 4.1: Circos diagram showing inter-chromosomal interactions detected through the endophenotype specific epistasis computations. Most significant interactions are depicted by one representative pair in red showing the chromosomal position, the corresponding SNP or gene name and the retrieved p-values. Further interactions with $p < 1 \times 10^{-10}$ are connected via black lines.

4.2.1 Single-word reading

The GWIA of SWR, one of the strongest dyslexia phenotypes, achieved the most significant result with $p \leq 8.5 \times 10^{-14}$ comprising the *FOXP2* intronic variant rs56253958 (7q31.1) and rs1409679 (9q31.3), a variant in *L1M1* a long interspersed nuclear element (LINE) region. Significant interactions or at least interactions with a trend towards significance (Fig. 4.3) were detectable over the adjacent LD-regions ($r^2 \geq 0.7$) of both loci. The SNP rs1409679 displayed a sparse LD density with just two proxy SNPs at the defined r^2 level (Fig. 4.2), whereas the *FOXP2* SNP is located in a greater LD region with 24 SNPs, demonstrating interactive effects with the 9q31.3 markers exhibiting p-values $\leq 1 \times 10^{-10}$.

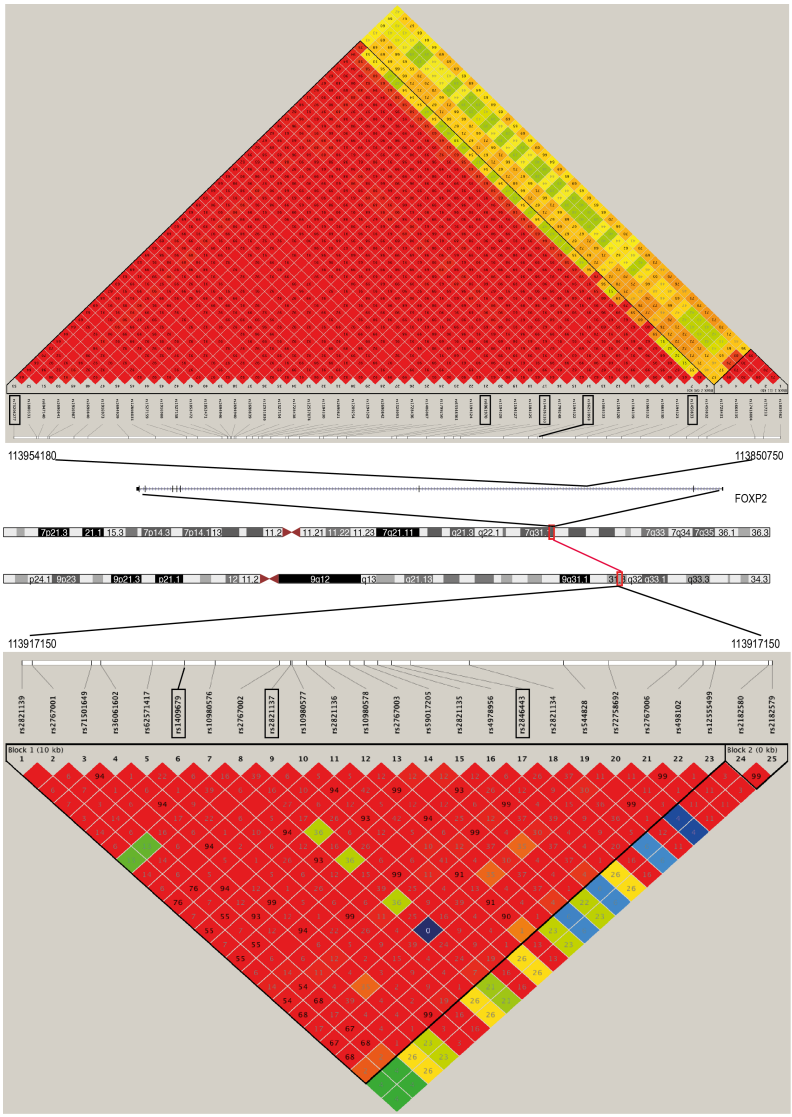


Figure 4.2: Genomic positions involved in interactions associated to single-word reading within the LD-structure of 7q31.1 and 9q31.3. Most significant interacting variants are highlighted with rectangles. Chromosomal positions of the top hits are shown on each ideogram and are connected in red. Color code indicate with red high LD and with blue low LD. *Reference: Haploview*

Effect detection and associations over the given LD-structures (Fig. 4.4) recorded association findings not being confounded by genotyping or imputation inaccuracies. Details of genome-wide significant results can be obtained in the according table 4.3.

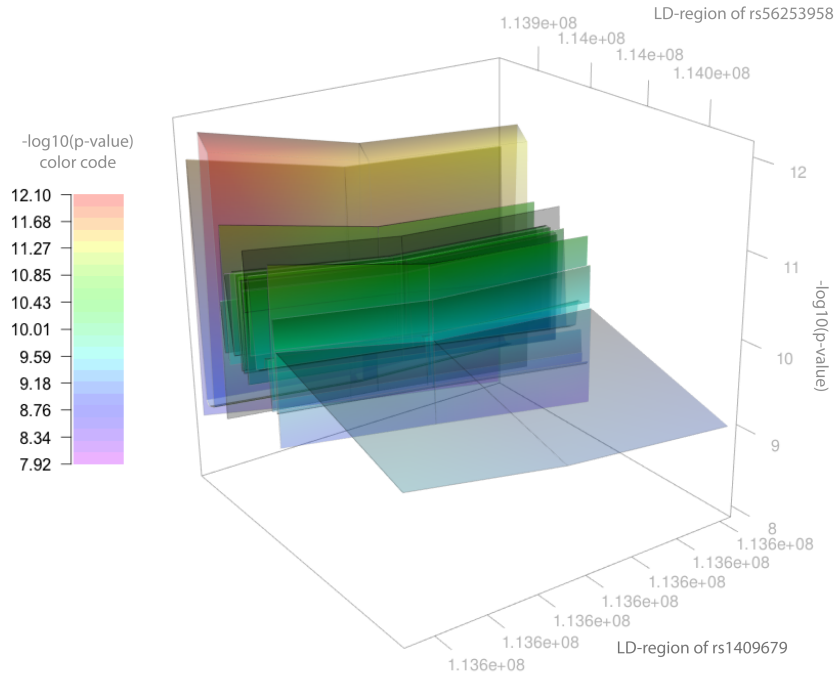


Figure 4.3: 3D plot of the 2-way associations to single-word reading over the genomic SNP regions. *LD-region SNP1* and *LD-region SNP2* depict the chromosomal positions of rs56253958 ($7q31.1$) and the interacting SNP rs1409679 ($9q31.3$) with their corresponding proxies SNP ($r^2 \geq 0.7$), respectively. The color legend indicates the detected $-\log_{10}(p)$-values.

To ensure correct assessment of rs1409679 on the genome an alignment of the variant plus 25 bp up- and downstream was performed via BLAT. The sequence alignment found $9q31.3$ as only match with 100% identity over the whole 51 nucleotides. Further full sequence uniqueness, and mappability on the reference genome (GRCh37/hg19) was given.
The interaction effect of the top hit is visualized in the boxplot (Fig. 4.4), illustrating the correlation of genotype combinations to reading scores. The correlation of reference/minor allele combination (rs56253958: C, rs1409679: T) to a decreasing reading score, exhibiting lessened reading abilities, was observed. Minor allele homozygotes for both markers displayed the weakest reading performance ("*risk*-alleles"). Corresponding contingency table for the SNP-pair genotype constellation is shown in figure 4.4. Single study effect size, a clear negative context of the effect direction (β) (Table 4.2), the

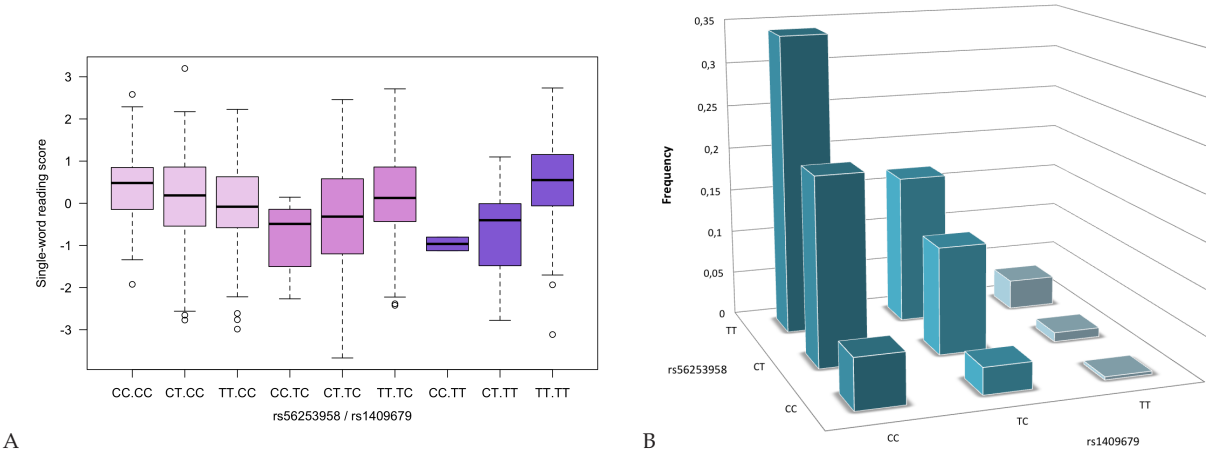


Figure 4.4: (A) Boxplot of genotypic associations through all samples with single-word reading score. Y-axis defines the range of the measured reading scores ($-3/+3$) normalized with a mean of 0 and a variance of 1. A score of -3 indicated the weakest reading performance. X-axis is the according genotypic combination of the SNPs rs56253958 (chromosome 7, minor allele C) and rs1409679 (chromosome 9, minor-allele T). The boxes define the interquartile range, the thick line the median, and single dots possible outliers. Box colors separate classes of genotype-combinations. **(B): Genotype-combination frequencies** of the SNPs rs56253958 and rs1409679 in the combined sample comprising all cohorts.

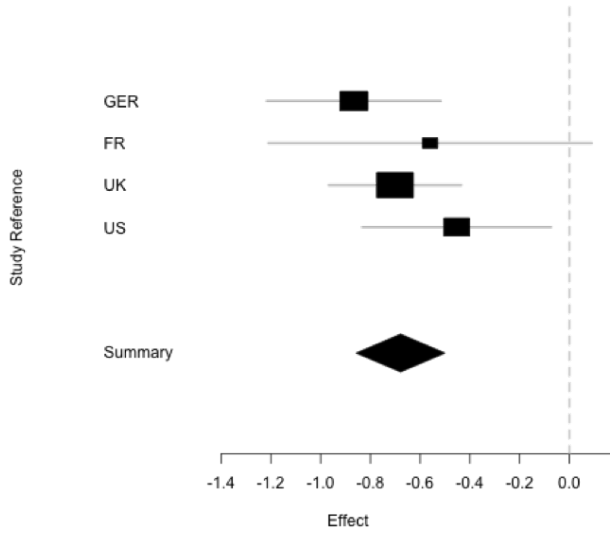


Figure 4.5: Meta-analysis of the SWR top hit. For each single cohort (German (GER), French (FR), British (UK), American (US)) the square and horizontal line show the estimated regression coefficient β and 95% confidence interval, representing the effect of each copy of the reference-allele on reading performance. The size of the square is inversely proportional to the SE of the estimated effect. Below the individual cohorts a summary diamond shows the random-effects when analyzing all four cohorts together as one single sample. Notably clear negative effect is present in all single sub-samples and in the combined sample.

confidence interval (CI), and the non-significant heterogeneity p-value of 0.454 (Table 4.3) proved homogeneity between the individual cohorts (Fig. 4.5).
As mentioned in the GWAS section no single main effects were existent, nevertheless

slightly marginal effect or rather a trend of the *FOXP2* variants for nominal association to SWR with $p \approx 1 \times 10^{-2}$ (4.1) could be observed.

SNP	Position	Allele	MAF	Pval	Beta	SE	CI_Low / CI_UP
rs56253958	7:113883531	C	0.209	1.6×10^{-2}	-0.143	0.059	-0.259 / -0.026
rs144261163	7:113893884	A	0.2115	3.1×10^{-2}	-0.129	0.06	-0.247 / -0.012
rs1450833	7:113865735	C	0.2132	7.7×10^{-3}	-0.156	0.058	-0.27 / -0.041
rs6961970	7:113901132	A	0.2139	0.107	-0.095	0.059	-0.209 / 0.02
rs73206277	7:113954180	G	0.2129	0.153	-0.084	0.058	-0.198 / 0.031
rs1409679	9:113586263	T	0.2665	0.664	-0.023	0.052	-0.125 / 0.079
rs2821137	9:113587820	A	0.2691	0.64	-0.025	0.052	-0.127 / 0.078
rs2846443	9:113589582	G	0.2661	0.798	-0.013	0.052	-0.115 / 0.089

Table 4.1: Univariate associations to single-word reading of interacting SNPs. Depicted are all SNPs involved in significant statistical interaction, or with a trend to significance. First column contain the individual "*SNP*" identifiers. Column "*Effect allele*" display the respective minor allele, followed by the corresponding frequency "*MAF*" in the sample, the calculated p-value "*Pval*" of the single association, the estimated regression coefficient "*Beta*", the standard error "*SE*", and the lower "*CI_Low*" and upper "*CI_UP*" bounds of a 95% confidence interval. All denoted SNPs on chromosome 7 are intronic variants of *FOXP2*. Variants on chromosome 9 are intergenic variants.

Cohort	SNP1	SNP2	Beta	SE	Pval
GER	rs56253958	rs1409679	-0.87	0.18	2.5×10^{-6}
FR	rs56253958	rs1409679	-0.56	0.59	9.5×10^{-2}
UK	rs56253958	rs1409679	-0.70	0.13	4.3×10^{-7}
US	rs56253958	rs1409679	-0.45	0.19	2.0×10^{-2}

Table 4.2: Cohort specific interactions of the single-word reading top hit. First column "*Cohort*" denote the corresponding cohort followed by the detected interacting variants under "*SNP1*" and "*SNP2*". Next column "*Beta*" represent the regression coefficient, followed by the standard error "*SE*" and the resulting p-value "*Pval*" for the interacting SNP-pair in each individual cohort.

SNP1	SNP2	Beta	SE	MPval	CI_Low	CI_UP	Het	GPval	Chr:Pos1	Chr:Pos2
rs56253958	rs1409679	-0.679	0.091	8.49×10^{-14}	-0.857	-0.5	0.454	7.9×10^{-13}	7:113883531	9:113586263
rs144261163	rs1409679	-0.673	0.091	1.51×10^{-13}	-0.852	-0.494	0.393	1.18×10^{-12}	7:113893884	9:113586263
rs56253958	rs2821137	-0.672	0.091	1.92×10^{-13}	-0.851	-0.493	0.396	2.36×10^{-12}	7:113883531	9:113587820
rs1450833	rs1409679	-0.66	0.09	2.23×10^{-13}	-0.836	-0.484	0.498	1.9×10^{-12}	7:113865735	9:113586263
rs56253958	rs2846443	-0.667	0.092	3.59×10^{-13}	-0.846	-0.487	0.388	3.50×10^{-12}	7:113883531	9:113589582
rs1450833	rs2821137	-0.653	0.09	4.98×10^{-13}	-0.83	-0.476	0.439	5.43×10^{-12}	7:113865735	9:113587820
rs6961970	rs1409679	-0.637	0.088	5.94×10^{-13}	-0.81	-0.463	0.68	6.73×10^{-12}	7:113901132	9:113586263
rs1450833	rs2846443	-0.648	0.09	6.85×10^{-13}	-0.825	-0.471	0.429	7.88×10^{-12}	7:113865735	9:113589582
rs73206277	rs1409679	-0.634	0.089	7.95×10^{-13}	-0.808	-0.461	0.562	9.14×10^{-12}	7:113954180	9:113586263
rs6961970	rs2821137	-0.63	0.089	1.27×10^{-12}	-0.804	-0.456	0.619	1.83×10^{-11}	7:113901132	9:113587820

Table 4.3: Genome-wide significant SNP interactions associated to single-word reading. Columns "*SNP1*" and "*SNP2*" denote detected interacting variants. Next column "*Beta*" represent the regression coefficient of the random effects meta-analyses, followed by the standard error "*SE*", p-value "*MPval*", lower "*CI_Low*" and upper "*CI_UP*" bounds of a 95% confidence interval, and the calculated heterogeneity p-value "*Het*" in-between the studies. Column "*GPval*" represent the computed p-values of GLIDE and final two columns "*Chr:Pos1*" and "*Chr:Pos2*" the respective SNP chromosome and position on the genome. All listed variants under *SNP1* are intron variants of the gene *FOXP2*.

4.2.2 Phonological awareness

This study comprised just a subset of the original sample, due to missing phenotypes. In total 645 individuals were subjected to the analysis, despite the exclusion of 217 probands we were able to observe highly significant statistical epistasis. The lowest p-value was retrieved with 8.05×10^{-13} comprising the variants rs620291 (11q23.2) and rs113400479 (18q11.2) (Table 4.6).

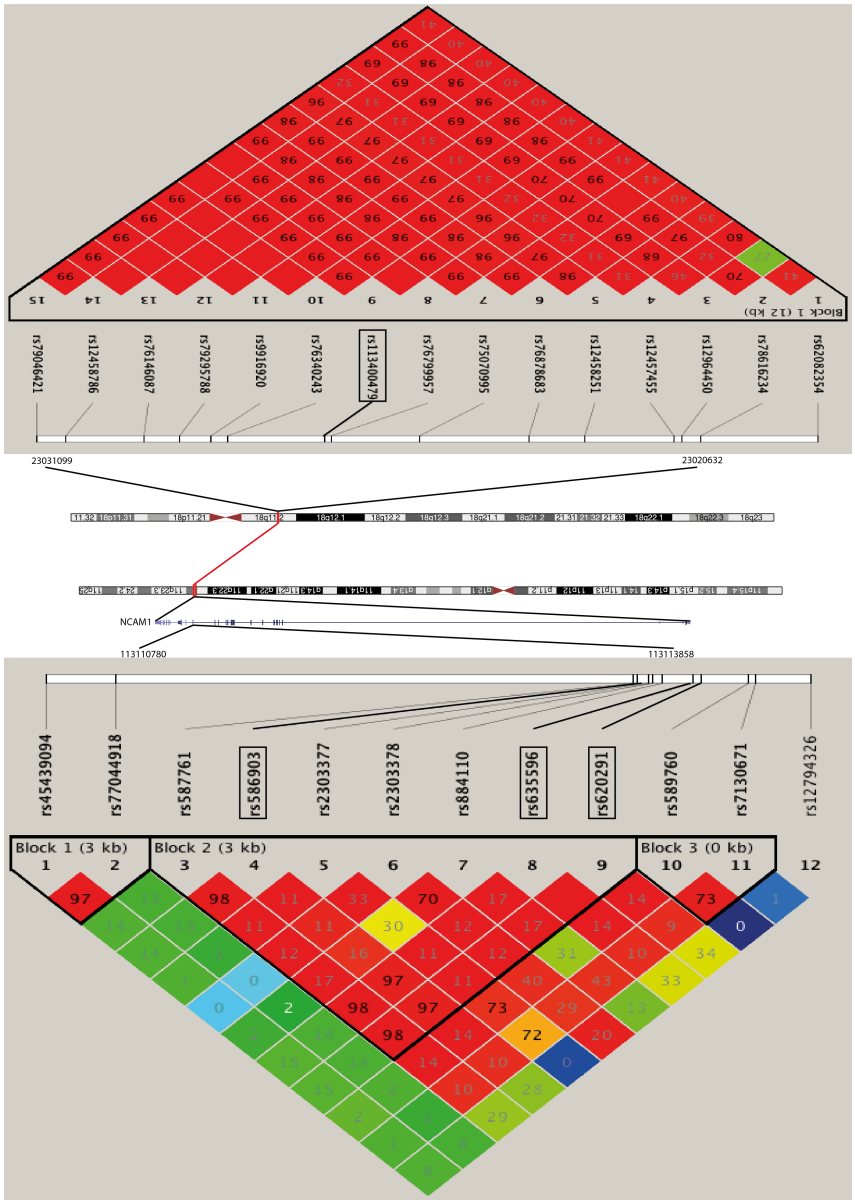


Figure 4.6: Genomic positions involved in interactions associated to phonological awareness within the LD-structure of 11q23.2 and 18q11.2. Most significant interacting variants are highlighted with rectangles. Chromosomal positions of the top hits are shown on each ideogram and are connected in red. Color code indicate with red high LD and with blue low LD.*Reference: Haploview*

Interaction associations were detachable over the respective LD regions of both candidate loci (Fig. 4.7).

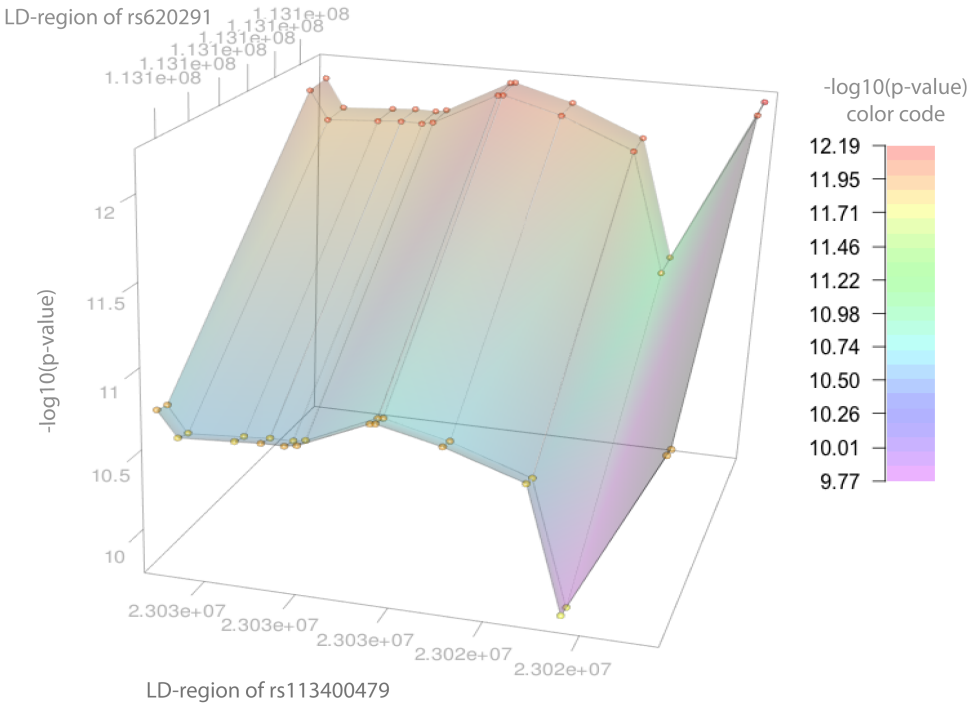


Figure 4.7: 3D plot of the 2-way associations to phonological awareness over the genomic SNP regions. *LD-region SNP1* and *LD-region SNP2* depict the chromosomal positions of rs620291 ($11q23.2$) and the interacting SNP rs113400479 ($18q11.2$) with their corresponding proxies SNP ($r^2 \geq 0.7$), respectively. The color legend indicates the detected $-\log_{10}(p)$-values.

The SNP rs620291 and its adjacent proxy SNPs are located in an intronic region of *NCAM1*, a gene involved in cell-adhesion, cell-migration and further neuronal functions [Deak et al., 2005]. The counterpart variants on $18q11.2$ are located by the DYX6 region, a known PA loci (Fig. 4.6). Although DYX6 is a popular PA QTL we could not detect any single main associations of the SNPs to the endophenotype. *NCAM1* variants showed even the weakest univariate associations (Table 4.4) compared to all other studies. Nevertheless the interactive correlation demonstrates (Fig. 4.9) a marginal coherence of the rs113400479 reference allele *C* with decreasing PA scores, but which seems to have an effect and be detectable only through the interaction with the *NCAM1* variants (rs620291 with *C* as reference allele).
We could observe that the combination of homozygous minor alleles appears in a single individual of the entire sample set, who also presented the weakest performance.

SNP	Position	Allele	MAF	Pval	Beta	SE	CI_Low / CI_UP
rs620291	11:113113858	C	0.1653	0.708	-0.032	0.086	-0.201 / 0.136
rs635596	11:113113515	G	0.1661	0.719	-0.031	0.086	-0.199 / 0.137
rs586903	11:113110946	C	0.1705	0.976	-0.003	0.086	-0.172 / 0.167
rs113400479	18:23026571	C	0.09206	0.451	-0.085	0.113	-0.306 / 0.136

Table 4.4: Univariate associations to phonological awareness of interacting SNPs. Depicted are all SNPs involved in significant statistical interaction, or with a trend to significance. First column contain the individual "*SNP*" identifiers. Column "*Effect allele*" display the respective minor allele, followed by the corresponding frequency "*MAF*" in the sample, the calculated p-value "*Pval*" of the single association, the estimated regression coefficient "*Beta*", the standard error "*SE*", and the lower "*CI_Low*" and upper "*CI_UP*" bounds of a 95% confidence interval. All denoted SNPs on chromosome 11 are intronic variants of *NCAM1*. Rs113400479 is an intergenic variant.

Excluding this individual tentatively from the sample and repeating the analysis the effect remains relatively robust with $p \leq 2.7 \times 10^{-12}$ compared to the original p-value of 8.05×10^{-13}. The exclusion of the next individual with an unique genotype combination for both markers had no effect on the p-value, explainable by a PA score close to zero (Fig. 4.9).

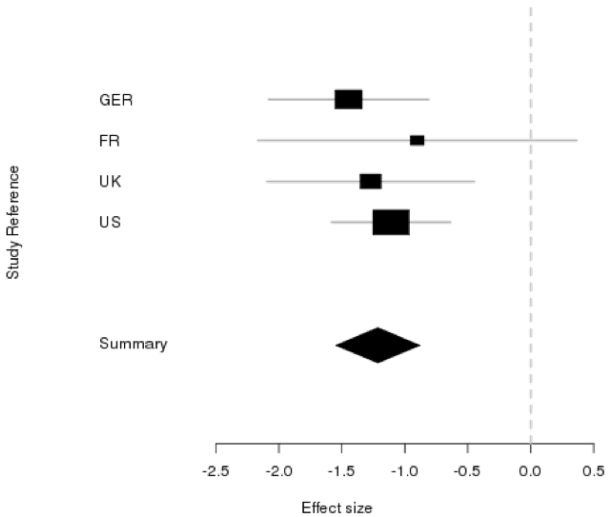


Figure 4.8: Meta-analysis of the phonological awareness top hit For each single cohort (German (GER), French (FR), British (UK), American (US)) the square and horizontal line show the estimated regression coefficient β and 95% confidence interval, representing the effect of each copy of the reference-allele on reading performance. The size of the square is inversely proportional to the standard error of the estimated effect. Below the individual cohorts a summary diamond shows the random-effects when analyzing all four cohorts together as one single sample. Notably clear negative effect is present in all single sub-samples and in the combined sample.

No heterogeneity between the cohorts ($p \geq 0.8$) was given (Fig. 4.8) and all cohorts demonstrated negative effect directions (Table 4.5).

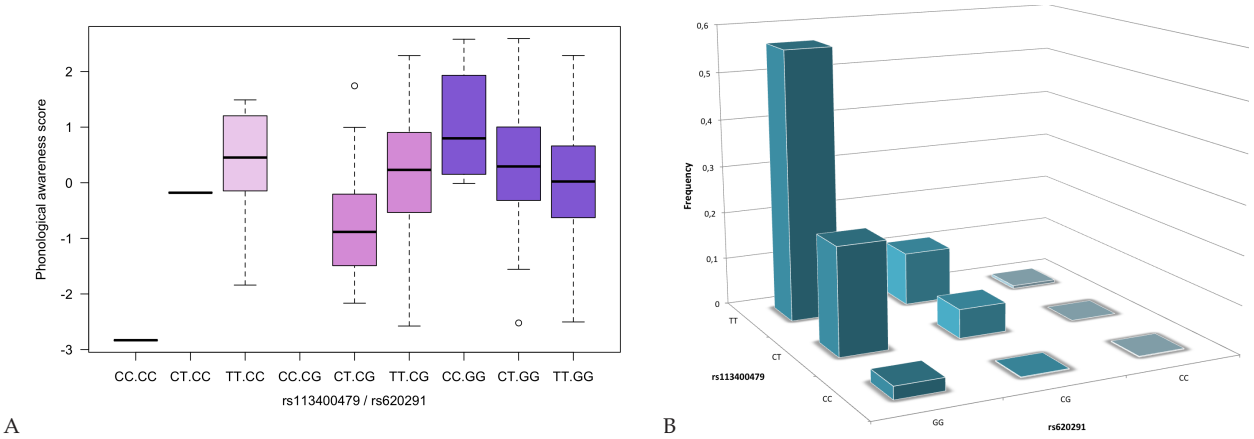


Figure 4.9: (A) Boxplot of genotypic associations through all samples with single-word reading score. Y-axis defines the range of the measured reading scores ($-3/+3$) normalized with a mean of 0 and a variance of 1. A score of -3 indicated the weakest phonological skills. X-axis is the according genotypic combination of the SNPs rs113400479 (chromosome 18, minor allele C) and rs620291 (chromosome 11, minor-allele C). The boxes define the interquartile range, the thick line the median, and single dots possible outliers. Box colors separate classes of genotype-combinations. **(B): Genotype-combination frequencies** of the SNPs rs113400479 and rs620291 in the combined sample comprising all cohorts.

Cohort	SNP1	SNP2	Beta	SE	Pval
GER	rs620291	rs113400479	-1.45	0.32	1.36×10^{-5}
FR	rs620291	rs113400479	-0.90	0.65	0.167
UK	rs620291	rs113400479	-1.27	0.42	2.8×10^{-3}
US	rs620291	rs113400479	-1.11	0.24	7.9×10^{-6}

Table 4.5: Cohort specific interactions of the phonological awareness top hit. First column "*Cohort*" denote the corresponding cohort followed by the detected interacting variants under "*SNP1*" and "*SNP2*". Next column "*Beta*" represent the regression coefficient, followed by the standard error "*SE*" and the resulting p-value "*Pval*" for the interacting SNP-pair in each individual cohort.

SNP1	SNP2	Beta	SE	MPval	CI_Low	CI_UP	Het	GPval	Chr:Pos1	Chr:Pos2
rs620291	rs113400479	-1.214	0.169	8.05×10^{-13}	-1.546	-0.881	0.812	6.65×10^{-13}	11:113113858	18:23026571
rs635596	rs113400479	-1.207	0.169	9.05×10^{-13}	-1.538	-0.875	0.809	7×10^{-13}	11:113113515	18:23026571
rs586903	rs113400479	-1.251	0.175	1.009×10^{-12}	-1.595	-0.907	0.393	1.64×10^{-11}	11:113110946	18:23026571

Table 4.6: Genome-wide significant SNP interactions associated to phonological awareness. Columns "*SNP1*" and "*SNP2*" denote detected interacting variants. Next column "*Beta*" represent the regression coefficient of the random effects meta-analyses, followed by the standard error "*SE*", p-value "*MPval*", lower "*CI_Low*" and upper "*CI_UP*" bounds of a 95% confidence interval, and the calculated heterogeneity p-value "*Het*" in-between the studies. Column "*GPval*" represent the computed p-values of GLIDE and final two columns "*Chr:Pos1*" and "*Chr:Pos2*" the respective SNP chromosome and position on the genome. All listed variants under *SNP1* are intron variants of the gene *NCAM1*.

4.2.3 Non-word reading

Epistasis study of the NWR endophenotype was the only case where we could not reach genome-wide significance, but a strong evidence for interaction with a p-value $p \leq 1.64 \times 10^{-12}$, barely missing the significance level of $p = 1.61 \times 10^{-12}$. The interaction was observed between the intronic SNP rs34981217 ($2q32.3$) of the gene *TMEFF2* and the singleton variant rs371164 (Fig. 4.10) located in a long terminal repeat (LTR) region (LTR67B) on $19p13.3$.

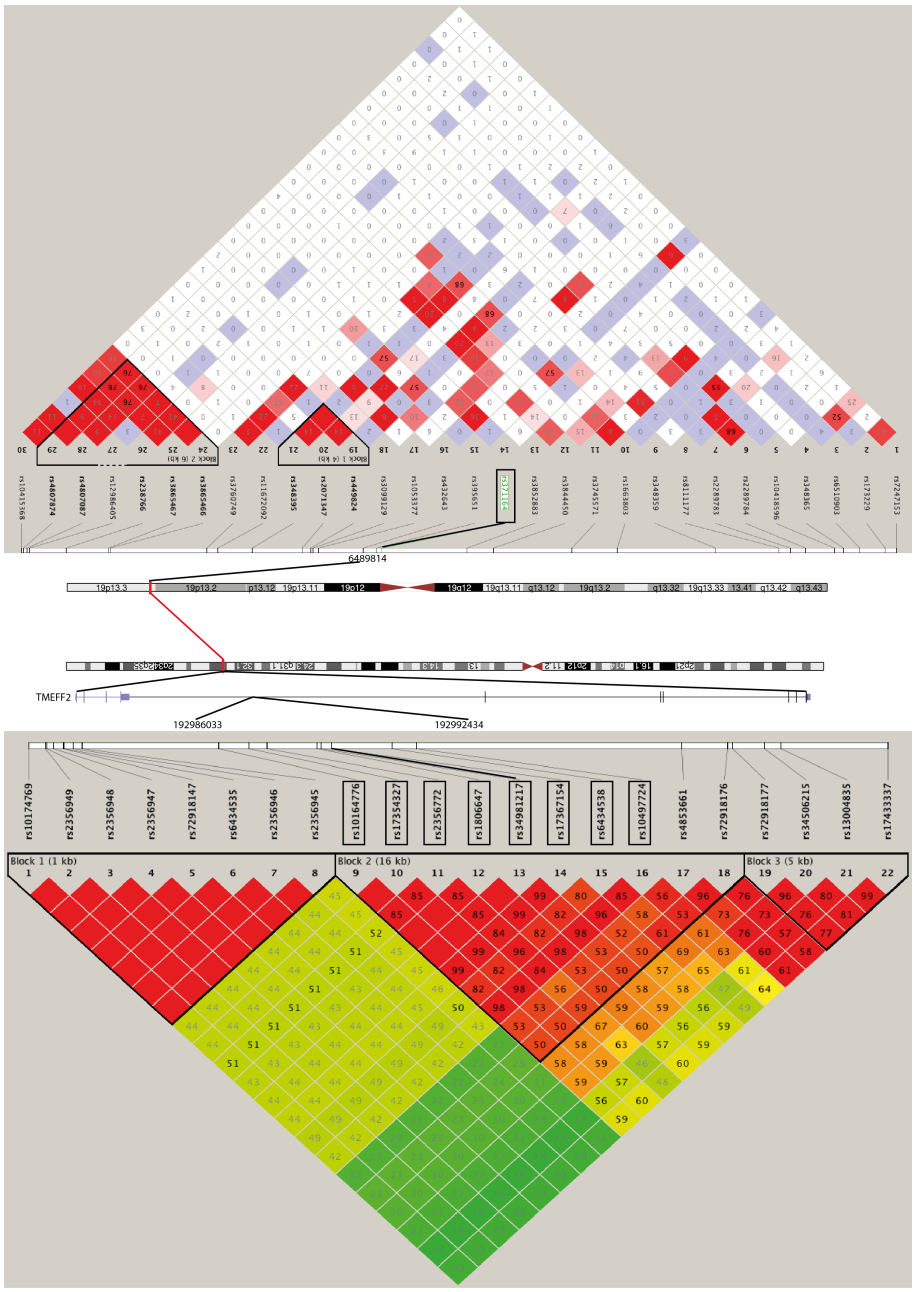


Figure 4.10: Genomic positions involved in interactions associated to non-word reading within the LD-structure of $19p13.3$ and $2q32.3$. Most significant interacting variants are highlighted with rectangles. Chromosomal positions of the top hits are shown on each ideogram and are connected in red. Color code indicate with red high LD and with blue low LD. *Reference: Haploview*

Likewise the case of the SWR study one of the interacting variants was located in a repetitive region, we executed a sequence alignment to ensure correct sequence mapping. Alignment was conducted with the usual procedure via BLAT, with 19p13.3 as the only match obtaining 100% identity over the entire 51bps. Moreover UCSC "Mappability" score indicated sequence uniqueness throughout the reference genome (hg19/GRCh37) on the region.
As rs371164 is a singleton, there were no interactions between further SNPs on 19p13.3, nonetheless we could observe interactions between proxy markers on 2q32.3 with rs371164 (see table 4.8).
Univariate association testing revealed again no marginal effects for any of the markers with p-values of 0.43 (rs371164) and 0.68 (rs34981217) (Table 4.7).

SNP	Position	Allele	MAF	Pval	Beta	SE	CI_Low / CI_UP
rs34981217	2:192989322	G	0.488	0.685	0.019	0.047	-0.073 / 0.112
rs17354327	2:192987014	T	0.49	0.636	0.022	0.047	-0.07 / 0.115
rs10164776	2:192986033	G	0.489	0.609	0.024	0.047	-0.069 / 0.117
rs17367154	2:192989322	T	0.484	0.682	0.019	0.047	-0.073 / 0.111
rs371164	19:6489814	G	0.19	0.433	-0.048	0.061	-0.168 / 0.072

Table 4.7: Univariate associations to non-word reading of interacting SNPs. Depicted are all SNPs involved in significant statistical interaction, or with a trend to significance. First column contain the individual "*SNP*" identifiers. Column "*Effect allele*" display the respective minor allele, followed by the corresponding frequency "*MAF*" in the sample, the calculated p-value "*Pval*" of the single association, the estimated regression coefficient "*Beta*", the standard error "*SE*", and the lower "*CI_Low*" and upper "*CI_UP*" bounds of a 95% confidence interval. All denoted SNPs on chromosome 2 are intronic variants of *TMEFF2*. Rs371164 is an intergenic singleton.

Interestingly this combination was unique in exhibiting a relatively huge difference between the MAFs of the interacting SNPs (Table 4.7). Verhoeven et al. mentioned in a publication [Verhoeven et al., 2010] the importance of allele frequency in epistatic models without marginal effects, which will be elucidated in the discussions (5).
The contingency table for the genotypes (Fig. 4.12) showed one of the most uniformly distributed constellations of our candidate SNP-pairs.
Apart from the fact that we barely missed the significance level in this case, we were able to observe a correlation of decreasing NWR scores, clearly under the mean, in dependency to the homozygous state of minor alleles GG/GG (Fig. 4.12). Again the negative interaction effect seemed homogenous through all participating cohorts (Fig. 4.11).

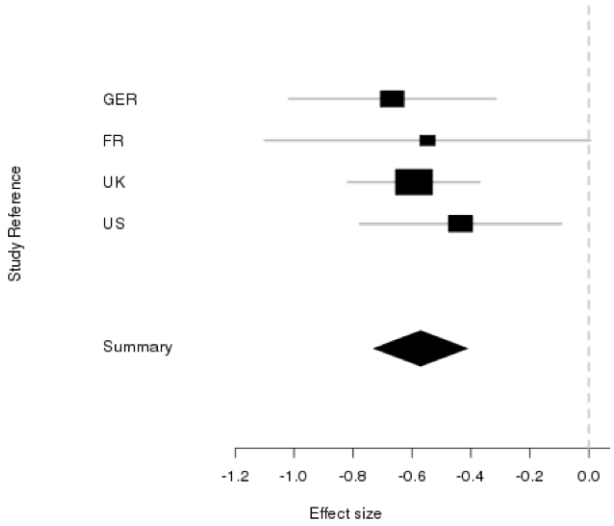
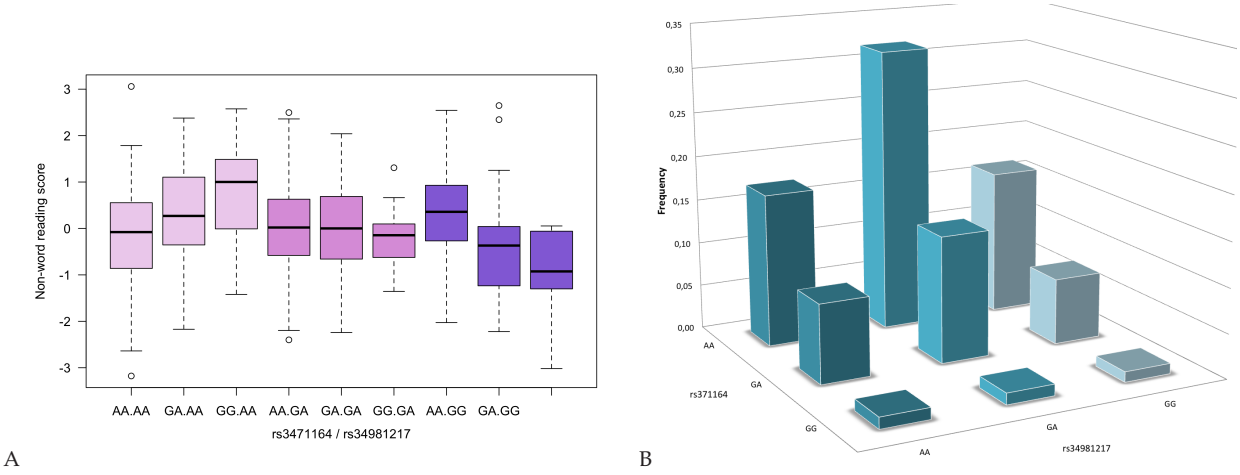


Figure 4.11: Meta-analysis of the non-word reading top hit For each single cohort (German (GER), French (FR), British (UK), American (US)) the square and horizontal line show the estimated regression coefficient β and 95% confidence interval, representing the effect of each copy of the reference-allele on reading performance. The size of the square is inversely proportional to the SE of the estimated effect. Below the individual cohorts a summary diamond shows the random-effects when analyzing all four cohorts together as one single sample. Notably clear negative effect is present in all single sub-samples and in the combined sample.



A B

Figure 4.12: (A) Boxplot of genotypic associations through all samples with single-word reading score. Y-axis defines the range of the measured reading scores ($-3/+3$) normalized with a mean of 0 and a variance of 1. A score of -3 indicated the weakest non-word reading performance. X-axis is the according genotypic combination of the SNPs rs34981217 (chromosome 2, minor allele G) and rs371164 (chromosome 19, minor-allele G). The boxes define the interquartile range, the thick line the median, and single dots possible outliers. Box colors separate classes of genotype-combinations. **(B): Genotype-combination frequencies** of the SNPs rs34981217 and rs371164 in the combined sample comprising all cohorts.

SNP1	SNP2	Beta	SE	MPval	CI_Low	CI_UP	Het	GPval	Chr:Pos1	Chr:Pos2
rs34981217	rs371164	-0.571	0.08	1.63×10^{-12}	-0.729	-0.412	0.818	9.26×10^{-13}	2:192989322	19:6489814
rs17354327	rs371164	-0.572	0.081	1.67×10^{-12}	-0.731	-0.4132	0.836	1.21×10^{-12}	2:192987014	19:6489814
rs10164776	rs371164	-0.572	0.081	1.76×10^{-12}	-0.731	-0.413	0.837	1.22×10^{-12}	2:192986033	19:6489814
rs17367154	rs371164	-0.562	0.08	2.97×10^{-12}	-0.719	-0.403	0.861	1.61×10^{-12}	2:192989689	19:6489814

Table 4.8: SNP interactions with strong evidence for association to non-word reading. Columns "*SNP1*" and "*SNP2*" denote detected interacting variants. Next column "*Beta*" represent the regression coefficient of the random effects meta-analyses, followed by the standard error "*SE*", p-value "*MPval*", lower "*CI_Low*" and upper "*CI_UP*" bounds of a 95% confidence interval, and the calculated heterogeneity p-value "*Het*" in-between the studies. Column "*GPval*" represent the computed p-values of GLIDE and final two columns "*Chr:Pos1*" and "*Chr:Pos2*" the respective SNP chromosome and position on the genome. All listed variants under *SNP1* are intron variants of the gene *TMEFF2*.

Cohort	SNP1	SNP2	Beta	SE	Pval
GER	rs34981217	rs371164	-0.67	0.18	2.6×10^{-4}
FR	rs34981217	rs371164	-0.55	0.28	5.5×10^{-2}
UK	rs34981217	rs371164	-0.59	0.11	3.4×10^{-7}
US	rs34981217	rs371164	-0.44	0.17	1.4×10^{-2}

Table 4.9: Cohort specific interactions of the non-word reading top hit. First column "*Cohort*" denote the corresponding cohort followed by the detected interacting variants under "*SNP1*" and "*SNP2*". Next column "*Beta*" represent the regression coefficient, followed by the standard error "*SE*" and the resulting p-value "*Pval*" for the interacting SNP-pair in each individual cohort.

4.2.4 Spelling

Significant ($p \approx 1.5 \times 10^{-12}$) epistasis was detected with the combination of the relatively *infrequent* alleles of rs59942153 (17p13.1) and rs7271609 (20q13.2) associated with spelling abilities of the probands (Table 4.12).

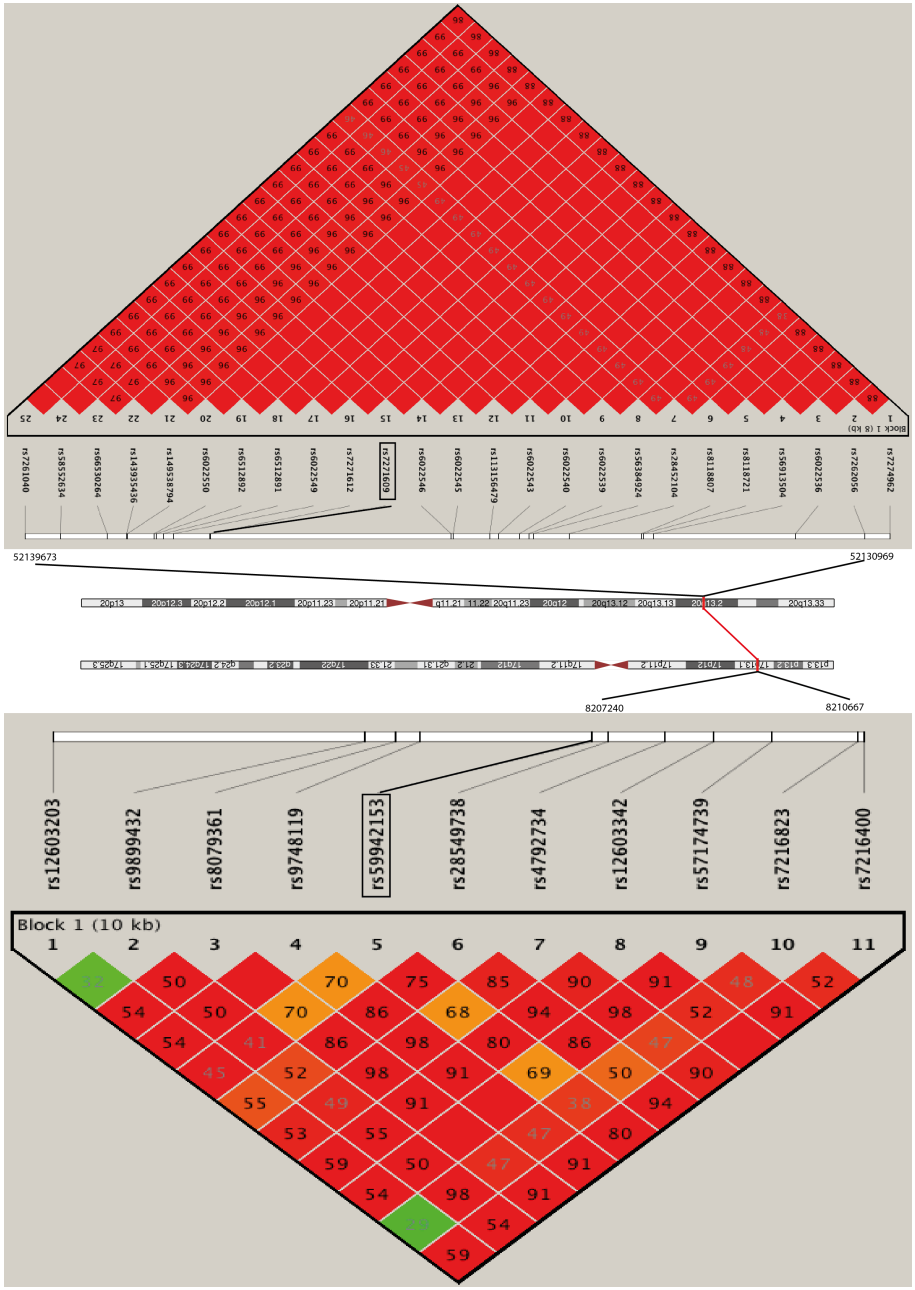


Figure 4.13: Genomic positions involved in interactions associated to spelling within the LD-structure of 17p13.1 and 20q13.2. Most significant interacting variants are highlighted with rectangles. Chromosomal positions of the top hits are shown on each ideogram and are connected in red. Color code indicate with red high LD and with blue low LD. *Reference: Haploview*

As visualized in the genotype contingency plot (Fig. 4.15) several genotype constellations were completely absent in the sample set. Neither any homozygous individual for both minor alleles *CC/CC*, nor any individual harboring a homozygous minor allele genotype for one of the loci in combination with a heterozygous genotype for the counterpart were available in our probands. Still associations to SP ability were detectable along the LD-regions (Fig. 4.14).

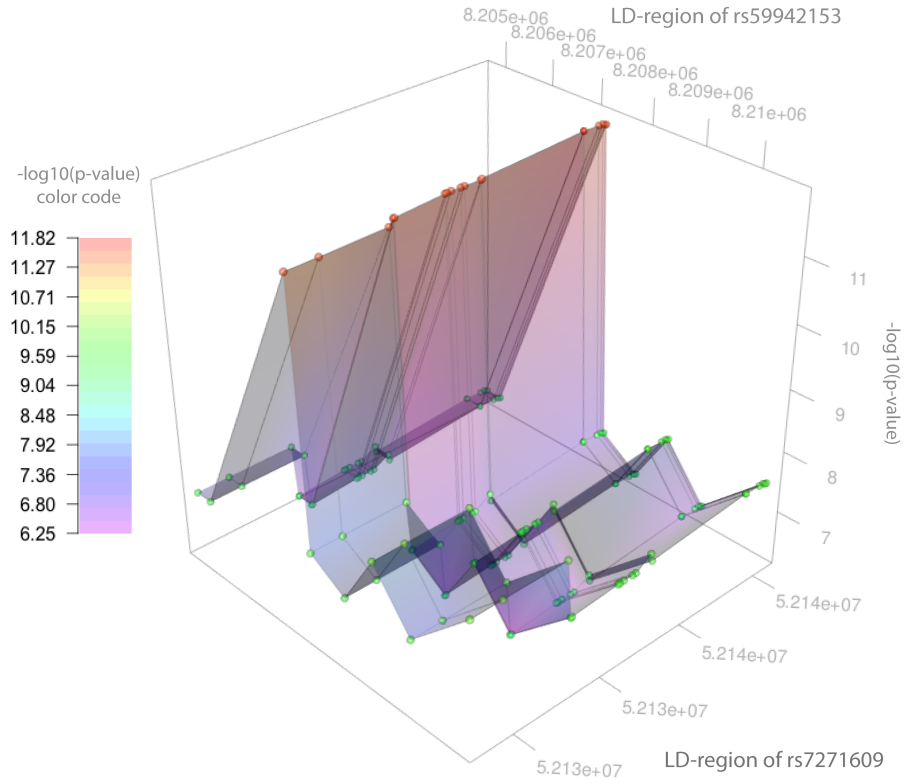


Figure 4.14: 3D plot of the 2-way associations to spelling over the genomic SNP regions. *LD-region SNP1* and *LD-region SNP2* depict the chromosomal positions of rs59942153 ($17p13.1$) and the interacting SNP rs7271609 ($20q13.2$) with their corresponding proxies SNP ($r^2 \geq 0.7$), respectively. The color legend indicates the detected $-\log_{10}(p)$-values.

Considering the boxplot (Fig. 4.15) we recognized an unusual pattern, in which the correlation between minor alleles and SP-score did not seem obviously, rather the distribution suggests just a context to weaker spelling ability, if minor alleles are present for both genetic variants. Other allele combinations hint to be distributed over the SP-score spectrum more uniformly with means around zero.
Regardless the missing genotype combinations the coefficient $\beta \approx -1.99$ (Fig. 4.16) of the meta-analysis indicates the clearest effect compared to the other studies and beyond that the cohorts displayed the least heterogeneity for the genotype combination with $p \approx 0.92$ (Table 4.12).

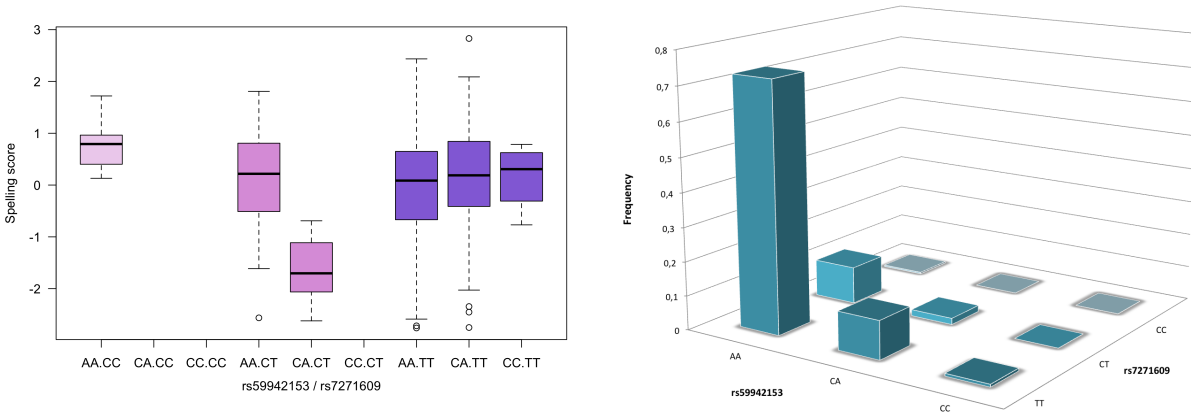


Figure 4.15: (A) Boxplot of genotypic associations through all samples with single-word reading score. Y-axis defines the range of the measured reading scores ($-3/+3$) normalized with a mean of 0 and a variance of 1. A score of -3 indicated the weakest spelling performance. X-axis is the according genotypic combination of the SNPs rs59942153 (chromosome 17, minor allele C) and rs7271609 (chromosome 20, minor-allele C). The boxes define the interquartile range, the thick line the median, and single dots possible outliers. Box colors separate classes of genotype-combinations. **(B): Genotype-combination frequencies** of the SNPs rs59942153 and rs7271609 in the combined sample comprising all cohorts.

No significant single loci correlations could be detected for rs59942153 or rs7271609 respectively (Table 4.10).

SNP	Position	Allele	MAF	Pval	Beta	SE	CI_Low / CI_UP
rs7271609	20:52137810	C	0.074	0.68	-0.037	0.091	-0.216 / 0.141
rs7271612	20:52137818	C	0.074	0.68	-0.037	0.091	-0.216 / 0.141
rs6022549	20:52138183	G	0.074	0.68	-0.037	0.091	-0.216 / 0.141
rs6512891	20:52138287	A	0.074	0.68	-0.037	0.091	-0.216 / 0.141
rs59942153	17:8207240	C	0.063	0.33	-0.098	0.1	-0.296 / 0.099

Table 4.10: Univariate associations to spelling of interacting SNPs. Depicted are all SNPs involved in significant statistical interaction, or with a trend to significance. First column contain the individual "*SNP*" identifiers. Column "*Effect allele*" display the respective minor allele, followed by the corresponding frequency "*MAF*" in the sample, the calculated p-value "*Pval*" of the single association, the estimated regression coefficient "*Beta*", the standard error "*SE*", and the lower "*CI_Low*" and upper "*CI_UP*" bounds of a 95% confidence interval.

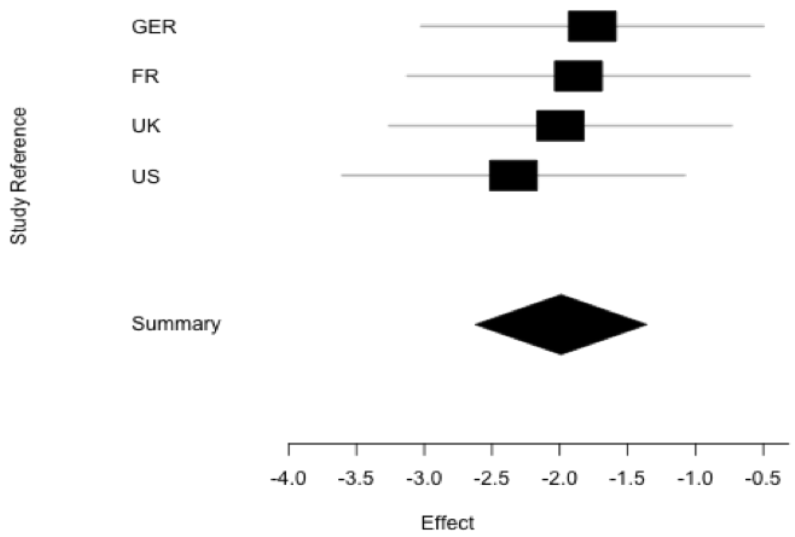


Figure 4.16: Meta-analysis of the spelling top hit For each single cohort (German (GER), French (FR), British (UK), American (US)) the square and horizontal line show the estimated regression coefficient β and 95% confidence interval, representing the effect of each copy of the reference-allele on reading performance. The size of the square is inversely proportional to the SE of the estimated effect. Below the individual cohorts a summary diamond shows the random-effects when analyzing all four cohorts together as one single sample. Notably clear negative effect is present in all single sub-samples and in the combined sample.

Cohort	SNP1	SNP2	Beta	SE	Pval
GER	rs59942153	rs7271609	-1.76	0.64	6.8×10^{-3}
FR	rs59942153	rs7271609	-1.86	0.59	2.3×10^{-3}
UK	rs59942153	rs7271609	-1.99	0.44	9.4×10^{-6}
US	rs59942153	rs7271609	-2.34	0.65	4.0×10^{-4}

Table 4.11: Cohort specific interactions of the spelling top hit. First column "*Cohort*" denote the corresponding cohort followed by the detected interacting variants under "*SNP1*" and "*SNP2*". Next column "*Beta*" represent the regression coefficient, followed by the standard error "*SE*" and the resulting p-value "*Pval*" for the interacting SNP-pair in each individual cohort.

SNP1	SNP2	Beta	SE	MPval	CI_Low	CI_UP	Het	GPval	Chr:Pos1	Chr:Pos2
rs59942153	rs7271609	-1.986	0.281	1.49×10^{-12}	-2.536	-1.436	0.926	1.16×10^{-11}	17:8207240	20:52137810
rs59942153	rs7271612	-1.986	0.281	1.5×10^{-12}	-2.536	-1.436	0.926	1.16×10^{-11}	17:8207240	20:52137818
rs59942153	rs6512891	-1.989	0.281	1.52×10^{-12}	-2.541	-1.438	0.926	1.19×10^{-11}	17:8207240	20:52138287
rs59942153	rs6022549	-1.989	0.281	1.52×10^{-12}	-2.541	-1.438	0.926	1.19×10^{-11}	17:8207240	20:52138183

Table 4.12: Genome-wide significant SNP interactions associated to spelling. Columns "*SNP1*" and "*SNP2*" denote detected interacting variants. Next column "*Beta*" represent the regression coefficient of the random effects meta-analyses, followed by the standard error "*SE*", p-value "*MPval*", lower "*CI_Low*" and upper "*CI_UP*" bounds of a 95% confidence interval, and the calculated heterogeneity p-value "*Het*" in-between the studies. Column "*GPval*" represent the computed p-values of GLIDE and final two columns "*Chr:Pos1*" and "*Chr:Pos2*" the respective SNP chromosome and position on the genome.

4.2.5 Validation of p-value distribution

To verify the distribution of our two-locus association test statistics obtained via GLIDE, we compared 100.000 randomly sampled interactions out of $\frac{50.000*(50.000-1)}{2}$ two-locus combinations (randomly sampled from whole-genome data and computed via GLIDE) against a normal distribution using `qqnorm` in R. The shown Q-Q plot (Fig. 4.17) displayed essentially no deviation from the unit slope. The plot indicated that findings were not due to any skewness of the distribution from normality and were distributed as expected.

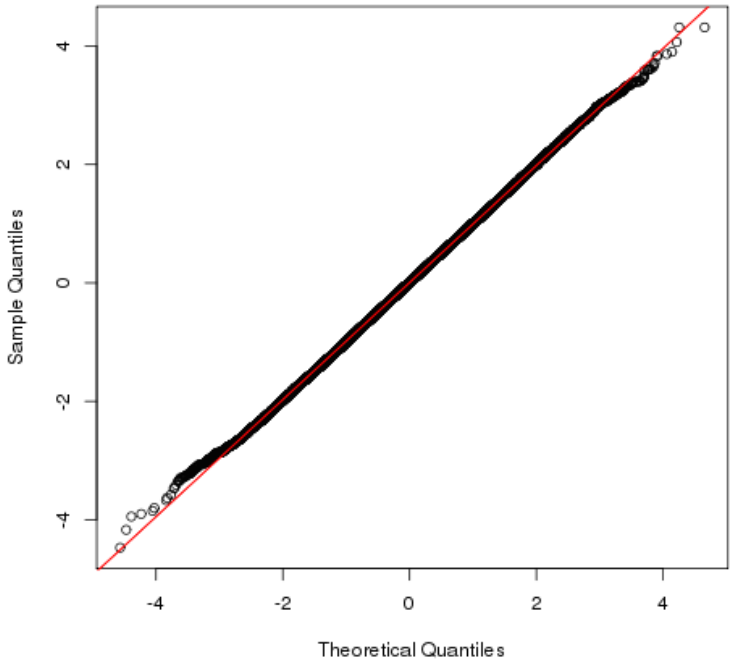


Figure 4.17: Q-Q plot of random sampling of the dyslexia study. Expected quantiles *Theoretical Quantiles* were plotted against the true *Sample Quantiles*. Distribution of data points (black) displayed no deviation from the unit slope (red).

4.2.6 Endophenotype correlation

Since one would expect similar cognitive requirements for our dyslexia endophenotypes (e. g., SWR and NWR), we calculated Pearson's product coefficients between the different endophenotype measurements to examine the correlation across studies. Additionally the association of each significant top SNP-pair, retrieved from the individual studies, to the other remaining three endophenotypes (Table 4.13) were computed.

Single-word reading top SNP-pair

As expected we obtained for the SWR top SNP-pair **rs1409679 and rs56253958** the most

conclusive association to the measurements of NWR with a p-value of 8.46×10^{-7}. The SNP combination received a p-value of 6.75×10^{-5} in association to SP, and 0.01 for the PA measurements. The association of the SNP-pair to each phenotype seemed to be compliant with the correlations in-between the phenotypes.

	SWR	NWR	SP	PA
SWR		0.53	0.51	0.38
NWR			0.21	0.39
SP				0.3

Table 4.13: Pearson's correlations between the endophenotypic measurements.

Phonological awareness top SNP-pair

The PA top hit **rs620291 and rs113400479** displayed stronger association to SWR with $p \leq 6 \times 10^{-3}$ and to NWR with $p \leq 6.5 \times 10^{-3}$, than to SP with a p-value of 1.2×10^{-2}, which again corresponds to detected correlations between the endophenotypes.

Spelling top SNP-pair

The association of the SP hit **rs59942153-rs7271609** to each endophenotype retrieved p-values of 7.38×10^{-4} for SWR, 0.02 for PA, and 0.57 for NWR.

Non-word reading top SNP-pair

The NWR top-pair **rs34981217-rs371164** revealed the slightest associations to the other endophenotypes with p-values of 0.059 for SWR, 0.07 for PA, and 0.64 for SP.

4.2.7 Functional analyses

Functional analyses were conducted via sequence mapping on the ENCODE functional annotation data. We were interested if any of our top-SNPs were located in an annotated chromosomal functional/regulatory site. All eight top-SNPs from the endophenotype specific analyses were mapped to annotated DNaseI HSs evidence, estimated from GM12878 blood cells and embryonic stem cells (H1-hESC). Furthermore we looked after annotated ChIP-seq data of NRSF binding sites, a neuronal repressor that function as a negative regulator of neurogenesis [Schoenherr and Anderson, 1995] in H1-neurons, TF binding sites estimated from GM12878 blood cells, and RNA polymerase II Pol2-4H8 binding sites estimated from neuronal cell lines derived from H1 embryonic stem cells [Birney et al., 2007].

Non-word reading – Functional sequence mapping

The singleton rs371164 is located in a region with evidence of chromatin accessibility for regulatory factors, indicated by enriched signals and a clear peak for DNaseI HS. Moreover we could observe an enrichment of binding sites for various TFs, identified by

ChIP-seq. The counterpart variant rs34981217 (*TMEFF2*) is located in a binding site for the TFs *p300*, a cellular transcriptional co-activator protein [Birney et al., 2007, Karolchik et al., 2014]. As shown in figure 4.18 approximately 200bp downstream and less then 2kilo base pairs (Kb) upstream of rs34981217 evidence for TFs and RNA polymerase II Pol2-4H8 binding sites are suggested by annotated ChIP-seq peaks. To retrieve a more complete picture we tested if any in UCSC annotated (hg19, GRCh37) *TMEFF2* variants with evidence for functionality were existent in our whole-genome data. We found rs6434538 (r^2 of 0.82 to rs34981217) mapping to a RNA polymerase II Pol2-4H8 binding site as only available SNP that withstood our applied whole-genome QC criteria and entered the two-locus analysis. The interaction between rs6434538 and rs371164 retrieved a p-value of 6.3×10^{-10} in the meta-analysis (see appendix).

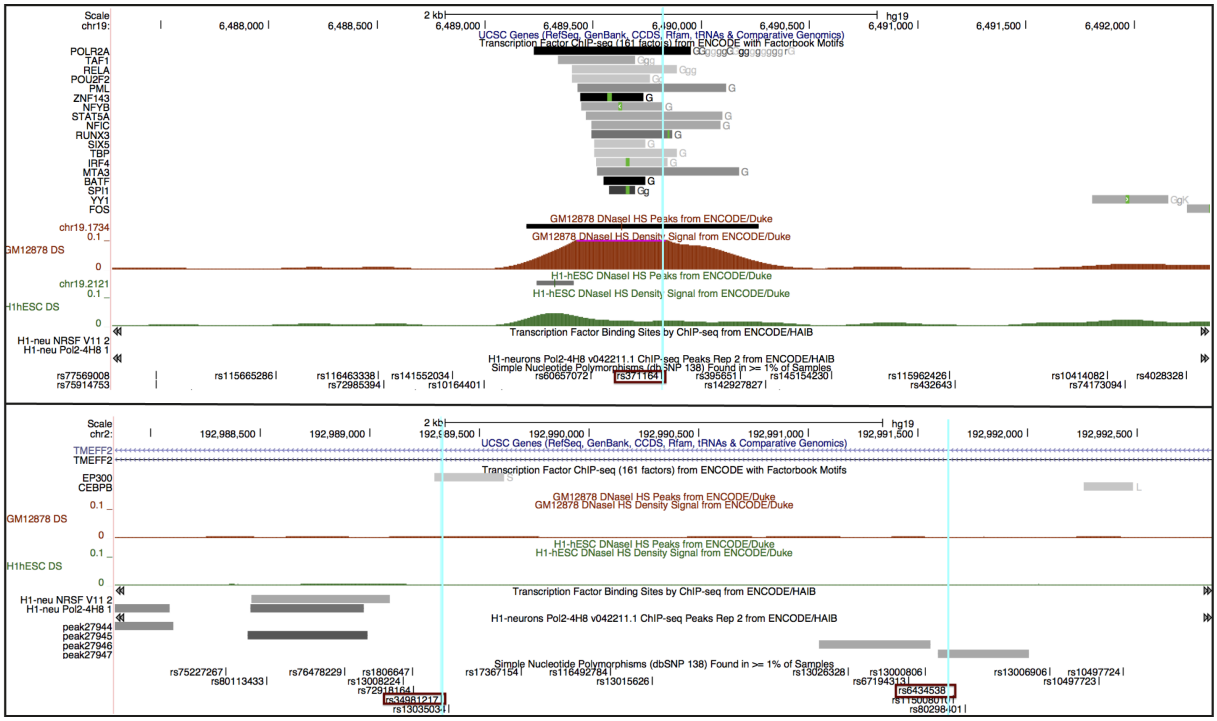


Figure 4.18: Non-word reading top hit SNPs mapped to UCSC browser (hg19,GRCh37). Mapped SNPs (upper panel: rs371164, lower panel: intronic SNPs (rs34981217,rs6434538) of *TMEFF2*) are declared in red rectangles and a vertical cyan line specify exact SNP position on the genome. DNaseI HS density signals of blood cells (GM12878) are colored red and density signals retrieved from embryonic stem cells (H1-hESC) in green. DNaseI HS peaks are either in black with a red line (GM12878) or grey with a green line (H1-hESC). Peaks indicating transcription factor binding sites and peaks for RNA polymerase II Pol2-4H8 binding sites identified via ChIP-seq are illustrated with grey and black bars.

Single-word reading – Functional sequence mapping

We observed evidence of TF binding sites for the chromosomal region of the SWR top hit SNP rs56253958 (*FOXP2*) (Fig. 4.19. Functional ENCODE annotation hinted to binding sites on the gene sequence of *FOXP2* for the TFs TCF7L2 and SMARCC1. None

of our remaining interacting *FOXP2* SNPs mapped to such regulatory sites, as the next annotated TF binding sites on the gene sequence are approximately 20 distant down- and upstream.

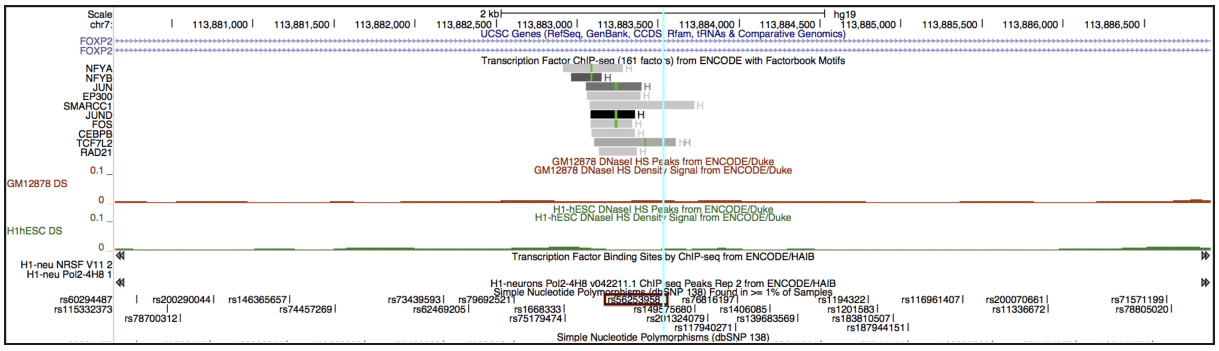


Figure 4.19: Single-word reading top hit SNPs mapped to UCSC browser (hg19,GRCh37). Mapped intronic *FOXP2* SNP (rs56253958) is declared in a red rectangle and a vertical cyan line specify exact SNP position on the genome. Peaks indicating transcription factor binding sites identified via ChIP-seq are illustrated with grey and black bars.

Phonological awareness – Functional sequence mapping

Two of the four *NCAM1* variants involved in significant interactions with variants on chromosome 18 showed to be located in functionally active sites. The closely located SNPs rs620291 and rs635596 mapped to denoted ChIP-seq peaks for various binding sites. Rs620291 is located in binding sites of the TFs EZH2 and YY1, whereas rs635596 mapped to a binding site of the TF CTCF, and additionally an annotated binding site of the RNA polymerase II Pol2-4H8 (Fig. 4.20).

Spelling - Functional sequence mapping

Even though the SP associated interacting variants were intergenic we could observe ChIP-seq identified functional sites. The SNPs rs6512891 and rs6022549 on $20q13.2$ are both located in different TF binding sites. Whereas for the interacting variant rs59942153 no peaks were detectable. Nevertheless we observed an enrichment of regulatory active sites in the surrounding region, so we extended the LD region of the variants to a $r^2 \geq 0.3$, as rs59942153 exhibited a sparse LD-density. As illustrated in figure 4.21 the variants rs9899432 ($r^2 = 0.412$, no interaction exceed the GLIDE t-value of 5) and rs4792734 ($r^2 = 0.675$, non-significant interaction with $20q13.2$ SNPs $p \leq 3.4 \times 10^{-7}$) could be observed in regions of highly enriched ENCODE signals. Interestingly rs4792734 maps among others to the same TF binding sites (USF2 and FOS) as its interacting variants on $20q13.2$. However since the interactions are not strong enough we can not make any assumptions about possible functionality of the interacting region.

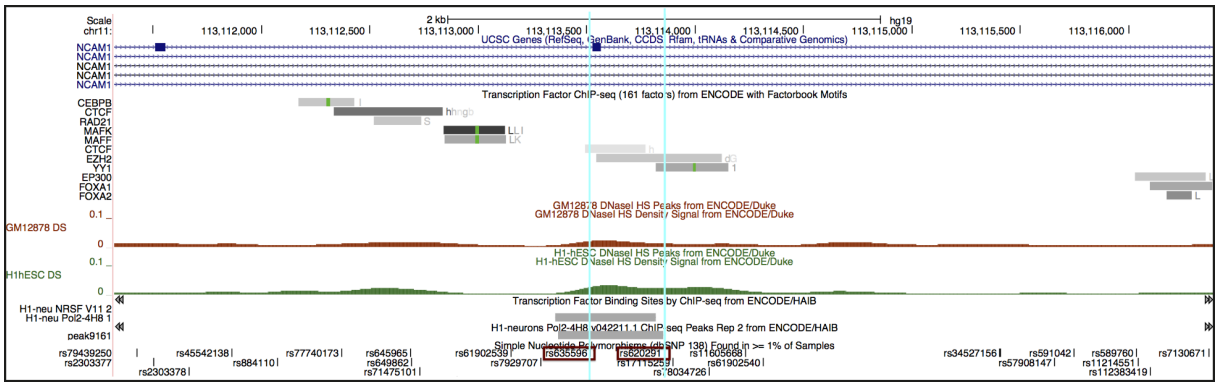


Figure 4.20: Non-word reading top hit SNPs mapped to UCSC browser (hg19,GRCh37). Mapped intronic *NCAM1* SNPs (rs620291, rs635596) are declared in red rectangles and a vertical cyan line specify exact SNP position on the genome. Peaks indicating transcription factor binding sites and peaks for RNA polymerase II Pol2-4H8 binding sites identified via ChIP-seq are illustrated with grey and black bars.

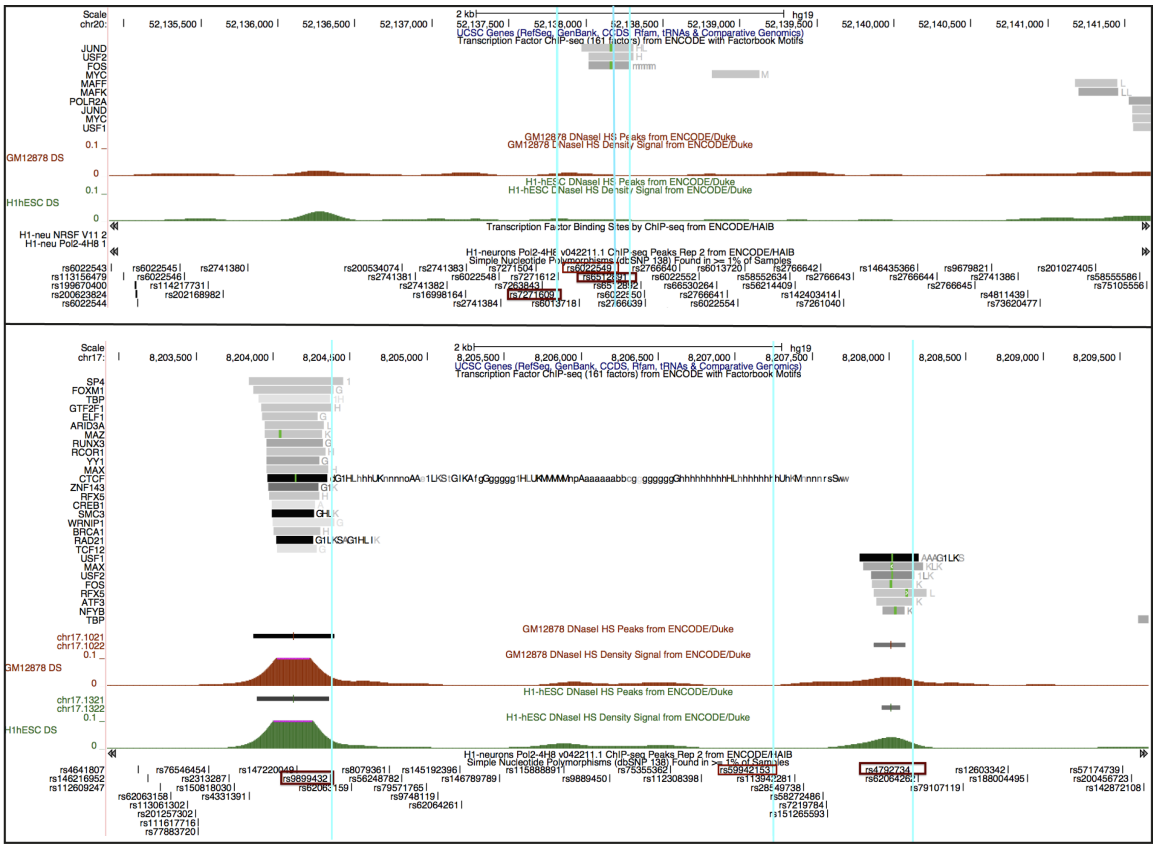


Figure 4.21: Spelling top hit SNPs mapped to UCSC browser (hg19,GRCh37). Mapped SNPs (upper panel: rs7271609,rs651289, lower panel: rs59942153, rs4792734, and rs9899432) are declared in red rectangles and a vertical cyan line specify exact SNP position on the genome. DNaseI HS density signals of blood cells (GM12878) are colored red, and density signals retrieved from embryonic stem cells (H1-hESC) in green. DNaseI HS peaks are either in black with a red line (GM12878) or grey with a green line (H1-hESC). Peaks indicating transcription factor binding sites identified via ChIP-seq are illustrated with grey and black bars.

4.2.8 Hippocampus candidate gene analyses

The concept of the hippocampus expression analyses in this thesis was to show a functional context of the dyslexia findings in the human pre-mortal brain. We wanted to elucidate whether a detected interaction causes a change in the expression level of the involved gene that in turn is linked to cognitive skills in the dyslexia pathogeneses. The ability to have insight into the expression behavior of an active human brain was almost unique. Unfortunately, some of the candidate genes of interest for our study were just poorly expressed in the sample of 138 epilepsy patients. None of the *FOXP2* transcripts withstood the expression QC criteria and just one transcript of *NCAM1* (ILMN_1676289) and *TMEFF2* (ILMN_1730645) showed an appropriate expression level. Nevertheless, we ran the regional based ($\pm$ 2Mb of the candidate SNPs) interaction analyses for all available candidate genes without any significant outcomes after correction for multiple testing. Bonferroni-Becker significance correction levels were calculated for each study individually in dependance of the SNPs number in the $\pm$ 2Mb regions.

SNP1	SNP2	p-value
rs12535428	rs1409679	1.6×10^{-2}
rs4117983	rs1409679	4.9×10^{-2}
rs2140615	rs1409679	4.9×10^{-2}
rs4730633	rs1409679	4.9×10^{-2}
rs12535428	rs2821137	1.6×10^{-2}
rs12535428	rs2846443	1.9×10^{-2}
rs4117983	rs2846443	4.8×10^{-2}
rs2140615	rs2846443	4.8×10^{-2}
rs4730633	rs2846443	4.8×10^{-2}

Table 4.14: SNP-pairs associated to the expression profile of the *FOXP2* transcript ILMN_1695355. Columns "*SNP1*" and "*SNP2*" denote detected interacting variants and "*Pval*" the computed interaction p-value.

The interacting variants of SWR and consequently markers in a region of $2Mb$ up- and downstream ($\pm$) of the original SNPs revealed interactions between markers of both chromosomes with $p \approx 10^{-7}$. Considering specifically markers located in *FOXP2* the best finding ($p = 2.3 \times 10^{-7}$) was the interaction between the *FOXP2* intronic variant rs1852471 and rs2762469 on chromosome 9 with a distance about 1.15Mb from the SWR top-SNPs rs1409679, rs2821137 and rs2846443 ($9q31.3$) on the expression of the FOXP2 transcript ILMN_1695355. Testing explicitly our $9q31.3$ loci (without any further proxy SNPs) with *FOXP2* variants we detected interactions with four intronic markers

exceeding the nominal significant threshold of 0.05. Regarding Bonferroni correction level ($p \leq 3.8 \times 10^{-10}$) we could not find any significant interactions consolidating the SWR results.
Interaction analysis between the $\pm$ 2Mb region on 11q23.2 and 18q11.2 on the expression of ILMN_1676289 (highest expressed *NCAM1* probe) revealed no significant SNP combinations regarding the significance correction level of $p \leq 5.21 \times 10^{-9}$. Concentrating explicitly on the PA top SNPs we could not observe any interactions on the expression of ILMN_1676289.
Most availing results included the closely ($\approx$ 2 Kb) to rs620291 located variant rs2303377 ($r^2 = 0.152$) in interaction with our dyslexia candidate SNPs on 18q11.2 (Table 4.15) associated to the poorly expressed *NCAM1* transcript ILMN_2398184. Calculated p-values ranged between 7×10^{-3} and 3×10^{-3}.

SNP1	SNP2	p-value
rs2303377	rs79046421	3.6×10^{-3}
rs2303377	rs12458786	4.2×10^{-3}
rs2303377	rs76146087	5.4×10^{-3}
rs2303377	rs79295788	5.9×10^{-3}
rs2303377	rs9916920	6.5×10^{-3}
rs2303377	rs76340243	6.9×10^{-3}
rs2303377	rs113400479	7.2×10^{-3}
rs2303377	rs76799957	8.3×10^{-3}

Table 4.15: SNP-pairs associated to the expression profile of the *NCAM1* transcript ILMN_2398184. Columns "*SNP1*" and "*SNP2*" denote detected interacting variants and "*Pval*" the computed interaction p-value.

A single Illumina probe was available for the *TMEFF2* gene, which exhibited a well expression profile in the human brain. Nonetheless, regarding the correction threshold of $p \leq 3.6 \times 10^{-5}$ for multiple testing no significant results were detectable. As the interacting dyslexia NWR pair involved the singleton SNP rs371164 on chromosome 19 the ability to search for interactions in the LD region of the variant was not given. However best interactions that could be observed were between the singleton rs371164 and variants in the *TMEFF2* sequence with p-values $\approx 3 \times 10^{-2}$.

SNP1	SNP2	p-value
rs73984908	rs371164	3.1×10^{-2}
rs73984907	rs371164	3.1×10^{-2}
rs13432794	rs371164	3.6×10^{-2}
rs4474831	rs371164	3.6×10^{-2}
rs73984904	rs371164	5×10^{-2}
rs16834268	rs371164	5×10^{-2}

Table 4.16: SNP-pairs associated to the expression profile of the *TMEFF2* transcript ILMN_1730645. Columns "*SNP1*" and "*SNP2*" denote detected interacting variants and "*Pval*" the computed interaction p-value.

4.2.9 Endophenotype specific pathway analysis

Additional surveys, to obtain a deeper insight in the biological relevance of our results, were fulfilled with an endophenotype specific and an overall findings pathway analysis. In the first instance we considered enriched pathways for every endophenotype specific gene-set of the epistasis analyses separately, utilizing MAGENTA [Segrè et al., 2010]. Gene-sets were determined by MAGENTA for every SNP of an interacting SNP-pair with a p-value $< 1 \times 10^{-10}$, which resulted in 716 genes in total. Pathways or groups with a GSEA p-value less then the nominal significance cutoff of ≤ 0.05 are depicted in the corresponding tables.
Pathway/groups occurring more than once in the results, due to the usage of multiple databases in MAGENTA, were listed in the corresponding tables by just one representative.

The GSEA of the NWR genes resulted in just three significant pathways as shown in table 4.17.

The GSEA of the PA genes disclosed an enrichment in neuronal activities as the most significant outcome with a p-value of 0.012 (Table 4.18). Further enriched pathways or groups respectively were *NCAM1* interactions, kinase activities, and pathways corresponding to the complement system that is involved in the immune system.

The GSEA of SWR genes revealed highly reasonable enrichments among others in the acetylcholin pathway ($p = 0.0436$), neuromuscular junction development ($p = 0.047$) and cerebral cortex development ($p = 0.0466$). Further significantly enriched

DB	Pathway/Group	*p*-value
Panther molecular function	Membrane-bound signaling molecule	0.0134
Gene Ontology	Nucleoplasm	0.0155
Reactome	Cell cycle Mitotic	0.0243

Table 4.17: Enriched pathways of the non-word reading gene-set analysis. The first column "*DB*" denotes the database in which the associated function "*Pathway/Group*", shown in the second column, was annotated. The third column denotes the corresponding enrichment "*p-value*".

DB	Pathway/Group	*p*-value
Panther biological process	Other neuronal activity	0.012
Reactome	RHO GTPase cycle	0.021
Gene Ontology	Protein serine/threonine kinase activity	0.041
Panther molecular function	Non-receptor serine/threonine protein kinase	0.044
KEGG	Complement and coagulation cascade	0.045
Gene Ontology	Eukaryotic cell surface binding	0.046
Gene Ontology	Response to oxidative stress	0.047
Gene Ontology	Receptor signaling protein activity	0.048
Gene Ontology	Acute-phase response	0.048
Reactome	Initial trigging of complemet	0.049
Reactome	*NCAM1* interactions	0.049
Panther	Endothelin signaling pathway	0.049
Gene Ontology	Cellular process	0.049
Gene Ontology	Complement activation, lectin pathway	0.05

Table 4.18: Enriched pathways of the phonological awareness gene-set analysis. The first column "*DB*" denotes the database in which the associated function "*Pathway/-Group*", shown in the second column, was annotated. The third column denotes the corresponding enrichment "*p-value*".

pathways and groups are listed in the according table 4.19. Most significant result were biological processes (BPs) involved in cell structure. Further we observed enrichments in embryogenesis and neurogenesis that barely missed the *p*-value cutoff.

DB	Pathway/Group	p-value
Panther biological process	Cell structure	0.039
Gene Ontology	Negative regulation of gene-specific transcription	0.041
Biocarta	Acetylcholin pathway	0.044
Panther molecular function	Tyrosine protein kinase receptor	0.044
Gene Ontology	N-acetylglucosamine metabolic process	0.044
Panther biological process	Receptor protein tyrosine kinase signaling pathway	0.045
Panther molecular function	Protein kinase	0.045
Gene Ontology	Single fertilization	0.045
Gene Ontology	Carbohydrate metabolic process	0.045
Gene Ontology	Muscle organ development	0.045
Biocarta	Argin Pathway	0.046
KEGG	Amino and nucleotide sugar metabolism	0.046
Panther biological process	Monosaccharide metabolism	0.046
Gene Ontology	Cerebral cortex development	0.047
Gene Ontology	Neuromuscular junction development	0.047
Gene Ontology	Generation of precursor metabolites and energy	0.047
Gene Ontology	Receptor activity	0.047
Panther molecular function	Other isomerase	0.048
Gene Ontology	Transmembrane receptor protein tyrosine kinase activity	0.048

Table 4.19: Enriched pathways of the single-word reading gene-set analysis. The first column "*DB*" denotes the database in which the associated function "*Pathway/-Group*", shown in the second column, was annotated. The third column denotes the corresponding enrichment "*p-value*".

The GSEA of the SP genes illustrated as most significant finding an enrichment of genes involved in developmental processes. Further enriched were pathways involved in axon guidance, known as an important disrupted mechanism of dyslexia pathogeneses (Table 4.20). Significantly enriched was further the GO term "microtubule-based flagellum", describing genes with mechanisms in cell movement among others.

DB	Pathway/Group	*p*-value
Gene Ontology	Multicellular organismal development	0.026
Panther biological process	Purine metabolism	0.047
Gene Ontology	Single-stranded DNA binding	0.047
Gene Ontology	Cell redox homeostasis	0.048
Ingenuity	Axonal guidance signaling	0.048
Panther molecular function	Synthase	0.048
Gene Ontology	Cytokinesis	0.049
Gene Ontology	Nuclear chromosome, telomeric region	0.049
Gene Ontology	Oligosaccharyltransferase complex	0.049
Gene Ontology	Dolichyl-diphosphooligosaccharide-protein glycotransferase activity	0.049
Gene Ontology	Microtubule-based flagellum	0.05
KEGG	N-Glycan biosynthesis	0.05

Table 4.20: Enriched pathways of the spelling gene-set analysis. The first column "*DB*" denotes the database in which the associated function "*Pathway/Group*", shown in the second column, was annotated. The third column denotes the corresponding enrichment "*p-value*".

4.2.10 Pathway analysis in dyslexia

The final GSEA was applied on the collective gene-set of all four endophenotype studies. With respect to the Bonferroni correction level (3.1×10^{-6}) we acquired 20 significantly enriched pathways or groups respectively (Table 4.21). The findings of the overall analysis were incredibly promising, as we could observe enrichments of genes acting in neuronal cells, axons or at the membrane surface, for example for signaling processes, transport activities, cell-cell signaling and adhesion. We observed also dyslexia reasonable BPs involved in locomotion, nervous system (NS) development, axon guidance and microtubule sliding.

Ontology domain	Pathway/Group	p-value
Biological process	Metabolic process	4.51×10^{-12}
Molecular function	Nucleotide binding	4.02×10^{-10}
Biological process	Transport	1.49×10^{-09}
Biological process	Locomotory behavior	3.35×10^{-9}
Biological process	Multicellular organismal development	5.09×10^{-9}
Cellular component	Membrane	1.45×10^{-8}
Biological process	Nervous system development	3.20×10^{-8}
Biological process	Transmembrane transport	4.11×10^{-8}
Cellular component	Plasma membrane	1.61×10^{-7}
Biological process	Axon guidance	2.16×10^{-7}
Cellular component	Cytoskeleton	2.20×10^{-7}
Cellular component	Neuronal cell body	2.24×10^{-7}
Molecular function	ATP binding	3.37×10^{-7}
Cellular component	Cytoplasm	3.5×10^{-7}
Biological process	Ion transport	5.73×10^{-7}
Cellular component	Cytosol	6×10^{-7}
Biological process	Signal transduction	7.43×10^{-7}
Biological process	Water transport	1.01×10^{-6}
Biological process	Ion transmembrane transport	1.61×10^{-6}
Biological process	Cerebral cortex development	3.10×10^{-6}

Table 4.21: Enriched pathways in the joint gene-set of the individual interaction analyses. First column *Ontology domain* addresses the domains of the annotated enriched pathway of the second column *Pathway/Group* and the third column the corresponding *p-value*.

5 Discussion

5.1 Epistasis and genome-wide interaction analyses

5.1.1 Interaction analyses are required for understanding complex diseases

Most biological phenotypes, from hair color to diseases, cannot be fully explained by single-locus genetic effects. However, the remaining variance can also not entirely be attributed to environmental contributions. In fact, interactions between two or more distinct genetic loci typically constitute a decisive factor in shaping the observed phenotypical variance. Such interplay between unlinked genetic loci is called epistasis [Bloom et al., 2013].
Currently, little is known about the role of epistasis in human biology, as, until now, the focus of research was mainly on single locus effects. Nevertheless, considering the network of biological and functional processes, most genetic candidates for complex traits are likely embedded within a broad network of interactions. When considering either biochemical and metabolic systems or the complexity of networks that robustly stabilize essential human systems, the ubiquity of epistasis in genetic systems becomes obvious [Moore and Williams, 2009].
Therefore, statistical and biological epistasis should be considered as just as important as the main, direct effects of candidate genes [Nicodemus et al., 2007]. Given the omnipresence of complexity in genetics, it has even been suggested to rephrase the current standard hypothesis for genetical research: *Instead of asking which SNP is associated with disease, we should be asking which combination of SNPs is associated with disease?* [Moore and Williams, 2009]. It is therefore clearly necessary to improve our knowledge regarding biological and statistical epistasis and its role in human health and disease. However, the field of multivariate interaction analyses has only recently been established. The acquisition of knowledge as well as the development of the necessary computational conditions are thus still in an early stage.
It constitutes a dilemma for medical research that single loci account only for a small fraction of common disease susceptibility. As a possible explanation for this lack of identified genetic contributions, Moore and colleagues argue that malign phenotypes could in fact be the result of multiple mutations in different parts of a network [Moore, 2003]. From this perspective, epistasis could even have a negative influence on human health: while positive epistasis might lead to higher fitness than is attributable to single

mutations, negative epistasis would, accordingly, incur lower fitness [Gros et al., 2009]. Such effects could be a consequence of evolutionary epistatic shielding,acting in either direction [Verhoeven et al., 2010, Moore, 2003]. In the context of positive interactions, *epistatic shielding* means that fitness is protected, whereas, in the case of detrimental interactions, harmful interactions persist. A possible explanation for the existence of epistatic shielding is that breaking a network of interaction partners is a far more complicated event than eliminating a single deleterious mutant. This notion is supported by the observation that the occurrence of double mutants is, in general, a rare event.
In addition, Moore et al. suggested genetic interactions to constitute a possible cause for unsuccessful replications in independent samples. If this hypothesis was correct, the consideration of combinatory allelic risks for human health could push the search for the origin of the often discussed *Missing heritability* significantly forward. However, while, from a theoretical point of view, the concept and relevance of epistasis detection is well established, obscurities as well as the diversity of hypotheses overshadow its realization in practice, as so often is the case in new fields in science.
Meanwhile, increasing amounts of studies concentrate on genetic interactions and thereby reveal a role of epistasis in the genetic control of complex phenotypes. At the beginning, most studies captured epistasis in Mendelian masking effects [Verhoeven et al., 2010], but currently, studies investigate interactions between polymorphisms that broadly contribute to multifactorial traits [Bloom et al., 2013, Marchini et al., 2005, Nicodemus et al., 2010b, Shao et al., 2008]. It became clear that epistasis correlates with genomic complexity: in the case of simpler mutational effects, epistasis rather illustrates a masking effect, whereas in more complex systems, mutations appear to aggravate each other, leading to synergetic effects [Van Steen, 2012].

5.1.2 Statistical power in interaction analyses

In quantitative traits, epistasis is considered as the deviation from additivity. Two studies have demonstrated that, in the presence of epistasis, statistical detection power can increase over the detection power for single-loci [Marchini et al., 2005, Verhoeven et al., 2010]. Morever, Verhoeven and colleagues illustrated a higher detection power of multiplicative interaction models, as compared to additive models. In the case of the results presented here, we could observe statistically highly significant epistatic effects with convincing detection power (Fig. 5.1) Detection power was computed using the software QUANTO [Gauderman, 2002] for each top hit, and was compared to either calculation power or detectable significant association effects of single SNPs (data not shown). Estimated detection power ranged from 44.7% in case of the endophenotype phonological awareness , to 62.8% for single-word reading SWR, using the respective study sample sizes (645 and 862). Prognoses for increasing sample sizes up to 1300 subjects predicted a power boost to approximately 99%.
Verhoeven et al. showed that, along with sample size, also allele frequencies representa-

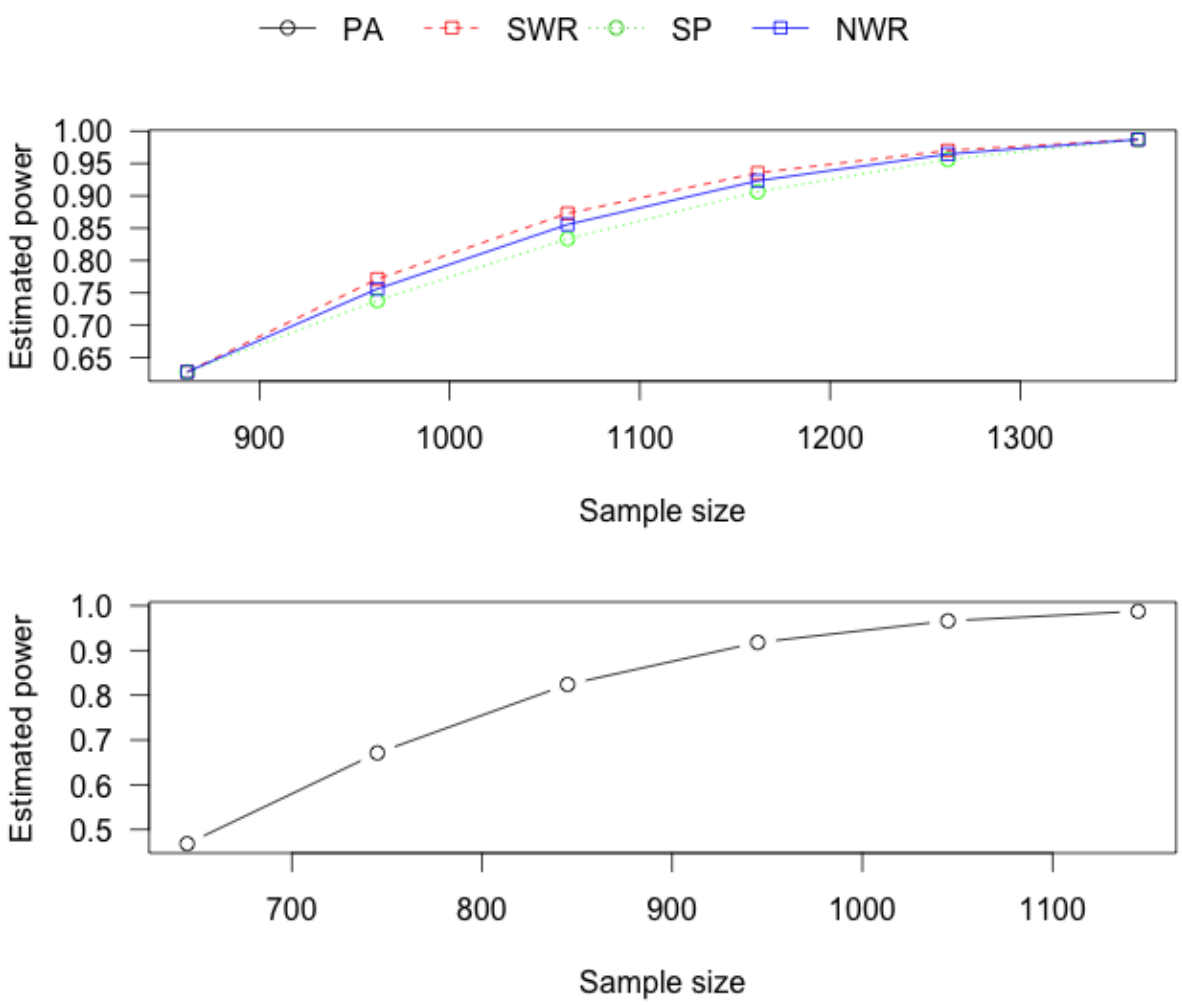


Figure 5.1: Detection power of study top hits. The first panel demonstrates the trend of the detection power (y-axis) starting at the given sample size (x-axis) of 862 individuals for the endophenotype specific analyses of SWR (red dashed line with squares), SP (green dotted line with circles), and NWR (blue solid line with squares). Computed prognoses indicate full power by a sample sizes over 1300 subjects. Second panel shows the calculation power for the PA top hit (black solid line with circles) starting at 645 and a power of approximately 45%. Full power can be reached in this case with a sample size over 1000 individuals.

tively influence power: the closer the MAFs of the involved SNPs are to each other, the higher is the power of the analysis and the lower the false positive rate, given an appropriate sample size of ≥ 500. In fact, we faced the same phenomenon in our data: three out of the four strong interactions that exceeded the significance level showed similar allele frequencies of the respective interacting markers. The only expectation falling just short of the significance threshold included two polymorphisms with unequal MAFs of 0.48 and 0.19 (SNPs associated to non-word reading NWR).

5.1.3 Marginal effects of interaction partners

During the last years, a variety of approaches and tools has been developed to capture SNP-SNP or gene-gene interactions [Culverhouse, 2012, Hemani et al., 2011, Hu et al., 2010, Kotti et al., 2007, Yung et al., 2011, Zhao et al., 2006]. In this context, different hypotheses have been established and tested. One of the most intensively discussed hypotheses is concerned with the necessity of marginal effects in an epistatic system. This motive implies that penetrance of the single polymorphisms is required in order to allow for any biologically relevant effect of an interaction on a trait [Ritchie, 2010]. However, in our GWAS results, we observed no marginal effects or only a trend towards them ($p \approx 10^{-2}$). In fact, several published studies proposed complex theoretical penetrance models that, as in our case, influence the trait only through the interaction of two or more genetic variants [Culverhouse et al., 2002, Frankel and Schork, 1996, Kotti et al., 2007, Musani et al., 2007]. These models are *purely epistatic* (Fig. 5.2). The prevalence of

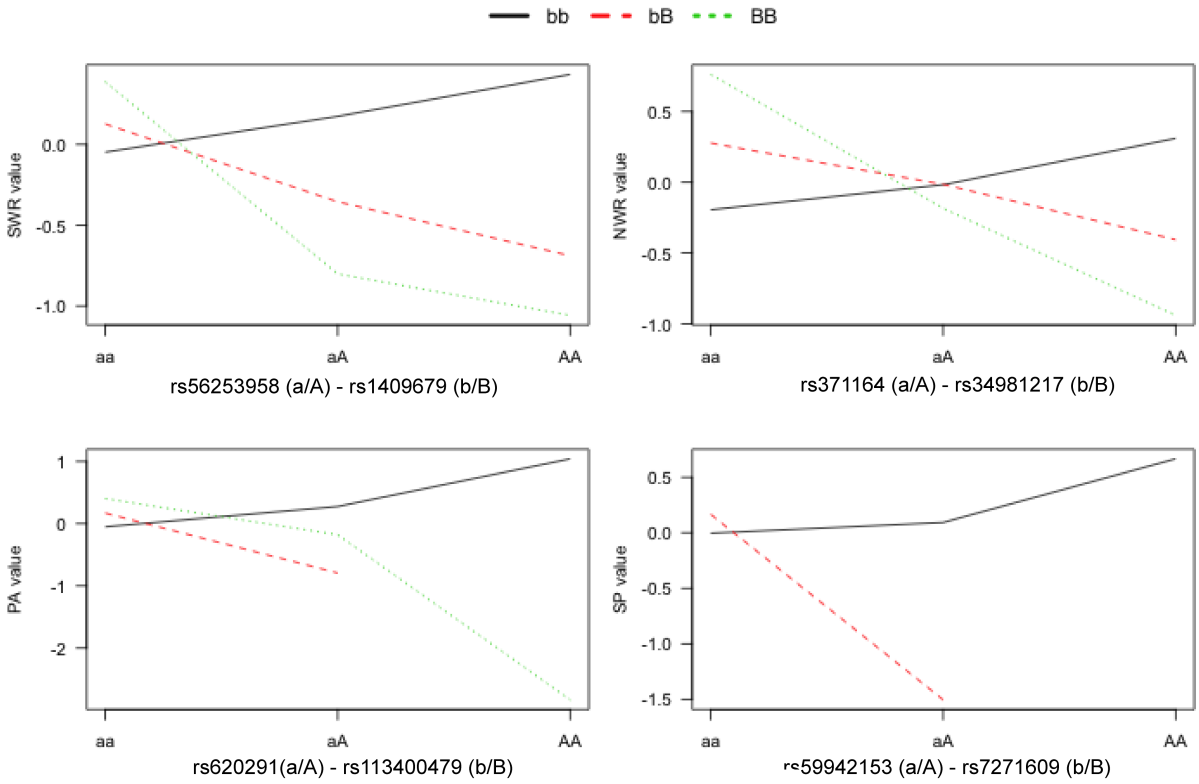


Figure 5.2: Molecular epistasis: two-locus models of all four top hits illustrate purely epistatic effects in the absence of main effects at either locus. The x-axis represent the three genotypes at one locus (aa (homozygote of major allele), aA (heterozygote), AA (homozygote of minor allele)) while the three lines indicate the different genotypes at the other locus with a black solid line for homozygote of major allele (bb), a red dashed line for heterozygote (bB), and a green dotted line for homozygote of minor allele (BB). The y-axis displays the respective measured endophenotype scores.

such disordinal interactions (non-marginal effects) is still unknown and debated con-

troversially. These interactions could either constitute rare events or be a consequence of compensatory evolution. During compensatory evolution, deleterious single effects are compensated by a second mutation of another allele, leading to double mutants. However, the observation of disordinal interactions might also simply be caused by a lack of detection ability [Hemani et al., 2014, Musani et al., 2007].
Interestingly, disordinal or pure epistasis has already been shown to affect diseases such as schizophrenia [Qin et al., 2005, Becker et al., 2005], and Alzheimer's [Martin et al., 2006], as well as many other phenotypes [Musani et al., 2007]. It is difficult to determine a specific, appropriate threshold for univariate associations in an epistasis model. Moreover, the question arises whether significant interactions should be rejected if no marginal effects are present. Since the computational abilities are given why not consider the whole genome.
As mentioned studies have demonstrated different genetic interactions with varying single-locus penetration models [Culverhouse et al., 2002, Musani et al., 2007]. They concluded that disregard of non-marginal effects leads to a substantial loss of power and information. Without non-marginal effects, small non-zero univariate trends affecting epistasis would never be considered and would therefore remain hidden. As a consequence, complex models, based on the interplay of genetic factors without any observable single effects that increase the susceptibility to disease, would have never been discovered, such as the aforementioned Alzheimer's disease, schizophrenia, or diabetes [Cox et al., 1999, Lee et al., 2004] and breast cancer [Abdi, 2006, Ritchie et al., 2001].
Given the current computational abilities, the whole genome can be considered for analysis. Recently, a hypothesis-free genome-wide interaction search on expression profiles of 846 individuals was published by Hemani et al., including a successful, subsequent replication in two additional sample sets. The authors detected significant cis-cis, cis-trans and trans-trans interactions on individual expression levels, with and without marginal effects of the involved loci. Compared to other quantitative traits, expression profiles are good candidates for such analyses, as they are highly biologically relevant, due to the massive effect size that explains large proportions of the genetic variance [Hemani et al., 2014].
However, the complexity of nature could harbor infinite concealed and branched interactions. Understanding the nature of genetic interactions and the possible effect of an allele in presence of other alleles in quantitative traits, is essential to uncover a more complete picture of the mechanisms of complex biological systems and their evolution.
In this work, a step towards unveiling the mystery of multifactorial heritability was undertaken, focusing on the highly heterogeneous phenotype dyslexia, leading to strikingly significant and promising results.

5.2 Statistical epistasis in dyslexia pathogenesis

Within this work, we present the first GWIA on endophenotypes associated with dyslexia. We performed four unbiased, hypothesis-free, and exhaustive epistasis scans. We refrained from limiting the scans to preselected genetic variants, as such an approach would strongly limit the possibilities for discovering completely new factors influencing the examined phenotypes, and facilitating new insights in the broad-sense heritability.
In a previous epistasis study on dyslexia-associated parameters, and enhancement of $1p36$ linkage was detected when linkage of $6p22.2$ was taken into account in dyslexic probands [Grigorenko et al., 2001]. These two regions constitute well-studied, dyslexia-associated candidate loci, showing both single and interaction effects.
Our analyses exposed several inter-chromosomal negative interactions displaying pure epistasis, which affected cognitive skills in dyslexic subjects. By contrast to the earlier study, the majority of our findings involved novel loci. We could not detect any significantly associated single main effects of these loci, but could allocate them to known susceptibility regions for either dyslexia or other comorbid disorders.
For example, variants on 18q11.2 could be assigned to an area containing putative dyslexia-associated loci in the vicinity of DYX6. Interestingly, variants on 7p31 could be allocated to the gene *FOXP2*, a factor that is well known in the field of psychiatric disorders, as it is involved in speech and language disorder (SLD) [MacDermot et al., 2005] and SLI [Smith, 2007]. Recent publications already indicated a role of *FOXP2* in dyslexia pathogenesis [Peter et al., 2010, Wilcke et al., 2011].
Developmental language and learning disorders, such as dyslexia, SLI, ADHD and autism spectrum disorder (ASD) are suggested to share a degree of genetic etiology. This is due to apprehended comorbidity, that influences the development of these phenotypes [Czamara et al., 2013, Germanò et al., 2010, Newbury and Monaco, 2010, Newbury et al., 2010, Reiersen and Todd, 2008, Smith, 2007]. Such cognitive disorders share a common genetic background or are linked to the same genomic regions. An example for such a shared aspect is the disturbance of neuronal migration [Smith, 2007]. In a study on the etiology of schizophrenia, combined effects of SNPs within the genes *GRIN1* and *GRIN2B* were discovered [Qin et al., 2005]. Polymorphisms of *GRIN2B* are also discovered to be associated with weak performance in vernal short-term memory in dyslexic children (see backgrounds 2.1.5.1). A relevant example to visualize the connectivity, the interplay and complexity of genetics as well as the possible comorbidity of psychiatric disorders.

5.2.1 Interactions involving the gene FOXP2

We detected a very strong interaction as well as a slight marginal effect of *FOXP2* variants on SWR. When taking a possible comorbidity of dyslexia with speech and language disturbances into account [Smith, 2007], the notion that *FOXP2* variants might influence

the development of dyslexia appears reasonable. The gene is highly conserved and is essential for the development of brain regions responsible for fine motor control [Newbury and Monaco, 2010]. In fact, it has been suggested that *FOXP2* -related disorders involve core deficits in learning, planning, and executing rapid movement sequences. *FOXP2* is also involved in language and learning processes and has frequently been studied in this context [Fisher and Scharff, 2009].
However, in the past it has proven difficult to discover associations of *FOXP2* to dyslexia by one-dimensional screening, which were significant on a genome-wide scale. Hence, *FOXP2* likely has no monogenic effect on dyslexia pathogenesis, as it is the case for SLD [Fisher and Scharff, 2009]. Instead, it is conceivable that a possible effect of the gene takes place through the interplay with other genetic factors or loci.
Our findings regarding statistical epistasis support the idea that *FOXP2* influences dyslexia via interactions with other loci. The results remained constant over the associated LD-sequence ($r^2 \geq 0.7$) on both chromosomes, indicating a robust outcome. Thus, it is unlikely that the data was confounded by genotyping inaccuracies or imputation artifacts.
To date, there are no studies describing the region on $9q31.3$ that we found to interact with *FOXP2*. Moreover, the top variant at this locus, rs1409679, is located in a LINE region (L1), a stretch of repetitive DNA. While this SNP seems to be assigned at the correct sequence position, a certain level of uncertainty always remains. The fraction of the genome showing poor confidence regarding the accuracy of the identified variations is called the "*"dark matter of mappability"* [Lee and Schatz, 2012]. Lee and Schatz reported that the majority of variations examined in the 1000 genomes project are well mapped, but that we have to keep in mind that approximately 50% of the human genome contain repetitive elements [Babushok and Kazazian, 2007, Lee and Schatz, 2012] . Nevertheless, in our case the detected loci appeared to be well mapped on the human genome.
Notably LINE1 elements are known to be the most active autonomous retrotransposons in mammalian genomes and variety in the number and activity levels of LINE1 contribute to variation in humans by affecting once the LINE1 mediated mutations, as well as the individual gene expression level [Babushok and Kazazian, 2007].
Notably, LINE1 elements are the most active autonomous retrotransposons in mammalian genomes. The variety in numbers and activity levels of LINE1 elements contribute to genetic variation in humans. They can mediate mutations and affect gene expression levels [Babushok and Kazazian, 2007].
However mapping our SWR SNPs to the functional ENCODE annotations we observed no DNaseI HS signals, but we could observe an enrichment of TF binding sites for the top *FOXP2* locus. This offers room for speculations regarding possible trans-cis regulations of *FOXP2*. A conceivable scenario is that through the allelic interaction of rs56253958 and rs1409679 transcriptional activities could be initiated or inhibited.

5.2.2 Interactions involving NCAM1

Another candidate gene identified in our analyses was *NCAM1*, coding for a transmembrane glycoprotein expressed in neurons. *NCAM1* regulates cell-cell adhesion, neurite outgrowth, synaptic plasticity, and memory processes [Markram et al., 2007, Schmid and Maness, 2008]. These processes play a crucial role in dyslexia [Dahdouh et al., 2009, Ludwig et al., 2010]. The gene might also influence genetics and pathophysiology of disorders like bipolarity and schizophrenia [Atz et al., 2007]. It has been suggested that schizophrenia and dyslexia share a common genetic background, but till date no association per se could be established. Studies have shown altered hemispheric asymmetry in schizophrenia patients, as well as in individuals with dyslexia [Heim et al., 2003]. In addition, GRINB2 affects both disorders, as mentioned before.
Interestingly the interaction we identified took place between *NCAM1* variants and polymorphisms at 18q11.2. The chromosomal region spanning the centromere from 18p11.2 − 8q12.2 harbors putative QTLs for PA and SWR [Fisher et al., 2002, Scerri et al., 2010], further supporting the validity of our finding.
When, mapping *NCAM1* SNPs to the functional ENCODE annotations, we observed an accumulation of the variants in presumed regulatory regions. The variants rs620291 and rs635596 appeared to be located within TF binding sites. Moreover, rs635596 maps to a ChIP-seq estimated binding site for RNA polymerase II Pol2-4H8. RNA polymerase II binds to the promoter of a gene to initiate transcription. The fact that our interacting variants are located within TF and polymerase binding sites supports the assumption that these SNPs have a regulatory function regarding the transcription of *NCAM1*. They might in turn be regulated by the interacting variants on 18q11.2.
Although we could not observe single associations (marginal effects) of these loci, our epistasis results are very robust. Furthermore, regarding the previously known functions or associations of each of the players, a relation between variants on 18q11.2 and *NCAM1*, which affects phonological abilities in dyslexic individuals, is likely.

5.2.3 Interactions involving TMEFF2

A third gene to be discovered in an interaction was *TMEFF2*, which is widely expressed throughout the brain. The protein promotes the survival of specific types of neurons, e. g., hippocampal and mesencephalic neurons [Horie et al., 2000]. However no evidence is yet present for a role of this gene in dyslexia. The interacting singleton SNP on 19p13.3 is located within a LTR element. It has been suggested that these elements have not been active in human over time, explainable by the lack of polymorphisms at these regions [Babushok and Kazazian, 2007]. Likewise, we could detect a lack of polymorphisms at the given region. However, we also observed DNaseI HS peaks and an enrichment of TF binding sites at the same locus, arguing in favor of a functional site.
Furthermore, the interacting intronic variant within the *TMEFF2* sequence also maps to a TF binding site. Of note, singleton SNPs are of a higher functional importance than

intronic or intergenic markers. They are assumed to be enriched in regions of higher functionality and recombination rate than non-singletons [Ke et al., 2008].
However, we examined the single locus association of rs371164 to expression of *TMEFF2* in the hippocampus without a significant outcome. Intriguingly, both regions are located either in-between or within loci that were suggested to be associated with autism [Smith, 2007]. As mentioned before, chromosome 2, besides chromosome 6, harbors the highes number of annotated loci related to reading disability RD.

5.2.4 Interaction involving two intergenic loci

Another inter-chromosomal connection was detected between intergenic variants on $17p13.1$ and $20q13.2$ which affected SP ability. No clear evidence linking the locus $20q13.2$ to dyslexia exists. However studies found deletions on $20q13.13 - 20q13.33$ to be associated with intellectual disabilities, seizures, ASD and expressive speech in a child with learning difficulties [Béna et al., 2007].
Two more studies detected effects of $17p13.1$ in dyslexia. Bates et al. reported of the region on fects irregular-word spelling [Bates et al., 2007]. A study found in a genome-wide screen associations of copy number variations (CNVs) at the region in ten dyslexic Indian families [Veerappa et al., 2013].
The variant rs59942153 is located in-between two DNaseI HS peaks, suggesting high functionality at this locus. Unfortunately, the variant demonstrated an insufficient LD-structure, so we were not able to test if any proxies with an adequate r^2 are located directly in a peak. The following SNP that could be mapped to annotated TF binding sites exhibited a r^2 of 0.67. The observation of the counterpart variants on $20q13.2$ were promising, they could be mapped to TF binding sites. An interpretation of the possible interaction scenario is difficult, as both variants are intergenic.

5.2.5 Expression levels of the identified genes

It is very promising that we could observe strong interactions between loci that are either connected to dyslexia, a comorbid disorder or to genes that are involved in functions of the NS. Yet although our results are encouraging, they are so far only descriptive, as the biological nature of the statistical interactions has to be validated. A first step in this direction was the annotation mapping on functional ENCODE annotations. This generally constitutes an option to examine whether loci of interest are located within any regulatory or functionally active sequence segments. With this approach, we were able to observe that six out of our eight candidate SNPs are located at TF binding sites and one SNP in a DNaseI HS indicating regulatory active regions.
We also attempted to examine the biological relevance of our findings by analyzing the association of the dyslexia candidate SNP-pairs to the gene expression within the hippocampus. This eQTL approach unfortunately failed. The *FOXP2* as well as three

NCAM1 probes all showed expression levels that did not pass quality control. One remaining *NCAM1* probe was expressed well (Illumina ID ILMN_1676289), yet was located more then 33 Kb away from the candidate SNPs. This could be an explanation for the absence of a detectable association between probe and SNP genotypes. However, we could detect a trend towards interaction between the genotypic marker rs2303377 and variants at the locus 18q11.2, which were associated with expression levels of the probe ILMN_2398184. This result is highly interesting, as this probe detects an adjacent transcript. In fact, this was the only probe showing at least a trend towards a connection between expression and genotypic interactions.
Nevertheless, we have to face the fact that, in general the measured expression levels and p-values were too weak to draw well-founded conclusions regarding possible epistatic effects, neither for *NCAM1* nor for *FOXP2* analyses.
Albeit expression levels of the *TMEFF2* transcript were sufficient, the candidate SNPs for the phenotype NWR were still not significantly associated. This could again be explained by the distance the loci and transcript ($\geq$ 220 Kb).
Therefore, our attempts failed to verify the dyslexia candidate SNPs by examining gene expression in the human hippocampus of patients receiving treatment against epilepsy. However, it is possible that *NCAM1* expression is down regulated in patients suffering from epilepsy, for instance due to medication. In fact, a recent publication indicated a role of *NCAM1* in epilepsy [Wang et al., 2012b]. Another study directly measured expression of the gene in the hippocampus of epileptic rats receiving medication, with the observation that the medication induced downregulation of *NCAM1*, with an anti-epileptic effect [Wang et al., 2010].
In the case of *FOXP2*, no sufficient expression profiles were detectable in the hippocampal tissue. We thus wondered whether, in general, *FOXP2* shows high expression levels in the hippocampus. Several studies have addressed this question [Lai, 2003, Takahashi et al., 2003], with the results that *FOXP2* seems to show either no expression at all or very low expression in the developing or mature hippocampus. Instead, expression appears to take place predominately in areas that are involved in motor control, for example in the thalamus or the cerebellum [Lai, 2003]. Lai et al. deduced that *FOXP2* transcription appears to be highly regulated during development of the central nervous system (CNS), both spatially and temporally.
We therefore conclude that, for our candidate genes, the hippocampal samples may not have been representative. Furthermore, the expression patterns of the genes may have been influenced by anti-epileptic medication.
A statistical explanation for not being able to identify associations in the expression analysis was the limited sample size of 138 individuals. We therefore computed the detection power for different sample sizes using [Gauderman, 2002]. As expected, our data offered too low power for an appropriate statistical epistasis analysis. The actually used sample size gave a power of 0.0006, power rose with sample size to up to 90% for a 10 fold larger sample.

5.2.6 Pathway analyses

The attempt to understand the elemental functional structure of our results using pathway analyses was more successful. In fact, we observed very reasonable pathways showing enrichment of our candidates (see results 4.2.10). For example, SWR-associated genes were enriched in pathways important for cerebral cortex development, a brain area showing *FOXP2* expression [Takahashi et al., 2003] and involved in learning processes. Furthermore, we observed an overrepresentation of genes involved in acetylcholine and argin pathways. Acetylcholine, a major excitatory neurotransmitter, plays an important role during learning and memory processes [Hasselmo, 2006]. The transmitter functions through the activation of mechanisms for persistent spiking of cortical neurons, a vital process for the maintenance of novel information and its encoding into long-term memory, as well as by stimulating synaptic plasticity. Moreover, components of acetylcholine pathways are highly expressed in brain areas such as the cortex and hippocampus, which are important for attention, learning, and memory [Hasselmo, 2006]. Argin, in turn, is a factor important for postsynaptic differentiation at neuromuscular junction (NMJ) and interacts with acetylcholine receptors [Bezakova and Ruegg, 2003].
We also observed an over-presentation of pathways connected to tyrosine kinase signaling and corresponding receptors, which are called neurotrophins. These growth factors are involved in the survival, development, and function of neurons among others in the CNS [Reichardt, 2006] and are also attributed to have a function in learning and memory [Yamada and Nabeshima, 2003].
Also a connection of our candidates to sugar metabolism was found, as we could observe enrichments of mocosaccharide-, N-acetylglucosamine, and carbohydrate metabolic processes. The latter is involved in energy metabolism during neuronal activation [Richards et al., 1999]. Interestingly, previous publications already indicated a role of a neuronal glucose transporter in dyslexia phenotypes [Roeske et al., 2011]. Furthermore, Richards and colleagues measured differences in lactate concentration, important for carbohydrate metabolism, between dyslexic cases and controls during a phonological task.
In contrast to the enrichment of tyrosine kinase activities for the phenotype SWR, we observed an over-representation of serine/threonine kinase activities for the analyses of PA. Gene-set enrichment analyses (GSEA) indicated their functions to lie within inflammation-related processes, e.g. the complement cascade, the lectin pathway, and response reactions to acute-phase stress as well as to oxidative stress. Serine/threonine protein kinases are found in high concentrations in neural tissues and are important for signal transduction, neurotransmitter synthesis, ion channel properties, synaptic plasticity, and ATP phosphorylation [Scott and Soderling, 1991]. It is, however, unclear to which extent inflammation and serine/threonine protein kinase activities might contribute to dyslexia. At the moment, we can merely speculate that unusual inflammations might exist in dyslexic brains, as is the case in autism, where local inflammation

contributes to the pathogenesis [Manto and Jissendi, 2012]. In addition, a shift of serine/threonine protein kinase activities could have a possible impact on subjects by modifying either ion channels or neurotransmitter synthesis, which, in turn, would affect cognitive abilities.
GSEA of the phenotype SP revealed a significant enrichment of processes related to axonal guidance signaling as well as of pathways within the microtubule-based flagellum, a specialization important for cell migration. The latter process is disturbed in dyslexia subjects [Smith, 2007].
When we finally considered all genes in a joint analysis, we obtained molecular functions and biological processes highly relevant in the context of dyslexia. Examples are genes involved in the development of the nervous system in general as well as of the cerebral cortex in particular, and genes relevant for locomotor behavior and for axonal guidance. In both the joint and the endophenotype-specific analyses, we observed a frequent appearance of components involved in signaling or transport activities. It is thus a possibility that transmembrane signaling plays a fundamental role in processes implicated in development and maintenance of cognitive skills. In general, enrichment analyses can help to illustrate whether certain candidate genes are enriched in pathways relevant in the context of the examined phenotypes. Nevertheless, it remains difficult to interpret the results of these analyses from a disease/disorder-focused perspective, especially in the case of epistasis.
Clearly, epistasis does exist and, with the recent computational progress, genome-wide search for interactions has become possible. However, it is still a huge challenge to validate the biological context of identified statistical connections, especially if opportunities for laboratory work are limited. We examined our statistically obtained gene-set using several approaches, having mixed success. Although we obtained highly relevant loci and pathways, it would require additional studies in order to be able to draw more reliable conclusions. Albeit our results should be regarded as descriptive, we nevertheless retrieved more significant and relevant hits for the pathogenesis of dyslexia than previous single association studies.
With this work we attempted to increase the knowledge on multifactorial genetics of dyslexia and its corresponding endophenotypes. We made striking and reasonable findings, which motivates us to continue with multivariate analyses for other questions in the wide field of complex traits.

5.3 Conclusion and outlook

With this thesis, we aimed to extend our understanding of statistical epistasis and to expose complex synergetic factors underlying susceptibility to developmental dyslexia. We conducted genome-wide exhaustive searches in a combined sample of several dyslexic cohorts using a hypothesis-free approach. From a statistical perspective on epistasis, we successfully uncovered substantially relevant interactions throughout the genome. However, from a biological perspective, we could not draw well-founded conclusions yet, even though the identified interacting loci appear to have biologically reasonable functions.
There is no doubt that epistasis plays an important role in nature. Yet although the computational burden is now approachable either with multicore machines, the advancement of efficient mathematical methodologies, or, as in our case, by employing parallelized graphics cards, the interpretation of the biological relevance of the thus obtained results remains to be a challenge.
An important next step in the study of genetical interplay is finding appropriate solutions for proving the relevant underlying biological epistasis. So far, few studies were able to provide biological evidence for their statistical discoveries with functional analyses [Gregersen et al., 2006, Nicodemus et al., 2010b, Van Steen, 2012].

Even though we were able to overcome several burdens from the computational side, such as the time required for the calculations, epistasis still harbors a wide field of unanswered and non-considered questions. For example, genetic factors in nature might not always interact in a linear manner and we might therefore not capture all possible forms of epistasis with our linear-regression approach. This is the case although the regression-based method is the most straightforward and natural approach available. Enhancements towards unveiling other possible pattern of interactions would require alternative approaches to capture nonlinearities, as for example support vector machine (SVM), multifactor dimension reduction (MDR), or decision tress. Unfortunately, these models are usually very time-consuming due to their computational intensity, or are partially not applicable to quantitative phenotypes [Musani et al., 2007].
Next to the incapacity of regression models to capture nonlinearity, another flaw of the approach could be the problem of overfitting, caused by incomplete data. To avoid such effects, quality control steps were performed especially accurately.
The analysis of epistasis performed here only takes interactions between two loci into account. Another extension of interaction models that is discussed and approached frequently, is the effort to capture higher-order interactions of three or even four loci. However, this would require for more sophisticated algorithms and hardware capabilities in order to deal with the increased computational demands.
Apart from computational considerations, an additional weak point of our study remains, as in most similar studies, the low sample size. This might constitute the reason

why many relevant interactions did not exceed the threshold for significance when correcting for multiple testing. As a solution for this issue, a replication study with a greater sample-set could possibly identify a larger number of interactions significant on a genome-wide scale and also confirm already identified loci. We see a high potential in both our approach and the examined phenotype of dyslexia, due to its heterogeneity and multifacetedness of cognitive manifestations. Therefore, we aim at the collection of a greater, polylingual sample for such a replication study, composed of thousands of cases.

In summary, our approach of exhaustive epistasis calculation is a promising route going beyond single loci detection, especially in such multifactorial phenotypes as dyslexia. The near future will offer state of the art capabilities in the wide area of interaction computation for any kind of omics data, which could help to improve clinical, personalized approaches. These could help to answer questions regarding functionality and shed light upon the often discussed necessity of single-locus penetrance in an interacting complex (marginal effects).

Summarized, the approach for exhaustive epistasis calculation, specially in such heterogeneous and multifactorial phenotypes such as dyslexia, is a promising way beyond single loci detection. The near future will offer state of the art capabilities in the wide area of interaction computation for any kind of omics data, which could help to improve clinical personalized targeted approaches, to answer the questions around functionality, and unveil the often discussed necessity of single-locus penetrance in an interacting complex.

As was the case for GWASs, also GWIAs will require time to get established and refined in its methodology as well as in its perception by scientists. We believe that, where the time required for the analysis is not essential, it is well worth to take a hypothesis-free look on the complex nature of genetics from an epistatic point of view. This might yield more elaborate results on the so-far still often unknown connections between genetics and human physiology.

Zusammenfassung

Ziel dieser Arbeit war die Identifizierung allelischer Interaktionen, die als genetische Suszeptibilitätsfaktoren mit der Entwicklung der Lese- und Rechtschreibstörung Dyslexie assoziiert sind. Dyslexie ist eine weitverbreitete Störung, die durch ein Defizit in Lese-und Rechtschreibfähigkeiten charakterisiert ist [Schulte-Körne et al., 2001]. Innerhalb des klinischen Bildes der Dyslexie können verschiedene Endophänotypen subkategorisiert werden. Aufgrund der teilweise hierarchisch aufgebauten Struktur dieser Endophenotypen [Carrion-Castillo et al., 2013] liegt die Vermutung nahe, dass die zugrunde liegende genetische Architektur eine große Heterogenität aufweist. Da bislang nur wenige genetische Suszeptibilitätsfaktoren identifiziert wurden, stellte sich die Frage, ob in diese Prozesse genetische Interaktionen involviert sein könnten.
Bislang wurden genetische Interaktionen bei der Erforschung der Dyslexie, so wie komplexer Erkrankungen im Allgemeinen, weitestgehend vernachlässigt. Frühere Studien konzentrierten sich meist auf die Identifizierung einzelner krankheitsassoziierter Polymorphismen. Die Effekte dieser Einzelloci erklären jedoch meist nur einen geringen Teil der Erblichkeit. Es wird daher vermutet, dass ein wesentlicher Bestandteil der Heredität durch Interaktionen zwischen genetischen Varianten erklärt werden könnte. Genetische Studien wurden in der Vergangenheit auf kandidatengenbasierte Ansätze limitiert bedingt durch die hohe Rechenzeit und -kapazität die Interaktionsstudien beanspruchen. Durch den Fortschritt in der Computertechnologie und die Entwicklung neuer Algorithmen wurde uns hier erstmalig die Möglichkeit zur Realisation einer genomweiten Interaktionsanalyse gegeben.
Die Dyslexie mit all ihren Endophänotypen erwies sich als sehr geeignet für die Suche nach epistatischen Effekten auf genomweiter Ebene. Das Genom von 862 dyslexischen Kindern unterschiedlicher ethnischer Zugehörigkeit (Deutschland, USA, Großbritannien und Frankreich) und die in verschiedenen psychometrischen Tests erfassten individuellen Endophänotypen Einzelwort-Lesen, Rechtschreibung, phonologische Wahrnehmung und Nicht-Wort- Lesen bildeten die Datenbasis einer Zwei-Locus Interaktionsanalyse mit der Software GLIDE [Kam-Thong et al., 2012].
Wir waren damit in der Lage, starke epistatische Effekte bei verschiedenen Endophänotypen der Dyslexie statistisch nachzuweisen. Die Ergebnisse enthielten höchst relevante Loci innerhalb des Genoms, die neue und bereits bekannte Dyslexiekandidaten repräsentieren. Ein Beispiel ist die Interaktion zwischen Varianten auf 18q11.2, einer Region der mit phonologischer Wahrnehmungsstörung und Einzelwort-Lesen assoziiert ist, und intronischen Varianten des *NCAM1* Gens, das an der Entwicklung des Nervensystems

beteiligt ist. Weitere statistisch signifikante Interaktionen, wurden zwischen intergenischen allelischen Varianten auf Chromosom 9 und intronischen Polymorphismen von *FOXP2* gefunden, einem Gen, das mit Sprachstörungen assoziiert ist und dessen Rolle in der Ätiologie von Dyslexie erst kürzlich beschrieben wurde [Wilcke et al., 2011].
In Anbetracht des komplexen Aufbaus von genetischen Prozessen und der Allgegenwärtigkeit von epistatischen Reaktionen sollten Multi-Locus-Effekte in der Zukunft noch stärker in den Fokus der Wissenschaft rücken, für ein besseres Verständnis von komplexen Phänotypen und genetischer Vererbung.

Bibliography

[Abdi, 2006] Abdi, H. The Bonferonni and Šidák Corrections for Multiple Comparisons. *In Salkind, N. J. Encyclopedia of Measurement and Statistics. Thousand Oaks, CA: Sage.* (2006) 1–9.

[Amaral et al., 2008] Amaral, A. J., Megens, H. J., Crooijmans, R. P. M. A., Heuven, H. C. M., and Groenen, M. A. M. Linkage Disequilibrium Decay and Haplotype Block Structure in the Pig. *Genetics*, 179 (2008) 569–579.

[Anthoni et al., 2007] Anthoni, H., Zucchelli, M., Matsson, H., Müller-Myhsok, B., Fransson, I., Schumacher, J., Massinen, S., Onkamo, P., Warnke, A., Griesemann, H., Hoffmann, P., Nopola-Hemmi, J., Lyytinen, H., Schulte-Körne, G., Kere, J., Nöthen, M. M., and Peyrard-Janvid, M. A locus on 2p12 containing the co-regulated MRPL19 and C2ORF3 genes is associated to dyslexia. *Human Molecular Genetics*, 16 (2007) 667–677.

[Ashworth et al., 2011] Ashworth, A., Lord, C. J., and Reis-Filho, J. S. Genetic Interactions in Cancer Progression and Treatment. *Cell*, 145 (2011) 30–38.

[Atz et al., 2007] Atz, M. E., Rollins, B., and Vawter, M. P. NCAM1 association study of bipolar disorder and schizophrenia: polymorphisms and alternatively spliced isoforms lead to similarities and differences. *Psychiatric genetics*, 17 (2007) 55.

[Aulchenko et al., 2007] Aulchenko, Y. S., Ripke, S., Isaacs, A., and van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics (Oxford, England)*, 23 (2007) 1294–1296.

[Babushok and Kazazian, 2007] Babushok, D. V. and Kazazian, H. H. Progress in understanding the biology of the human mutagen LINE-1. *Human Mutation*, 28 (2007) 527–539.

[Bakwin, 2008] Bakwin, H. Reading Disability in Twins. *Developmental Medicine & Child Neurology*, 15 (2008) 184–187.

[Bates et al., 2007] Bates, T. C., Luciano, M., Castles, A., Coltheart, M., Wright, M. J., and Martin, N. G. Replication of reported linkages for dyslexia and spelling and suggestive evidence for novel regions on chromosomes 4 and 17. *European Journal of Human Genetics*, 15 (2007) 194–203.

[Bateson, 1909] Bateson. Discussion on the Influence of Heredity on Disease, with special Reference to Tuberculosis, Cancer, and Diseases of the Nervous System: Introductory Address. *Proceedings of the Royal Society of Medicine*, 2 (1909) 22–30.

[Becker et al., 2010] Becker, T., Herold, C., Meesters, C., Mattheisen, M., and Baur, M. P. Significance Levels in Genome-Wide Interaction Analysis (GWIA). *Annals of Human Genetics*, 75 (2010) 29–35.

[Becker et al., 2005] Becker, T., Schumacher, J., Cichon, S., Baur, M. P., and Knapp, M. Haplotype interaction analysis of unlinked regions. *Genetic Epidemiology*, 29 (2005) 313–322.

[Béna et al., 2007] Béna, F., Bottani, A., Marcelli, F., Sizonenko, L. D., Conrad, B., and Dahoun, S. A de novo 1.1-1.6 Mb subtelomeric deletion of chromosome 20q13.33 in a patient with learning difficulties but without obvious dysmorphic features. *American Journal of Medical Genetics Part A*, 143A (2007) 1894–1899.

[Benjamini and Hochberg, 1995] Benjamini, Y. and Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* (1995) 289–300.

[Bezakova and Ruegg, 2003] Bezakova, G. and Ruegg, M. A. New insights into the roles of agrin. *Nature reviews. Molecular cell biology*, 4 (2003) 295–308.

[Birney et al., 2007] Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Stamatoyannopoulos, J. A., Thurman, R. E., Kuehn, M. S., Taylor, C. M., Neph, S., Koch, C. M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J. A., Andrews, R. M., Flicek, P., Boyle, P. J., Cao, H., Carter, N. P., Clelland, G. K., Davis, S., Day, N., Dhami, P., Dillon, S. C., Dorschner, M. O., Fiegler, H., Giresi, P. G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K. D., Johnson, B. E., Johnson, E. M., Frum, T. T., Rosenzweig, E. R., Karnani, N., Lee, K., Lefebvre, G. C., Navas, P. A., Neri, F., Parker, S. C. J., Sabo, P. J., Sandstrom, R., Shafer, A., Vetrie, D., Weaver, M., Wilcox, S., Yu1, M., Collins, F. S., Dekker, J., Lieb, J. D., Tullius, T. D., Crawford, G. E., Sunyaev, S., Noble, W. S., Dunham, I., Dutta, A., Guigó, R., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I. L., Baertsch, R., Keefe, D., Flicek, P., Dike, S., Cheng, J., Hirsch, H. A., Sekinger, E. A., Lagarde, J., Abril, J. F., Shahab, A., Flamm, C., Fried, C., Hackermüller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J. S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M. C., Thomas, D. J., Weirauch, M. T., Gilbert, J., Drenkow, J., Bell, I., Zhao, X., Srinivasan, K. G., Sung, W.-K., Ooi, H. S., Chiu, K. P., Foissac, S., Alioto, T., Brent, M., Pachter, L., Tress, M. L., Valencia, A., Choo, S. W., Choo, C. Y., Ucla, C., Manzano, C., Wyss, C., Cheung, E., Clark, T. G., Brown, J. B., Ganesh, M., Patel,

S., Tammana, H., Chrast, J., Henrichsen, C. N., Kai, C., Kawai, J., Nagalakshmi, U., Wu, J., Lian, Z., Lian, J., Newburger, P., Zhang, X., Bickel, P., Mattick, J. S., Carninci, P., Hayashizaki, Y., Weissman, S., Dermitzakis, E. T., Margulies, E. H., Hubbard, T., Myers, R. M., Rogers, J., Stadler, P. F., Lowe, T. M., Wei, C.-L., Snyder, M., Snyder, M., Birney, E., Struhl, K., Gerstein, M., Antonarakis, S. E., Gingeras, T. R., Brown, J. B., Flicek, P., Fu, Y., Keefe, D., Birney, E., Denoeud, F., Gerstein, M., Green, E. D., Kapranov, P., Karaöz, U., Myers, R. M., Noble, W. S., Reymond, A., Rozowsky, J., Struhl, K., Siepel, A., Stamatoyannopoulos, J. A., Taylor, C. M., Taylor, J., Thurman, R. E., Tullius, T. D., Washietl, S., Zheng, D., Liefer, L. A., Wetterstrand, K. A., Good, P. J., Feingold, E. A., Guyer, M. S., Collins, F. S., Margulies, E. H., Cooper, G. M., Asimenos, G., Thomas, D. J., Dewey, C. N., Siepel, A., Birney, E., Keefe, D., Hou, M., Taylor, J., Nikolaev, S., Montoya-Burgos, J. I., Löytynoja, A., Whelan, S., Pardi, F., Massingham, T., Brown, J. B., Huang, H., Zhang, N. R., Bickel, P., Holmes, I., Mullikin, J. C., Ureta-Vidal, A., Paten, B., Seringhaus, M., Church, D., Rosenbloom, K., Kent, W. J., Stone, E. A., Sequencing Program, N. C., Human Genome Sequencing Center, B. C. o. M., Genome Sequencing Center, W. U., Institute, B., Oakland Research Institute, C. H., Gerstein, M., Antonarakis, S. E., Batzoglou, S., Goldman, N., Hardison, R. C., Haussler, D., Miller, W., Pachter, L., Green, E. D., Sidow, A., Weng, Z., Trinklein, N. D., Fu, Y., Zhang, Z. D., Karaöz, U., Barrera, L., Stuart, R., Zheng, D., Ghosh, S., Flicek, P., King, D. C., Taylor, J., Ameur, A., Enroth, S., Bieda, M. C., Koch, C. M., Hirsch, H. A., Wei, C.-L., Cheng, J., Kim, J., Bhinge, A. A., Giresi, P. G., Jiang, N., Liu, J., Yao, F., Sung, W.-K., Chiu, K. P., Vega, V. B., Lee, C. W. H., Ng, P., Shahab, A., Sekinger, E. A., Yang, A., Moqtaderi, Z., Zhu, Z., Xu, X., Squazzo, S., Oberley, M. J., Inman, D., Singer, M. A., Richmond, T. A., Munn, K. J., Rada-Iglesias, A., Wallerman, O., Komorowski, J., Clelland, G. K., Wilcox, S., Dillon, S. C., Andrews, R. M., Fowler, J. C., Couttet, P., James, K. D., Lefebvre, G. C., Bruce, A. W., Dovey, O. M., Ellis, P. D., Dhami, P., Langford, C. F., Carter, N. P., Vetrie, D., Kapranov, P., Nix, D. A., Bell, I., Patel, S., Rozowsky, J., Euskirchen, G., Hartman, S., Lian, J., Wu, J., Urban, A. E., Kraus, P., Van Calcar, S., Heintzman, N., Hoon Kim, T., Wang, K., Qu, C., Hon, G., Luna, R., Glass, C. K., Rosenfeld, M. G., Aldred, S. F., Cooper, S. J., Halees, A., Lin, J. M., Shulha, H. P., Zhang, X., Xu, M., Haidar, J. N. S., Yu, Y., Birney, E., Weissman, S., Ruan, Y., Lieb, J. D., Iyer, V. R., Green, R. D., Gingeras, T. R., Wadelius, C., Dunham, I., Struhl, K., Hardison, R. C., Gerstein, M., Farnham, P. J., Myers, R. M., Ren, B., Snyder, M., Thomas, D. J., Rosenbloom, K., Harte, R. A., Hinrichs, A. S., Trumbower, H., Clawson, H., Hillman-Jackson, J., Zweig, A. S., Smith, K., Thakkapallayil, A., Barber, G., Kuhn, R. M., Karolchik, D., Haussler, D., Kent, W. J., Dermitzakis, E. T., Armengol, L., Bird, C. P., Clark, T. G., Cooper, G. M., de Bakker, P. I. W., Kern, A. D., Lopez-Bigas, N., Martin, J. D., Stranger, B. E., Thomas, D. J., Woodroffe, A., Batzoglou, S., Davydov, E., Dimas, A., Eyras, E., Hallgrímsdóttir, I. B., Hardison, R. C., Huppert, J., Sidow, A., Taylor, J., Trumbower, H., Zody, M. C., Guigó, R., Mullikin, J. C., Abecasis, G. R., Estivill, X., Birney, E., Bouffard, G. G., Guan, X., Hansen, N. F., Idol, J. R., Maduro, V.

V. B., Maskeri, B., McDowell, J. C., Park, M., Thomas, P. J., Young, A. C., Blakesley, R. W., Muzny, D. M., Sodergren, E., Wheeler, D. A., Worley, K. C., Jiang, H., Weinstock, G. M., Gibbs, R. A., Graves, T., Fulton, R., Mardis, E. R., Wilson, R. K., Clamp, M., Cuff, J., Gnerre, S., Jaffe, D. B., Chang, J. L., Lindblad-Toh, K., Lander, E. S., Koriabine, M., Nefedov, M., Osoegawa, K., Yoshinaga, Y., Zhu, B., and de Jong, P. J. Identification and analysis of functional elements in 1human genome by the ENCODE pilot project. *Nature*, 447 (2007) 799–816.

[Bloom et al., 2013] Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T.-L. V., and Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature*, 494 (2013) 234–237.

[Buckner et al., 2010] Buckner, J., Wilson, J., Seligman, M., Athey, B., Watson, S., and Meng, F. The gputools package enables GPU computing in R. *Bioinformatics (Oxford, England)*, 26 (2010) 134–135.

[Carrion-Castillo et al., 2013] Carrion-Castillo, A., Franke, B., and Fisher, S. E. Molecular genetics of dyslexia: an overview. *Dyslexia (Chichester, England)*, 19 (2013) 214–240.

[Clayton, 2009] Clayton, D. G. Prediction and interaction in complex disease genetics: experience in type 1 diabetes. *PLoS genetics*, 5 (2009) e1000540.

[Compton et al., 2001] Compton, D. L., DeFries, J. C., and Olson, R. K. Are RAN- and phonological awareness-deficits additive in children with reading disabilities? *Dyslexia (Chichester, England)*, 7 (2001) 125–149.

[Cordell, 2009] Cordell, H. J. Detecting gene-gene interactions that underlie human diseases. *Nature Reviews Genetics*, 10 (2009) 392–404.

[Core, 2008] Core, T. R. D. (2008). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

[Cox et al., 1999] Cox, N. J., Frigge, M., Nicolae, D. L., Concannon, P., Hanis, C. L., Bell, G. I., and Kong, A. Loci on chromosomes 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans. *Nature genetics*, 21 (1999) 213–215.

[Culverhouse et al., 2002] Culverhouse, R., Suarez, B. K., Lin, J., and Reich, T. A perspective on epistasis: limits of models displaying no main effect. *American journal of human genetics*, 70 (2002) 461–471.

[Culverhouse, 2012] Culverhouse, R. C. A Comparison of Methods Sensitive to Interactions With Small Main Effects. *Genetic Epidemiology*, 36 (2012) 303–311.

[Czamara et al., 2013] Czamara, D., Tiesler, C. M. T., Kohlböck, G., Berdel, D., Hoffmann, B., Bauer, C.-P., Koletzko, S., Schaaf, B., Lehmann, I., Herbarth, O., von Berg, A.,

Müller-Myhsok, B., Schulte-Körne, G., and Heinrich, J. Children with ADHD symptoms have a higher risk for reading, spelling and math difficulties in the GINIplus and LISAplus cohort studies. *PloS one*, 8 (2013) e63859.

[Dahdouh et al., 2009] Dahdouh, F., Anthoni, H., Tapia-Páez, I., Peyrard-Janvid, M., Schulte-Körne, G., Warnke, A., Remschmidt, H., Ziegler, A., Kere, J., Müller-Myhsok, B., Nöthen, M. M., Schumacher, J., and Zucchelli, M. Further evidence for DYX1C1 as a susceptibility factor for dyslexia. *Psychiatric genetics*, 19 (2009) 59–63.

[de Kovel et al., 2004] de Kovel, C. G. F., Hol, F. A., Heister, J. G. A. M., Willemen, J. J. H. T., Sandkuijl, L. A., Franke, B., and Padberg, G. W. Genomewide scan identifies susceptibility locus for dyslexia on Xq27 in an extended Dutch family. *Journal of Medical Genetics*, 41 (2004) 652–657.

[Deak et al., 2005] Deak, K. L., Boyles, A. L., Etchevers, H. C., Melvin, E. C., Siegel, D. G., Graham, F. L., Slifer, S. H., Enterline, D. S., George, T. M., Vekemans, M., McClay, D., Bassuk, A. G., Kessler, J. A., Linney, E., Gilbert, J. R., and Speer, M. C. SNPs in the neural cell adhesion molecule 1 gene (NCAM1) may be associated with human neural tube defects. *Human genetics*, 117 (2005) 133–142.

[Fagerheim et al., 1999] Fagerheim, T., Raeymaekers, P., Tønnessen, F. E., Pedersen, M., Tranebjaerg, L., and Lubs, H. A. A new gene (DYX3) for dyslexia is located on chromosome 2. *Journal of Medical Genetics*, 36 (1999) 664–669.

[Fields, 2007] Fields, S. Molecular biology. Site-seeing by sequencing. *Science (New York, N.Y.)*, 316 (2007) 1441–1442.

[Fisher and DeFries, 2002] Fisher, S. E. and DeFries, J. C. Developmental dyslexia: genetic dissection of a complex cognitive trait. *Nature reviews. Neuroscience*, 3 (2002) 767–780.

[Fisher et al., 2002] Fisher, S. E., Francks, C., Marlow, A. J., MacPhie, I. L., Newbury, D. F., Cardon, L. R., Ishikawa-Brush, Y., Richardson, A. J., Talcott, J. B., Gayán, J., Olson, R. K., Pennington, B. F., Smith, S. D., DeFries, J. C., Stein, J. F., and Monaco, A. P. Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. *Nature genetics*, 30 (2002) 86–91.

[Fisher and Scharff, 2009] Fisher, S. E. and Scharff, C. FOXP2 as a molecular window into speech and language. *Trends in Genetics*, 25 (2009) 166–177.

[Fisher, 1918] Fisher, S. R. A. (1918). The Correlation Between Relatives on the Supposition of Mendelian Inheritance. Royal Society of Edinb.

[Francks et al., 2004] Francks, C., Paracchini, S., Smith, S. D., Richardson, A. J., Scerri, T. S., Cardon, L. R., Marlow, A. J., MacPhie, I. L., Walter, J., Pennington, B. F., Fisher,

S. E., Olson, R. K., DeFries, J. C., Stein, J. F., and Monaco, A. P. A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. *The American Journal of Human Genetics*, 75 (2004) 1046–1058.

[Frankel and Schork, 1996] Frankel, W. N. and Schork, N. J. Who's afraid of epistasis? *Nature genetics*, 14 (1996) 371–373.

[Galaburda, 1999] Galaburda, A. M. Developmental dyslexia: A multilevel syndrome. *Dyslexia (Chichester, England)*, *Dyslexia (Chichester, England)* (1999).

[Garret et al., 1997] Garret, A. S., Wood, F. B., Flowers, D. L., and Absher, J. R. Glucose metabolism in the inferior temporal cortex is related to accuracy of performance on a letter recognition task. *Unpublished manuscript* (1997).

[Gauderman, 2002] Gauderman, W. J. Sample Size Requirements for Association Studies of Gene-Gene Interaction. *American Journal of Epidemiology*, 155 (2002) 478–484.

[Gayán and Olson, 2003] Gayán, J. and Olson, R. K. Genetic and environmental influences on individual differences in printed word recognition. *Journal of Experimental Child Psychology*, 84 (2003) 97–123.

[Gentleman et al., 2004] Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y., and Zhang, J. Genome Biology | Full text | Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*, 5 (2004) R80.

[Germanò et al., 2010] Germanò, E., Gagliano, A., and Curatolo, P. Comorbidity of ADHD and dyslexia. *Developmental neuropsychology*, 35 (2010) 475–493.

[Gerstein et al., 2012] Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K.-K., Cheng, C., Mu, X. J., Khurana, E., Rozowsky, J., Alexander, R., Min, R., Alves, P., Abyzov, A., Addleman, N., Bhardwaj, N., Boyle, A. P., Cayting, P., Charos, A., Chen, D. Z., Cheng, Y., Clarke, D., Eastman, C., Euskirchen, G., Frietze, S., Fu, Y., Gertz, J., Grubert, F., Harmanci, A., Jain, P., Kasowski, M., Lacroute, P., Leng, J., Lian, J., Monahan, H., O'Geen, H., Ouyang, Z., Partridge, E. C., Patacsil, D., Pauli, F., Raha, D., Ramirez, L., Reddy, T. E., Reed, B., Shi, M., Slifer, T., Wang, J., Wu, L., Yang, X., Yip, K. Y., Zilberman-Schapira, G., Batzoglou, S., Sidow, A., Farnham, P. J., Myers, R. M., Weissman, S. M., and Snyder, M. Architecture of the human regulatory network derived from ENCODE data. *Nature*, 489 (2012) 91–100.

[Gibson and Gruen, 2008] Gibson, C. J. and Gruen, J. R. The human lexinome: Genes of language and reading. *Journal of Communication Disorders*, 41 (2008) 409–420.

[Gottesman and Gould, 2003] Gottesman, I. I. and Gould, T. D. The endophenotype concept in psychiatry: etymology and strategic intentions. *American Journal of Psychiatry*, 160 (2003) 636–645.

[Gregersen et al., 2006] Gregersen, J. W., Kranc, K. R., Ke, X., Svendsen, P., Madsen, L. S., Thomsen, A. R., Cardon, L. R., Bell, J. I., and Fugger, L. Functional epistasis on a common MHC haplotype associated with multiple sclerosis. *Nature*, 443 (2006) 574–577.

[Gretton et al., 2007] Gretton, A., Borgwardt, K. M., Rasch, M., Schölkopf, B., and Smola, A. J. A kernel method for the two-sample-problem. *Journal of Machine Learning Research 1*, 19 (2007) 513.

[Gretton et al., 2005] Gretton, A., Bousquet, O., Smola, A., and Schölkopf, B. (2005). Measuring Statistical Dependence with Hilbert-Schmidt Norms. Springer Berlin Heidelberg.

[Grigorenko, 2001] Grigorenko, E. L. Developmental dyslexia: an update on genes, brains, and environments. *Journal of child psychology and psychiatry, and allied disciplines*, 42 (2001) 91–125.

[Grigorenko, 2005] Grigorenko, E. L. A Conservative Meta-Analysis of Linkage and Linkage-Association Studies of Developmental Dyslexia. *Scientific Studies of Reading*, 9 (2005) 285–316.

[Grigorenko et al., 2001] Grigorenko, E. L., Wood, F. B., Meyer, M. S., Pauls, J. E., Hart, L. A., and Pauls, D. L. Linkage studies suggest a possible locus for developmental dyslexia on chromosome 1p. *American journal of medical genetics*, 105 (2001) 120–129.

[Gros et al., 2009] Gros, P.-A., Le Nagard, H., and Tenaillon, O. The evolution of epistasis and its links with genetic robustness, complexity and drift in a phenotypic model of adaptation. *Genetics*, 182 (2009) 277–293.

[Hannula-Jouppi et al., 2005] Hannula-Jouppi, K., Kaminen-Ahola, N., Taipale, M., Eklund, R., Nopola-Hemmi, J., Kääriäinen, H., and Kere, J. The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS genetics*, 1 (2005) e50–e50.

[Harold et al., 2006] Harold, D., Paracchini, S., Scerri, T., Dennis, M., Cope, N., Hill, G., Moskvina, V., Walter, J., Richardson, A. J., Owen, M. J., Stein, J. F., Green, E. D., O'Donovan, M. C., Williams, J., and Monaco, A. P. Further evidence that the KIAA0319 gene confers susceptibility to developmental dyslexia. *Molecular psychiatry*, 11 (2006) 1085–1061.

[Hasselmo, 2006] Hasselmo, M. E. The role of acetylcholine in learning and memory. *Current Opinion in Neurobiology*, 16 (2006) 710–715.

[Heim et al., 2003] Heim, S. S., Kissler, J. J., Elbert, T. T., and Rockstroh, B. B. Cerebral lateralization in schizophrenia and dyslexia: neuromagnetic responses to auditory stimuli. *Neuropsychologia*, 42 (2003) 6–6.

[Hemani et al., 2014] Hemani, G., Shakhbazov, K., Westra, H.-J., Esko, T., Henders, A. K., McRae, A. F., Yang, J., Gibson, G., Martin, N. G., Metspalu, A., Franke, L., Montgomery, G. W., Visscher, P. M., and Powell, J. E. Detection and replication of epistasis influencing transcription in humans. *Nature*, 508 (2014) 249–253.

[Hemani et al., 2011] Hemani, G., Theocharidis, A., Wei, W., and Haley, C. EpiGPU: exhaustive pairwise epistasis scans parallelized on consumer level graphics cards. *Bioinformatics (Oxford, England)*, 27 (2011) 1462–1465.

[Hermann, 1959] Hermann, K. *Reading disability: A medical study of word-blindness and related handicaps*. *Munksgaard* (1959).

[Ho and Crabtree, 2010] Ho, L. and Crabtree, G. R. Chromatin remodelling during development. *Nature*, 463 (2010) 474–484.

[Horie et al., 2000] Horie, M., Mitsumoto, Y., Kyushiki, H., Kanemoto, N., Watanabe, A., Taniguchi, Y., Nishino, N., Okamoto, T., Kondo, M., Mori, T., Noguchi, K., Nakamura, Y., Takahashi, E.-i., and Tanigami, A. Identification and Characterization of TMEFF2, a Novel Survival Factor for Hippocampal and Mesencephalic Neurons. *Genomics*, 67 (2000) 146–152.

[Howie et al., 2012] Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature genetics*, 44 (2012) 955–959.

[Howie et al., 2009] Howie, B. N., Donnelly, P., and Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics*, 5 (2009) e1000529.

[Hsiung et al., 2004] Hsiung, G.-Y. R., Kaplan, B. J., Petryshen, T. L., Lu, S., and Field, L. L. A dyslexia susceptibility locus (DYX7) linked to dopamine D4 receptor (DRD4) region on chromosome 11p15.5. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 125B (2004) 112–119.

[Hu et al., 2010] Hu, X., Liu, Q., Zhang, Z., Li, Z., Wang, S., He, L., and Shi, Y. SHEsisEpi, a GPU-enhanced genome-wide SNP-SNP interaction scanning algorithm, efficiently reveals the risk genetic epistasis in bipolar disorder. *Cell research*, 20 (2010) 854–857.

[Huber et al., 2002] Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A., and Vingron, M. Variance stabilization applied to microarray data calibration and to the

quantification of differential expression. *Bioinformatics (Oxford, England)*, 18 Suppl 1 (2002) S96–104.

[Humphreys et al., 1990] Humphreys, P., Kaufmann, W. E., and Galaburda, A. M. Developmental dyslexia in women: neuropathological findings in three patients. *Annals of Neurology*, 28 (1990) 727–738.

[Johnson et al., 2007] Johnson, D. S., Mortazavi, A., Myers, R. M., and Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science (New York, N.Y.)*, 316 (2007) 1497–1502.

[Kam-Thong, 2012] Kam-Thong, T. (2012). *Massive Parallelisierung der kombinatorischen statistischen Genetik-Analysen mit Methoden des maschinellen Lernens auf graphics processing units (GPU)*. PhD thesis, Berlin.

[Kam-Thong et al., 2012] Kam-Thong, T., Azencott, C.-A., Cayton, L., Pütz, B., Altmann, A., Karbalai, N., Sämann, P. G., Schölkopf, B., Müller-Myhsok, B., and Borgwardt, K. M. GLIDE: GPU-based linear regression for detection of epistasis. *Human heredity*, 73 (2012) 220–236.

[Kam-Thong et al., 2010] Kam-Thong, T., Czamara, D., Tsuda, K., Borgwardt, K., Lewis, C. M., Erhardt-Lehmann, A., Hemmer, B., Rieckmann, P., Daake, M., Weber, F., Wolf, C., Ziegler, A., tz, B. P. u., Holsboer, F., lkopf, B. S. o., and ller Myhsok, B. M. u. EPIBLASTER-fast exhaustive two-locus epistasis detection strategy using graphical processing units. *European Journal of Human Genetics*, 19 (2010) 465–471.

[Kam-Thong et al., 2011] Kam-Thong, T., Pütz, B., Karbalai, N., Müller-Myhsok, B., and Borgwardt, K. Epistasis detection on quantitative phenotypes by exhaustive enumeration using GPUs. *Bioinformatics (Oxford, England)*, 27 (2011) i214–21.

[Kaminen et al., 2003] Kaminen, N., Hannula-Jouppi, K., Kestilä, M., Lahermo, P., Muller, K., Kaaranen, M., Myllyluoma, B., Voutilainen, A., Lyytinen, H., Nopola-Hemmi, J., and Kere, J. A genome scan for developmental dyslexia confirms linkage to chromosome 2p11 and suggests a new locus on 7q32. *Journal of Medical Genetics*, 40 (2003) 340–345.

[Karolchik et al., 2014] Karolchik, D., Barber, G. P., Casper, J., Clawson, H., Cline, M. S., Diekhans, M., Dreszer, T. R., Fujita, P. A., Guruvadoo, L., Haeussler, M., Harte, R. A., Heitner, S., Hinrichs, A. S., Learned, K., Lee, B. T., Li, C. H., Raney, B. J., Rhead, B., Rosenbloom, K. R., Sloan, C. A., Speir, M. L., Zweig, A. S., Haussler, D., Kuhn, R. M., and Kent, W. J. The UCSC Genome Browser database: 2014 update. *Nucleic acids research*, 42 (2014) D764–70.

[Ke et al., 2008] Ke, X., Taylor, M. S., and Cardon, L. R. Singleton SNPs in the human genome and implications for genome-wide association studies. *European Journal of Human Genetics*, 16 (2008) 506–515.

[Kent et al., 2002] Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., and Haussler, D. The Human Genome Browser at UCSC. *Cold Spring Harbor Laboratory Press*, *Cold Spring Harbor Laboratory Press* (2002).

[König et al., 2011] König, I. R., Schumacher, J., Hoffmann, P., Kleensang, A., Ludwig, K. U., Grimm, T., Neuhoff, N., Preis, M., Roeske, D., Warnke, A., Propping, P., Remschmidt, H., Nöthen, M. M., Ziegler, A., Müller-Myhsok, B., and Schulte-Körne, G. Mapping for dyslexia and related cognitive trait loci provides strong evidence for further risk genes on chromosome 6p21. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 156B (2011) 36–43.

[Kotti et al., 2007] Kotti, S., Bickeb ouml ller, H., and Clerget-Darpoux, F. Strategy for Detecting Susceptibility Genes with Weak or No Marginal Effect. *Human heredity*, 63 (2007) 85–92.

[Kurt et al., 2012] Kurt, S., Fisher, S. E., and Ehret, G. Foxp2 Mutations Impair Auditory-Motor Association Learning. *PloS one*, 7 (2012) e33130.

[Kwok, 2003] Kwok, P.-Y. (2003). *Single Nucleotide Polymorphisms*. Methods and Protocols. Springer.

[Lai et al., 2001] Lai, C. S. C., Fisher, S. E. S., Hurst, J. A. J., Vargha-Khadem, F. F., and Monaco, A. P. A. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, 413 (2001) 519–523.

[Lai, 2003] Lai, C. S. L. FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain : a journal of neurology*, 126 (2003) 2455–2462.

[Lee and Schatz, 2012] Lee, H. and Schatz, M. C. Genomic dark matter: the reliability of short read mapping illustrated by the genome mappability score. *Bioinformatics (Oxford, England)*, 28 (2012) 2097–2105.

[Lee et al., 2004] Lee, H. K., Park, K. S., Cho, Y. M., Ritchie, M. D., Moore, J. H., Park, J. Y., Lee, K. U., and Shin, H. D. Multifactor-dimensionality reduction shows a two-locus interaction associated with Type 2 diabetes mellitus. *Diabetologia*, 47 (2004) 549–554.

[Li and Durbin, 2009] Li, H. and Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25 (2009) 1754–1760.

[Li et al., 2003] Li, S. S., Khalid, N., Carlson, C., and Zhao, L. P. Estimating haplotype frequencies and standard errors for multiple single nucleotide polymorphisms. *Biostatistics (Oxford, England)*, 4 (2003) 513–522.

[Ludwig et al., 2010] Ludwig, K. U., Roeske, D., Herms, S., Schumacher, J., Warnke, A., Plume, E., Neuhoff, N., Bruder, J., Remschmidt, H., Schulte-Körne, G., Müller-Myhsok, B., Nöthen, M. M., 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, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 153B (2010) 503–511.

[Ludwig et al., 2008] Ludwig, K. U., Roeske, D., Schumacher, J., Schulte-Körne, G., König, I. R., Warnke, A., Plume, E., Ziegler, A., Remschmidt, H., Müller-Myhsok, B., Nöthen, M. M., and Hoffmann, P. Investigation of interaction between DCDC2 and KIAA0319 in a large German dyslexia sample. *Journal of neural transmission (Vienna, Austria : 1996)*, 115 (2008) 1587–1589.

[Lumley, 2013] Lumley, T. (2013). *Package 'rmeta'*. University of Washington.

[Maag and Behrens, 1989] Maag, J. W. and Behrens, J. T. Depression And Cognitive Self-Statements Of Learning Disabled And Seriously Emotionally Disturbed Adolescents. *The Journal of Special Education*, 23 (1989) 17–27.

[MacDermot et al., 2005] MacDermot, K. D., Bonora, E., Sykes, N., Coupe, A.-M., Lai, C. S. L., Vernes, S. C., Vargha-Khadem, F., McKenzie, F., Smith, R. L., Monaco, A. P., and Fisher, S. E. Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *American journal of human genetics*, 76 (2005) 1074–1080.

[Manolio et al., 2009] Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F. C., McCarroll, S. A., and Visscher, P. M. Finding the missing heritability of complex diseases. *Nature*, 461 (2009) 747–753.

[Manto and Jissendi, 2012] Manto, M. U. and Jissendi, P. Cerebellum: links between development, developmental disorders and motor learning. *Frontiers in neuroanatomy*, 6 (2012) 1.

[Marchini et al., 2005] Marchini, J., Donnelly, P., and Cardon, L. R. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nature genetics*, 37 (2005) 413–417.

[Marino et al., 2012] Marino, C., Meng, H., Mascheretti, S., Rusconi, M., Cope, N., Giorda, R., Molteni, M., and Gruen, J. R. DCDC2 genetic variants and susceptibility to developmental dyslexia. *Psychiatric genetics*, 22 (2012) 25–30.

[Markram et al., 2007] Markram, K., Gerardy-Schahn, R., and Sandi, C. Selective learning and memory impairments in mice deficient for polysialylated NCAM in adulthood. *Neuroscience*, 144 (2007) 788–796.

[Martin et al., 2006] Martin, E. R., Ritchie, M. D., Hahn, L., Kang, S., and Moore, J. H. A novel method to identify gene–gene effects in nuclear families: the MDR-PDT. *Genetic Epidemiology*, 30 (2006) 111–123.

[Meng et al., 2005] Meng, H., Smith, S. D., Hager, K., Held, M., Liu, J., Olson, R. K., Pennington, B. F., DeFries, J. C., Gelernter, J., O'Reilly-Pol, T., Somlo, S., Skudlarski, P., Shaywitz, S. E., Shaywitz, B. A., Marchione, K., Wang, Y., Paramasivam, M., LoTurco, J. J., Page, G. P., and Gruen, J. R. DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proceedings of the National Academy of Sciences of the United States of America*, 102 (2005) 17053–17058.

[Miko, 2008] Miko, I. Epistasis: Gene interaction and phenotype effects. *Nature Education* (2008).

[Moore, 2003] Moore, J. H. The Ubiquitous Nature of Epistasis in Determining Susceptibility to Common Human Diseases. *Human heredity*, 56 (2003) 73–82.

[Moore and Williams, 2009] Moore, J. H. and Williams, S. M. Epistasis and its implications for personal genetics. *American journal of human genetics*, 85 (2009) 309–320.

[Mosby, Inc, 2009] Mosby, Inc (2009). *Mosby's Dictionary of Medicine, Nursing and Health Professions*. Mosby Incorporated.

[Moser and Moser, 1998] Moser, M. B. and Moser, E. I. Functional differentiation in the hippocampus. *Hippocampus*, 8 (1998) 608–619.

[Musani et al., 2007] Musani, S. K., Shriner, D., Liu, N., Feng, R., Coffey, C. S., Yi, N., Tiwari, H. K., and Allison, D. B. Detection of Gene Gene Interactions in Genome-Wide Association Studies of Human Population Data. *Human heredity*, 63 (2007) 67–84.

[Newbury and Monaco, 2010] Newbury, D. F. and Monaco, A. P. Genetic advances in the study of speech and language disorders. *Neuron*, 68 (2010) 309–320.

[Newbury et al., 2010] Newbury, D. F., Paracchini, S., Scerri, T. S., Winchester, L., Addis, L., Richardson, A. J., Walter, J., Stein, J. F., Talcott, J. B., and Monaco, A. P. Investigation of Dyslexia and SLI Risk Variants in Reading- and Language-Impaired Subjects. *Behavior Genetics*, 41 (2010) 90–104.

[Nicodemus et al., 2010a] Nicodemus, K. K., Callicott, J. H., Higier, R. G., Luna, A., Nixon, D. C., Lipska, B. K., Vakkalanka, R., Giegling, I., Rujescu, D., St Clair, D., Muglia, P., Shugart, Y. Y., and Weinberger, D. R. Evidence of statistical epistasis between DISC1, CIT and NDEL1 impacting risk for schizophrenia: biological validation with functional neuroimaging. *Human genetics*, 127 (2010a) 441–452.

[Nicodemus et al., 2007] Nicodemus, K. K., Kolachana, B. S., Vakkalanka, R., Straub, R. E., Giegling, I., Egan, M. F., Rujescu, D., and Weinberger, D. R. Evidence for statistical epistasis between catechol-O-methyltransferase (COMT) and polymorphisms in RGS4, G72 (DAOA), GRM3, and DISC1: influence on risk of schizophrenia. *Human genetics*, 120 (2007) 889–906.

[Nicodemus et al., 2010b] Nicodemus, K. K., Law, A. J., Radulescu, E., Luna, A., Kolachana, B., Vakkalanka, R., Rujescu, D., Giegling, I., Straub, R. E., McGee, K., Gold, B., Dean, M., Muglia, P., Callicott, J. H., Tan, H.-Y., and Weinberger, D. R. Biological validation of increased schizophrenia risk with NRG1, ERBB4, and AKT1 epistasis via functional neuroimaging in healthy controls. *Archives of general psychiatry*, 67 (2010b) 991–1001.

[Nikitin et al., 2003] Nikitin, A., Egorov, S., Daraselia, N., and Mazo, I. Pathway studio–the analysis and navigation of molecular networks. *Bioinformatics (Oxford, England)*, 19 (2003) 2155–2157.

[Nopola-Hemmi et al., 2001] Nopola-Hemmi, J., Myllyluoma, B., Haltia, T., Taipale, M., Ollikainen, V., Ahonen, T., Voutilainen, A., Kere, J., and Widén, E. A dominant gene for developmental dyslexia on chromosome 3. *Journal of Medical Genetics*, 38 (2001) 658–664.

[Nopola-Hemmi et al., 2000] Nopola-Hemmi, J., Taipale, M., Haltia, T., Lehesjoki, A. E., Voutilainen, A., and Kere, J. Two translocations of chromosome 15q associated with dyslexia. *Journal of Medical Genetics*, 37 (2000) 771–775.

[Olson et al., 1999] Olson, R. K., Datta, H., Gayán, J., and DeFries, J. C. A behavioral-genetic analysis of reading disabilities and component processes. *Converging methods for understanding reading and dyslexia*, *Converging methods for understanding reading and dyslexia* (1999) 133–153.

[Pammer, 2014] Pammer, K. Review ArticleBrain Mechanisms and Reading Remediation: More Questions Than Answers. *Scientifica*, *Scientifica* (2014) 1–9.

[Paracchini et al., 2008] Paracchini, S., Steer, C. D., Buckingham, L.-L., Morris, A. P., Ring, S., Scerri, T., Stein, J., Pembrey, M. E., Ragoussis, J., Golding, J., and Monaco, A. P. Association of the KIAA0319 dyslexia susceptibility gene with reading skills in the general population. *American Journal of Psychiatry*, 165 (2008) 1576–1584.

[Paracchini et al., 2006] Paracchini, S., Thomas, A., Castro, S., Lai, C., Paramasivam, M., Wang, Y., Keating, B. J., Taylor, J. M., Hacking, D. F., Scerri, T., Francks, C., Richardson, A. J., Wade-Martins, R., Stein, J. F., Knight, J. C., Copp, A. J., Loturco, J., and Monaco, A. P. The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. *Human Molecular Genetics*, 15 (2006) 1659–1666.

[Paracchini, 2011] Paracchini, S. S. Dissection of genetic associations with language-related traits in population-based cohorts. *Journal of Neurodevelopmental Disorders*, 3 (2011) 365–373.

[Peter et al., 2010] Peter, B., Raskind, W. H., Matsushita, M., Lisowski, M., Vu, T., Berninger, V. W., Wijsman, E. M., and Brkanac, Z. Replication of CNTNAP2 association with nonword repetition and support for FOXP2 association with timed reading and motor activities in a dyslexia family sample. *Journal of Neurodevelopmental Disorders*, 3 (2010) 39–49.

[Petryshen et al., 2001] Petryshen, T. L., Kaplan, B. J., Fu Liu, M., de French, N. S., Tobias, R., Hughes, M. L., and Field, L. L. Evidence for a susceptibility locus on chromosome 6q influencing phonological coding dyslexia. *American journal of medical genetics*, 105 (2001) 507–517.

[Phillips, 2008] Phillips, P. C. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, 9 (2008) 855–867.

[Pinel et al., 2012] Pinel, P., Fauchereau, F., Moreno, A., Barbot, A., Lathrop, M., Zelenika, D., Le Bihan, D., Poline, J. B., Bourgeron, T., and Dehaene, S. Genetic Variants of FOXP2 and KIAA0319/TTRAP/THEM2 Locus Are Associated with Altered Brain Activation in Distinct Language-Related Regions. *Journal of Neuroscience*, 32 (2012) 817–825.

[Purcell et al., 2007] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., and Sham, P. C. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, 81 (2007) 559–575.

[Pütz et al., 2013] Pütz, B., Kam-Thong, T., Karbalai, N., Altmann, A., and Müller-Myhsok, B. Cost-effective GPU-grid for genome-wide epistasis calculations. *Methods of information in medicine*, 52 (2013) 91–95.

[Qin et al., 2005] Qin, S., Zhao, X., Pan, Y., Liu, J., Feng, G., Fu, J., Bao, J., Zhang, Z., and He, L. An association study of the N-methyl-D-aspartate receptor NR1 subunit gene (GRIN1) and NR2B subunit gene (GRIN2B) in schizophrenia with universal DNA microarray. *European Journal of Human Genetics*, 13 (2005) 807–814.

[Rack et al., 1992] Rack, J. P., Snowling, M. J., and Olson, R. K. JSTOR: Reading Research Quarterly, Vol. 27, No. 1 (Winter, 1992), pp. 28-53. *Reading Research Quarterly* (1992).

[Reichardt, 2006] Reichardt, L. F. Neurotrophin-regulated signalling pathways. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361 (2006) 1545–1564.

[Reiersen and Todd, 2008] Reiersen, A. M. and Todd, R. D. Co-occurrence of ADHD and autism spectrum disorders: phenomenology and treatment. *Expert review of neurotherapeutics*, 8 (2008) 657–669.

[Richards et al., 1999] Richards, T. L., Dager, S. R., Corina, D., Serafini, S., Heide, A. C., Steury, K., Strauss, W., Hayes, C. E., Abbott, R. D., Craft, S., Shaw, D., Posse, S., and Berninger, V. W. Dyslexic children have abnormal brain lactate response to reading-related language tasks. *AJNR. American journal of neuroradiology*, 20 (1999) 1393–1398.

[Ritchie et al., 2001] Ritchie, M. D., Hahn, L. W., Roodi, N., Bailey, L. R., Dupont, W. D., Parl, F. F., and Moore, J. H. Multifactor-Dimensionality Reduction Reveals High-Order Interactions among Estrogen-Metabolism Genes in Sporadic Breast Cancer. *The American Journal of Human Genetics*, 69 (2001) 138–147.

[Ritchie, 2010] Ritchie, M. D. M. Using biological knowledge to uncover the mystery in the search for epistasis in genome-wide association studies. *Annals of Human Genetics*, 75 (2010) 172–182.

[Roeske et al., 2011] Roeske, D., Ludwig, K. U., Neuhoff, N., Becker, J., Bartling, J., Bruder, J., Brockschmidt, F. F., Warnke, A., Remschmidt, H., Hoffmann, P., Müller-Myhsok, B., Nöthen, M. M., and Schulte-Körne, G. First genome-wide association scan on neurophysiological endophenotypes points to trans-regulation effects on SLC2A3 in dyslexic children. *Molecular psychiatry*, 16 (2011) 97–107.

[Rohlfs et al., 2010] Rohlfs, R. V., Swanson, W. J., and Weir, B. S. AR TICLEDetecting Coevolution through Allelic Association between Physically Unlinked Loci. *The American Journal of Human Genetics*, 86 (2010) 674–685.

[Rosenbloom et al., 2013] Rosenbloom, K. R., Sloan, C. A., Malladi, V. S., Dreszer, T. R., Learned, K., Kirkup, V. M., Wong, M. C., Maddren, M., Fang, R., Heitner, S. G., Lee, B. T., Barber, G. P., Harte, R. A., Diekhans, M., Long, J. C., Wilder, S. P., Zweig, A. S., Karolchik, D., Kuhn, R. M., Haussler, D., and Kent, W. J. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic acids research*, 41 (2013) D56–63.

[Salyakina, 2007] Salyakina, D. (2007). *Candidate gene association testing in the dissection of genetic causes for depressive disorders and the response to antidepressant treatment*. PhD thesis, Dissertation.

[Scerri et al., 2012] Scerri, T. S., Darki, F., Newbury, D. F., Whitehouse, A. J. O., Peyrard-Janvid, M., Matsson, H., Ang, Q. W., Pennell, C. E., Ring, S., Stein, J., Morris, A. P., Monaco, A. P., Kere, J., Talcott, J. B., Klingberg, T., and Paracchini, S. The dyslexia candidate locus on 2p12 is associated with general cognitive ability and white matter structure. *PloS one*, 7 (2012) e50321–e50321.

[Scerri et al., 2011] Scerri, T. S., Morris, A. P., Buckingham, L. L., Newbury, D. F., Miller, L. L., Monaco, A. P., Bishop, D. V. M., and Paracchini, S. DCDC2, KIAA0319 and CMIP Are Associated with Reading-Related Traits. *BPS*, 70 (2011) 237–245.

[Scerri et al., 2010] Scerri, T. S., Paracchini, S., Morris, A., MacPhie, I. L., Talcott, J., Stein, J., Smith, S. D., Pennington, B. F., Olson, R. K., DeFries, J. C., Monaco, A. P., and Richardson, A. J. Identification of candidate genes for dyslexia susceptibility on chromosome 18. *PloS one*, 5 (2010) e13712.

[Schmid and Maness, 2008] Schmid, R. S. and Maness, P. F. L1 and NCAM adhesion molecules as signaling coreceptors in neuronal migration and process outgrowth. *Current Opinion in Neurobiology*, 18 (2008) 245–250.

[Schoenherr and Anderson, 1995] Schoenherr, C. J. and Anderson, D. J. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science (New York, N.Y.)*, 267 (1995) 1360–1363.

[Schulte-Körne et al., 2001] Schulte-Körne, G., Deimel, W., and Remschmidt, H. [Diagnosis of reading and spelling disorder]. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 29 (2001) 113–116.

[Schulte-Körne et al., 2007] Schulte-Körne, G., Ziegler, A., Deimel, W., Schumacher, J., Plume, E., Bachmann, C., Kleensang, A., Propping, P., Nöthen, M. M., Warnke, A., Remschmidt, H., and König, I. R. Interrelationship and familiality of dyslexia related quantitative measures. *Annals of Human Genetics*, 71 (2007) 160–175.

[Schumacher et al., 2005] Schumacher, J., Anthoni, H., Dahdouh, F., König, I. R., Hillmer, A. M., Kluck, N., Manthey, M., Plume, E., Warnke, A., Remschmidt, H., Hülsmann, J., Cichon, S., Lindgren, C. M., Propping, P., Zucchelli, M., Ziegler, A., Peyrard-Janvid, M., Schulte-Körne, G., Nöthen, M. M., and Kere, J. Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. *The American Journal of Human Genetics*, 78 (2005) 52–62.

[Schumacher et al., 2007] Schumacher, J., Hoffmann, P., Schmäl, C., Schulte-Körne, G., and Nöthen, M. M. Genetics of dyslexia: the evolving landscape. *Journal of Medical Genetics*, 44 (2007) 289–297.

[Scott and Soderling, 1991] Scott, J. D. and Soderling, T. R. Serine/threonine protein kinases. *Current Opinion in Neurobiology*, 2 (1991) 289–295.

[Segrè et al., 2010] Segrè, A. V., Groop, L., Mootha, V. K., Daly, M. J., and Altshuler, D. PLOS Genetics: Common Inherited Variation in Mitochondrial Genes Is Not Enriched for Associations with Type 2 Diabetes or Related Glycemic Traits. *PLoS genetics* (2010).

[Shao et al., 2008] Shao, H., Burrage, L. C., Sinasac, D. S., Hill, A. E., Ernest, S. R., O'Brien, W., Courtland, H.-W., Jepsen, K. J., Kirby, A., Kulbokas, E. J., Daly, M. J., Broman, K. W., Lander, E. S., and Nadeau, J. H. Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. *Proceedings of the National Academy of Sciences of the United States of America*, 105 (2008) 19910–19914.

[Shaywitz et al., 1990] Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., and Escobar, M. D. Prevalence of reading disability in boys and girls. Results of the Connecticut Longitudinal Study. *JAMA : the journal of the American Medical Association*, 264 (1990) 998–1002.

[Smith, 2007] Smith, S. D. Genes, language development, and language disorders. *Mental Retardation and Developmental Disabilities Research Reviews*, 13 (2007) 96–105.

[Smith et al., 1983] Smith, S. D., Kimberling, W. J., Pennington, B. F., and Lubs, H. A. Specific reading disability: identification of an inherited form through linkage analysis. *Science (New York, N.Y.)*, 219 (1983) 1345–1347.

[Smith, 1991] Smith, S. L. Succeeding against the odds: How the learning-disabled can realize their promise. *New York : G.P Putnam's Sons* (1991).

[Stegle et al., 2012] Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nature Protocols*, 7 (2012) 500–507.

[Stein et al., 2004] Stein, C. M., Schick, J. H., Gerry Taylor, H., Shriberg, L. D., Millard, C., Kundtz-Kluge, A., Russo, K., Minich, N., Hansen, A., Freebairn, L. A., Elston, R. C., Lewis, B. A., and Iyengar, S. K. Pleiotropic effects of a chromosome 3 locus on speech-sound disorder and reading. *The American Journal of Human Genetics*, 74 (2004) 283–297.

[Takahashi et al., 2003] Takahashi, K., Liu, F.-C., Hirokawa, K., and Takahashi, H. Expression of Foxp2, a gene involved in speech and language, in the developing and adult striatum. *Journal of neuroscience research*, 73 (2003) 61–72.

[Tammimies et al., 2013] Tammimies, K., Vitezic, M., Matsson, H., Le Guyader, S., Bürglin, T. R., Öhman, T., Strömblad, S., Daub, C. O., Nyman, T. A., Kere, J., and Tapia-Páez, I. Molecular Networks of DYX1C1 Gene Show Connection to Neuronal Migration Genes and Cytoskeletal Proteins. *Biological Psychiatry*, 73 (2013) 583–590.

[Thorleifsson et al., 2007] Thorleifsson, G., Magnusson, K. P., Sulem, P., Walters, G. B., Gudbjartsson, D. F., Stefansson, H., Jonsson, T., Jonasdottir, A., Jonasdottir, A., Stefansdottir, G., Masson, G., Hardarson, G. A., Petursson, H., Arnarsson, A., Motallebipour, M., Wallerman, O., Wadelius, C., Gulcher, J. R., Thorsteinsdottir, U., Kong, A., Jonasson, F., and Stefansson, K. Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. *Science (New York, N.Y.)*, 317 (2007) 1397–1400.

[Tzenova et al., 2004] Tzenova, J., Kaplan, B. J., Petryshen, T. L., and Field, L. L. Confirmation of a dyslexia susceptibility locus on chromosome 1p34-p36 in a set of 100 Canadian families. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 127B (2004) 117–124.

[Van Steen, 2012] Van Steen, K. Travelling the world of gene-gene interactions. *Briefings in Bioinformatics*, 13 (2012) 1–19.

[Veerappa et al., 2013] Veerappa, A. M., Saldanha, M., Padakannaya, P., and Ramachandra, N. B. Family-based genome-wide copy number scan identifies five new genes of dyslexia involved in dendritic spinal plasticity. *Journal of Human Genetics*, 58 (2013) 539–547.

[Verhoeven et al., 2010] Verhoeven, K., Casella, G., and McIntyre, L. M. Epistasis: Obstacle or Advantage for Mapping Complex Traits? *PloS one* (2010).

[Vernes et al., 2008] Vernes, S. C., Newbury, D. F., Abrahams, B. S., Winchester, L., Nicod, J., Groszer, M., Alarcón, M., Oliver, P. L., Davies, K. E., Geschwind, D. H., Monaco, A. P., and Fisher, S. E. A Functional Genetic Link between Distinct Developmental Language Disorders. *New England Journal of Medicine*, 359 (2008) 2337–2345.

[Vernes et al., 2011] Vernes, S. C., Oliver, P. L., Spiteri, E., Lockstone, H. E., Puliyadi, R., Taylor, J. M., Ho, J., Mombereau, C., Brewer, A., Lowy, E., Nicod, J., Groszer, M., Baban, D., Sahgal, N., Cazier, J.-B., Ragoussis, J., Davies, K. E., Geschwind, D. H., and Fisher, S. E. Foxp2 Regulates Gene Networks Implicated in Neurite Outgrowth in the Developing Brain. *PLoS genetics*, 7 (2011) e1002145.

[Wade et al., 2001] Wade, M. J., Winther, R. G., Agrawal, A. F., and Goodnight, C. J. Alternative definitions of epistasis: dependence and interaction. *Trends in Ecology & Evolution*, 16 (2001) 498–504.

[Wang et al., 2012a] Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T. W., Greven, M. C., Pierce, B. G., Dong, X., Kundaje, A., Cheng, Y., Rando, O. J., Birney, E., Myers, R. M., Noble, W. S., Snyder, M., and Weng, Z. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome research*, 22 (2012a) 1798–1812.

[Wang et al., 2013] Wang, J., Zhuang, J., Iyer, S., Lin, X.-Y., Greven, M. C., Kim, B.-H., Moore, J., Pierce, B. G., Dong, X., Virgil, D., Birney, E., Hung, J.-H., and Weng, Z. Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic acids research*, 41 (2013) D171–6.

[Wang et al., 2010] Wang, S.-Q., Kang, Y.-M., Qi, Z.-C., Ma, X.-R., Liang, Y.-F., Li, D.-G., and Zhang, B. Effects of ganoderma lucidum spores powder on expression of

NCAM-1 and NCAM-L1 in hippocampus of rats with epilepsy. *The FASEB Journal* (2010).

[Wang et al., 2012b] Wang, W., Wang, L., Luo, J., Xi, Z., Wang, X., Chen, G., and Chu, L. Role of a neural cell adhesion molecule found in cerebrospinal fluid as a potential biomarker for epilepsy. *Neurochemical Research*, 37 (2012b) 819–825.

[Wigginton et al., 2005] Wigginton, J. E., Cutler, D. J., and Abecasis, G. R. A note on exact tests of Hardy-Weinberg equilibrium. *American journal of human genetics*, 76 (2005) 887–893.

[Wilcke et al., 2011] Wilcke, A., Ligges, C., Burkhardt, J., Alexander, M., Wolf, C., Quente, E., Ahnert, P., Hoffmann, P., Becker, A., ller Myhsok, B. M. u., Cichon, S., Boltze, J., and Kirsten, H. Imaging genetics of FOXP2 in dyslexia. *European Journal of Human Genetics*, 20 (2011) 224–229.

[Wimmer, 1996] Wimmer, H. The Nonword Reading Deficit in Developmental Dyslexia: Evidence from Children Learning to Read German. *Journal of Experimental Child Psychology*, 61 (1996) 80–90.

[Wolff and Melngailis, 1994] Wolff, P. H. and Melngailis, I. Family patterns of developmental dyslexia: clinical findings. *American journal of medical genetics*, 54 (1994) 122–131.

[Wood et al., 1991] Wood, F., Flowers, L., Buchsbaum, M., and Tallal, P. Investigation of abnormal left temporal functioning in dyslexia through rCBF, auditory evoked potentials, and positron emisson tomography. *Reading and Writing*, 3 (1991) 379–393.

[Yamada and Nabeshima, 2003] Yamada, K. and Nabeshima, T. Brain-derived neurotrophic factor/TrkB signaling in memory processes. *Journal of pharmacological sciences*, 91 (2003) 267–270.

[Yung et al., 2011] Yung, L. S., Yang, C., Wan, X., and Yu, W. GBOOST: a GPU-based tool for detecting gene-gene interactions in genome-wide case control studies. *Bioinformatics (Oxford, England)*, 27 (2011) 1309–1310.

[Zerbin-Rüdin, 1967] Zerbin-Rüdin, E. Kongenitale wortblindheit oder spezifische dyslexie (Congenital word-blindness). *Annals of Dyslexia*, 17 (1967) 47–54.

[Zhang et al., 2008] Zhang, Y., Liu, T., Meyer, C. A., and Eeckhoute, J. Model-based analysis of ChIP-Seq (MACS). *Genome . . . , Genome . . .* (2008).

[Zhao et al., 2006] Zhao, J., Jin, L., and Xiong, M. Test for interaction between two unlinked loci. *American journal of human genetics*, 79 (2006) 831–845.

[Zuk et al., 2012] Zuk, O. O., Hechter, E. E., Sunyaev, S. R. S., and Lander, E. S. E. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences of the United States of America*, 109 (2012) 1193–1198.

Single-word reading

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs56253958	7:113883531	FOXP2	rs1409679	9:113586263	Intergenic	-0.679	0.091	8.49e-14	-0.857	-0.5	0.454	7.90e-13
rs144261163	7:113893884	FOXP2	rs1409679	9:113586263	Intergenic	-0.673	0.091	1.51e-13	-0.852	-0.494	0.393	1.18e-12
rs56253958	7:113883531	FOXP2	rs2821137	9:113587820	Intergenic	-0.672	0.091	1.92e-13	-0.851	-0.493	0.396	2.36e-12
rs1450833	7:113865735	FOXP2	rs1409679	9:113586263	Intergenic	-0.66	0.09	2.23e-13	-0.836	-0.484	0.498	1.90e-12
rs56253958	7:113883531	FOXP2	rs2846443	9:113589582	Intergenic	-0.667	0.092	3.59e-13	-0.846	-0.487	0.388	3.50e-12
rs1450833	7:113865735	FOXP2	rs2821137	9:113587820	Intergenic	-0.653	0.09	4.98e-13	-0.83	-0.476	0.439	5.44e-12
rs6961970	7:113901132	FOXP2	rs1409679	9:113586263	Intergenic	-0.637	0.088	5.94e-13	-0.81	-0.463	0.68	6.73e-12
rs1450833	7:113865735	FOXP2	rs2846443	9:113589582	Intergenic	-0.648	0.09	6.86e-13	-0.825	-0.471	0.429	7.89e-12
rs73206277	7:113954180	FOXP2	rs1409679	9:113586263	Intergenic	-0.634	0.089	7.95e-13	-0.808	-0.461	0.562	9.14e-12
rs6961970	7:113901132	FOXP2	rs2821137	9:113587820	Intergenic	-0.63	0.089	1.27e-12	-0.804	-0.456	0.619	1.84e-11
rs73206277	7:113954180	FOXP2	rs2821137	9:113587820	Intergenic	-0.628	0.089	1.68e-12	-0.802	-0.454	0.497	2.49e-11
rs6961970	7:113901132	FOXP2	rs2846443	9:113589582	Intergenic	-0.625	0.089	1.77e-12	-0.799	-0.452	0.615	2.68e-11
rs63348361	7:113903189	FOXP2	rs1409679	9:113586263	Intergenic	-0.577	0.082	1.98e-12	-0.737	-0.416	0.914	2.34e-11
rs12537074	7:113921127	FOXP2	rs1409679	9:113586263	Intergenic	-0.575	0.082	1.98e-12	-0.736	-0.415	0.922	1.97e-11
rs549449	1:170610710	Intergenic	rs6570245	6:138915745	Intergenic	0.572	0.081	2.03e-12	0.413	0.732	0.341	1.38e-11
rs1528093	7:113912913	FOXP2	rs1409679	9:113586263	Intergenic	-0.576	0.082	2.06e-12	-0.736	-0.415	0.917	1.91e-11
rs73206277	7:113954180	FOXP2	rs2846443	9:113589582	Intergenic	-0.623	0.089	2.34e-12	-0.797	-0.449	0.499	3.64e-11
rs1852471	7:113925317	FOXP2	rs1409679	9:113586263	Intergenic	-0.566	0.081	2.96e-12	-0.725	-0.407	0.882	4.68e-11
rs1406087	7:113910505	FOXP2	rs1409679	9:113586263	Intergenic	-0.57	0.082	3.05e-12	-0.731	-0.41	0.907	2.11e-11
rs1194327	7:113896089	FOXP2	rs1409679	9:113586263	Intergenic	-0.582	0.083	3.07e-12	-0.745	-0.418	0.826	5.34e-11
rs12537074	7:113921127	FOXP2	rs2821137	9:113587820	Intergenic	-0.571	0.082	3.44e-12	-0.731	-0.41	0.908	3.64e-11
rs63348361	7:113903189	FOXP2	rs2821137	9:113587820	Intergenic	-0.572	0.082	3.44e-12	-0.733	-0.411	0.9	4.30e-11
rs1852472	7:113925350	FOXP2	rs1409679	9:113586263	Intergenic	-0.57	0.082	3.53e-12	-0.731	-0.41	0.904	2.70e-11
rs79874450	6:2636108	Intergenic	rs73024913	19:29778736	LOC284395	-1.388	0.2	3.53e-12	-1.779	-0.997	0.69	2.49e-12
rs1528093	7:113912913	FOXP2	rs2821137	9:113587820	Intergenic	-0.571	0.082	3.58e-12	-0.732	-0.41	0.902	3.52e-11
rs2694946	7:113925004	FOXP2	rs1409679	9:113586263	Intergenic	-0.568	0.082	3.80e-12	-0.729	-0.408	0.912	2.63e-11
rs63348361	7:113903189	FOXP2	rs2846443	9:113589582	Intergenic	-0.569	0.082	3.97e-12	-0.73	-0.409	0.89	5.56e-11
rs12537074	7:113921127	FOXP2	rs2846443	9:113589582	Intergenic	-0.568	0.082	3.98e-12	-0.728	-0.407	0.899	4.71e-11
rs2690839	7:113924502	FOXP2	rs1409679	9:113586263	Intergenic	-0.567	0.082	4.06e-12	-0.728	-0.407	0.915	2.15e-11
rs1528093	7:113912913	FOXP2	rs2846443	9:113589582	Intergenic	-0.568	0.082	4.14e-12	-0.729	-0.408	0.893	4.56e-11
rs1527154	7:113922817	FOXP2	rs1409679	9:113586263	Intergenic	-0.567	0.082	4.15e-12	-0.727	-0.407	0.914	2.10e-11
rs1527159	7:113930918	FOXP2	rs1409679	9:113586263	Intergenic	-0.564	0.082	4.35e-12	-0.724	-0.405	0.873	1.68e-11
rs1728436	7:113912412	FOXP2	rs1409679	9:113586263	Intergenic	-0.567	0.082	4.45e-12	-0.727	-0.406	0.91	1.99e-11

Single-word reading

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs2690842	7:113915642	FOXP2	rs1409679	9:113586263	Intergenic	-0.566	0.082	4.57e-12	-0.726	-0.406	0.912	1.95e-11
rs1209254	7:113917539	FOXP2	rs1409679	9:113586263	Intergenic	-0.566	0.082	4.62e-12	-0.726	-0.405	0.913	1.92e-11
rs549449	1:170610710	Intergenic	rs6918682	6:138909231	Intergenic	0.563	0.081	4.81e-12	0.403	0.722	0.61	2.19e-11
rs75270819	5:141484933	Intergenic	rs180655	10:117763073	Intergenic	-1.01	0.146	4.83e-12	-1.297	-0.724	0.844	1.13e-11
rs1852471	7:113925317	FOXP2	rs2821137	9:113587820	Intergenic	-0.562	0.081	4.90e-12	-0.721	-0.403	0.87	8.28e-11
rs77557441	6:2668406	MYLK4	rs73024913	19:29778736	LOC284395	-1.365	0.198	4.99e-12	-1.752	-0.977	0.773	3.59e-12
rs75270819	5:141484933	Intergenic	rs180662	10:117758764	Intergenic	-1.02	0.148	5.07e-12	-1.31	-0.731	0.519	1.21e-11
rs1406087	7:113910505	FOXP2	rs2821137	9:113587820	Intergenic	-0.566	0.082	5.21e-12	-0.727	-0.405	0.89	3.84e-11
rs1194327	7:113896089	FOXP2	rs2821137	9:113587820	Intergenic	-0.577	0.084	5.32e-12	-0.741	-0.413	0.784	9.69e-11
rs9392409	6:2637106	C6orf195	rs73024913	19:29778736	LOC284395	-1.375	0.199	5.35e-12	-1.765	-0.984	0.712	2.74e-12
rs1194330	7:113920688	FOXP2	rs1409679	9:113586263	Intergenic	-0.568	0.082	5.41e-12	-0.729	-0.406	0.916	2.25e-11
rs549449	1:170610710	Intergenic	rs3861392	6:138910102	Intergenic	0.562	0.082	5.49e-12	0.402	0.722	0.503	2.27e-11
rs74735747	6:2666614	MYLK4	rs73024913	19:29778736	LOC284395	-1.36	0.197	5.52e-12	-1.747	-0.973	0.789	4.24e-12
rs1527158	7:113926120	FOXP2	rs1409679	9:113586263	Intergenic	-0.561	0.081	5.53e-12	-0.721	-0.402	0.856	9.39e-12
rs549449	1:170610710	Intergenic	rs4521605	6:138908662	Intergenic	0.579	0.084	5.73e-12	0.414	0.744	0.534	5.61e-11
rs1406087	7:113910505	FOXP2	rs2846443	9:113589582	Intergenic	-0.563	0.082	5.87e-12	-0.724	-0.403	0.882	4.94e-11
rs1852472	7:113925350	FOXP2	rs2821137	9:113587820	Intergenic	-0.566	0.082	6.04e-12	-0.727	-0.404	0.889	4.92e-11
rs17135412	6:2666428	MYLK4	rs73024913	19:29778736	LOC284395	-1.358	0.197	6.09e-12	-1.745	-0.971	0.778	4.48e-12
rs2694946	7:113925004	FOXP2	rs2821137	9:113587820	Intergenic	-0.564	0.082	6.51e-12	-0.724	-0.403	0.897	4.79e-11
rs1852472	7:113925350	FOXP2	rs2846443	9:113589582	Intergenic	-0.563	0.082	6.87e-12	-0.724	-0.402	0.881	6.35e-11
rs2690839	7:113924502	FOXP2	rs2821137	9:113587820	Intergenic	-0.563	0.082	6.95e-12	-0.723	-0.402	0.9	3.94e-11
rs6961519	7:113958453	FOXP2	rs1409679	9:113586263	Intergenic	-0.607	0.089	6.96e-12	-0.781	-0.434	0.654	7.95e-11
rs1527154	7:113922817	FOXP2	rs2821137	9:113587820	Intergenic	-0.562	0.082	7.09e-12	-0.723	-0.402	0.899	3.84e-11
rs549449	1:170610710	Intergenic	rs6923972	6:138909661	Intergenic	0.559	0.081	7.23e-12	0.399	0.718	0.624	3.06e-11
rs2694946	7:113925004	FOXP2	rs2846443	9:113589582	Intergenic	-0.561	0.082	7.39e-12	-0.721	-0.4	0.889	6.18e-11
rs549449	1:170610710	Intergenic	rs7765298	6:138907520	Intergenic	0.561	0.082	7.41e-12	0.4	0.721	0.601	3.57e-11
rs1728436	7:113912412	FOXP2	rs2821137	9:113587820	Intergenic	-0.562	0.082	7.60e-12	-0.723	-0.401	0.893	3.64e-11
rs1527159	7:113930918	FOXP2	rs2821137	9:113587820	Intergenic	-0.559	0.082	7.66e-12	-0.72	-0.399	0.85	3.07e-11
rs11903584	2:75672583	Intergenic	rs2619111	10:118966996	KCNK18	-0.954	0.139	7.70e-12	-1.228	-0.681	0.872	7.51e-11
rs2690842	7:113915642	FOXP2	rs2821137	9:113587820	Intergenic	-0.561	0.082	7.79e-12	-0.722	-0.4	0.896	3.56e-11
rs2690839	7:113924502	FOXP2	rs2846443	9:113589582	Intergenic	-0.56	0.082	7.87e-12	-0.72	-0.4	0.892	5.10e-11
rs1209254	7:113917539	FOXP2	rs2821137	9:113587820	Intergenic	-0.561	0.082	7.90e-12	-0.722	-0.4	0.897	3.52e-11
rs12068805	1:170601313	Intergenic	rs9385834	6:138908551	Intergenic	0.654	0.096	7.96e-12	0.467	0.842	0.484	2.08e-11

Single-word reading

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs12566725	1:170601935	Intergenic	rs9385834	6:138908551	Intergenic	0.653	0.096	7.98e-12	0.466	0.84	0.491	2.14e-11
rs1527154	7:113922817	FOXP2	rs2846443	9:113589582	Intergenic	-0.56	0.082	8.04e-12	-0.72	-0.399	0.891	4.98e-11
rs12068805	1:170601313	Intergenic	rs4629702	6:138909908	Intergenic	0.651	0.095	8.08e-12	0.464	0.838	0.459	1.68e-11
rs12566725	1:170601935	Intergenic	rs4629702	6:138909908	Intergenic	0.65	0.095	8.10e-12	0.464	0.836	0.466	1.73e-11
rs12068805	1:170601313	Intergenic	rs111809728	6:138908234	Intergenic	0.654	0.096	8.11e-12	0.467	0.842	0.484	2.13e-11
rs12566725	1:170601935	Intergenic	rs111809728	6:138908234	Intergenic	0.653	0.096	8.13e-12	0.466	0.841	0.491	2.19e-11
rs12068805	1:170601313	Intergenic	rs9389594	6:138907678	Intergenic	0.654	0.096	8.44e-12	0.467	0.842	0.485	2.21e-11
rs12566725	1:170601935	Intergenic	rs9389594	6:138907678	Intergenic	0.653	0.096	8.46e-12	0.466	0.841	0.492	2.28e-11
rs1728436	7:113912412	FOXP2	rs2846443	9:113589582	Intergenic	-0.559	0.082	8.62e-12	-0.72	-0.399	0.886	4.72e-11
rs1527159	7:113930918	FOXP2	rs2846443	9:113589582	Intergenic	-0.557	0.082	8.69e-12	-0.717	-0.397	0.844	3.97e-11
rs1527153	7:113957092	FOXP2	rs1409679	9:113586263	Intergenic	-0.558	0.082	8.70e-12	-0.718	-0.398	0.875	3.46e-11
rs2690842	7:113915642	FOXP2	rs2846443	9:113589582	Intergenic	-0.559	0.082	8.84e-12	-0.719	-0.398	0.888	4.62e-11
rs1209254	7:113917539	FOXP2	rs2846443	9:113589582	Intergenic	-0.558	0.082	8.94e-12	-0.719	-0.398	0.889	4.56e-11
rs1194330	7:113920688	FOXP2	rs2821137	9:113587820	Intergenic	-0.563	0.083	9.10e-12	-0.725	-0.401	0.903	3.98e-11
rs12068805	1:170601313	Intergenic	rs9389593	6:138907676	Intergenic	0.653	0.096	9.41e-12	0.465	0.841	0.47	2.22e-11
rs12566725	1:170601935	Intergenic	rs9389593	6:138907676	Intergenic	0.652	0.096	9.43e-12	0.465	0.84	0.477	2.28e-11
rs12068805	1:170601313	Intergenic	rs58003101	6:138906215	Intergenic	0.654	0.096	9.45e-12	0.466	0.842	0.486	2.48e-11
rs12566725	1:170601935	Intergenic	rs58003101	6:138906215	Intergenic	0.653	0.096	9.47e-12	0.465	0.841	0.493	2.55e-11
rs691686	1:87264545	Intergenic	rs10790477	11:98989176	CNTN5	-1.069	0.157	9.52e-12	-1.377	-0.762	0.51	4.02e-11
rs1527158	7:113926120	FOXP2	rs2821137	9:113587820	Intergenic	-0.556	0.082	9.67e-12	-0.716	-0.396	0.829	1.73e-11
rs797978	1:87263948	Intergenic	rs10790477	11:98989176	CNTN5	-1.069	0.157	9.73e-12	-1.377	-0.761	0.508	4.13e-11
rs12068805	1:170601313	Intergenic	rs35802064	6:138905615	Intergenic	0.654	0.096	9.89e-12	0.466	0.843	0.485	2.61e-11
rs12566725	1:170601935	Intergenic	rs35802064	6:138905615	Intergenic	0.653	0.096	9.91e-12	0.465	0.841	0.492	2.68e-11
rs12068805	1:170601313	Intergenic	rs9389592	6:138906732	Intergenic	0.653	0.096	1.01e-11	0.465	0.841	0.47	2.39e-11
rs12566725	1:170601935	Intergenic	rs9389592	6:138906732	Intergenic	0.652	0.096	1.01e-11	0.464	0.84	0.477	2.46e-11
rs533603	1:170608668	Intergenic	rs6918682	6:138909231	Intergenic	0.555	0.082	1.03e-11	0.395	0.714	0.364	7.13e-11
rs1194330	7:113920688	FOXP2	rs2846443	9:113589582	Intergenic	-0.56	0.082	1.10e-11	-0.722	-0.399	0.899	5.56e-11
rs1527158	7:113926120	FOXP2	rs2846443	9:113589582	Intergenic	-0.554	0.082	1.10e-11	-0.713	-0.394	0.824	2.24e-11
rs578928	1:170609743	Intergenic	rs6918682	6:138909231	Intergenic	0.554	0.082	1.13e-11	0.394	0.713	0.39	6.11e-11
rs1916978	7:113957196	FOXP2	rs1409679	9:113586263	Intergenic	-0.571	0.084	1.21e-11	-0.736	-0.406	0.752	8.39e-11
rs144261163	7:113893884	FOXP2	rs2821137	9:113587820	Intergenic	-0.662	0.098	1.28e-11	-0.854	-0.471	0.346	3.46e-12
rs1527153	7:113957092	FOXP2	rs2821137	9:113587820	Intergenic	-0.554	0.082	1.30e-11	-0.715	-0.394	0.851	6.33e-11
rs644784	1:170606086	Intergenic	rs9385834	6:138908551	Intergenic	0.646	0.096	1.31e-11	0.459	0.834	0.526	2.14e-11

Single-word reading

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	Het	Glide.Pval		
rs656539	1:170606354	Intergenic	rs9385834	6:138908551	Intergenic	0.646	0.096	1.32e-11	0.459	0.833	0.526	2.16e-11
rs644784	1:170606086	Intergenic	rs4629702	6:138909908	Intergenic	0.643	0.095	1.33e-11	0.457	0.829	0.498	1.73e-11
rs644784	1:170606086	Intergenic	rs111809728	6:138908234	Intergenic	0.647	0.096	1.33e-11	0.459	0.834	0.526	2.19e-11
rs656539	1:170606354	Intergenic	rs4629702	6:138909908	Intergenic	0.643	0.095	1.34e-11	0.457	0.829	0.498	1.74e-11
rs12068805	1:170601313	Intergenic	rs34202255	6:138907488	Intergenic	0.653	0.097	1.34e-11	0.464	0.842	0.486	3.08e-11
rs656539	1:170606354	Intergenic	rs111809728	6:138908234	Intergenic	0.646	0.096	1.34e-11	0.459	0.834	0.526	2.20e-11
rs12566725	1:170601935	Intergenic	rs34202255	6:138907488	Intergenic	0.652	0.096	1.34e-11	0.463	0.841	0.493	3.17e-11
rs1916977	7:113957306	FOXP2	rs1409679	9:113586263	Intergenic	-0.569	0.084	1.35e-11	-0.735	-0.404	0.758	7.62e-11
rs644784	1:170606086	Intergenic	rs9389594	6:138907678	Intergenic	0.647	0.096	1.39e-11	0.459	0.834	0.526	2.28e-11
rs656539	1:170606354	Intergenic	rs9389594	6:138907678	Intergenic	0.646	0.096	1.40e-11	0.459	0.834	0.526	2.30e-11
rs12074059	1:170595118	Intergenic	rs9385834	6:138908551	Intergenic	0.645	0.096	1.44e-11	0.458	0.832	0.57	2.13e-11
rs12074059	1:170595118	Intergenic	rs4629702	6:138909908	Intergenic	0.642	0.095	1.45e-11	0.456	0.828	0.541	1.71e-11
rs12074059	1:170595118	Intergenic	rs111809728	6:138908234	Intergenic	0.645	0.096	1.46e-11	0.458	0.833	0.57	2.17e-11
rs533603	1:170608668	Intergenic	rs9385834	6:138908551	Intergenic	0.644	0.095	1.51e-11	0.457	0.831	0.525	2.49e-11
rs12074059	1:170595118	Intergenic	rs9389594	6:138907678	Intergenic	0.645	0.096	1.52e-11	0.458	0.833	0.57	2.26e-11
rs533603	1:170608668	Intergenic	rs4629702	6:138909908	Intergenic	0.641	0.095	1.53e-11	0.455	0.827	0.495	2.01e-11
rs533603	1:170608668	Intergenic	rs6923972	6:138909661	Intergenic	0.55	0.082	1.54e-11	0.39	0.71	0.375	9.88e-11
rs533603	1:170608668	Intergenic	rs111809728	6:138908234	Intergenic	0.644	0.096	1.54e-11	0.457	0.831	0.525	2.54e-11
rs644784	1:170606086	Intergenic	rs9389593	6:138907676	Intergenic	0.645	0.096	1.54e-11	0.458	0.833	0.512	2.28e-11
rs644784	1:170606086	Intergenic	rs58003101	6:138906215	Intergenic	0.646	0.096	1.55e-11	0.458	0.834	0.528	2.55e-11
rs656539	1:170606354	Intergenic	rs9389593	6:138907676	Intergenic	0.645	0.096	1.56e-11	0.458	0.833	0.511	2.30e-11
rs656539	1:170606354	Intergenic	rs58003101	6:138906215	Intergenic	0.646	0.096	1.56e-11	0.458	0.834	0.527	2.57e-11
rs533603	1:170608668	Intergenic	rs9389594	6:138907678	Intergenic	0.644	0.096	1.60e-11	0.457	0.832	0.526	2.65e-11
rs644784	1:170606086	Intergenic	rs35802064	6:138905615	Intergenic	0.646	0.096	1.62e-11	0.458	0.834	0.527	2.68e-11
rs656539	1:170606354	Intergenic	rs35802064	6:138905615	Intergenic	0.646	0.096	1.63e-11	0.458	0.834	0.526	2.70e-11
rs644784	1:170606086	Intergenic	rs9389592	6:138906732	Intergenic	0.645	0.096	1.66e-11	0.457	0.833	0.511	2.46e-11
rs656539	1:170606354	Intergenic	rs9389592	6:138906732	Intergenic	0.645	0.096	1.67e-11	0.457	0.833	0.511	2.48e-11
rs549449	1:170610710	Intergenic	rs4257875	6:138905710	Intergenic	0.556	0.083	1.68e-11	0.394	0.717	0.615	3.52e-11
rs578928	1:170609743	Intergenic	rs6923972	6:138909661	Intergenic	0.549	0.082	1.69e-11	0.389	0.709	0.402	8.48e-11
rs12074059	1:170595118	Intergenic	rs9389593	6:138907676	Intergenic	0.644	0.096	1.69e-11	0.457	0.832	0.555	2.26e-11
rs12074059	1:170595118	Intergenic	rs58003101	6:138906215	Intergenic	0.645	0.096	1.70e-11	0.457	0.833	0.571	2.53e-11
rs1527153	7:113957092	FOXP2	rs2846443	9:113589582	Intergenic	-0.55	0.082	1.71e-11	-0.711	-0.39	0.855	8.07e-11
rs578928	1:170609743	Intergenic	rs7765298	6:138907520	Intergenic	0.551	0.082	1.74e-11	0.391	0.712	0.384	9.92e-11

Single-word reading

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs79874450	6:2636108	Intergenic	rs73022990	19:29748852	Intergenic	-1.304	0.194	1.75e-11	-1.685	-0.924	0.775	1.52e-11
rs533603	1:170608668	Intergenic	rs9389593	6:138907676	Intergenic	0.643	0.096	1.78e-11	0.456	0.831	0.511	2.65e-11
rs12074059	1:170595118	Intergenic	rs35802064	6:138905615	Intergenic	0.645	0.096	1.78e-11	0.457	0.833	0.571	2.66e-11
rs533603	1:170608668	Intergenic	rs58003101	6:138906215	Intergenic	0.644	0.096	1.78e-11	0.456	0.832	0.527	2.96e-11
rs578928	1:170609743	Intergenic	rs9385834	6:138908551	Intergenic	0.642	0.095	1.79e-11	0.455	0.829	0.511	2.58e-11
rs79874450	6:2636108	Intergenic	rs138099966	19:29760811	Intergenic	-1.265	0.188	1.80e-11	-1.634	-0.896	0.741	1.18e-11
rs12074059	1:170595118	Intergenic	rs9389592	6:138906732	Intergenic	0.644	0.096	1.82e-11	0.456	0.832	0.555	2.44e-11
rs578928	1:170609743	Intergenic	rs111809728	6:138908234	Intergenic	0.642	0.096	1.83e-11	0.455	0.829	0.511	2.64e-11
rs578928	1:170609743	Intergenic	rs4629702	6:138909908	Intergenic	0.638	0.095	1.83e-11	0.452	0.825	0.481	2.09e-11
rs77557441	6:2668406	MYLK4	rs138099966	19:29760811	Intergenic	-1.253	0.186	1.83e-11	-1.618	-0.887	0.859	1.51e-11
rs533603	1:170608668	Intergenic	rs35802064	6:138905615	Intergenic	0.644	0.096	1.86e-11	0.456	0.832	0.526	3.11e-11
rs578928	1:170609743	Intergenic	rs9389594	6:138907678	Intergenic	0.642	0.096	1.90e-11	0.454	0.829	0.511	2.74e-11
rs533603	1:170608668	Intergenic	rs9389592	6:138906732	Intergenic	0.643	0.096	1.91e-11	0.455	0.831	0.511	2.85e-11
rs79874450	6:2636108	Intergenic	rs73022996	19:29755516	Intergenic	-1.301	0.194	1.96e-11	-1.681	-0.921	0.775	1.69e-11
rs77557441	6:2668406	MYLK4	rs73022990	19:29748852	Intergenic	-1.285	0.192	1.96e-11	-1.661	-0.91	0.885	1.70e-11
rs79874450	6:2636108	Intergenic	rs73022999	19:29757113	Intergenic	-1.305	0.195	2.00e-11	-1.687	-0.924	0.787	1.60e-11
rs79874450	6:2636108	Intergenic	rs73023000	19:29757133	Intergenic	-1.305	0.195	2.01e-11	-1.687	-0.924	0.787	1.60e-11
rs74735747	6:2666614	MYLK4	rs73022990	19:29748852	Intergenic	-1.283	0.191	2.07e-11	-1.658	-0.907	0.898	1.98e-11
rs12518079	5:130539946	LYRM7	rs7168357	15:97018475	Intergenic	0.674	0.101	2.08e-11	0.477	0.871	0.891	1.58e-11
rs12518055	5:130539826	LYRM7	rs7168357	15:97018475	Intergenic	0.674	0.101	2.08e-11	0.477	0.871	0.891	1.58e-11
rs578928	1:170609743	Intergenic	rs58003101	6:138906215	Intergenic	0.642	0.096	2.11e-11	0.454	0.829	0.513	3.07e-11
rs578928	1:170609743	Intergenic	rs9389593	6:138907676	Intergenic	0.641	0.096	2.12e-11	0.453	0.828	0.496	2.75e-11
rs4706010	5:130546717	Intergenic	rs7168357	15:97018475	Intergenic	0.677	0.101	2.12e-11	0.479	0.875	0.89	1.49e-11
rs74735747	6:2666614	MYLK4	rs138099966	19:29760811	Intergenic	-1.247	0.186	2.13e-11	-1.611	-0.882	0.865	1.82e-11
rs79874450	6:2636108	Intergenic	rs74983087	19:29761422	Intergenic	-1.311	0.196	2.19e-11	-1.695	-0.927	0.811	1.48e-11
rs644784	1:170606086	Intergenic	rs34202255	6:138907488	Intergenic	0.645	0.096	2.20e-11	0.456	0.834	0.528	3.17e-11
rs578928	1:170609743	Intergenic	rs35802064	6:138905615	Intergenic	0.642	0.096	2.21e-11	0.454	0.83	0.512	3.23e-11
rs656539	1:170606354	Intergenic	rs34202255	6:138907488	Intergenic	0.645	0.096	2.21e-11	0.456	0.834	0.528	3.19e-11
rs79874450	6:2636108	Intergenic	rs73024908	19:29762766	Intergenic	-1.312	0.196	2.22e-11	-1.696	-0.927	0.813	1.48e-11
rs2961694	5:141543351	Intergenic	rs180662	10:117758764	Intergenic	-0.995	0.149	2.25e-11	-1.286	-0.703	0.57	6.92e-11
rs17135412	6:2666428	MYLK4	rs73022990	19:29748852	Intergenic	-1.28	0.191	2.25e-11	-1.656	-0.905	0.892	2.09e-11
rs578928	1:170609743	Intergenic	rs9389592	6:138906732	Intergenic	0.641	0.096	2.27e-11	0.453	0.828	0.496	2.96e-11
rs17135412	6:2666428	MYLK4	rs138099966	19:29760811	Intergenic	-1.244	0.186	2.34e-11	-1.609	-0.879	0.857	1.89e-11

Single-word reading

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs12074059	1:170595118	Intergenic	rs34202255	6:138907488	Intergenic	0.644	0.096	2.39e-11	0.455	0.833	0.572	3.14e-11
rs2961694	5:141543351	Intergenic	rs180655	10:117763073	Intergenic	-0.993	0.149	2.40e-11	-1.285	-0.702	0.772	6.40e-11
rs9392409	6:2637106	C6orf195	rs73022990	19:29748852	Intergenic	-1.293	0.194	2.47e-11	-1.673	-0.913	0.801	1.65e-11
rs4770836	13:26037909	ATP8A2	rs12961650	18:27431670	Intergenic	-0.5	0.075	2.48e-11	-0.646	-0.353	0.478	6.96e-11
rs77557441	6:2668406	MYLK4	rs73022996	19:29755516	Intergenic	-1.278	0.192	2.50e-11	-1.654	-0.903	0.891	2.25e-11
rs533603	1:170608668	Intergenic	rs34202255	6:138907488	Intergenic	0.643	0.096	2.53e-11	0.454	0.831	0.528	3.67e-11
rs9392409	6:2637106	C6orf195	rs138099966	19:29760811	Intergenic	-1.255	0.188	2.54e-11	-1.623	-0.886	0.767	1.25e-11
rs4770837	13:26037960	ATP8A2	rs12961650	18:27431670	Intergenic	-0.499	0.075	2.54e-11	-0.646	-0.353	0.479	6.91e-11
rs77557441	6:2668406	MYLK4	rs73022999	19:29757113	Intergenic	-1.283	0.192	2.58e-11	-1.66	-0.906	0.897	2.17e-11
rs77557441	6:2668406	MYLK4	rs73023000	19:29757133	Intergenic	-1.283	0.192	2.59e-11	-1.66	-0.906	0.897	2.17e-11
rs549449	1:170610710	Intergenic	rs9385834	6:138908551	Intergenic	0.636	0.095	2.63e-11	0.449	0.824	0.471	3.73e-11
rs549449	1:170610710	Intergenic	rs111809728	6:138908234	Intergenic	0.637	0.096	2.67e-11	0.449	0.824	0.472	3.80e-11
rs549449	1:170610710	Intergenic	rs4629702	6:138909908	Intergenic	0.633	0.095	2.69e-11	0.447	0.819	0.431	3.03e-11
rs74735747	6:2666614	MYLK4	rs73022996	19:29755516	Intergenic	-1.275	0.191	2.70e-11	-1.65	-0.9	0.902	2.63e-11
rs9392409	6:2637106	C6orf195	rs73022996	19:29755516	Intergenic	-1.29	0.194	2.75e-11	-1.669	-0.91	0.8	1.83e-11
rs549449	1:170610710	Intergenic	rs9389594	6:138907678	Intergenic	0.637	0.096	2.77e-11	0.449	0.824	0.472	3.96e-11
rs74735747	6:2666614	MYLK4	rs73022999	19:29757113	Intergenic	-1.28	0.192	2.79e-11	-1.656	-0.903	0.907	2.54e-11
rs74735747	6:2666614	MYLK4	rs73023000	19:29757133	Intergenic	-1.28	0.192	2.79e-11	-1.656	-0.903	0.908	2.54e-11
rs77557441	6:2668406	MYLK4	rs74983087	19:29761422	Intergenic	-1.289	0.194	2.82e-11	-1.669	-0.91	0.907	2.04e-11
rs9392409	6:2637106	C6orf195	rs73022999	19:29757113	Intergenic	-1.294	0.194	2.83e-11	-1.675	-0.913	0.81	1.73e-11
rs9392409	6:2637106	C6orf195	rs73023000	19:29757133	Intergenic	-1.294	0.194	2.83e-11	-1.675	-0.913	0.811	1.73e-11
rs77557441	6:2668406	MYLK4	rs73024908	19:29762766	Intergenic	-1.29	0.194	2.85e-11	-1.67	-0.91	0.907	2.03e-11
rs17135412	6:2666428	MYLK4	rs73022996	19:29755516	Intergenic	-1.273	0.191	2.95e-11	-1.648	-0.898	0.896	2.77e-11
rs578928	1:170609743	Intergenic	rs34202255	6:138907488	Intergenic	0.64	0.096	3.00e-11	0.451	0.829	0.514	3.81e-11
rs17135412	6:2666428	MYLK4	rs73022999	19:29757113	Intergenic	-1.277	0.192	3.05e-11	-1.654	-0.9	0.901	2.67e-11
rs17135412	6:2666428	MYLK4	rs73023000	19:29757133	Intergenic	-1.277	0.192	3.05e-11	-1.654	-0.9	0.901	2.67e-11
rs74735747	6:2666614	MYLK4	rs74983087	19:29761422	Intergenic	-1.286	0.194	3.05e-11	-1.665	-0.906	0.916	2.37e-11
rs9392409	6:2637106	C6orf195	rs74983087	19:29761422	Intergenic	-1.3	0.196	3.08e-11	-1.683	-0.916	0.832	1.58e-11
rs549449	1:170610710	Intergenic	rs58003101	6:138906215	Intergenic	0.636	0.096	3.08e-11	0.449	0.824	0.474	4.41e-11
rs74735747	6:2666614	MYLK4	rs73024908	19:29762766	Intergenic	-1.287	0.194	3.09e-11	-1.666	-0.907	0.916	2.36e-11
rs549449	1:170610710	Intergenic	rs9389593	6:138907676	Intergenic	0.635	0.096	3.09e-11	0.448	0.823	0.458	3.96e-11
rs9392409	6:2637106	C6orf195	rs73024908	19:29762766	Intergenic	-1.3	0.196	3.12e-11	-1.684	-0.917	0.834	1.58e-11
rs507012	13:26034360	ATP8A2	rs12961650	18:27431670	Intergenic	-0.499	0.075	3.16e-11	-0.646	-0.351	0.468	9.33e-11

Single-word reading

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs72787011	5:130538051	LYRM7	rs7168357	15:97018475	Intergenic	0.668	0.101	3.21e-11	0.471	0.865	0.876	2.00e-11
rs549449	1:170610710	Intergenic	rs35802064	6:138905615	Intergenic	0.636	0.096	3.21e-11	0.448	0.824	0.473	4.64e-11
rs549449	1:170610710	Intergenic	rs9389592	6:138906732	Intergenic	0.635	0.096	3.31e-11	0.448	0.823	0.458	4.26e-11
rs17135412	6:2666428	MYLK4	rs74983087	19:29761422	Intergenic	-1.283	0.194	3.34e-11	-1.662	-0.904	0.91	2.50e-11
rs17135412	6:2666428	MYLK4	rs73024908	19:29762766	Intergenic	-1.284	0.194	3.38e-11	-1.664	-0.905	0.91	2.49e-11
rs10113869	9:79600226	Intergenic	rs9548110	13:38500040	Intergenic	-0.61	0.092	3.53e-11	-0.79	-0.429	0.992	6.96e-11
rs10869859	9:79600922	Intergenic	rs9548110	13:38500040	Intergenic	-0.61	0.092	3.59e-11	-0.79	-0.429	0.992	7.09e-11
rs10869854	9:79599582	Intergenic	rs9548110	13:38500040	Intergenic	-0.61	0.092	3.62e-11	-0.791	-0.43	0.99	7.17e-11
rs10869855	9:79599614	Intergenic	rs9548110	13:38500040	Intergenic	-0.61	0.092	3.62e-11	-0.791	-0.43	0.991	7.17e-11
rs113372431	5:130537926	LYRM7	rs7168357	15:97018475	Intergenic	0.647	0.098	3.83e-11	0.455	0.839	0.734	3.85e-11
rs578928	1:170609743	Intergenic	rs4257875	6:138905710	Intergenic	0.546	0.083	3.89e-11	0.384	0.708	0.397	9.68e-11
rs10488529	7:92653334	Intergenic	rs2181324	10:98145284	TLL2	-1.039	0.157	4.00e-11	-1.348	-0.731	0.947	2.75e-11
rs7851647	9:79597761	Intergenic	rs9548110	13:38500040	Intergenic	-0.61	0.092	4.15e-11	-0.791	-0.429	0.986	8.30e-11
rs2377819	9:79597199	Intergenic	rs9548110	13:38500040	Intergenic	-0.61	0.092	4.19e-11	-0.792	-0.429	0.984	8.40e-11
rs10488529	7:92653334	Intergenic	rs17111762	10:98141748	TLL2	-1.037	0.157	4.25e-11	-1.345	-0.729	0.948	2.72e-11
rs10488529	7:92653334	Intergenic	rs56094817	10:98142008	TLL2	-1.037	0.157	4.25e-11	-1.345	-0.729	0.948	2.72e-11
rs977368	11:43210233	Intergenic	rs608519	18:75368523	Intergenic	-0.711	0.108	4.25e-11	-0.923	-0.5	0.953	3.21e-12
rs144261163	7:113893884	FOXP2	rs2846443	9:113589582	Intergenic	-0.657	0.1	4.28e-11	-0.852	-0.462	0.333	5.16e-12
rs10869852	9:79596796	Intergenic	rs9548110	13:38500040	Intergenic	-0.61	0.093	4.38e-11	-0.791	-0.429	0.984	8.76e-11
rs549449	1:170610710	Intergenic	rs34202255	6:138907488	Intergenic	0.635	0.096	4.38e-11	0.446	0.824	0.476	5.47e-11
rs7144982	14:103540728	Intergenic	rs8045102	16:49958285	Intergenic	0.613	0.093	4.53e-11	0.431	0.796	0.515	5.57e-11
rs28479678	5:130533902	LYRM7	rs7168357	15:97018475	Intergenic	0.657	0.1	4.71e-11	0.461	0.853	0.91	1.75e-11
rs4854375	2:763618	Intergenic	rs114158060	6:90261115	ANKRD6	-1.008	0.154	5.29e-11	-1.309	-0.707	0.528	8.48e-11
rs10056240	5:156972294	ADAM19	rs113043616	18:26967794	Intergenic	-0.756	0.115	5.54e-11	-0.981	-0.53	0.749	9.93e-11
rs74751208	3:147377956	Intergenic	rs9301999	13:95457166	Intergenic	0.962	0.147	5.96e-11	0.674	1.25	0.514	9.30e-11
rs2037054	12:101644517	Intergenic	rs6029105	20:39041594	Intergenic	-0.817	0.125	6.19e-11	-1.062	-0.572	0.473	8.68e-11
rs7719543	5:130534556	LYRM7	rs7168357	15:97018475	Intergenic	0.635	0.097	6.33e-11	0.445	0.826	0.78	3.38e-11
rs6879803	5:130534325	LYRM7	rs7168357	15:97018475	Intergenic	0.636	0.097	6.36e-11	0.445	0.827	0.775	3.29e-11
rs4706009	5:130536044	LYRM7	rs7168357	15:97018475	Intergenic	0.635	0.097	6.37e-11	0.444	0.825	0.779	3.37e-11
rs10488529	7:92653334	Intergenic	rs2093557	10:98140110	TLL2	-1.029	0.158	6.96e-11	-1.338	-0.719	0.942	4.07e-11
rs79874450	6:2636108	Intergenic	rs116561825	19:29740738	Intergenic	-1.276	0.196	7.41e-11	-1.66	-0.892	0.823	7.01e-11
rs77557441	6:2668406	MYLK4	rs116561825	19:29740738	Intergenic	-1.26	0.194	7.64e-11	-1.639	-0.88	0.925	7.24e-11
rs9930290	16:50370543	BRD7	rs6029105	20:39041594	Intergenic	1.068	0.164	7.65e-11	0.747	1.39	0.436	6.71e-11

Single-word reading

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs79874450	6:2636108	Intergenic	rs112289248	19:29744225	Intergenic	-1.274	0.196	7.71e-11	-1.658	-0.89	0.815	7.03e-11
rs74735747	6:2666614	MYLK4	rs116561825	19:29740738	Intergenic	-1.257	0.193	7.99e-11	-1.636	-0.878	0.935	8.26e-11
rs8059133	16:50370703	BRD7	rs6029105	20:39041594	Intergenic	1.064	0.164	8.13e-11	0.743	1.385	0.455	7.08e-11
rs77557441	6:2668406	MYLK4	rs112289248	19:29744225	Intergenic	-1.258	0.194	8.13e-11	-1.637	-0.878	0.921	7.49e-11
rs4595801	16:50369888	BRD7	rs6029105	20:39041594	Intergenic	1.063	0.164	8.42e-11	0.742	1.384	0.452	7.35e-11
rs74735747	6:2666614	MYLK4	rs112289248	19:29744225	Intergenic	-1.255	0.193	8.54e-11	-1.634	-0.876	0.931	8.56e-11
rs17135412	6:2666428	MYLK4	rs116561825	19:29740738	Intergenic	-1.255	0.193	8.66e-11	-1.634	-0.876	0.931	8.67e-11
rs79874450	6:2636108	Intergenic	rs57732835	19:29737939	Intergenic	-1.27	0.196	8.85e-11	-1.654	-0.886	0.815	8.22e-11
rs10491723	9:100927632	CORO2A	rs7130200	11:111211691	Intergenic	-0.461	0.071	8.88e-11	-0.6	-0.321	0.89	8.63e-11
rs77557441	6:2668406	MYLK4	rs57732835	19:29737939	Intergenic	-1.255	0.194	8.97e-11	-1.634	-0.876	0.919	8.40e-11
rs17135412	6:2666428	MYLK4	rs112289248	19:29744225	Intergenic	-1.253	0.193	9.27e-11	-1.632	-0.874	0.926	8.99e-11
rs74735747	6:2666614	MYLK4	rs57732835	19:29737939	Intergenic	-1.252	0.193	9.38e-11	-1.631	-0.873	0.929	9.58e-11
rs2079511	5:130560128	Intergenic	rs7168357	15:97018475	Intergenic	0.658	0.102	1.00e-10	0.459	0.858	0.922	2.90e-11

Phonological awareness

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs620291	11:113113858	NCAM1	rs113400479	18:23026571	Intergenic	-1.214	0.169	8.06e-13	-1.546	-0.881	0.812	6.65e-13
rs635596	11:113113515	NCAM1	rs113400479	18:23026571	Intergenic	-1.207	0.169	9.05e-13	-1.538	-0.876	0.81	7.01e-13
rs586903	11:113110946	NCAM1	rs113400479	18:23026571	Intergenic	-1.251	0.175	1.01e-12	-1.595	-0.907	0.393	1.64e-11
rs587761	11:113110780	NCAM1	rs78616234	18:23020632	Intergenic	-1.229	0.174	1.85e-12	-1.571	-0.887	0.49	1.46e-11
rs586903	11:113110946	NCAM1	rs78616234	18:23020632	Intergenic	-1.223	0.174	1.87e-12	-1.563	-0.883	0.538	1.53e-11
rs586903	11:113110946	NCAM1	rs76799957	18:23026460	Intergenic	-1.219	0.173	1.89e-12	-1.559	-0.88	0.529	1.62e-11
rs587761	11:113110780	NCAM1	rs76799957	18:23026460	Intergenic	-1.226	0.174	1.89e-12	-1.567	-0.884	0.482	1.55e-11
rs586903	11:113110946	NCAM1	rs79046421	18:23031099	Intergenic	-1.211	0.173	2.22e-12	-1.549	-0.873	0.512	1.96e-11
rs587761	11:113110780	NCAM1	rs79046421	18:23031099	Intergenic	-1.218	0.173	2.22e-12	-1.558	-0.878	0.463	1.86e-11
rs586903	11:113110946	NCAM1	rs75070995	18:23025063	Intergenic	-1.212	0.173	2.39e-12	-1.55	-0.873	0.537	1.99e-11
rs587761	11:113110780	NCAM1	rs75070995	18:23025063	Intergenic	-1.217	0.174	2.40e-12	-1.558	-0.877	0.491	1.89e-11
rs587761	11:113110780	NCAM1	rs113400479	18:23026571	Intergenic	-1.257	0.179	2.41e-12	-1.608	-0.905	0.379	1.57e-11
rs620291	11:113113858	NCAM1	rs76799957	18:23026460	Intergenic	-1.177	0.168	2.44e-12	-1.507	-0.848	0.864	6.59e-13
rs620291	11:113113858	NCAM1	rs78616234	18:23020632	Intergenic	-1.177	0.168	2.55e-12	-1.506	-0.847	0.871	6.48e-13
rs635596	11:113113515	NCAM1	rs76799957	18:23026460	Intergenic	-1.171	0.167	2.71e-12	-1.499	-0.843	0.872	6.94e-13
rs635596	11:113113515	NCAM1	rs78616234	18:23020632	Intergenic	-1.173	0.168	2.73e-12	-1.502	-0.844	0.886	6.81e-13
rs620291	11:113113858	NCAM1	rs79046421	18:23031099	Intergenic	-1.171	0.168	2.77e-12	-1.499	-0.843	0.844	7.84e-13
rs587761	11:113110780	NCAM1	rs76340243	18:23028090	Intergenic	-1.219	0.175	3.03e-12	-1.562	-0.877	0.509	2.40e-11
rs620291	11:113113858	NCAM1	rs75070995	18:23025063	Intergenic	-1.17	0.168	3.07e-12	-1.499	-0.841	0.869	8.30e-13
rs635596	11:113113515	NCAM1	rs79046421	18:23031099	Intergenic	-1.164	0.167	3.13e-12	-1.491	-0.837	0.851	8.25e-13
rs587761	11:113110780	NCAM1	rs79295788	18:23028857	Intergenic	-1.217	0.175	3.13e-12	-1.56	-0.875	0.506	2.48e-11
rs587761	11:113110780	NCAM1	rs9916920	18:23028354	Intergenic	-1.218	0.175	3.14e-12	-1.56	-0.875	0.513	2.48e-11
rs586903	11:113110946	NCAM1	rs76340243	18:23028090	Intergenic	-1.212	0.174	3.17e-12	-1.553	-0.871	0.563	2.53e-11
rs587761	11:113110780	NCAM1	rs76146087	18:23029422	Intergenic	-1.217	0.175	3.19e-12	-1.559	-0.874	0.504	2.54e-11
rs586903	11:113110946	NCAM1	rs79295788	18:23028857	Intergenic	-1.21	0.174	3.28e-12	-1.551	-0.87	0.559	2.62e-11

Phonological awareness

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs586903	11:113110946	NCAM1	rs9916920	18:23028354	Intergenic	-1.21	0.174	3.30e-12	-1.551	-0.87	0.567	2.61e-11
rs587761	11:113110780	NCAM1	rs12458786	18:23030650	Intergenic	-1.215	0.174	3.33e-12	-1.557	-0.873	0.5	2.67e-11
rs586903	11:113110946	NCAM1	rs76146087	18:23029422	Intergenic	-1.209	0.174	3.35e-12	-1.55	-0.869	0.558	2.68e-11
rs635596	11:113113515	NCAM1	rs75070995	18:23025063	Intergenic	-1.163	0.167	3.40e-12	-1.491	-0.836	0.881	8.73e-13
rs587761	11:113110780	NCAM1	rs76878683	18:23023338	Intergenic	-1.216	0.175	3.43e-12	-1.558	-0.873	0.531	2.71e-11
rs586903	11:113110946	NCAM1	rs12458786	18:23030650	Intergenic	-1.207	0.174	3.49e-12	-1.547	-0.867	0.555	2.83e-11
rs586903	11:113110946	NCAM1	rs76878683	18:23023338	Intergenic	-1.209	0.174	3.57e-12	-1.55	-0.868	0.583	2.87e-11
rs620291	11:113113858	NCAM1	rs76340243	18:23028090	Intergenic	-1.173	0.169	3.74e-12	-1.504	-0.842	0.882	1.09e-12
rs620291	11:113113858	NCAM1	rs79295788	18:23028857	Intergenic	-1.172	0.169	3.83e-12	-1.503	-0.841	0.879	1.12e-12
rs620291	11:113113858	NCAM1	rs9916920	18:23028354	Intergenic	-1.172	0.169	3.86e-12	-1.503	-0.841	0.883	1.13e-12
rs620291	11:113113858	NCAM1	rs76146087	18:23029422	Intergenic	-1.171	0.169	3.90e-12	-1.502	-0.84	0.877	1.15e-12
rs620291	11:113113858	NCAM1	rs12458786	18:23030650	Intergenic	-1.17	0.169	4.03e-12	-1.5	-0.839	0.872	1.21e-12
rs635596	11:113113515	NCAM1	rs76340243	18:23028090	Intergenic	-1.166	0.168	4.22e-12	-1.496	-0.836	0.89	1.14e-12
rs620291	11:113113858	NCAM1	rs76878683	18:23023338	Intergenic	-1.169	0.169	4.30e-12	-1.5	-0.838	0.892	1.29e-12
rs635596	11:113113515	NCAM1	rs79295788	18:23028857	Intergenic	-1.164	0.168	4.34e-12	-1.494	-0.835	0.887	1.17e-12
rs635596	11:113113515	NCAM1	rs9916920	18:23028354	Intergenic	-1.165	0.168	4.36e-12	-1.494	-0.835	0.892	1.18e-12
rs635596	11:113113515	NCAM1	rs76146087	18:23029422	Intergenic	-1.164	0.168	4.42e-12	-1.493	-0.834	0.885	1.20e-12
rs635596	11:113113515	NCAM1	rs12458786	18:23030650	Intergenic	-1.162	0.168	4.59e-12	-1.491	-0.833	0.88	1.26e-12
rs635596	11:113113515	NCAM1	rs76878683	18:23023338	Intergenic	-1.163	0.168	4.77e-12	-1.492	-0.833	0.906	1.34e-12
rs10038365	5:108602269	Intergenic	rs10140868	14:62616479	Intergenic	0.582	0.084	5.42e-12	0.417	0.748	0.395	4.16e-10
rs58008430	5:108595048	Intergenic	rs10140868	14:62616479	Intergenic	0.582	0.085	5.85e-12	0.416	0.747	0.39	4.42e-10
rs6885647	5:108593449	Intergenic	rs10140868	14:62616479	Intergenic	0.582	0.085	5.90e-12	0.416	0.747	0.389	4.46e-10
rs12655632	5:108592797	Intergenic	rs10140868	14:62616479	Intergenic	0.579	0.084	6.86e-12	0.413	0.744	0.399	6.85e-10
rs10039717	5:108577839	Intergenic	rs9671388	14:62626952	Intergenic	0.62	0.091	1.01e-11	0.442	0.799	0.802	4.04e-10
rs10038277	5:108602064	Intergenic	rs10140868	14:62616479	Intergenic	0.576	0.085	1.01e-11	0.41	0.742	0.412	7.36e-10

Phonological awareness

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs12519126	5:108602037	Intergenic	rs10140868	14:62616479	Intergenic	0.576	0.085	1.02e-11	0.41	0.742	0.413	7.41e-10
rs6866681	5:108593535	Intergenic	rs10140868	14:62616479	Intergenic	0.576	0.085	1.08e-11	0.41	0.742	0.409	7.74e-10
rs620291	11:113113858	NCAM1	rs12458251	18:23022456	Intergenic	1.173	0.173	1.11e-11	0.835	1.512	0.762	7.75e-12
rs2600168	3:9376490	Intergenic	rs206514	18:10411194	Intergenic	-0.665	0.098	1.11e-11	-0.857	-0.473	0.595	2.03e-10
rs2600168	3:9376490	Intergenic	rs206515	18:10411039	Intergenic	-0.665	0.098	1.11e-11	-0.857	-0.473	0.594	2.03e-10
rs1833570	5:108601482	Intergenic	rs10140868	14:62616479	Intergenic	0.577	0.085	1.12e-11	0.411	0.744	0.409	7.66e-10
rs6882785	5:108601897	Intergenic	rs10140868	14:62616479	Intergenic	0.575	0.085	1.18e-11	0.409	0.741	0.398	1.25e-09
rs635596	11:113113515	NCAM1	rs12458251	18:23022456	Intergenic	1.17	0.173	1.19e-11	0.832	1.508	0.745	8.09e-12
rs2430802	8:22837121	Intergenic	rs11014020	10:24598415	KIAA1217	-0.726	0.107	1.23e-11	-0.936	-0.516	0.601	1.23e-09
rs10038365	5:108602269	Intergenic	rs9671388	14:62626952	Intergenic	0.582	0.086	1.24e-11	0.414	0.751	0.767	1.96e-10
rs58008430	5:108595048	Intergenic	rs9671388	14:62626952	Intergenic	0.582	0.086	1.26e-11	0.414	0.751	0.765	2.01e-10
rs6885647	5:108593449	Intergenic	rs9671388	14:62626952	Intergenic	0.582	0.086	1.33e-11	0.413	0.75	0.764	2.09e-10
rs11241001	5:108608418	Intergenic	rs10140868	14:62616479	Intergenic	0.581	0.086	1.34e-11	0.412	0.749	0.4	2.87e-10
rs60191921	5:108606053	Intergenic	rs10140868	14:62616479	Intergenic	0.573	0.085	1.45e-11	0.406	0.739	0.42	3.53e-10
rs12655632	5:108592797	Intergenic	rs9671388	14:62626952	Intergenic	0.579	0.086	1.48e-11	0.411	0.748	0.78	3.45e-10
rs2016908	5:108575251	Intergenic	rs9671388	14:62626952	Intergenic	0.614	0.091	1.65e-11	0.435	0.793	0.757	8.97e-10
rs587761	11:113110780	NCAM1	rs12458251	18:23022456	Intergenic	-1.195	0.178	1.65e-11	-1.543	-0.847	0.403	1.62e-10
rs587761	11:113110780	NCAM1	rs12458251	18:23022456	Intergenic	1.195	0.177	1.67e-11	0.847	1.543	0.41	1.62e-10
rs586903	11:113110946	NCAM1	rs12458251	18:23022456	Intergenic	1.189	0.177	1.70e-11	0.843	1.535	0.449	1.69e-10
rs586903	11:113110946	NCAM1	rs12458251	18:23022456	Intergenic	-1.189	0.177	1.71e-11	-1.535	-0.842	0.449	1.69e-10
rs2466236	8:22836788	Intergenic	rs11014020	10:24598415	KIAA1217	-0.725	0.108	1.76e-11	-0.936	-0.513	0.602	1.62e-09
rs620291	11:113113858	NCAM1	rs12458251	18:23022456	Intergenic	-1.147	0.171	2.00e-11	-1.482	-0.812	0.745	7.75e-12
rs635596	11:113113515	NCAM1	rs12458251	18:23022456	Intergenic	-1.143	0.171	2.18e-11	-1.478	-0.808	0.757	8.09e-12
rs10038277	5:108602064	Intergenic	rs9671388	14:62626952	Intergenic	0.576	0.086	2.22e-11	0.408	0.745	0.795	3.62e-10
rs2466237	8:22835860	Intergenic	rs11014020	10:24598415	KIAA1217	-0.723	0.108	2.23e-11	-0.934	-0.511	0.608	2.06e-09

Phonological awareness

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs6866681	5:108593535	Intergenic	rs9671388	14:62626952	Intergenic	0.577	0.086	2.23e-11	0.408	0.746	0.792	3.60e-10
rs12519126	5:108602037	Intergenic	rs9671388	14:62626952	Intergenic	0.576	0.086	2.33e-11	0.407	0.745	0.795	3.73e-10
rs1833570	5:108601482	Intergenic	rs9671388	14:62626952	Intergenic	0.577	0.086	2.38e-11	0.408	0.747	0.791	3.50e-10
rs2600168	3:9376490	Intergenic	rs404472	18:10412186	Intergenic	-0.663	0.099	2.38e-11	-0.857	-0.468	0.709	2.90e-10
rs75588101	7:134710913	AGBL3	rs927829	10:53253804	PRKG1	-1.348	0.202	2.45e-11	-1.743	-0.952	0.442	6.71e-11
rs113450037	7:134716098	AGBL3	rs927829	10:53253804	PRKG1	-1.35	0.202	2.46e-11	-1.746	-0.954	0.436	6.50e-11
rs144367847	7:134753328	AGBL3	rs927829	10:53253804	PRKG1	-1.352	0.202	2.47e-11	-1.749	-0.955	0.431	6.57e-11
rs11978136	7:134776093	AGBL3	rs927829	10:53253804	PRKG1	-1.353	0.203	2.56e-11	-1.751	-0.956	0.438	6.43e-11
rs76261347	7:134762945	AGBL3	rs927829	10:53253804	PRKG1	-1.353	0.203	2.57e-11	-1.751	-0.956	0.437	6.46e-11
rs17231184	7:134785709	AGBL3	rs927829	10:53253804	PRKG1	-1.352	0.203	2.58e-11	-1.75	-0.955	0.438	6.44e-11
rs78006222	7:134783732	AGBL3	rs927829	10:53253804	PRKG1	-1.352	0.203	2.58e-11	-1.75	-0.955	0.438	6.44e-11
rs11970908	7:134787934	AGBL3	rs927829	10:53253804	PRKG1	-1.352	0.203	2.58e-11	-1.75	-0.955	0.438	6.45e-11
rs78463102	7:134788570	AGBL3	rs927829	10:53253804	PRKG1	-1.352	0.203	2.58e-11	-1.75	-0.955	0.438	6.45e-11
rs76864211	7:134803527	AGBL3	rs927829	10:53253804	PRKG1	-1.352	0.203	2.58e-11	-1.749	-0.954	0.437	6.44e-11
rs77913513	7:134795841	AGBL3	rs927829	10:53253804	PRKG1	-1.352	0.203	2.59e-11	-1.749	-0.955	0.437	6.46e-11
rs148321733	7:134810967	AGBL3	rs927829	10:53253804	PRKG1	-1.351	0.203	2.59e-11	-1.748	-0.954	0.438	6.45e-11
rs112521537	7:134804487	AGBL3	rs927829	10:53253804	PRKG1	-1.351	0.203	2.59e-11	-1.749	-0.954	0.437	6.47e-11
rs113954552	7:134808002	AGBL3	rs927829	10:53253804	PRKG1	-1.351	0.203	2.60e-11	-1.748	-0.954	0.437	6.48e-11
rs7732292	5:39641154	Intergenic	rs4910588	11:3963675	STIM1	-0.717	0.108	2.66e-11	-0.928	-0.506	0.865	3.18e-09
rs6843046	4:124400993	Intergenic	rs4480666	13:36713946	Intergenic	-0.938	0.141	2.80e-11	-1.214	-0.662	0.508	4.84e-11
rs6843046	4:124400993	Intergenic	rs9565875	13:36708040	Intergenic	-0.937	0.141	2.83e-11	-1.213	-0.661	0.511	5.34e-11
rs6843046	4:124400993	Intergenic	rs1328652	13:36709527	Intergenic	-0.937	0.141	2.83e-11	-1.213	-0.661	0.511	5.34e-11
rs6843046	4:124400993	Intergenic	rs9565874	13:36707981	Intergenic	-0.937	0.141	2.83e-11	-1.213	-0.661	0.511	5.35e-11
rs6843046	4:124400993	Intergenic	rs9575380	13:36708286	Intergenic	-0.937	0.141	2.84e-11	-1.213	-0.661	0.511	5.35e-11
rs17321024	2:45525440	Intergenic	rs10809384	9:11200242	Intergenic	-1.397	0.21	2.86e-11	-1.808	-0.985	0.973	2.24e-10

Phonological awareness

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs10039717	5:108577839	Intergenic	rs10140868	14:62616479	Intergenic	0.589	0.089	2.87e-11	0.416	0.763	0.461	4.85e-09
rs6843046	4:124400993	Intergenic	rs1928017	13:36726009	Intergenic	-0.943	0.142	3.01e-11	-1.221	-0.665	0.415	4.92e-11
rs6882785	5:108601897	Intergenic	rs9671388	14:62626952	Intergenic	0.573	0.086	3.17e-11	0.404	0.742	0.801	6.88e-10
rs6843046	4:124400993	Intergenic	rs9575407	13:36716901	Intergenic	-0.936	0.141	3.26e-11	-1.213	-0.66	0.483	5.26e-11
rs2600168	3:9376490	Intergenic	rs12964263	18:10410256	Intergenic	-0.653	0.098	3.27e-11	-0.846	-0.46	0.722	5.23e-10
rs60191921	5:108606053	Intergenic	rs9671388	14:62626952	Intergenic	0.572	0.086	3.28e-11	0.403	0.742	0.813	1.93e-10
rs6843046	4:124400993	Intergenic	rs2149422	13:36724392	Intergenic	-0.937	0.141	3.34e-11	-1.214	-0.66	0.431	5.51e-11
rs2600168	3:9376490	Intergenic	rs12963926	18:10410275	Intergenic	-0.653	0.098	3.36e-11	-0.846	-0.46	0.712	4.44e-10
rs6843046	4:124400993	Intergenic	rs55785621	13:36718445	Intergenic	-0.935	0.141	3.44e-11	-1.212	-0.659	0.474	5.42e-11
rs6848895	4:158626629	Intergenic	rs1390204	12:71587160	Intergenic	0.593	0.09	3.49e-11	0.418	0.769	0.793	3.06e-11
rs492943	10:111629030	XPNPEP1	rs10824779	10:54488855	Intergenic	0.604	0.091	3.68e-11	0.425	0.783	0.366	1.75e-11
rs492943	10:111629030	XPNPEP1	rs11003084	10:54488686	Intergenic	0.603	0.091	3.71e-11	0.425	0.782	0.364	2.00e-11
rs35502738	5:39629430	Intergenic	rs4910588	11:3963675	STIM1	-0.744	0.113	3.71e-11	-0.965	-0.524	0.996	2.63e-09
rs1592805	5:108621580	Intergenic	rs10140868	14:62616479	Intergenic	0.564	0.085	3.72e-11	0.397	0.732	0.387	6.19e-10
rs62338559	5:3322058	Intergenic	rs3815802	16:58296545	CCDC113	-0.579	0.088	3.75e-11	-0.751	-0.408	0.99	4.81e-10
rs76579103	7:134745973	AGBL3	rs927829	10:53253804	PRKG1	-1.36	0.206	3.82e-11	-1.764	-0.957	0.479	1.39e-10
rs77845976	7:134814984	AGBL3	rs927829	10:53253804	PRKG1	-1.336	0.202	3.83e-11	-1.732	-0.94	0.401	8.74e-11
rs6843046	4:124400993	Intergenic	rs11839407	13:36721457	Intergenic	-0.934	0.141	3.84e-11	-1.211	-0.657	0.454	5.74e-11
rs1354128	5:39634963	Intergenic	rs4910588	11:3963675	STIM1	-0.715	0.108	3.88e-11	-0.927	-0.503	0.878	2.54e-09
rs35695483	7:136817670	LOC349160	rs833828	12:49352257	ARF3	0.701	0.106	3.93e-11	0.493	0.908	0.542	2.36e-10
rs11982570	7:136817016	LOC349160	rs833828	12:49352257	ARF3	0.7	0.106	3.94e-11	0.493	0.908	0.543	2.36e-10
rs36110794	7:136835778	LOC349160	rs833828	12:49352257	ARF3	0.703	0.106	3.99e-11	0.494	0.911	0.53	3.33e-10
rs17168934	7:136808272	LOC349160	rs833828	12:49352257	ARF3	0.7	0.106	4.04e-11	0.492	0.908	0.545	2.36e-10
rs67751562	7:136833387	LOC349160	rs833828	12:49352257	ARF3	0.702	0.106	4.05e-11	0.494	0.911	0.528	3.37e-10
rs73450817	7:136810307	LOC349160	rs833828	12:49352257	ARF3	0.7	0.106	4.05e-11	0.492	0.908	0.545	2.37e-10

Phonological awareness

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs73450814	7:136810300	LOC349160	rs833828	12:49352257	ARF3	0.7	0.106	4.05e-11	0.492	0.908	0.545	2.37e-10
rs713333	1:81725122	Intergenic	rs7081305	10:28112405	ARMC4	-0.606	0.092	4.05e-11	-0.786	-0.426	0.672	1.26e-09
rs12707345	7:136831401	LOC349160	rs833828	12:49352257	ARF3	0.702	0.106	4.11e-11	0.494	0.911	0.527	3.42e-10
rs35119148	7:136829311	LOC349160	rs833828	12:49352257	ARF3	0.702	0.106	4.11e-11	0.494	0.911	0.527	3.42e-10
rs112227646	7:134823966	LOC653739	rs927829	10:53253804	PRKG1	-1.331	0.202	4.21e-11	-1.727	-0.936	0.392	9.33e-11
rs11977188	7:134825994	LOC653739	rs927829	10:53253804	PRKG1	-1.331	0.202	4.26e-11	-1.727	-0.935	0.392	9.39e-11
rs75952361	7:134831638	LOC653739	rs927829	10:53253804	PRKG1	-1.33	0.202	4.29e-11	-1.725	-0.935	0.39	9.46e-11
rs35655262	7:136826384	LOC349160	rs833828	12:49352257	ARF3	0.701	0.106	4.30e-11	0.493	0.91	0.529	3.45e-10
rs11971038	7:136842197	LOC349160	rs833828	12:49352257	ARF3	0.702	0.106	4.33e-11	0.493	0.911	0.544	3.46e-10
rs10488605	7:136839540	LOC349160	rs833828	12:49352257	ARF3	0.702	0.106	4.33e-11	0.493	0.911	0.544	3.47e-10
rs13226089	7:136843583	LOC349160	rs833828	12:49352257	ARF3	0.702	0.106	4.34e-11	0.493	0.911	0.544	3.47e-10
rs79098801	7:134837554	Intergenic	rs927829	10:53253804	PRKG1	-1.329	0.202	4.36e-11	-1.724	-0.934	0.388	9.50e-11
rs1592804	5:108621680	Intergenic	rs10140868	14:62616479	Intergenic	0.562	0.085	4.38e-11	0.395	0.729	0.394	7.03e-10
rs11973120	7:134837944	Intergenic	rs927829	10:53253804	PRKG1	-1.328	0.202	4.40e-11	-1.723	-0.933	0.387	9.49e-11
rs11241001	5:108608418	Intergenic	rs9671388	14:62626952	Intergenic	0.576	0.087	4.47e-11	0.405	0.747	0.783	1.95e-10
rs113174456	7:134838585	Intergenic	rs927829	10:53253804	PRKG1	-1.326	0.201	4.49e-11	-1.721	-0.932	0.384	9.51e-11
rs9991802	4:158627103	Intergenic	rs1390204	12:71587160	Intergenic	0.591	0.09	4.65e-11	0.415	0.767	0.793	6.24e-11
rs6893599	5:39654978	Intergenic	rs11030584	11:4025021	STIM1	-0.692	0.105	4.79e-11	-0.898	-0.486	0.856	1.07e-09
rs6987672	8:58254694	Intergenic	rs4886105	13:53587879	0.20955	-1.273	0.194	4.79e-11	-1.652	-0.894	0.559	6.49e-11
rs7379603	5:3320814	Intergenic	rs3815802	16:58296545	CCDC113	-0.572	0.087	4.85e-11	-0.743	-0.402	0.995	4.65e-10
rs7445446	5:3320906	Intergenic	rs3815802	16:58296545	CCDC113	-0.572	0.087	4.94e-11	-0.743	-0.401	0.995	4.74e-10
rs10824777	10:54487249	Intergenic	rs492943	10:111629030	XPNPEP1	0.604	0.092	5.06e-11	0.424	0.785	0.354	3.25e-11
rs492943	10:111629030	XPNPEP1	rs10824777	10:54487249	Intergenic	0.604	0.092	5.06e-11	0.424	0.785	0.354	3.25e-11
rs17168947	7:136830478	LOC349160	rs833828	12:49352257	ARF3	0.698	0.106	5.16e-11	0.49	0.906	0.49	4.08e-10
rs11980336	7:136848097	LOC349160	rs833828	12:49352257	ARF3	0.7	0.107	5.26e-11	0.491	0.909	0.558	3.79e-10

Phonological awareness

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs6893599	5:39654978	Intergenic	rs2860454	11:3977591	STIM1	-0.689	0.105	5.27e-11	-0.895	-0.483	0.9	1.56e-09
rs2600168	3:9376490	Intergenic	rs206517	18:10410087	Intergenic	-0.647	0.099	5.31e-11	-0.84	-0.453	0.648	7.84e-10
rs2600168	3:9376490	Intergenic	rs206521	18:10409505	Intergenic	-0.647	0.099	5.32e-11	-0.84	-0.454	0.632	6.84e-10
rs1363215	5:108572971	Intergenic	rs9671388	14:62626952	Intergenic	0.594	0.091	5.40e-11	0.417	0.772	0.713	2.39e-09
rs62338558	5:3321736	Intergenic	rs3815802	16:58296545	CCDC113	-0.571	0.087	5.46e-11	-0.741	-0.4	0.995	5.18e-10
rs6987672	8:58254694	Intergenic	rs7991400	13:53587569	Intergenic	-1.27	0.194	5.48e-11	-1.649	-0.89	0.562	5.95e-11
rs12082409	1:51514557	Intergenic	rs72782094	2:19664012	Intergenic	1.252	0.191	5.62e-11	0.877	1.626	0.452	2.15e-09
rs11978828	7:136811432	LOC349160	rs833828	12:49352257	ARF3	0.691	0.105	5.65e-11	0.484	0.898	0.544	2.79e-10
rs1364716	7:136804753	LOC349160	rs833828	12:49352257	ARF3	0.691	0.105	5.71e-11	0.484	0.898	0.545	2.78e-10
rs35695483	7:136817670	LOC349160	rs3825184	12:49298340	CCDC65	0.689	0.105	5.72e-11	0.483	0.896	0.496	1.75e-10
rs11982570	7:136817016	LOC349160	rs3825184	12:49298340	CCDC65	0.689	0.105	5.74e-11	0.483	0.896	0.496	1.75e-10
rs7732292	5:39641154	Intergenic	rs10742194	11:3953708	STIM1	-0.707	0.108	5.77e-11	-0.918	-0.495	0.849	2.61e-09
rs73450859	7:136846425	LOC349160	rs833828	12:49352257	ARF3	0.697	0.106	5.77e-11	0.488	0.905	0.545	4.06e-10
rs2600168	3:9376490	Intergenic	rs206518	18:10409931	Intergenic	-0.645	0.099	5.79e-11	-0.838	-0.452	0.637	7.27e-10
rs7804134	7:136820679	LOC349160	rs833828	12:49352257	ARF3	0.699	0.107	5.81e-11	0.49	0.908	0.545	3.73e-10
rs33914231	7:136851039	LOC349160	rs833828	12:49352257	ARF3	0.701	0.107	5.84e-11	0.491	0.911	0.533	4.08e-10
rs17168934	7:136808272	LOC349160	rs3825184	12:49298340	CCDC65	0.689	0.105	5.87e-11	0.483	0.895	0.498	1.75e-10
rs36110794	7:136835778	LOC349160	rs3825184	12:49298340	CCDC65	0.691	0.106	5.88e-11	0.484	0.898	0.485	2.53e-10
rs73450817	7:136810307	LOC349160	rs3825184	12:49298340	CCDC65	0.689	0.105	5.88e-11	0.483	0.895	0.498	1.75e-10
rs73450814	7:136810300	LOC349160	rs3825184	12:49298340	CCDC65	0.689	0.105	5.88e-11	0.483	0.895	0.498	1.75e-10
rs10068611	5:39655406	Intergenic	rs11030584	11:4025021	STIM1	-0.688	0.105	5.94e-11	-0.894	-0.482	0.851	1.09e-09
rs67751562	7:136833387	LOC349160	rs3825184	12:49298340	CCDC65	0.691	0.106	5.96e-11	0.484	0.898	0.484	2.56e-10
rs7732292	5:39641154	Intergenic	rs2860454	11:3977591	STIM1	-0.702	0.107	6.00e-11	-0.913	-0.492	0.826	3.99e-09
rs12707345	7:136831401	LOC349160	rs3825184	12:49298340	CCDC65	0.691	0.106	6.04e-11	0.484	0.898	0.482	2.59e-10
rs35119148	7:136829311	LOC349160	rs3825184	12:49298340	CCDC65	0.691	0.106	6.05e-11	0.484	0.898	0.482	2.59e-10

Phonological awareness

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs7732292	5:39641154	Intergenic	rs11030584	11:4025021	STIM1	-0.703	0.107	6.11e-11	-0.914	-0.492	0.814	8.34e-10
rs12472602	2:45528055	Intergenic	rs10809384	9:11200242	Intergenic	-1.368	0.209	6.11e-11	-1.778	-0.958	0.925	5.46e-10
rs16926382	8:61597689	CHD7	rs7194067	16:87988440	BANP	-1.26	0.193	6.14e-11	-1.638	-0.882	0.792	2.61e-10
rs10068611	5:39655406	Intergenic	rs2860454	11:3977591	STIM1	-0.686	0.105	6.21e-11	-0.892	-0.48	0.887	1.56e-09
rs35655262	7:136826384	LOC349160	rs3825184	12:49298340	CCDC65	0.69	0.106	6.30e-11	0.483	0.897	0.484	2.63e-10
rs11971038	7:136842197	LOC349160	rs3825184	12:49298340	CCDC65	0.691	0.106	6.37e-11	0.484	0.898	0.499	2.62e-10
rs10488605	7:136839540	LOC349160	rs3825184	12:49298340	CCDC65	0.691	0.106	6.38e-11	0.484	0.898	0.499	2.63e-10
rs13226089	7:136843583	LOC349160	rs3825184	12:49298340	CCDC65	0.691	0.106	6.38e-11	0.484	0.898	0.499	2.63e-10
rs62338559	5:3322058	Intergenic	rs10852563	16:58332076	Intergenic	-0.565	0.086	6.42e-11	-0.735	-0.396	0.906	5.22e-10
rs6987672	8:58254694	Intergenic	rs1891943	13:53587172	Intergenic	-1.265	0.194	6.44e-11	-1.645	-0.886	0.567	5.47e-11
rs62338559	5:3322058	Intergenic	rs4784030	16:58303599	CCDC113	-0.558	0.086	6.72e-11	-0.726	-0.391	0.937	5.21e-10
rs6893599	5:39654978	Intergenic	rs4379839	11:4038256	STIM1	-0.685	0.105	6.75e-11	-0.89	-0.479	0.861	1.54e-09
rs6893599	5:39654978	Intergenic	rs2959080	11:4028447	STIM1	-0.684	0.105	6.96e-11	-0.889	-0.478	0.851	1.24e-09
rs6987672	8:58254694	Intergenic	rs1936387	13:53586746	Intergenic	-1.262	0.194	7.16e-11	-1.641	-0.882	0.57	5.25e-11
rs73836233	4:92499795	CCSER1	rs9542856	13:72699635	Intergenic	-0.901	0.138	7.17e-11	-1.172	-0.63	0.874	4.88e-11
rs6848895	4:158626629	Intergenic	rs1705244	12:71589246	Intergenic	0.583	0.089	7.18e-11	0.407	0.758	0.835	3.66e-11
rs7732292	5:39641154	Intergenic	rs4379839	11:4038256	STIM1	-0.698	0.107	7.28e-11	-0.908	-0.488	0.799	1.28e-09
rs35695483	7:136817670	LOC349160	rs7304942	12:49306149	CCDC65	0.686	0.105	7.29e-11	0.48	0.893	0.486	4.25e-10
rs11982570	7:136817016	LOC349160	rs7304942	12:49306149	CCDC65	0.686	0.105	7.31e-11	0.48	0.893	0.487	4.25e-10
rs35502738	5:39629430	Intergenic	rs2860454	11:3977591	STIM1	-0.73	0.112	7.31e-11	-0.95	-0.511	0.998	3.12e-09
rs17168934	7:136808272	LOC349160	rs7304942	12:49306149	CCDC65	0.686	0.105	7.49e-11	0.479	0.892	0.488	4.25e-10
rs36110794	7:136835778	LOC349160	rs7304942	12:49306149	CCDC65	0.688	0.106	7.50e-11	0.481	0.895	0.476	6.02e-10
rs73450817	7:136810307	LOC349160	rs7304942	12:49306149	CCDC65	0.686	0.105	7.50e-11	0.479	0.892	0.488	4.27e-10
rs73450814	7:136810300	LOC349160	rs7304942	12:49306149	CCDC65	0.686	0.105	7.50e-11	0.479	0.892	0.488	4.27e-10
rs67751562	7:136833387	LOC349160	rs7304942	12:49306149	CCDC65	0.688	0.106	7.60e-11	0.481	0.895	0.474	6.10e-10

Phonological awareness

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs6893599	5:39654978	Intergenic	rs727152	11:4019765	STIM1	-0.682	0.105	7.68e-11	-0.887	-0.476	0.849	1.28e-09
rs6893599	5:39654978	Intergenic	rs10835500	11:4039401	STIM1	-0.684	0.105	7.69e-11	-0.89	-0.478	0.826	1.90e-09
rs11980336	7:136848097	LOC349160	rs3825184	12:49298340	CCDC65	0.689	0.106	7.70e-11	0.481	0.897	0.514	2.87e-10
rs12707345	7:136831401	LOC349160	rs7304942	12:49306149	CCDC65	0.688	0.106	7.71e-11	0.481	0.895	0.473	6.18e-10
rs35119148	7:136829311	LOC349160	rs7304942	12:49306149	CCDC65	0.688	0.106	7.72e-11	0.481	0.895	0.473	6.18e-10
rs11978828	7:136811432	LOC349160	rs3825184	12:49298340	CCDC65	0.681	0.105	7.77e-11	0.476	0.886	0.494	2.06e-10
rs1364716	7:136804753	LOC349160	rs3825184	12:49298340	CCDC65	0.681	0.105	7.85e-11	0.475	0.886	0.495	2.06e-10
rs17168947	7:136830478	LOC349160	rs3825184	12:49298340	CCDC65	0.686	0.105	7.86e-11	0.479	0.893	0.442	3.21e-10
rs35502738	5:39629430	Intergenic	rs10742194	11:3953708	STIM1	-0.733	0.113	7.88e-11	-0.954	-0.512	0.997	2.18e-09
rs35655262	7:136826384	LOC349160	rs7304942	12:49306149	CCDC65	0.687	0.106	8.04e-11	0.48	0.894	0.474	6.25e-10
rs73450859	7:136846425	LOC349160	rs3825184	12:49298340	CCDC65	0.686	0.106	8.07e-11	0.479	0.893	0.5	3.06e-10
rs1197282	3:133052473	TMEM108	rs4607084	3:41021872	Intergenic	0.855	0.132	8.09e-11	0.597	1.112	0.394	1.90e-09
rs4607084	3:41021872	Intergenic	rs1197282	3:133052473	TMEM108	0.855	0.132	8.09e-11	0.597	1.112	0.394	1.90e-09
rs10068611	5:39655406	Intergenic	rs4379839	11:4038256	STIM1	-0.681	0.105	8.23e-11	-0.887	-0.476	0.853	1.57e-09
rs7379603	5:3320814	Intergenic	rs10852563	16:58332076	Intergenic	-0.558	0.086	8.25e-11	-0.727	-0.39	0.933	4.97e-10
rs7732292	5:39641154	Intergenic	rs2959080	11:4028447	STIM1	-0.696	0.107	8.25e-11	-0.905	-0.486	0.788	9.30e-10
rs34307840	2:49085269	Intergenic	rs7960668	12:27000650	Intergenic	-0.762	0.117	8.29e-11	-0.992	-0.532	0.53	1.64e-10
rs4257682	4:64885239	Intergenic	rs11075392	16:18007223	Intergenic	0.619	0.095	8.31e-11	0.432	0.806	0.853	9.30e-10
rs1354128	5:39634963	Intergenic	rs10742194	11:3953708	STIM1	-0.704	0.108	8.39e-11	-0.916	-0.492	0.862	2.08e-09
rs7445446	5:3320906	Intergenic	rs10852563	16:58332076	Intergenic	-0.558	0.086	8.42e-11	-0.726	-0.39	0.933	5.07e-10
rs16926382	8:61597689	CHD7	rs4843289	16:87999896	BANP	-1.244	0.192	8.43e-11	-1.619	-0.868	0.771	3.96e-10
rs7804134	7:136820679	LOC349160	rs3825184	12:49298340	CCDC65	0.688	0.106	8.46e-11	0.48	0.895	0.5	2.83e-10
rs62338559	5:3322058	Intergenic	rs2550344	16:58299711	CCDC113	-0.566	0.087	8.48e-11	-0.737	-0.395	0.978	4.75e-10
rs33914231	7:136851039	LOC349160	rs3825184	12:49298340	CCDC65	0.69	0.106	8.49e-11	0.482	0.899	0.489	3.05e-10
rs10068611	5:39655406	Intergenic	rs2959080	11:4028447	STIM1	-0.68	0.105	8.55e-11	-0.885	-0.474	0.845	1.25e-09

Phonological awareness

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs142447549	4:64816631	Intergenic	rs11075392	16:18007223	Intergenic	0.618	0.095	8.56e-11	0.432	0.805	0.592	1.39e-10
rs7732292	5:39641154	Intergenic	rs10835500	11:4039401	STIM1	-0.696	0.107	8.57e-11	-0.906	-0.486	0.74	1.57e-09
rs1354128	5:39634963	Intergenic	rs11030584	11:4025021	STIM1	-0.701	0.108	8.70e-11	-0.912	-0.489	0.835	6.46e-10
rs16926382	8:61597689	CHD7	rs76736497	16:87998561	BANP	-1.242	0.191	8.79e-11	-1.617	-0.867	0.771	3.88e-10
rs7732292	5:39641154	Intergenic	rs727152	11:4019765	STIM1	-0.694	0.107	9.05e-11	-0.903	-0.484	0.783	9.53e-10
rs1197282	3:133052473	TMEM108	rs13064610	3:41021704	Intergenic	0.852	0.132	9.28e-11	0.594	1.11	0.39	2.08e-09
rs13064610	3:41021704	Intergenic	rs1197282	3:133052473	TMEM108	0.852	0.132	9.28e-11	0.594	1.11	0.39	2.08e-09
rs10068611	5:39655406	Intergenic	rs727152	11:4019765	STIM1	-0.678	0.105	9.38e-11	-0.883	-0.473	0.841	1.29e-09
rs12643891	4:64810034	Intergenic	rs11075392	16:18007223	Intergenic	0.609	0.094	9.51e-11	0.425	0.793	0.591	1.77e-10
rs13115659	4:64810536	Intergenic	rs11075392	16:18007223	Intergenic	0.609	0.094	9.52e-11	0.424	0.793	0.591	1.78e-10
rs12511573	4:64810380	Intergenic	rs11075392	16:18007223	Intergenic	0.609	0.094	9.53e-11	0.424	0.793	0.591	1.78e-10
rs76914853	7:134818706	AGBL3	rs927829	10:53253804	PRKG1	-1.32	0.204	9.54e-11	-1.72	-0.921	0.428	3.02e-10
rs9991802	4:158627103	Intergenic	rs1705244	12:71589246	Intergenic	0.58	0.09	9.58e-11	0.405	0.756	0.836	7.42e-11
rs10068611	5:39655406	Intergenic	rs10835500	11:4039401	STIM1	-0.68	0.105	9.63e-11	-0.886	-0.474	0.814	1.94e-09
rs79112226	8:58254272	Intergenic	rs1936387	13:53586746	Intergenic	-1.262	0.195	9.71e-11	-1.644	-0.88	0.572	6.56e-11
rs6856100	4:64884721	Intergenic	rs11075392	16:18007223	Intergenic	0.612	0.095	9.75e-11	0.427	0.798	0.66	5.54e-10

Spelling

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs59942153	17:8207240	Intergenic	rs7271609	20:52137810	Intergenic	-1.986	0.281	1.50e-12	-2.537	-1.436	0.926	1.16e-11
rs59942153	17:8207240	Intergenic	rs7271612	20:52137818	Intergenic	-1.986	0.281	1.50e-12	-2.537	-1.436	0.926	1.16e-11
rs59942153	17:8207240	Intergenic	rs6512891	20:52138287	Intergenic	-1.99	0.281	1.53e-12	-2.541	-1.438	0.927	1.19e-11
rs59942153	17:8207240	Intergenic	rs6022549	20:52138183	Intergenic	-1.99	0.281	1.53e-12	-2.541	-1.438	0.927	1.19e-11
rs59942153	17:8207240	Intergenic	rs6512892	20:52138357	Intergenic	-1.992	0.282	1.63e-12	-2.544	-1.439	0.922	1.24e-11
rs59942153	17:8207240	Intergenic	rs6022550	20:52138379	Intergenic	-1.991	0.282	1.63e-12	-2.544	-1.439	0.922	1.25e-11
rs59942153	17:8207240	Intergenic	rs6022540	20:52134700	Intergenic	-1.983	0.281	1.84e-12	-2.534	-1.431	0.93	1.42e-11
rs59942153	17:8207240	Intergenic	rs6022539	20:52134609	Intergenic	-1.983	0.281	1.85e-12	-2.534	-1.431	0.93	1.44e-11
rs59942153	17:8207240	Intergenic	rs6022546	20:52135393	Intergenic	-1.983	0.281	1.85e-12	-2.534	-1.431	0.93	1.45e-11
rs59942153	17:8207240	Intergenic	rs56384924	20:52134563	Intergenic	-1.983	0.281	1.86e-12	-2.534	-1.431	0.93	1.44e-11
rs59942153	17:8207240	Intergenic	rs6022543	20:52134914	Intergenic	-1.982	0.281	1.86e-12	-2.534	-1.431	0.93	1.44e-11
rs59942153	17:8207240	Intergenic	rs113156479	20:52134996	Intergenic	-1.982	0.281	1.86e-12	-2.534	-1.431	0.929	1.45e-11
rs59942153	17:8207240	Intergenic	rs6022545	20:52135365	Intergenic	-1.985	0.282	1.91e-12	-2.537	-1.432	0.931	1.09e-11
rs59942153	17:8207240	Intergenic	rs8118807	20:52133473	Intergenic	-1.985	0.282	2.08e-12	-2.538	-1.431	0.932	1.61e-11
rs59942153	17:8207240	Intergenic	rs8118721	20:52133451	Intergenic	-1.985	0.282	2.08e-12	-2.538	-1.431	0.932	1.61e-11
rs59942153	17:8207240	Intergenic	rs56913504	20:52133353	Intergenic	-1.977	0.283	2.59e-12	-2.531	-1.423	0.932	2.91e-11
rs59942153	17:8207240	Intergenic	rs6022536	20:52131921	Intergenic	-1.976	0.283	2.69e-12	-2.53	-1.422	0.931	2.99e-11
rs59942153	17:8207240	Intergenic	rs7262056	20:52131227	Intergenic	-1.977	0.283	2.71e-12	-2.531	-1.423	0.931	3.03e-11
rs9293566	5:90622516	Intergenic	rs1987901	8:15322608	Intergenic	0.519	0.074	2.88e-12	0.373	0.664	0.756	2.13e-11
rs6878402	5:90631972	Intergenic	rs1987901	8:15322608	Intergenic	0.518	0.074	3.17e-12	0.373	0.664	0.895	9.80e-12
rs62523288	8:103585618	Intergenic	rs17114091	10:90666952	STAMBPL1	-0.796	0.115	3.77e-12	-1.02	-0.571	0.778	8.63e-11
rs10942630	5:90631384	Intergenic	rs1987901	8:15322608	Intergenic	0.513	0.074	3.91e-12	0.368	0.658	0.872	1.58e-11
rs62373944	5:90632625	Intergenic	rs1987901	8:15322608	Intergenic	0.512	0.074	4.89e-12	0.367	0.657	0.89	1.84e-11
rs62523288	8:103585618	Intergenic	rs10509557	10:90663839	STAMBPL1	-0.805	0.117	5.13e-12	-1.033	-0.576	0.795	1.44e-10
rs10037293	5:90628300	Intergenic	rs1987901	8:15322608	Intergenic	0.508	0.074	5.14e-12	0.364	0.652	0.792	5.06e-11
rs58734007	5:90633072	Intergenic	rs1987901	8:15322608	Intergenic	0.511	0.074	5.30e-12	0.366	0.657	0.897	1.94e-11
rs10042981	5:90628789	Intergenic	rs1987901	8:15322608	Intergenic	0.508	0.074	5.42e-12	0.363	0.652	0.801	5.45e-11
rs7702782	5:90638506	Intergenic	rs1987901	8:15322608	Intergenic	0.513	0.075	8.15e-12	0.366	0.661	0.96	7.38e-11

Spelling

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs9293566	5:90622516	Intergenic	rs1390051	8:15322770	Intergenic	0.505	0.074	8.85e-12	0.36	0.65	0.66	1.80e-11
rs6878402	5:90631972	Intergenic	rs1390051	8:15322770	Intergenic	0.505	0.074	9.80e-12	0.36	0.65	0.874	7.97e-12
rs10942630	5:90631384	Intergenic	rs1390051	8:15322770	Intergenic	0.5	0.074	1.21e-11	0.355	0.644	0.816	1.29e-11
rs6998091	8:103588880	Intergenic	rs17114091	10:90666952	STAMBPL1	-0.775	0.114	1.23e-11	-0.999	-0.551	0.615	1.26e-10
rs62523288	8:103585618	Intergenic	rs6586154	10:90661135	STAMBPL1	-0.791	0.117	1.30e-11	-1.02	-0.562	0.727	4.07e-10
rs62373944	5:90632625	Intergenic	rs1390051	8:15322770	Intergenic	0.498	0.074	1.51e-11	0.354	0.643	0.837	1.50e-11
rs10037293	5:90628300	Intergenic	rs1390051	8:15322770	Intergenic	0.494	0.073	1.55e-11	0.351	0.638	0.713	4.16e-11
rs62523288	8:103585618	Intergenic	rs1998629	10:90664283	STAMBPL1	-0.785	0.117	1.59e-11	-1.014	-0.557	0.751	5.05e-10
rs62523288	8:103585618	Intergenic	rs1968028	10:90662086	STAMBPL1	-0.788	0.117	1.62e-11	-1.017	-0.558	0.732	4.98e-10
rs58734007	5:90633072	Intergenic	rs1390051	8:15322770	Intergenic	0.498	0.074	1.63e-11	0.353	0.643	0.845	1.58e-11
rs10042981	5:90628789	Intergenic	rs1390051	8:15322770	Intergenic	0.494	0.073	1.63e-11	0.35	0.638	0.724	4.49e-11
rs6998091	8:103588880	Intergenic	rs10509557	10:90663839	STAMBPL1	-0.782	0.116	1.80e-11	-1.01	-0.554	0.631	2.10e-10
rs7702782	5:90638506	Intergenic	rs1390051	8:15322770	Intergenic	0.5	0.075	2.46e-11	0.353	0.646	0.921	5.97e-11
rs10516497	4:103942714	SLC9B1	rs3741869	12:31148187	TSPAN11	0.902	0.136	3.34e-11	0.635	1.168	0.99	6.06e-11
rs7829010	8:103589295	Intergenic	rs17114091	10:90666952	STAMBPL1	-0.757	0.114	3.77e-11	-0.981	-0.533	0.553	3.58e-10
rs62523288	8:103585618	Intergenic	rs10788618	10:90661893	STAMBPL1	-0.772	0.117	4.26e-11	-1.001	-0.542	0.72	1.19e-09
rs59942153	17:8207240	Intergenic	rs143935436	20:52138657	Intergenic	-1.928	0.293	5.02e-11	-2.503	-1.353	0.978	2.23e-10
rs59942153	17:8207240	Intergenic	rs58552634	20:52139323	Intergenic	-1.91	0.291	5.08e-11	-2.479	-1.34	0.979	2.62e-10
rs59942153	17:8207240	Intergenic	rs7261040	20:52139673	Intergenic	-1.911	0.291	5.17e-11	-2.482	-1.341	0.979	2.66e-10
rs59942153	17:8207240	Intergenic	rs66530264	20:52138847	Intergenic	-1.907	0.29	5.17e-11	-2.476	-1.338	0.979	2.68e-10
rs6998091	8:103588880	Intergenic	rs1998629	10:90664283	STAMBPL1	-0.762	0.116	5.68e-11	-0.99	-0.534	0.593	7.60e-10
rs7829010	8:103589295	Intergenic	rs10509557	10:90663839	STAMBPL1	-0.762	0.116	5.82e-11	-0.99	-0.534	0.565	6.11e-10
rs9817226	3:132885668	TMEM108	rs11055024	12:12858969	Intergenic	-0.599	0.092	6.15e-11	-0.778	-0.419	0.426	1.18e-09
rs6998091	8:103588880	Intergenic	rs6586154	10:90661135	STAMBPL1	-0.762	0.117	6.23e-11	-0.991	-0.534	0.537	8.57e-10
rs59942153	17:8207240	Intergenic	rs149538794	20:52138649	Intergenic	-1.906	0.292	6.86e-11	-2.478	-1.333	0.965	3.46e-10
rs11777481	8:67219528	Intergenic	rs75226807	19:19026851	COPE	0.8	0.123	6.94e-11	0.56	1.04	0.818	1.74e-09
rs11987288	8:67218784	Intergenic	rs75226807	19:19026851	COPE	0.8	0.123	6.94e-11	0.56	1.04	0.818	1.74e-09
rs4737782	8:67218257	Intergenic	rs75226807	19:19026851	COPE	0.8	0.123	6.94e-11	0.56	1.04	0.818	1.74e-09

Spelling

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs2357826	8:67214877	Intergenic	rs75226807	19:19026851	COPE	0.8	0.123	7.28e-11	0.559	1.04	0.819	1.84e-09
rs7827371	8:67213402	Intergenic	rs75226807	19:19026851	COPE	0.8	0.123	7.29e-11	0.559	1.04	0.819	1.84e-09
rs7843247	8:67213340	Intergenic	rs75226807	19:19026851	COPE	0.8	0.123	7.29e-11	0.559	1.04	0.819	1.84e-09
rs6998091	8:103588880	Intergenic	rs1968028	10:90662086	STAMBPL1	-0.76	0.117	7.38e-11	-0.988	-0.531	0.557	9.90e-10
rs72672330	1:62101780	Intergenic	rs1101362	6:72142135	Intergenic	-0.738	0.113	7.43e-11	-0.96	-0.516	0.602	6.55e-10
rs13385547	2:99548333	KIAA1211L	rs2669429	8:105463690	DPYS	-0.564	0.087	7.58e-11	-0.734	-0.394	0.874	2.03e-10
rs12144911	1:176566124	PAPPA2	rs2585514	13:78699007	RNF219-AS1	0.637	0.098	7.59e-11	0.445	0.828	0.569	6.41e-11
rs1450055	3:132891280	TMEM108	rs11055024	12:12858969	Intergenic	-0.591	0.091	7.62e-11	-0.769	-0.413	0.445	1.06e-09
rs2342702	2:45411188	UNQ6975	rs11560388	7:130026288	CPA1	0.671	0.103	7.71e-11	0.469	0.874	0.53	5.70e-10
rs72672322	1:62093026	Intergenic	rs1101362	6:72142135	Intergenic	-0.74	0.114	7.81e-11	-0.963	-0.517	0.668	4.94e-10
rs72672320	1:62092944	Intergenic	rs1101362	6:72142135	Intergenic	-0.74	0.114	7.83e-11	-0.963	-0.517	0.667	4.95e-10
rs72672318	1:62092821	Intergenic	rs1101362	6:72142135	Intergenic	-0.74	0.114	7.86e-11	-0.963	-0.517	0.668	4.95e-10
rs111765756	1:62091597	Intergenic	rs1101362	6:72142135	Intergenic	-0.74	0.114	7.86e-11	-0.963	-0.517	0.669	5.18e-10
rs78488132	1:62091529	Intergenic	rs1101362	6:72142135	Intergenic	-0.74	0.114	7.86e-11	-0.963	-0.517	0.669	5.18e-10
rs112109175	1:62090170	Intergenic	rs1101362	6:72142135	Intergenic	-0.74	0.114	7.92e-11	-0.963	-0.517	0.669	5.16e-10
rs4433428	1:62090912	Intergenic	rs1101362	6:72142135	Intergenic	-0.74	0.114	7.92e-11	-0.963	-0.517	0.67	5.19e-10
rs1495951	1:62090587	Intergenic	rs1101362	6:72142135	Intergenic	-0.74	0.114	7.95e-11	-0.963	-0.517	0.67	5.19e-10
rs1495950	1:62090706	Intergenic	rs1101362	6:72142135	Intergenic	-0.74	0.114	7.96e-11	-0.963	-0.517	0.67	5.20e-10
rs2342702	2:45411188	UNQ6975	rs34339293	7:130017077	Intergenic	0.7	0.108	7.98e-11	0.489	0.911	0.439	5.07e-10
rs72672314	1:62090174	Intergenic	rs1101362	6:72142135	Intergenic	-0.74	0.114	8.01e-11	-0.963	-0.517	0.67	5.22e-10
rs60883941	1:62092020	Intergenic	rs1101362	6:72142135	Intergenic	-0.74	0.114	8.01e-11	-0.963	-0.517	0.668	5.02e-10
rs112477095	1:62091863	Intergenic	rs1101362	6:72142135	Intergenic	-0.74	0.114	8.02e-11	-0.963	-0.517	0.668	5.02e-10
rs2342702	2:45411188	UNQ6975	rs3823581	7:130017787	Intergenic	0.699	0.108	8.08e-11	0.488	0.909	0.446	5.06e-10
rs2342702	2:45411188	UNQ6975	rs10954269	7:130018863	CPA1	0.696	0.107	8.21e-11	0.486	0.906	0.449	5.02e-10
rs16878585	7:15722249	MEOX2	rs6967370	7:38600176	AMPH	-1.327	0.204	8.27e-11	-1.728	-0.927	0.97	7.44e-10
rs6967370	7:38600176	AMPH	rs16878585	7:15722249	MEOX2	-1.327	0.204	8.27e-11	-1.728	-0.927	0.97	7.44e-10
rs72672330	1:62101780	Intergenic	rs852949	6:72140542	Intergenic	-0.736	0.113	8.89e-11	-0.958	-0.513	0.588	8.94e-10
rs72672330	1:62101780	Intergenic	rs852948	6:72140419	Intergenic	-0.736	0.113	8.92e-11	-0.958	-0.513	0.588	8.97e-10

Spelling

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs10798465	1:176596408	PAPPA2	rs2585514	13:78699007	RNF219-AS1	0.635	0.098	9.33e-11	0.443	0.827	0.567	9.96e-11
rs72672322	1:62093026	Intergenic	rs852949	6:72140542	Intergenic	-0.738	0.114	9.38e-11	-0.961	-0.515	0.659	6.72e-10
rs72672320	1:62092944	Intergenic	rs852949	6:72140542	Intergenic	-0.738	0.114	9.40e-11	-0.961	-0.515	0.659	6.73e-10
rs72672322	1:62093026	Intergenic	rs852948	6:72140419	Intergenic	-0.738	0.114	9.40e-11	-0.961	-0.515	0.659	6.73e-10
rs72672320	1:62092944	Intergenic	rs852948	6:72140419	Intergenic	-0.738	0.114	9.42e-11	-0.961	-0.515	0.659	6.75e-10
rs72672318	1:62092821	Intergenic	rs852949	6:72140542	Intergenic	-0.738	0.114	9.43e-11	-0.961	-0.515	0.659	6.73e-10
rs111765756	1:62091597	Intergenic	rs852949	6:72140542	Intergenic	-0.738	0.114	9.44e-11	-0.961	-0.514	0.661	7.04e-10
rs78488132	1:62091529	Intergenic	rs852949	6:72140542	Intergenic	-0.738	0.114	9.44e-11	-0.961	-0.514	0.661	7.04e-10
rs72672318	1:62092821	Intergenic	rs852948	6:72140419	Intergenic	-0.738	0.114	9.46e-11	-0.961	-0.514	0.659	6.75e-10
rs111765756	1:62091597	Intergenic	rs852948	6:72140419	Intergenic	-0.738	0.114	9.46e-11	-0.961	-0.514	0.661	7.06e-10
rs78488132	1:62091529	Intergenic	rs852948	6:72140419	Intergenic	-0.738	0.114	9.46e-11	-0.961	-0.514	0.661	7.06e-10
rs112109175	1:62090170	Intergenic	rs852949	6:72140542	Intergenic	-0.737	0.114	9.51e-11	-0.961	-0.514	0.661	7.01e-10
rs4433428	1:62090912	Intergenic	rs852949	6:72140542	Intergenic	-0.737	0.114	9.51e-11	-0.961	-0.514	0.662	7.06e-10
rs112109175	1:62090170	Intergenic	rs852948	6:72140419	Intergenic	-0.737	0.114	9.53e-11	-0.961	-0.514	0.661	7.03e-10
rs4433428	1:62090912	Intergenic	rs852948	6:72140419	Intergenic	-0.737	0.114	9.54e-11	-0.961	-0.514	0.662	7.08e-10
rs1495951	1:62090587	Intergenic	rs852949	6:72140542	Intergenic	-0.737	0.114	9.54e-11	-0.961	-0.514	0.662	7.06e-10
rs1495950	1:62090706	Intergenic	rs852949	6:72140542	Intergenic	-0.737	0.114	9.55e-11	-0.961	-0.514	0.662	7.07e-10
rs1495951	1:62090587	Intergenic	rs852948	6:72140419	Intergenic	-0.737	0.114	9.57e-11	-0.961	-0.514	0.662	7.08e-10
rs1495950	1:62090706	Intergenic	rs852948	6:72140419	Intergenic	-0.737	0.114	9.58e-11	-0.961	-0.514	0.662	7.09e-10
rs3795320	1:176591291	PAPPA2	rs2585514	13:78699007	RNF219-AS1	0.636	0.098	9.58e-11	0.443	0.828	0.572	1.20e-10
rs72672314	1:62090174	Intergenic	rs852949	6:72140542	Intergenic	-0.737	0.114	9.62e-11	-0.961	-0.514	0.662	7.09e-10
rs60883941	1:62092020	Intergenic	rs852949	6:72140542	Intergenic	-0.737	0.114	9.62e-11	-0.961	-0.514	0.659	6.82e-10
rs112477095	1:62091863	Intergenic	rs852949	6:72140542	Intergenic	-0.737	0.114	9.63e-11	-0.961	-0.514	0.66	6.82e-10
rs72672314	1:62090174	Intergenic	rs852948	6:72140419	Intergenic	-0.737	0.114	9.65e-11	-0.96	-0.514	0.662	7.11e-10
rs60883941	1:62092020	Intergenic	rs852948	6:72140419	Intergenic	-0.737	0.114	9.65e-11	-0.961	-0.514	0.659	6.84e-10
rs112477095	1:62091863	Intergenic	rs852948	6:72140419	Intergenic	-0.737	0.114	9.65e-11	-0.961	-0.514	0.659	6.84e-10

Non-word reading

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs34981217	2:192989322	TMEFF2	rs371164	19:6489814	Intergenic	-0.571	0.081	1.64e-12	-0.729	-0.412	0.817	9.27e-13
rs17354327	2:192987014	TMEFF2	rs371164	19:6489814	Intergenic	-0.572	0.081	1.67e-12	-0.731	-0.413	0.836	1.21e-12
rs10164776	2:192986033	TMEFF2	rs371164	19:6489814	Intergenic	-0.572	0.081	1.76e-12	-0.731	-0.413	0.837	1.22e-12
rs17367154	2:192989689	TMEFF2	rs371164	19:6489814	Intergenic	-0.562	0.08	2.97e-12	-0.719	-0.404	0.861	1.61e-12
rs1806647	2:192989191	TMEFF2	rs371164	19:6489814	Intergenic	-0.564	0.082	5.32e-12	-0.724	-0.404	0.821	1.08e-11
rs17426817	3:111664064	PHLDB2	rs7674025	4:136013609	Intergenic	1.038	0.151	5.78e-12	0.743	1.335	0.691	1.17e-10
rs11714432	3:126929153	Intergenic	rs13062221	3:89778009	Intergenic	-0.63	0.092	5.98e-12	-0.809	-0.45	0.423	9.15e-09
rs13062221	3:89778009	Intergenic	rs11714432	3:126929153	Intergenic	-0.63	0.092	5.98e-12	-0.809	-0.45	0.423	9.15e-09
rs17426817	3:111664064	PHLDB2	rs7678960	4:136013912	Intergenic	1.039	0.151	6.30e-12	0.742	1.335	0.685	1.24e-10
rs582780	3:172121443	Intergenic	rs4729987	7:103907516	Intergenic	-0.486	0.071	9.86e-12	-0.626	-0.346	0.485	2.35e-10
rs13143126	4:120637804	Intergenic	rs74879864	22:29705881	GAS2L1	-1.074	0.158	1.11e-11	-1.384	-0.764	0.55	5.13e-11
rs13143126	4:120637804	Intergenic	rs56037813	22:29704662	GAS2L1	-1.074	0.158	1.11e-11	-1.384	-0.764	0.55	5.14e-11
rs13138402	4:120622789	Intergenic	rs74879864	22:29705881	GAS2L1	-1.05	0.155	1.38e-11	-1.355	-0.746	0.468	4.15e-11
rs13138402	4:120622789	Intergenic	rs56037813	22:29704662	GAS2L1	-1.05	0.155	1.39e-11	-1.355	-0.746	0.468	4.15e-11
rs13144584	4:120625313	Intergenic	rs74879864	22:29705881	GAS2L1	-1.05	0.155	1.39e-11	-1.355	-0.746	0.467	4.17e-11
rs13144584	4:120625313	Intergenic	rs56037813	22:29704662	GAS2L1	-1.05	0.155	1.40e-11	-1.355	-0.746	0.467	4.18e-11
rs2583610	4:120624185	Intergenic	rs74879864	22:29705881	GAS2L1	-1.05	0.155	1.41e-11	-1.354	-0.745	0.464	6.18e-11
rs2583610	4:120624185	Intergenic	rs56037813	22:29704662	GAS2L1	-1.05	0.155	1.41e-11	-1.354	-0.745	0.464	6.19e-11
rs2714959	4:120631509	Intergenic	rs74879864	22:29705881	GAS2L1	-1.047	0.155	1.58e-11	-1.352	-0.743	0.454	4.93e-11
rs2714959	4:120631509	Intergenic	rs56037813	22:29704662	GAS2L1	-1.047	0.155	1.59e-11	-1.351	-0.742	0.453	4.94e-11
rs1009077	4:120627421	Intergenic	rs74879864	22:29705881	GAS2L1	-1.047	0.155	1.59e-11	-1.352	-0.743	0.453	5.06e-11
rs1009077	4:120627421	Intergenic	rs56037813	22:29704662	GAS2L1	-1.047	0.155	1.59e-11	-1.352	-0.743	0.453	5.07e-11
rs582780	3:172121443	Intergenic	rs7357226	7:103907234	Intergenic	-0.476	0.071	1.88e-11	-0.615	-0.337	0.441	3.62e-10
rs2389901	4:120635908	Intergenic	rs74879864	22:29705881	GAS2L1	-1.047	0.156	1.90e-11	-1.352	-0.741	0.471	4.54e-11
rs2389901	4:120635908	Intergenic	rs56037813	22:29704662	GAS2L1	-1.046	0.156	1.91e-11	-1.352	-0.741	0.47	4.54e-11
rs13143126	4:120637804	Intergenic	rs72547435	22:29683558	EWSR1	-1.08	0.161	1.94e-11	-1.396	-0.765	0.559	9.77e-11

Non-word reading

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs582780	3:172121443	Intergenic	rs4729988	7:103908170	Intergenic	-0.475	0.071	2.06e-11	-0.614	-0.336	0.448	3.86e-10
rs17426817	3:111664064	PHLDB2	rs6835201	4:135991382	Intergenic	1.01	0.151	2.10e-11	0.714	1.305	0.76	1.90e-10
rs582780	3:172121443	Intergenic	rs10265838	7:103908726	Intergenic	-0.475	0.071	2.14e-11	-0.614	-0.336	0.452	3.93e-10
rs2892875	4:120616193	Intergenic	rs74879864	22:29705881	GAS2L1	-1.046	0.156	2.19e-11	-1.352	-0.74	0.522	5.21e-11
rs2892875	4:120616193	Intergenic	rs56037813	22:29704662	GAS2L1	-1.046	0.156	2.19e-11	-1.352	-0.739	0.522	5.22e-11
rs13138402	4:120622789	Intergenic	rs72547435	22:29683558	EWSR1	-1.055	0.158	2.51e-11	-1.364	-0.745	0.478	8.07e-11
rs13144584	4:120625313	Intergenic	rs72547435	22:29683558	EWSR1	-1.054	0.158	2.52e-11	-1.364	-0.745	0.477	8.11e-11
rs2583610	4:120624185	Intergenic	rs72547435	22:29683558	EWSR1	-1.054	0.158	2.55e-11	-1.364	-0.744	0.474	1.18e-10
rs2356772	2:192987584	TMEFF2	rs371164	19:6489814	Intergenic	-0.543	0.082	2.71e-11	-0.703	-0.383	0.731	2.57e-11
rs2356772	2:192987584	TMEFF2	rs371164	19:6489814	Intergenic	-0.543	0.082	2.71e-11	-0.703	-0.383	0.731	2.57e-11
rs2714959	4:120631509	Intergenic	rs72547435	22:29683558	EWSR1	-1.051	0.158	2.85e-11	-1.361	-0.742	0.464	9.50e-11
rs1009077	4:120627421	Intergenic	rs72547435	22:29683558	EWSR1	-1.051	0.158	2.88e-11	-1.361	-0.742	0.464	9.77e-11
rs2389901	4:120635908	Intergenic	rs72547435	22:29683558	EWSR1	-1.051	0.159	3.39e-11	-1.362	-0.74	0.48	8.73e-11
rs855969	10:119437857	Intergenic	rs6054462	20:6671666	Intergenic	-0.57	0.086	3.54e-11	-0.739	-0.401	0.567	3.39e-10
rs855969	10:119437857	Intergenic	rs6054463	20:6671823	Intergenic	-0.57	0.086	3.54e-11	-0.738	-0.401	0.57	3.39e-10
rs855969	10:119437857	Intergenic	rs6038585	20:6671980	Intergenic	-0.569	0.086	3.67e-11	-0.737	-0.4	0.58	3.36e-10
rs10497724	2:192992434	TMEFF2	rs371164	19:6489814	Intergenic	-0.553	0.084	3.70e-11	-0.717	-0.389	0.885	1.49e-11
rs10497724	2:192992434	TMEFF2	rs371164	19:6489814	Intergenic	-0.553	0.084	3.70e-11	-0.717	-0.389	0.885	1.49e-11
rs60347052	2:20353631	Intergenic	rs10906819	10:15109377	OLAH	1.248	0.189	3.83e-11	0.878	1.619	0.822	2.06e-10
rs1862182	5:95999526	CAST	rs9988716	10:82715404	Intergenic	0.62	0.094	3.85e-11	0.436	0.804	0.949	5.49e-11
rs60347052	2:20353631	Intergenic	rs11259456	10:15106649	OLAH	1.249	0.189	3.88e-11	0.878	1.619	0.824	2.00e-10
rs2892875	4:120616193	Intergenic	rs72547435	22:29683558	EWSR1	-1.05	0.159	3.93e-11	-1.361	-0.738	0.531	1.02e-10
rs60347052	2:20353631	Intergenic	rs10796255	10:15106379	OLAH	1.249	0.189	3.95e-11	0.878	1.619	0.825	2.01e-10
rs60347052	2:20353631	Intergenic	rs7075914	10:15106281	OLAH	1.249	0.189	3.97e-11	0.878	1.619	0.826	2.01e-10
rs855969	10:119437857	Intergenic	rs6038583	20:6670473	Intergenic	-0.565	0.086	4.05e-11	-0.733	-0.397	0.621	3.85e-10
rs61933476	12:115385249	Intergenic	rs11665940	19:19830928	ZNF14	-0.877	0.133	4.85e-11	-1.138	-0.615	0.571	1.14e-10

Non-word reading

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs60347052	2:20353631	Intergenic	rs11259459	10:15116780	Intergenic	1.243	0.189	4.88e-11	0.873	1.614	0.814	2.68e-10
rs61933476	12:115385249	Intergenic	rs73006763	19:19827948	ZNF14	-0.875	0.133	5.02e-11	-1.137	-0.614	0.572	1.24e-10
rs61933476	12:115385249	Intergenic	rs73006760	19:19827898	ZNF14	-0.875	0.133	5.03e-11	-1.137	-0.614	0.572	1.25e-10
rs61933476	12:115385249	Intergenic	rs56167115	19:19827724	ZNF14	-0.875	0.133	5.06e-11	-1.136	-0.614	0.572	1.27e-10
rs60347052	2:20353631	Intergenic	rs113240764	10:15112925	OLAH	1.24	0.189	5.46e-11	0.869	1.61	0.818	3.18e-10
rs855969	10:119437857	Intergenic	rs6038586	20:6673320	Intergenic	-0.569	0.087	5.60e-11	-0.739	-0.399	0.581	5.50e-10
rs60347052	2:20353631	Intergenic	rs11259474	10:15129402	ACBD7	1.285	0.196	5.63e-11	0.901	1.67	0.896	1.76e-09
rs855969	10:119437857	Intergenic	rs6117384	20:6673542	Intergenic	-0.569	0.087	5.65e-11	-0.739	-0.399	0.579	5.40e-10
rs60347052	2:20353631	Intergenic	rs6602808	10:15129880	ACBD7	1.285	0.196	5.73e-11	0.901	1.67	0.896	1.80e-09
rs1552572	1:79335977	Intergenic	rs4406444	8:72636606	Intergenic	-0.909	0.139	5.73e-11	-1.181	-0.637	0.578	1.38e-10
rs62357516	5:35333842	Intergenic	rs11022262	11:12260355	Intergenic	-0.61	0.093	5.94e-11	-0.793	-0.427	0.78	1.87e-10
rs1552572	1:79335977	Intergenic	rs113927247	8:72641419	Intergenic	-0.913	0.139	5.99e-11	-1.186	-0.639	0.591	1.26e-10
rs13382464	2:223412143	SGPP2	rs16969089	19:28425579	Intergenic	-0.616	0.094	6.38e-11	-0.8	-0.431	0.748	1.22e-10
rs60347052	2:20353631	Intergenic	rs11259470	10:15126900	ACBD7	1.274	0.195	6.71e-11	0.891	1.657	0.922	2.41e-09
rs717246	1:79329349	Intergenic	rs4406444	8:72636606	Intergenic	-0.907	0.139	6.76e-11	-1.179	-0.635	0.57	1.79e-10
rs60347052	2:20353631	Intergenic	rs7911265	10:15127063	ACBD7	1.274	0.195	6.80e-11	0.891	1.657	0.921	2.42e-09
rs17426817	3:111664064	PHLDB2	rs7664664	4:135989375	Intergenic	0.975	0.149	6.83e-11	0.682	1.268	0.679	3.32e-10
rs60347052	2:20353631	Intergenic	rs58877824	10:15129337	ACBD7	1.276	0.196	6.90e-11	0.892	1.659	0.929	2.94e-09
rs1178829	12:119928497	CCDC60	rs4646579	15:58329528	ALDH1A2	-0.42	0.064	6.92e-11	-0.547	-0.294	0.394	1.63e-11
rs12336386	9:118124074	DEC1	rs497234	11:100817012	ARHGAP42	0.841	0.129	6.99e-11	0.588	1.094	0.924	2.61e-12
rs10817758	9:118143933	DEC1	rs497234	11:100817012	ARHGAP42	0.849	0.13	7.00e-11	0.594	1.104	0.829	4.69e-12
rs855967	10:119437923	Intergenic	rs6054462	20:6671666	Intergenic	-0.567	0.087	7.04e-11	-0.737	-0.397	0.671	2.88e-10
rs717246	1:79329349	Intergenic	rs113927247	8:72641419	Intergenic	-0.911	0.14	7.05e-11	-1.184	-0.637	0.583	1.63e-10
rs113706158	1:79340958	Intergenic	rs4406444	8:72636606	Intergenic	-0.906	0.139	7.06e-11	-1.179	-0.634	0.577	1.76e-10
rs1552572	1:79335977	Intergenic	rs113584153	8:72639552	Intergenic	-0.905	0.139	7.10e-11	-1.177	-0.633	0.571	1.96e-10
rs855967	10:119437923	Intergenic	rs6054463	20:6671823	Intergenic	-0.567	0.087	7.11e-11	-0.737	-0.396	0.675	2.88e-10

Non-word reading

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs72480678	1:79339882	Intergenic	rs4406444	8:72636606	Intergenic	-0.906	0.139	7.34e-11	-1.178	-0.633	0.59	1.67e-10
rs113706158	1:79340958	Intergenic	rs113927247	8:72641419	Intergenic	-0.91	0.14	7.36e-11	-1.184	-0.636	0.59	1.60e-10
rs855967	10:119437923	Intergenic	rs6038585	20:6671980	Intergenic	-0.566	0.087	7.43e-11	-0.736	-0.395	0.684	2.83e-10
rs28816175	1:79340628	Intergenic	rs4406444	8:72636606	Intergenic	-0.905	0.139	7.63e-11	-1.178	-0.633	0.587	1.73e-10
rs72480678	1:79339882	Intergenic	rs113927247	8:72641419	Intergenic	-0.909	0.14	7.66e-11	-1.183	-0.635	0.603	1.52e-10
rs28846376	1:79340794	Intergenic	rs4406444	8:72636606	Intergenic	-0.905	0.139	7.69e-11	-1.177	-0.632	0.588	1.74e-10
rs28816175	1:79340628	Intergenic	rs113927247	8:72641419	Intergenic	-0.909	0.14	7.95e-11	-1.183	-0.635	0.6	1.57e-10
rs11587713	1:79341461	Intergenic	rs4406444	8:72636606	Intergenic	-0.904	0.139	7.96e-11	-1.177	-0.632	0.586	1.79e-10
rs28846376	1:79340794	Intergenic	rs113927247	8:72641419	Intergenic	-0.909	0.14	8.01e-11	-1.182	-0.635	0.601	1.58e-10
rs1862182	5:95999526	CAST	rs11187785	10:82710004	Intergenic	0.611	0.094	8.07e-11	0.427	0.795	0.98	1.38e-10
rs11587713	1:79341461	Intergenic	rs113927247	8:72641419	Intergenic	-0.908	0.14	8.30e-11	-1.182	-0.634	0.599	1.63e-10
rs717246	1:79329349	Intergenic	rs113584153	8:72639552	Intergenic	-0.903	0.139	8.37e-11	-1.175	-0.63	0.564	2.54e-10
rs566676	12:119927705	CCDC60	rs4646579	15:58329528	ALDH1A2	-0.419	0.065	8.56e-11	-0.546	-0.292	0.401	2.06e-11
rs113706158	1:79340958	Intergenic	rs113584153	8:72639552	Intergenic	-0.902	0.139	8.79e-11	-1.175	-0.629	0.57	2.50e-10
rs143911136	12:115384277	Intergenic	rs11665940	19:19830928	ZNF14	-0.862	0.133	8.91e-11	-1.122	-0.601	0.732	2.01e-10
rs61933475	12:115384431	Intergenic	rs11665940	19:19830928	ZNF14	-0.862	0.133	8.91e-11	-1.122	-0.601	0.732	2.01e-10
rs1862182	5:95999526	CAST	rs10786160	10:82736947	Intergenic	0.61	0.094	8.95e-11	0.426	0.795	0.686	1.80e-10
rs10010558	4:173185798	GALNTL6	rs9654631	6:122136335	Intergenic	0.631	0.097	9.08e-11	0.44	0.822	0.7	1.05e-09
rs60347052	2:20353631	Intergenic	rs7896927	10:15126964	ACBD7	1.292	0.199	9.10e-11	0.901	1.683	0.937	5.40e-09
rs72480678	1:79339882	Intergenic	rs113584153	8:72639552	Intergenic	-0.901	0.139	9.13e-11	-1.174	-0.629	0.584	2.36e-10
rs143911136	12:115384277	Intergenic	rs73006763	19:19827948	ZNF14	-0.861	0.133	9.22e-11	-1.121	-0.6	0.733	2.21e-10
rs61933475	12:115384431	Intergenic	rs73006763	19:19827948	ZNF14	-0.86	0.133	9.22e-11	-1.121	-0.6	0.733	2.21e-10
rs143911136	12:115384277	Intergenic	rs73006760	19:19827898	ZNF14	-0.861	0.133	9.23e-11	-1.121	-0.6	0.733	2.22e-10
rs61933475	12:115384431	Intergenic	rs73006760	19:19827898	ZNF14	-0.86	0.133	9.24e-11	-1.121	-0.6	0.733	2.22e-10
rs1862182	5:95999526	CAST	rs7088863	10:82714182	Intergenic	0.611	0.094	9.27e-11	0.426	0.795	0.961	2.19e-10
rs143911136	12:115384277	Intergenic	rs56167115	19:19827724	ZNF14	-0.86	0.133	9.31e-11	-1.121	-0.6	0.733	2.26e-10

Non-word reading

SNP1	Chr1:Pos1	Gen1	SNP2	Chr2:Pos2	Gen2	Beta	SE	Meta.Pval	CI_Low	CI_UP	Het	Glide.Pval
rs61933475	12:115384431	Intergenic	rs56167115	19:19827724	ZNF14	-0.86	0.133	9.31e-11	-1.121	-0.6	0.733	2.26e-10
rs1862182	5:95999526	CAST	rs12219230	10:82736598	Intergenic	0.61	0.094	9.43e-11	0.425	0.795	0.686	1.82e-10
rs60347052	2:20353631	Intergenic	rs11259476	10:15130315	ACBD7	1.269	0.196	9.45e-11	0.885	1.653	0.939	3.91e-09
rs28816175	1:79340628	Intergenic	rs113584153	8:72639552	Intergenic	-0.901	0.139	9.49e-11	-1.174	-0.628	0.58	2.45e-10
rs28846376	1:79340794	Intergenic	rs113584153	8:72639552	Intergenic	-0.901	0.139	9.57e-11	-1.173	-0.628	0.581	2.46e-10
rs1862182	5:95999526	CAST	rs10882418	10:82731800	Intergenic	0.609	0.094	9.59e-11	0.425	0.794	0.691	1.95e-10
rs1862182	5:95999526	CAST	rs10882414	10:82729687	Intergenic	0.609	0.094	9.61e-11	0.425	0.794	0.692	1.95e-10
rs1862182	5:95999526	CAST	rs10882417	10:82731700	Intergenic	0.609	0.094	9.62e-11	0.425	0.794	0.691	1.95e-10
rs1862182	5:95999526	CAST	rs10786157	10:82730435	Intergenic	0.609	0.094	9.62e-11	0.425	0.794	0.692	1.95e-10
rs1862182	5:95999526	CAST	rs11187880	10:82735722	Intergenic	0.61	0.094	9.64e-11	0.425	0.794	0.686	1.86e-10
rs1862182	5:95999526	CAST	rs10159799	10:82733412	Intergenic	0.61	0.094	9.70e-11	0.425	0.794	0.687	1.87e-10
rs7521586	1:79323362	Intergenic	rs4406444	8:72636606	Intergenic	-0.901	0.139	9.74e-11	-1.174	-0.628	0.54	4.12e-10
rs1862182	5:95999526	CAST	rs10882434	10:82736052	Intergenic	0.612	0.095	9.75e-11	0.426	0.797	0.69	1.60e-10
rs78739610	1:79324284	Intergenic	rs4406444	8:72636606	Intergenic	-0.903	0.14	9.76e-11	-1.177	-0.63	0.559	4.20e-10
rs10982719	9:118141673	DEC1	rs497234	11:100817012	ARHGAP42	0.841	0.13	9.85e-11	0.586	1.096	0.855	4.68e-12
rs11587713	1:79341461	Intergenic	rs113584153	8:72639552	Intergenic	-0.9	0.139	9.92e-11	-1.173	-0.627	0.579	2.54e-10
rs6434538	2:192991640	TMEFF2	rs371164	19:6489814	Intergenic	-0.514	0.083	6.30e-10	-0.677	-0.351	0.808	4.84e-10

Acknowledgment

First of all, I would like to thank my advisors Bertram Müller-Myhsok and Hans-Werner Mewes for giving me the opportunity for this PhD thesis. My especial thanks to Bertram Müller-Myhsok for supporting me especially in the most difficult time I faced during my thesis. In addition, I would like to thank my fabulous colleagues Benno Pütz, Darina Czamara, Christiane Wolf, Tony Kam-Thong, Till Andlauer and Janine Arloth for always having an answer and an advise.
My thanks to Rainer Malik for his help in conducting the phenotype specific pathway analyses. Many thanks to Gerd Schulte-Körne and Kistina Moll (Klinik fr Kinder- und Jugendpsychiatrie, Ludwig-Maximilian Universit, Simon Fisher, Clyde Francks, and Alessandro Gialluisi (MPI for Psycholinguistic), Silvia Paracchini (University of St. Andrews, School of Medicine), Markus Nthen, Johannes Schumacher, Per Hoffmann and Kerstin Ludwig (Institut fr Humangenetik, Universitonn), Anthony Monaco, Andrew Morris and Thomas Scerri (Wellcome Trust Centre for Human Genetics), Tom Bourgeron, Franck Ramus, Fabien Fauchereau and Guillaume Huguet (Centre National de la Recherche Scientifique, Institut Pasteur) for sharing your data and made this survey possible.

To my wonderful family and friends. Thank you for your incredible support.
Most of all I want to thank my family and Christoph Schreiber.

Dedication

I want to dedicate this thesis to those that I lost during my PhD time.
To my aunt and to my lovely uncles.
Most of all I want to thank and dedicate this thesis to the light of my life, the most beautiful, graceful and warm-hearted mother.
Thank you for giving us more than anyone could ever expect. All I am, all I ever did and will is because of you.
I love and miss you.