THE UNIVERSITY
of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Discovery of Genetic Factors for Reading Ability and Dyslexia

Catherine Doust

PhD in Psychology

The University of Edinburgh

2023

CONTENTS

ABSTRACT **1**

LAY SUMMARY **4**

DECLARATION **8**

ACKNOWLEDGEMENTS **9**

LIST OF TABLES **10**

LIST OF FIGURES **15**

LIST OF ABBREVIATIONS **25**

CHAPTER 1: INTRODUCTION **28**

Language and literacy 28

Reading ability and dyslexia 29

Introduction to reading ability and dyslexia 29

Measuring reading ability and identifying dyslexia 33

The relationship of dyslexia to reading ability 34

The causes of variation in reading ability and dyslexia 35

Genetic research of reading ability and dyslexia 38

Introduction to the genetics of reading ability and dyslexia 38

Hypothesis-driven association analyses 43

Genome-wide association analyses 44

Thesis aims 47

CHAPTER 2: DISCOVERING GENETIC FACTORS ASSOCIATED WITH DYSLEXIA ... 50

Introduction ... 50

Supplementary Methods ... 65

23andMe Genotyping and Imputation ... 65

Chinese Reading Study Sample ... 67

Participants ... 67

Phenotypic Measures ... 67

Genotype Quality Control, Imputation, and Analysis ... 67

Biological Annotations ... 68

Partitioned Heritability ... 71

Evolutionary Analysis ... 71

Genlang Quantitative Trait Consortium Acknowledgements ... 73

UK Dyslexia ... 81

Supplementary References ... 82

Extended Data ... 89

Conclusion ... 101

CHAPTER 3: INVESTIGATING THE USE OF UNSELECTED ADULT SAMPLES FOR DISCOVERY OF GENETIC FACTORS ASSOCIATED WITH READING ABILITY ... 102

Introduction ... 102

Conclusion ... 113

CHAPTER 4: INVESTIGATING THE USE OF A PROXY PHENOTYPE FOR DISCOVERY OF GENETIC FACTORS ASSOCIATED WITH READING ABILITY 114

Introduction 114

Materials and Methods 117

Participants 117

Measures 118

Phenotypic analyses 123

Genotyping 124

Genetic analyses 124

Results 127

Phenotypic analyses 127

Genetic analyses 135

Discussion 136

Conclusion 139

CHAPTER 5: DISCUSSION 140

Objective 1: To identify novel genetic markers associated with dyslexia 140

Objective 2: To investigate whether any of the novel genetic markers for dyslexia occur in candidate genes and/or pathways for reading ability and/or dyslexia ... 142

Objective 3: To investigate which traits are genetically correlated with dyslexia 145

Objective 4a: To investigate whether unselected adult cohorts are valid for increasing power to detect variants associated with quantitative reading skill 146

Objective 4b: To investigate whether proxy measures of reading ability are valid for increasing power to detect variants associated with quantitative reading skill.... 148

Implications of the findings for our understanding of the genetics of dyslexia and reading ability 150

What does GWAS tell us about candidate gene findings? 150

What do the genetic findings tell us about the relationship between reading ability and dyslexia? 151

What more do we need to know to determine the relationship between reading ability and dyslexia? 152

What do the genetic findings tell us about the relationships between dyslexia and other developmental disorders? 154

What more do we need to know to determine the relationship between dyslexia and other developmental disorders? 156

How can understanding the genetics of reading ability and dyslexia help refine dyslexia identification criteria? 157

Conclusion 158

REFERENCES 160

ABSTRACT

The ability to read is critical to access wider learning and achieve qualifications, for accessing employment, and for adult life skills. Approximately one in ten individuals are affected by dyslexia, a learning difficulty which primarily impacts word reading and spelling. Specifically, phonological processing (the ability to decode phonemes) is impaired in dyslexia. Whilst some believe dyslexia represents the extreme end of a continuum of reading ability, others have suggested it is a distinct trait.

Variation in reading ability is a highly heritable (possibly 70%) complex trait caused by many genetic variants with a small effect size. However, the genetic architecture of reading ability and dyslexia is largely unknown due to a lack of quantitative genetic studies with sufficient statistical power to detect such small effect sizes. Previously, most genetic studies of reading ability have been conducted using samples of children with dyslexia, which tend to be modest in size. Whilst large samples of genotyped unselected adults have been collected (for example UK Biobank), phenotypic data on reading or language skills is rarely prioritised.

The overall aim of this thesis is to discover genetic variants associated with dyslexia and variation in reading skill in order to better understand the aetiology of reading difficulties, which in turn, may inform prediction, identification and intervention strategies in the future. Firstly, I will conduct a genome-wide association (GWA) study of over 50,000 adults with a self-reported dyslexia diagnosis and over 1 million controls to identify associated single nucleotide polymorphisms (SNPs). I will also explore ways to improve power for discovering genetic factors associated with reading ability. To do this, I will first investigate whether unselected adult samples are valid as a means to identify genetic factors associated with reading skill through a candidate gene

approach. Secondly, I will investigate whether proxy reading phenotypes are also a means to gain power through large cohorts that have no quantitative measure of reading ability. Such samples may be informative for future GWA meta-analysis of quantitative reading ability.

In Chapter 1, I will first introduce reading ability and dyslexia. I will discuss how reading ability is a quantitative trait and how it can be measured before discussing how dyslexia is identified. Then, I will consider how dyslexia may relate to reading ability: whether it represents the extreme end of a continuum of reading or whether it is a distinct trait. I will then introduce the known causes of variation in reading ability and dyslexia, which includes both environmental and genetic factors. Next, I will present the history of genetic studies of reading ability and dyslexia and their limitations. Finally, I will discuss the current state of genetic research into reading ability and introduce the aims of my thesis in detail.

Chapter 2 is a publication in Nature Genetics entitled *'Discovery of 42 genome-wide significant loci associated with dyslexia'* which includes GWA analysis of over 1 million 23andMe, Inc participants reporting on dyslexia diagnosis. I identify 42 independent genome-wide significant loci, 15 of which are in genes previously linked to cognitive ability and/or educational attainment, and 27 of which are novel and may be more specific to dyslexia. Extensive downstream biological analysis is performed alongside genetic correlations with other traits and dyslexia polygenic score prediction of quantitative reading scores.

Chapter 3 is a publication in Twin Research and Human Genetics on 'The association of dyslexia and developmental speech and language disorder candidate genes with reading and language abilities in adults' which analyses an adult population cohort

with quantitative measures of reading and language ability to replicate previous associations of candidate genes and biological pathways with dyslexia. I demonstrate that unselected adult populations are a valid means by which to identify genes which have previously been associated with dyslexia and/or speech and language disorder.

Chapter 4 is a research chapter in which I construct a proxy reading phenotype from measures of reading frequency in an unselected adult sample for whom a quantitative measure of reading ability is not available. I find that a dyslexia polygenic score constructed from the dyslexia GWA analysis in Chapter 3 cannot explain variation in the proxy phenotype suggesting that book reading is not a sufficient substitute for reading ability.

Finally, in Chapter 5, I integrate and discuss my research findings. I highlight the discovery of 42 variants associated with dyslexia through GWAS, in addition to the discovery of new genes and biological pathways which may form part of the biological basis of dyslexia. Following this, I consider what GWAS tells us about candidate gene findings. I discuss traits which are genetically correlated with dyslexia, including quantitative reading skills and ADHD. I consider the relationship between dyslexia and reading ability, and how genetic studies can help us to understand this better. I also consider the relationship between dyslexia and other developmental disorders, and how genetic studies can help us to understand this better. Lastly, I discuss methods to boost power for GWAS of reading ability.

LAY SUMMARY

In modern society, the ability to read is critical. We need to read to progress through school and achieve qualifications that enable us to access employment as adults. We need to read to access services, such as healthcare and transport, and to understand financial and legal documents. Without the ability the read, we limit our access to the rich culture available through the books, newspapers, magazines, and the internet.

How good we are at reading varies from person to person. Some find it harder than others and approximately one in ten individuals have a reading disability called dyslexia. Those with dyslexia find it difficult to read and spell words. Words are made up of individual units of sounds called phonemes. For example, the word 'chip' is made up of three phonemes, 'ch', 'i' and 'p'. To read words, we need to identify the individual phonemes and put them together to make a whole word. This process is known as phonological awareness and it is impaired in dyslexia. Because of this, it takes longer for children with dyslexia to learn how to read and they need to rely more on memorising whole words than other children.

How good we are at reading and whether or not we have dyslexia depends on our genes and the environment we grow up in. It is thought that up to 70% of the differences between people in their reading ability is determined by genes. However, reading ability and dyslexia are not controlled by a single gene like some inherited traits. They are controlled by many thousands of genes which each have a very small effect on our reading ability. Because of this, it is challenging for genetic scientists to identify which genes are responsible.

Understanding which genes are responsible for reading ability and dyslexia is important for two reasons. Firstly, because each gene has a unique function in the

body, this information helps neuroscientists to understand how the brain functions to enable us to read, and which parts of the brain are involved. Secondly, once we know which genes are responsible, we could develop a genetic test for infants, based on a polygenic score (PGS), which predicts how likely they are to develop dyslexia. Currently, dyslexia is not usually identified until years after children have begun learning to read, and sometimes not even until teenage years or adulthood. If we can identify children who are at risk from birth, we can provide far greater support from the start.

To find out which genes cause a particular trait, geneticists use a method called a genome-wide association study (GWAS). This method uses a computer to analyse the complete set of DNA code, called the genome, of every person in a large group. They look for pieces of DNA code which are more similar in one set of people than in another. In this case, geneticists are looking for pieces of DNA code which are more common in people with dyslexia, or with a lower reading ability, than in people without dyslexia, or those with a higher reading ability. These pieces of DNA code are likely to be responsible for how good we are at reading or whether we have dyslexia.

Groups of people who volunteer to have their information collected for such studies are called cohorts. GWAS is very effective when cohorts are very large or when just a few genes have a big impact on the trait being investigated. Most of the existing cohorts with information about reading ability are formed of small groups of children or adolescents identified as having dyslexia, which are unfortunately not big enough for GWAS to be very effective. Cohorts of adults from the general population tend to be much bigger because it is easier to recruit adult volunteers to donate their DNA. Unfortunately, very few adult cohorts have information about participants' reading ability. However, most fortunately, the personal genetics company 23andMe asked

their customers (over 1 million people) whether they had been diagnosed with dyslexia.

In my research, I had the opportunity to use 23andMe Inc.'s data. I aimed to discover new genetic factors that are responsible for dyslexia. To do this, I analysed the genomes of their customers, which included over 50,000 adults with dyslexia, compared with over 1 million adults without dyslexia. This is the largest cohort of its kind and the largest genetic study of dyslexia to date. I used GWAS to identify parts of the DNA code that are more common in those with dyslexia, finding 42 new genetic factors. From this, I could create a PGS for dyslexia, which can be used to test whether individuals in other cohorts have similar DNA, by comparing their genomes with the dyslexia PGS.

An additional aim of my research was to investigate ways to improve our ability to find genes that are responsible for reading ability. It is important to investigate the genetics of reading ability to better understand how dyslexia works, because the two are thought to have a closely related biological basis. Very large cohorts like 23andMe Inc. are not available with proper reading tests administered. Therefore, firstly, I investigated whether we can make more use of adult cohorts from the general population, since these are more readily available in greater numbers than cohorts of children. Using a new cohort of adults called the Brisbane Adult Reading Study, tested on psychometric measures of reading and spelling, I replicated genes that have previously been linked to dyslexia and/or reading and language ability. This was a positive indicator that adult cohorts are a valuable untapped resource for genetic studies of reading ability.

Secondly, I investigated whether it is possible to use alternative traits to represent reading ability in cohorts that have no measure of reading ability. Some traits, such as how often we read books, are closely related to our reading ability. These can be used instead of the original trait and are called proxies. Many cohorts that have no measure of reading ability do have other closely related information that could be used a proxy. This would enable geneticists to make use of many large cohorts and potentially discover new genes responsible for reading ability. I used a cohort of adults from the general population called the Health & Retirement Study (HRS) to investigate this. In this cohort, information is available on how much time participants spend reading. I created a dyslexia PGS for each individual in the HRS cohort and tested whether this PGS could predict their reading proxy score. In this case, the proxy was not a valid replacement for reading ability, however, the use of proxies should be refined and further investigated due to the benefits they could bring.

DECLARATION

This thesis has been composed by myself with the exception of two multiple-authored publications, to which I made the primary contribution as the first author in both cases. The publications are clearly identified as such throughout the thesis. This work has not been submitted for any other degree or professional qualification.

The publications included within this thesis:

Doust, C., Gordon, S. D., Garden, N., Fisher, S. E., Martin, N. G., Bates, T. C., & Luciano, M. (2020). The Association of Dyslexia and Developmental Speech and Language Disorder Candidate Genes with Reading and Language Abilities in Adults. In *Twin Research and Human Genetics* (Cambridge University Press; 2020/04/06 ed., Vol. 23, Issue 1, pp. 23–32).

Doust, C., Fontanillas, P., Eising, E., Gordon, S. D., Wang, Z., Alagöz, G., Molz, B., Aslibekyan, S., Auton, A., Babalola, E., Bell, R. K., Bielenberg, J., Bryc, K., Bullis, E., Coker, D., Partida, G. C., Dhamija, D., Das, S., Elson, S. L., … Quantitative Trait Working Group of the GenLang Consortium. (2022). Discovery of 42 genome-wide significant loci associated with dyslexia. *Nature Genetics*, *54*(11), 1621–1629. https://doi.org/10.1038/s41588-022-01192-y

Signed

Catherine Doust

ACKNOWLEDGEMENTS

Thank you to my co-authors of our publications for their contributions and feedback. Thank you also to the participants in the Brisbane Adults Reading Study, 23andMe Inc, the GenLang Consortium, and the Health and Retirement Study.

Thank you to the School of Philosophy, Psychology and Language Sciences at the University of Edinburgh for my scholarship. Lastly, thank you to my supervisors Michelle Luciano and Riccardo Marioni for your time and support.

LIST OF TABLES

Chapter 1

Table 1	Summary of Studies Investigating Loci Associated with Dyslexia and/or Reading Phenotypes	Page 40

Chapter 2

Table 1	New SNP Associations with Dyslexia, Including Gene-Based Results, EQTL Status, Expression in Brain and Validation in Three Independent Cohorts (GenLang Consortium, CRS and Neurodys	Page 55

Supplementary Tables

Available at https://www.nature.com/articles/s41588-022-01192-y#Sec31

Table S1	Genome-Wide Significant (p < 5x10-8) and Suggestive (p < 1x10-6) Index Variants in the GWAS of Self-Reported Dyslexia Diagnosis from 23andMe
Table S2	Genome-Wide Significant and Suggestive Index Variants in the 23andMe GWAS of Self-Reported Dyslexia Diagnosis Reported in the GWAS Catalog at a Genome-Wide Significant Level
Table S3	Validation of the 23andMe Significant Dyslexia GWAS Variants within the GenLang GWAS Meta-Analysis of Reading Ability
Table S4	Validation of the 23andMe Significant Dyslexia GWAS Variants within the GenLang GWAS Meta-Analysis of Spelling Ability

Table S5	Validation of the 23andMe Significant Dyslexia GWAS Variants Within a Meta-Analysis GWAS of 2,274 Dyslexia Cases by Gialluisi et al. (2021) - NeuroDys (English, German, Finnish, French, Dutch and Hungarian languages)
Table S6	Validation of the 23andMe Significant Dyslexia GWAS Variants within the Chinese Reading Study Measured on Reading Accuracy and Fluency
Table S7	Genes for which the Combined Association of all Variants is Significantly Associated with Self-Reported Dyslexia Diagnosis from 23andMe as Estimated in MAGMA
Table S8	The Association of Biological Pathways with Self-Reported Dyslexia Diagnosis from 23andMe as Estimated in MAGMA
Table S9	Most Probable Gene Target for the Independent Index Variants Significantly Associated with Self-Reported Dyslexia Diagnosis from 23andMe as identified by Opentargets (https://genetics.opentargets.org/)
Table S10	Functional Annotation of Genome-Wide Significant Index Variants Associated with Self-Reported Dyslexia Diagnosis from 23andMe
Table S11	Descriptions of the Functions of Genes (where known) that are Significantly Associated with Self-Reported Dyslexia Diagnosis from 23andMe in Genome-wide Gene-based Tests
Table S12	Functional Annotation of Genome-Wide Significant Genes Associated with Self-Reported Dyslexia Diagnosis from 23andMe
Table S13	Gene Property Analysis of the GWA Results for Self-Reported Dyslexia Diagnosis from 23andMe for General Tissues

Table S14	Gene Property Analysis of the GWA Results for Self-Reported Dyslexia Diagnosis from 23andMe for Brain Tissues
Table S15	Brain Tissue-Specific Gene Expression of Genome-wide Associated Genes in the GWAS of Self-Reported Dyslexia Diagnosis in 23andMe
Table S16	Heritability Enrichment Estimates Based on the GWA Results for Self-Reported Dyslexia Diagnosis from 23andMe for Major Functional Annotations
Table S17	Heritability of Self-Reported Dyslexia Diagnosis from 23andMe Partitioned by Brain Tissue Gene Expression
Table S18	Heritability of Self-Reported Dyslexia Diagnosis from 23andMe Partitioned by Brain Cell Type
Table S19	Heritability of Self-Reported Dyslexia Diagnosis from 23andMe Partitioned by Cell Type-Specific H3K4me1
Table S20	Heritability of Self-Reported Dyslexia Diagnosis from 23andMePartitioned by Tissue-Specific H3K4me3
Table S21	Heritability of Self-Reported Dyslexia Diagnosis from 23andMe Partitioned by Evolutionary Annotations
Table S22	Genetic Correlations Between Self-Reported Dyslexia Diagnosis from 23andMe and Other Traits Estimated by LDSC
Table S23	Genetic Correlations Between Self-Reported Dyslexia Diagnosis from 23andMe and UK Biobank Brain Imaging Measures Estimated by LDSC

Table S24	Dyslexia Polygenic Score Prediction of Quantitative Reading, Spelling, and Nonword Repetition Test Scores in Four Independent Cohorts
Table S25	Look-Up of SNP Variants Previously Reported to be Significantly Associated with Dyslexia or Reading Ability in the GWAS of Self-Reported Dyslexia Diagnosis from 23andMe
Table S26	Look-Up of Dyslexia Candidate Genes from the Literature in the MAGMA Gene-Based Test Results from the GWAS of Self-Reported Dyslexia Diagnosis from 23andMe
Table S27	Results for Three Biological Pathways Previously Implicated in Dyslexia from MAGMA Gene Set Analysis of the GWAS of Self-Reported Dyslexia Diagnosis from 23andMe

Chapter 3

Table 1	Full Assessment Battery Administered to the Brisbane Adult Cohort	Page 105
Table 2	Intercorrelations (Pearson’s R) Between Reading, Spelling and Language Measures and Covariates	Page 107
Table 3	Summary Statistics of the Main Continuous Variables Including Reading, Spelling and Language Measures	Page 107
Table 4	Percentage Frequencies of Major Discrete Variables	Page 108
Table 5	Gene-Based Analysis of Dyslexia and Speech/Language Disorder Candidate Genes for Association with Measures of Reading, Spelling and Language	Page 109

Table 6	Gene-Set Analysis of Dyslexia and Speech/Language Disorder Candidate Gene Sets and Neuron Migration and Axon Guidance Pathways for Association with Measures of Reading, Spelling and Language	Page 109

Chapter 4

Table 1	Variables Included from the Health and Retirement Study	Page 121
Table 2	Descriptive Statistics of the Variables	Page 128
Table 3	Correlation Between Variables (Pearson’s *r*)	Page 131
Table 4	Linear Regression Parameters for Age-Adjusted Variables	Page 132
Table 5	Standardised Loadings of the Variables from the Principal Components Analysis	Page 134
Table 6	Dyslexia Polygenic Score Prediction of Proxy Reading Measures in 1822 Unselected Adults	Page 135

LIST OF FIGURES

Chapter 1

Figure 1	The Simple View of Reading	Page 30
Figure 2	The Dual Route Cascaded Model of Reading	Page 31

Chapter 2

Figure 1	Manhattan Plot of the Genome-Wide Association Analysis of Dyslexia	Page 54
Figure 2	Genetic Correlation of Dyslexia with Other Phenotypes	Page 56
Figure 3	Genetic Correlations Between Dyslexia and Measures of Reading, Language and Nonverbal IQ	Page 57

Supplementary Figures

Extended Data Figure 1	QQ Plot of Dyslexia GWAS Results	Page 89
Extended Data Figure 2	Manhattan Plot of Dyslexia GWAS Results for Females	Page 91
Extended Data Figure 3	Manhattan Plot of Dyslexia GWAS Results for Males	Page 92
Extended Data Figure 4	Variant Effect Predictor Summary for the Credible Set of Variants Significantly Associated with Dyslexia	Page 93

Extended Data Figure 5	Enrichment Estimates for Major Functional Annotations	Page 94
Extended Data Figure 6	Heritability of Dyslexia Partitioned by Brain Tissue Gene Expression	Page 95
Extended Data Figure 7	Heritability of Dyslexia Partitioned by Brain Cell Type	Page 96
Extended Data Figure 8	Heritability of Dyslexia Partitioned by Cell-Type Specific H3k4me1	Page 97
Extended Data Figure 9	Heritability of Dyslexia Partitioned by Cell-Type Specific H3k4me3	Page 99

Available in Supplementary Information at https://www.nature.com/articles/s41588-022-01192-y#Sec30

Supplementary Figure 1.i	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr9p22.3 esv3619796 structural variant nearby rs3122702
Supplementary Figure 1.ii	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr3q22.3 rs13082684
Supplementary Figure 1.iii	Regional association plot which includes information on the credible variant set and known genes in the region

	for the significant association between dyslexia and chr20q13.13 rs11393101
Supplementary Figure 1.iv	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr17q12 rs34349354
Supplementary Figure 1.v	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr9q34.11 rs9696811
Supplementary Figure 1.vi	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr20q11.21 rs4911257
Supplementary Figure 1.vii	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr11q23.1 rs138127836
Supplementary Figure 1.viii	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr17q12 rs12453682
Supplementary Figure 1.ix	Regional association plot which includes information on the credible variant set and known genes in the region

	for the significant association between dyslexia and chr2q33.1 rs72916919
Supplementary Figure 1.x	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr11p14.1 rs676217
Supplementary Figure 1.xi	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr1p32.1 rs12737449
Supplementary Figure 1.xii	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr4q31.3 rs4696277
Supplementary Figure 1.xiii	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr12q24.12 rs7310615
Supplementary Figure 1.xiv	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr10q24.2 rs10786387
Supplementary Figure 1.xv	Regional association plot which includes information on the credible variant set and known genes in the region

	for the significant association between dyslexia and chr5q34 rs41012
Supplementary Figure 1.xvi	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr7q22.3 rs3839821
Supplementary Figure 1.xvii	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr3p12.1 rs10511073
Supplementary Figure 1.xviii	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr1q21.3 rs4845687
Supplementary Figure 1.xix	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr3p24.3 rs373178590
Supplementary Figure 1.xx	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr3p13 rs13097431
Supplementary Figure 1.xxi	Regional association plot which includes information on the credible variant set and known genes in the region

	for the significant association between dyslexia and chr2p22.1 rs906549
Supplementary Figure 1.xxii	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr1p13.3 rs2091329
Supplementary Figure 1.xxiii	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr2q22.3 rs497418
Supplementary Figure 1.xxiv	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr3p21.31 rs2624839
Supplementary Figure 1.xxv	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chrXq27.3 rs5904158
Supplementary Figure 1.xxvi	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr10q24.33 rs34732054
Supplementary Figure 1.xxvii	Regional association plot which includes information on the credible variant set and known genes in the region

	for the significant association between dyslexia and chr3q26.33 rs7625418
Supplementary Figure 1.xxviii	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr2q33.1 rs6435017
Supplementary Figure 1.xxix	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr17q23.3 rs72841395
Supplementary Figure 1.xxx	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr13q12.13 rs375018025
Supplementary Figure 1.xxxi	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr14q32.2 rs35131341
Supplementary Figure 1.xxxii	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr5q35.1 rs59261790
Supplementary Figure 1.xxxiii	Regional association plot which includes information on the credible variant set and known genes in the region

	for the significant association between dyslexia and chr9p22.3 rs3122702
Supplementary Figure 1.xxxiv	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr2q12.1 rs367982014
Supplementary Figure 1.xxxv	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr12q24.31 rs4767921
Supplementary Figure 1.xxxvi	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr5q33.3 rs867009
Supplementary Figure 1.xxxvii	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr19q13.2 rs60963584
Supplementary Figure 1.xxxviii	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr7q11.22 rs77059784
Supplementary Figure 1.xxxix	Regional association plot which includes information on the credible variant set and known genes in the region

	for the significant association between dyslexia and chr2p23.2 rs1969131
Supplementary Figure 1.xl	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr7p14.1 rs62453457
Supplementary Figure 1.xli	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr6p22.3 rs2876430
Supplementary Figure 1.xlii	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr1q41 rs35570426
Supplementary Figure 1.xliii	Regional association plot which includes information on the credible variant set and known genes in the region for the significant association between dyslexia and chr7q11.22 rs3735260

Chapter 4

Figure 1	Frequency Distribution of the Variables	Page 129
Figure 2	Scree Plot of Eigenvalues of the Components from a Principal Components Analysis	Page 133

Figure 3	Component Loadings (> .2) of the First Two Rotated Components from the Principal Components Analysis	Page 134
Figure 4	Scatter Plot of the Self-Reported Dyslexia PGS Against the HRS Proxy Reading Ability Measure	Page 136

LIST OF ABBREVIATIONS

ADHD	Attention deficit hyperactivity disorder
CADD	Combined annotation dependent depletion
CAMS	Consumption and Activities Mail Survey
CNS	Central nervous system
DRC	Dual Route Cascaded
DSM	Diagnostic and Statistical Manual of Mental Disorders
GWA	Genome-wide association
GWAS	Genome-wide association study
HRS	Health & Retirement Study
HWE	Hardy-Weinberg equilibrium
IBD	Identity-by-descent
IQ	Intelligence quotient
MAF	Minor allele frequency
ncCADD	Non-coding combined annotation dependent depletion
ncGERP	Non-coding genomic evolutionary rate profiling
ncGWAVA	Non-coding genome-wide annotation of variants
ncRVIS	Non-coding residual variation intolerance score

PC	Principal component
PCA	Principal components analysis
pcGERP	Protein-coding genomic evolutionary rate profiling
PGS	Polygenic score
pLI	Probability of loss-of-function intolerance
QC	Quality control
QQ	Quantile-quantile
QTL	Quantitative trait locus
RDB	RegulomeDB category
RNA	Ribonucleic acid
SEM	Structural equation modelling
SES	Socio-economic status
SLI	Specific language impairment
SNP	Single nucleotide polymorphism
SVR	Simple View of Reading
TES	Transcription end site
TPM	Transcripts per million
TSS	Transcription start site

UTR	Untranslated region

CHAPTER 1: INTRODUCTION

Language and literacy

Language is a uniquely human ability that evolved in the lineage leading to humans (Tallerman & Gibson, 2011). By 50,000 to 150,000 years ago, *Homo sapiens* had developed distinct cognitive and social abilities (Tilot et al., 2020). Writing is a recent cultural invention that draws on our capacity for spoken language (Corballis, 2017). It was first invented over 4000 years ago by the Sumerians, and was independently invented by several other civilisations before the Common Era (Yushu, 2010). Writing gave structure to trade, finances, governance, law, medicine, and religion. With writing, came reading, the ability to interpret symbols by sight or touch.

In today's society, the ability to write and read words is an essential skill known as literacy. Functional literacy expands upon this and refers to, "*The capacity of a person to engage in all those activities in which literacy is required for effective function of his or her group and community and also for enabling him or her to continue to use reading, writing and calculation for his or her own and the community's development.*" (UNESCO Institute of Statistics, 2020, webpage). Functional literacy can be further classified into sub-types, for example: digital literacy (the ability to access, evaluate and create digital information including websites) (UNESCO Institute of Statistics, 2018), health literacy (the ability to retrieve and evaluate healthcare information to make informed health-related decisions) (Chenxi Liu et al., 2020), or financial literacy (the ability to retrieve, interpret, and organise financial information to make sound financial decisions) (OECD, 2022). A lack of these capacities, known as functional illiteracy, impacts social, economic and physical wellbeing. It prohibits wider learning, educational achievement, access to employment, access to services and health,

financial, and legal information, and access to cultural texts and social communications. Further, the lower academic achievement and higher unemployment rates that can result from poor literacy skill affect self-esteem and increase the likelihood of individuals being subject to discrimination (Schumacher et al., 2007). Individuals with lower literacy skill are also therefore more susceptible to internalising problems such as anxiety and depression (Francis et al., 2019).

Reading ability and dyslexia

Introduction to reading ability and dyslexia

The process of reading requires a combination of cognitive functions. Phonological awareness is required to identify individual graphemes (the written units of words) and their corresponding phonemes (the sound units), in order to recognise whole words (APA, 2023). Language comprehension is required to understand the meaning of the words and sentences. Also required is morphological awareness, the recognition and interpretation of morphemes (units of words that affect meaning such as ‘ex’ or ‘ful’), and vocabulary knowledge. The ability to read varies largely between individuals, and individuals can be poorer at some aspects of reading than others.

The Simple View of Reading (SVR) theory proposes that reading ability (*R*) is the product of decoding (*D*) and comprehension (*C*), that $R = D \times C$ (Gough & Tunmer, 1986). Thus, according to this theory, an individual who can decode but not comprehend cannot read, and an individual who can comprehend but cannot decode cannot read. This can also be visualised as a plot of two dimensions in which individuals can exist anywhere in the four quadrants (Figure 1). Whilst both decoding and comprehension are integral to reading, the model fails to account for other aspects of reading including morphological awareness, vocabulary and fluency. Further, this

model assumes decoding and comprehension are distinct processes, however evidence shows they are bridged by important sub-processes. For example, morphological awareness and vocabulary awareness affect both word recognition and language comprehension (Kearns & Al Ghanem, 2019).

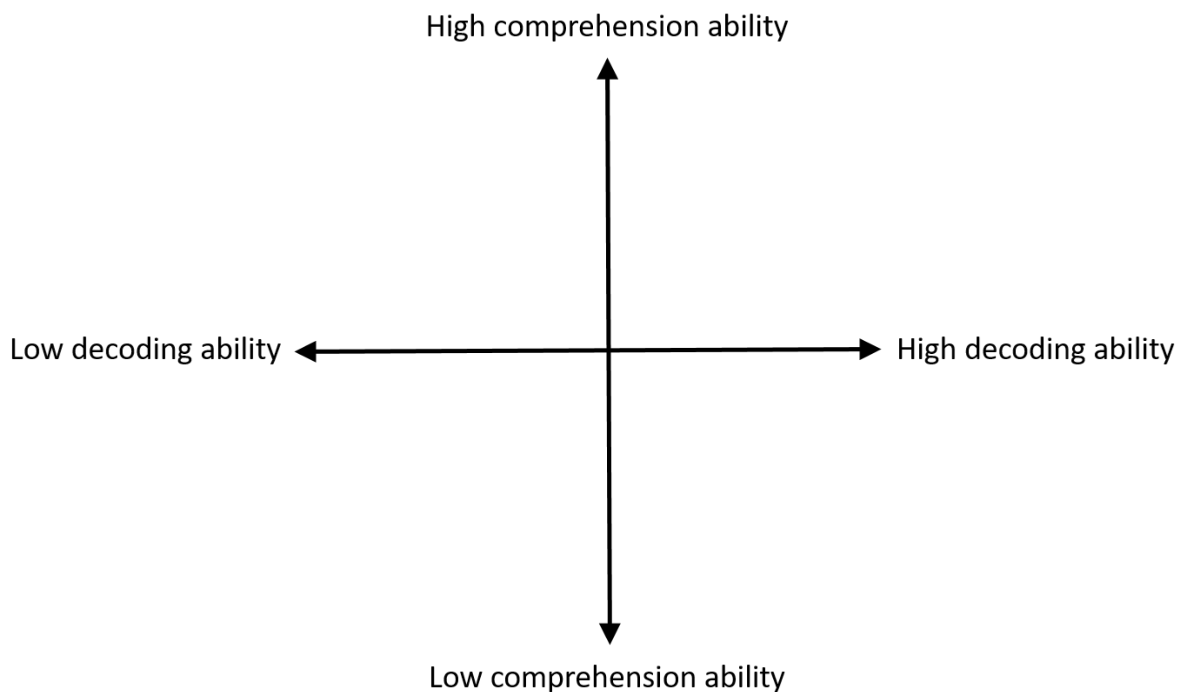


Figure 1. The Simple View of Reading (adapted from Rose, 2009)

In comparison, the Dual Route Cascaded (DRC) model of reading (Figure 2) proposes that the process of reading occurs via two routes: nonlexical and lexical (Coltheart et al., 2001). The nonlexical route is comparable to decoding within the SVR, relying on grapheme-phoneme correspondence rules to decode regular words and nonwords. The lexical route involves retrieving a word from a mental lexicon, and is used to read irregular words which cannot be decoded via the nonlexical route. The lexical route may be a direct path from written word to speech, or it may incorporate semantic processing, for example, interpreting morphemes within words (units that affect meaning such as ‘s’ or ‘ed’). The DRC model offers a more comprehensive explanation

of reading than the SVR and computational modelling has comparable outcomes to human testing.

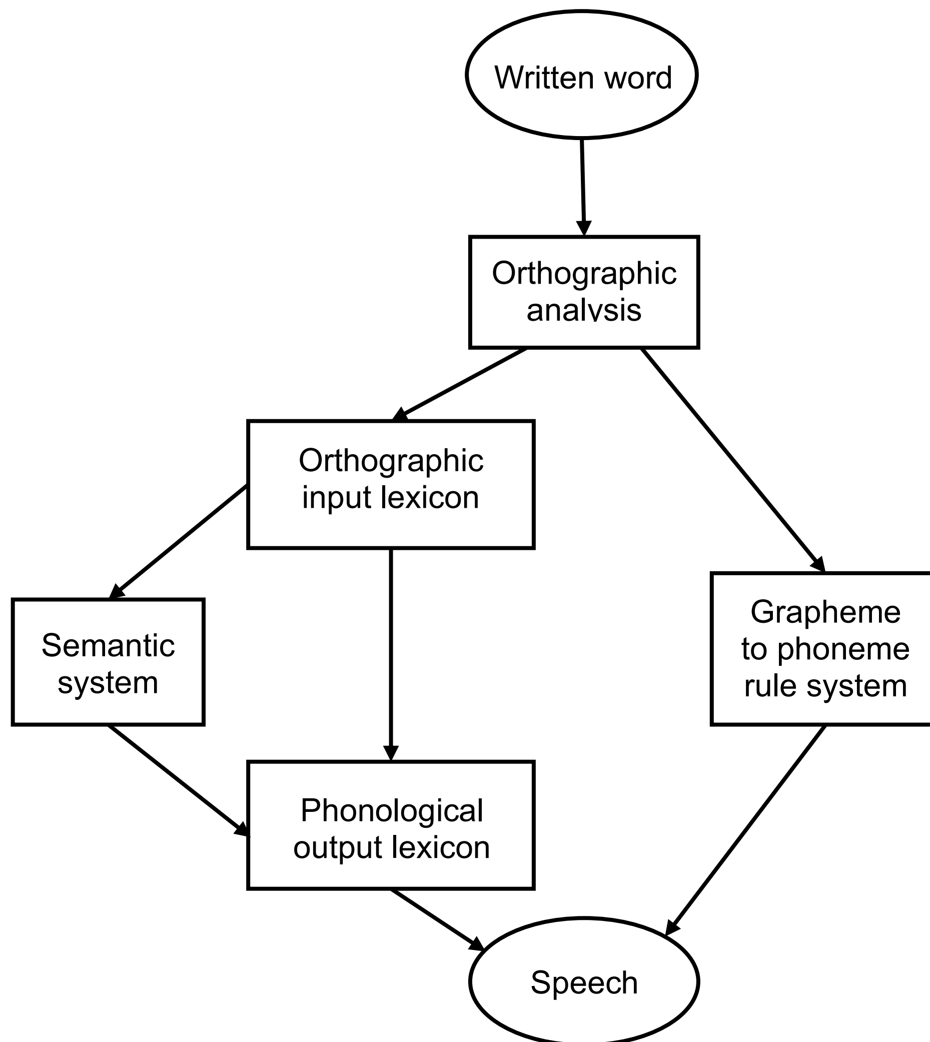


Figure 2. The Dual Route Cascaded Model of Reading (adapted from Coltheart et al. 2001)

The reading disability dyslexia was first described in Germany in 1877 as wordblindheit (word blindness), a difficulty of reading words despite being able to see them clearly. In the early 1900s, word blindness was described in literature, particularly by Hinshelwood, both as a condition that was acquired in adulthood following brain injury or disease (Hinshelwood, 1895, 1902) and as a condition affecting healthy children who were otherwise intelligent (Hinshelwood, 1900). Hinshelwood proposed that word

blindness was caused by a defective visual memory for letters and words. He hypothesised that the region of the brain responsible for visual memory was the angular and supra-marginal gyri on the left-side of right-handed individuals, because this was the region that was damaged in the adults with acquired word blindness.

As a result of Hinshelwood and others' case studies of 'bright and intelligent' children with dyslexia, for many decades, dyslexia was only diagnosed when an individual's reading skill was markedly discrepant from what would be expected based on their general cognitive ability (Critchley, 1970), known as the discrepancy diagnostic model.

Now, in a widely accepted definition of dyslexia in the UK, dyslexia is defined as a learning difficulty that occurs across a range of intellectual abilities which affects accurate word reading and spelling (Rose, 2009), specifically, the phonological awareness aspect of reading. The American Diagnostic and Statistical Manual of Mental Disorders 5 (DSM-5) similarly defines dyslexia as a neurocognitive disorder of impaired word reading and spelling that is not a result of other cognitive impairments or an adverse environment (American Psychiatric Association, 2013). Similarly, within the DRC model of reading, components of the nonlexical route are impaired in those with dyslexia, explaining the difficulty in reading nonwords (*The Science of Reading*, 2005). Whilst the lexical route remains intact, enabling individuals with dyslexia to store regular and irregular words in their lexicon, this route is slower than the nonlexical route. As a result, children with dyslexia have difficulty with letter recognition, identifying phonemes, and rapid naming of familiar symbols (Hulme & Snowling, 2016). Dyslexia also affects verbal memory and verbal processing speed. In the Rose definition, dyslexia is considered a continuum as opposed to a distinct category, in which individuals can exhibit a range of reading skill scores below an arbitrary cut-off. However, dyslexia is identified categorically based on a specified threshold. In addition

to impaired reading and spelling, individuals with dyslexia may also have difficulties with language, motor coordination and mental mathematics. Depending on the diagnostic criteria, dyslexia may affect five to 15% of the population (Katusic et al., 2001; Shaywitz et al., 1990).

Measuring reading ability and identifying dyslexia

Reading ability is measured using standardised tests which can include assessing reading of regular words (those which follow phoneme-correspondence rules), nonwords (made up words that follow regular phoneme-grapheme correspondence rules, e.g., swad) and irregular words (words which do not follow regular phoneme-grapheme correspondence rules), as well as comprehension. In line with the DRC model of reading, irregular words, also known as sight words, are learned by memorisation, stored in lexical memory, and recalled as a whole word, and therefore do not require phonological awareness. Regular words can either be decoded or stored in lexical memory. Only nonwords specifically test phonological processing because they have not previously been stored in the lexicon. Assessment of reading comprehension is broader and can take into account word meaning, grammar, and integrating new information in a text with existing knowledge. Reading ability can be measured in terms of standard deviations above or below the population mean for a given age.

Dyslexia is usually identified through reading and spelling tests, particularly, letter recognition (in young children), phoneme awareness, rapid naming, nonword reading, and single word spelling (Simpson & Everatt, 2005). Children with language difficulties upon school entry are at greater risk of developing dyslexia (Thompson et al., 2015), as are children with a parent or sibling with dyslexia. Further, dyslexia is often comorbid with speech/language disorder (Snowling et al., 2020) and attention deficit

hyperactivity disorder (ADHD). Specific language impairment (SLI), now known as developmental language delay, occurs in less than 10% of children (Tomblin et al., 1997), however studies have estimated that over 50% of children with SLI meet criteria for dyslexia (summarised by Adlof, 2017). ADHD also occurs in less than 10% of the general population but an estimated 25-40% of children with either a reading disability or ADHD meet the criteria for both disorders (e.g., August & Garfinkel, 1990; Willcutt & Pennington, 2000).

The relationship of dyslexia to reading ability

There are different theories as to how dyslexia relates to reading ability. Reading ability is a quantitative trait that is measured as a continuous variable and can include a range of skills including phonological awareness and comprehension. Dyslexia is a dichotomous categorisation usually based upon an agreed threshold, for example, a reading accuracy score of 1.5 standard deviations or more below the population mean for a given age (Peterson & Pennington, 2012).

Some argue that dyslexia is distinct phenotype from the lower end of a continuum of reading ability. Influentially, Rutter & Yule (1975) distinguished 'specific reading retardation' from 'general reading backwardness'. The former group's reading ability was discrepant from general cognitive ability whilst the latter represented those at the lower end of a continuum of reading ability and poor general cognitive ability. In line with this view, the SVR originally proposed that individuals with dyslexia (those in the specific retardation group) were specifically poor at decoding, but not comprehension (Hoover & Gough, 1990). It is now recognised that individuals with poor decoding ability and any comprehension ability may have dyslexia (those in either left quadrant in the SVR (Figure 1, Rose, 2009). Further, the National Health Service definition of dyslexia in the UK states that dyslexia can occur across a range of intellectual abilities

(NHS, 2018), indicating the irrelevance of the discrepancy diagnostic model today and broadening the net of dyslexia to include those who would have formally be identified as simply poor readers. Further, functional magnetic resonance imaging has demonstrated that the neurobiological basis of dyslexia is independent of IQ (Tanaka et al., 2011).

Others argue that dyslexia is the extreme end of a continuum of reading ability, which can be quantified using a normal distribution model (Rodgers, 1983; Shaywitz et al., 1992). They argue that any threshold along the continuum can be used to categorise disability and the cut-off is therefore arbitrary. This aligns with how dyslexia is commonly assessed as a specific agreed upon deviation from mean score in quantitative tests of ability. In the multiple deficit model of developmental disorders, environmental and genetic risk factors interact to alter development and determine the variation in performance on such quantitative tests (Pennington, 2006). The model was derived from a drive to understand the aforementioned high comorbidity of developmental disorders, including dyslexia and ADHD. These disorders share aetiological risk resulting from a combination of multiple factors. This model is in contrast to the prevailing cognitive understanding of dyslexia: that it is caused by a single phonological deficit, because it suggests no single aetiological factor is sufficient. Instead, the liability distribution is continuous amongst the population and influenced by additive and interactive effects.

The causes of variation in reading ability and dyslexia

Variation in reading ability is caused by multiple factors. Dyslexia has substantial familial transmission and heritability of dyslexia (e.g., DeFries et al., 1987; Olson et al., 1989) and reading ability (e.g., Bates et al., 2007; Harlaar et al., 2005; Hayiou-Thomas et al., 2010) has been estimated at 30-80%. Heritability is the proportion of

phenotypic variation in a population that is due to genetic factors (W. G. Hill, 2013). In the case of dyslexia, a categorical trait, it indicates that a considerable proportion of an individual's propensity for developing dyslexia is a result of their genes. It is important to note, however, that heritability estimates are specific to the population in which they have been estimated, hence the wide range in findings. The environment also has influence over individuals' reading skill and propensity for developing dyslexia. I will first discuss the environmental factors which influence reading, then discuss our understanding of the genetic mechanisms.

Both peri-natal and childhood factors influence reading ability. Pre-term birth (Soleimani et al., 2014) and very low birthweight (Liu et al., 2016; Samuelsson et al., 2006) are risk factors for reading disability due to adverse neurodevelopment. Additionally, male sex is a higher risk factor for reading difficulties than female sex. The ratio of low reading scores/dyslexia in boys to girls has been demonstrated to range from 1.6:1 to 3:1 (e.g., Arnett et al., 2017; Harlaar et al., 2005; Liu et al., 2016; Quinn & Wagner, 2015; Rutter et al., 2004). It is unclear with the sex difference has biological or environmental origins or a combination of the two (Granocchio et al., 2023). Neurological differences in dyslexia between the sexes have been investigated with mixed results (Krafnick & Evans, 2019), whilst a likely environmental factor is that boys' reading performance is more susceptible to disadvantaged socioeconomic backgrounds than girls' (OECD, 2019).

Of early-childhood factors, socio-economic status (SES) and home literacy environment (Hamilton et al., 2016) are predictive of reading skill. Home literacy environment encompasses multiple factors including availability of books, family reading habits and parental literacy (Mascheretti et al., 2018). However, home literacy environments are primarily created by parents, each of whom are differently equipped

to create an advantageous environment for learning to read depending on factors including intelligence and SES, which have a strong genetic influence (Deary et al., 2006; W. D. Hill et al., 2019). Therefore, the correlation between home literacy environment and reading skill is likely in part due to the shared genes of parents and children.

Parental age at birth and level of parental education are also associated with reading ability (Mascheretti et al., 2015). This may be in part because higher maternal age is associated with a more verbally stimulating environment (Fergusson & Woodward, 1999) and similarly, more educated mothers are more likely to talk to their children in ways which are positively associated with language development (Hoff & Tian, 2005). This aligns with the finding that language skill in early childhood effects later phonological awareness and is a risk factor for dyslexia (NICHD Early Child Care Research Network, 2005). Again, gene-environment correlations are likely to contribute here. Indeed, reading ability, which is highly heritable, correlates with level of education (M. C. Smith, 1990).

Understanding environmental risk factors can improve prediction and inform intervention strategies. Factors such as pre-term birth and SES could inform prediction of those who are more likely to develop dyslexia. Intervention strategies can enhance known protective factors such an advantageous home literacy environment in early childhood. For example, the Scottish Government targets vulnerable families with their Play Talk Read early years literacy and language programme (Scottish Government, 2010). Additionally, since the environment can moderate genetic effects (Friend et al., 2008; Kremen et al., 2005; Rosenberg et al., 2012), it is important to understand environmental risk factors in order to untangle gene by environment interactions. Importantly, understanding the genetic mechanisms underlying reading ability and

dyslexia will help us to understand the biological processes involved and could inform predictive tests in the future. Additionally, it may inform a more refined definition of dyslexia to enable more accurate identification and perhaps even subtypes that may result from subtly different neurocognitive pathways.

Genetic research of reading ability and dyslexia

Introduction to the genetics of reading ability and dyslexia

It was first proposed that dyslexia has a genetic basis in 1907, when observations of 'word-blindness' occurring in the same families were recorded (Hinshelwood, 1907; Stephenson, 1907). Since then, twin and adoption studies have demonstrated that genetic factors have a stronger influence on both reading ability (Stevenson et al., 1987; Wadsworth et al., 2002) and propensity for dyslexia (DeFries et al., 1987; Olson et al., 1989) than a shared environment. These studies reported high heritability, with findings indicating continuity between disability and the normal distribution (Harlaar et al., 2005). Further, variation in reading ability has been shown to be genetically influenced across different languages and cultures (e.g., Christopher et al., 2013).

The high heritability of dyslexia led to a drive to seek the causative gene(s). Originally, it was thought that just a few genes had large effects on phenotypes. Before genome-wide association technology became widely available, linkage analysis was used in an attempt to identify a specific locus or loci that were linked to individuals with dyslexia in multiple generation pedigrees using DNA markers that segregate with the trait. Linkage analysis has a low resolution and low sensitivity. Therefore, the regions identified tend to be quite broad, as opposed to a specific gene or section of a gene, and the technique is only sensitive to genetic factors with large effect sizes. However, given the lack of significant linkage findings, it became apparent that dyslexia was

more likely to be a complex trait caused by many genetic variations very small effect sizes. Despite the limitations of linkage analysis, nine genetic loci reported to be associated with dyslexia in early linkage analyses were named by the Human Gene Nomenclature Committee (Seal et al., 2023) as DYX1 to DYX9 (Table 1). For example, a locus at 15q15-21, which became known as *DYX1*, was linked to reading disability in several studies (Cardon et al., 1994; Chapman et al., 2004; Grigorenko et al., 1997; Nöthen et al., 1999; Schumacher et al., 2008). Additionally, in the first successful linkage analysis of a cognitive trait, a quantitative trait locus (QTL) on chromosome 6 was associated with reading performance in sibling pairs (Cardon et al., 1994). An attempt was made to replicate these loci in independent samples in a range of languages, however the results were inconsistent. For example, despite linkage of the *DYX1* locus being reported in several studies, many more were unable to replicate the association (Bisgaard et al., 1987; de Kovel et al., 2004; Fagerheim et al., 1999; Grigorenko et al., 2001; Kaminen et al., 2003; Kaplan et al., 2002; Marlow et al., 2003; Nopola-Hemmi et al., 2001). Similarly inconsistent findings were reported for the other candidate loci (Table 1). Linkage analyses tend to be limited in size, and findings could be a type one error, producing inflated effect sizes, which would explain the difficulty in reproducing results (Button et al., 2013). However, linkage analyses studies did reveal the polygenic nature of dyslexia and the need for more refined, higher-powered genetic analysis of reading ability and dyslexia.

Table 1

Summary of studies investigating loci associated with dyslexia and/or reading phenotypes

Locus	Region	Studies which support the association	Studies which do not support the association	Candidate genes
DYX1	15q21	(Fulker et al., 1991) (S. D. Smith et al., 1991) (Grigorenko et al., 1997) (Schulte-Körne et al., 1998) (Nöthen et al., 1999) (Morris et al., 2000) (Chapman et al., 2004) (Marino et al., 2004) (Schumacher et al., 2008)	(Bisgaard et al., 1987) (Rabin et al., 1993) (Sawyer et al., 1998) (Fagerheim et al., 1999) (Norton et al., 2000) (Nopola-Hemmi et al., 2001) (Fisher et al., 2002) (Kaminen et al., 2003) (Marlow et al., 2003) (de Kovel et al., 2004)	*DYX1C1*
DYX2	6p21-22	(S. D. Smith et al., 1991) (Cardon et al., 1994) (Grigorenko et al., 1997) (Fisher et al., 1999) (Gayán et al., 1999) (Grigorenko et al., 2000) (Fisher et al., 2002) (Kaplan et al., 2002) (Grigorenko et al., 2003) (Marlow et al., 2003) (Turic et al., 2003)	(Field & Kaplan, 1998) (Sawyer et al., 1998) (Schulte-Körne et al., 1998) (Fagerheim et al., 1999) (Nöthen et al., 1999) (Fagerheim et al., 2000) (Norton et al., 2000) (Petryshen et al., 2000) (Nopola-Hemmi et al., 2001) (Fagerheim et al., 2002) (Kaminen et al., 2003) (Chapman et al., 2004) (de Kovel et al., 2004) (Raskind et al., 2005)	*DCDC2* *KIAA0319*

			(Igo et al., 2006)	
DYX3	2p15-16	(Fagerheim et al., 1999) (Fagerheim et al., 2000) (Fagerheim et al., 2002) (Fisher et al., 2002) (Francks et al., 2002) (Petryshen et al., 2002) (Marlow et al., 2003)	(Norton et al., 2000) (Nopola-Hemmi et al., 2001) (Chapman et al., 2004) (de Kovel et al., 2004) (Raskind et al., 2005) (Igo et al., 2006)	*MRPL19* *C2ORF3*
DYX4	6q11-12	(Petryshen et al., 2001)	(Fagerheim et al., 1999) (Norton et al., 2000) (Nopola-Hemmi et al., 2001) (Fisher et al., 2002) (Kaminen et al., 2003) (Marlow et al., 2003) (de Kovel et al., 2004) (Raskind et al., 2005) (Igo et al., 2006)	
DYX5	3p12-12	(Nopola-Hemmi et al., 2001) (Fisher et al., 2002) (Hannula-Jouppi et al., 2005)	(Fagerheim et al., 1999) (Norton et al., 2000) (Kaminen et al., 2003) (de Kovel et al., 2004) (Raskind et al., 2005) (Igo et al., 2006)	*ROBO1*
DYX6	18p11	(Fisher et al., 2002) (Marlow et al., 2003)	(Fagerheim et al., 1999) (Norton et al., 2000) (Nopola-Hemmi et al., 2001) (Kaminen et al., 2003) (Chapman et al., 2004) (de Kovel et al., 2004) (Raskind et al., 2005)	

			(Igo et al., 2006) (Schumacher et al., 2006)	
DYX7	11p15	(Hsiung et al., 2004)	(Fagerheim et al., 1999) (Norton et al., 2000) (Nopola-Hemmi et al., 2001) (Fisher et al., 2002) (Kaminen et al., 2003) (de Kovel et al., 2004) (Raskind et al., 2005) (Igo et al., 2006)	
DYX8	1p34-36	(Rabin et al., 1993) (Grigorenko et al., 2001) (Tzenova et al., 2004)	(Sawyer et al., 1998) (Fagerheim et al., 1999) (Fagerheim et al., 2000) (Norton et al., 2000) (Nopola-Hemmi et al., 2001) (Fagerheim et al., 2002) (Fisher et al., 2002) (Kaminen et al., 2003) (Marlow et al., 2003) (de Kovel et al., 2004) (Raskind et al., 2005) (Igo et al., 2006)	*KIA0319L*
DYX9	Xq27-28	(de Kovel et al., 2004)	(Norton et al., 2000) (Nopola-Hemmi et al., 2001) (Fisher et al., 2002) (Kaminen et al., 2003) (de Kovel et al., 2004)	

Note. Adapted from Scerri & Schulte-Korne (2010)

Hypothesis-driven association analyses

As genetic sequencing technology improved, fine mapping of the proposed dyslexia loci and hypothesis-driven candidate gene association analyses were conducted in attempt to verify candidate genes within the loci identified through linkage studies. In gene-based tests, genetic variation of pre-defined genomic regions is analysed in case-control cohorts. Many of these were small samples of children, adolescents and young adults. This led to identification of several dyslexia candidate genes including *DYX1C1*, *DCDC2*, *KIAA0319*, *ROBO1*, *FOXP2* and *CNTNAP2* (Deffenbacher et al., 2004; Hannula-Jouppi et al., 2005; Morgan et al., 1993; Taipale et al., 2003; Vernes et al., 2008; Wilcke et al., 2012), but again, the findings proved difficult to replicate consistently in independent samples (Mascheretti et al., 2014; Tran et al., 2013). A major problem is the small sample sizes which are underpowered to detect the small individual effect of variants and may lead to Type 1 errors. In a GWA meta-analysis of over 30,000 individuals, which is well powered compared to previous studies, variation in *DCDC2* was associated with nonword reading after correction for multiple testing (Eising et al., 2022), suggesting this gene may be a valid candidate.

Similarly, studies have sought to replicate associations with biological pathways that have been suggested to be involved in reading and dyslexia through gene-set analyses. Gene-set tests for overrepresentation of significantly-associated SNPs within a specific set of genes. Biological pathways previously proposed for involvement in dyslexia include the axon guidance pathway (GO:0007411: ‘chemotaxis process that directs the migration of an axon growth cone to a specific target site’; 216 genes) and the neuron migration pathway (GO:0001764: ‘movement of an immature neuron from germinal zones to specific positions where they will reside as they mature’; 214 genes) (Galaburda et al., 2006; Paracchini et al., 2007; Poelmans et al., 2011),

although these proposals have been criticised for a lack of robust evidence (Guidi et al., 2018).

Genome-wide association analyses

In recent years, advances in genotyping technology has enabled hypothesis-free genome-wide association (GWA) analyses in which single nucleotide polymorphisms (SNPs) which associate with the trait are identified. As with candidate gene association analysis, a large sample size is required to have sufficient statistical power to avoid generating Type 1 errors. Few samples have been collected with quantitative measures of reading ability or information on dyslexia diagnosis which are sufficiently powered for discovery of novel genetic factors. One approach to improve power is performing GWA meta-analyses, in which multiple samples are analysed together to boost statistical power for detecting SNPs of small effect size (e.g. Eising et al., 2022; Gialluisi et al., 2019). Eising et al. presented a well-powered GWA meta-analysis of quantitative reading and language abilities in tens of thousands of children and young adults from the international GenLang network. This incorporated 22 cohorts, identifying a genome-wide significant association of the rs11208009 SNP on chromosome 1 with word reading. All five reading and language traits used demonstrated robust SNP heritability which accounted for up to 26% of trait variation. This approach highlights the value of larger sample sizes and the need to re-evaluate candidate gene studies, in light of this more robust, systematic technique. However, whilst this study was reasonably well powered, sample sizes must increase further still to uncover the genetic factors underlying reading ability and dyslexia which are now know to be polygenic, that is, influenced by many genes of small effect.

Complications of genetic studies of dyslexia

In addition to the limitation of sample size, several factors complicate the genetic study of dyslexia. Firstly, samples are likely to be heterogeneous and include individuals with other disorders with a genetic basis, because developmental disorder co-occur with each more frequently than in the general population (Gidziela et al., 2023). In fact, an estimated 40% to 60% of children with dyslexia have an additional disorder (Moll et al., 2020). Studies have demonstrated genetic correlations between developmental disorders (Gialluisi et al., 2020; Wadsworth et al., 2015). Genetic correlation is the proportion of phenotypic variation that two traits share due to genetic causes, indicating shared biological pathways and/or causal relationships (van Rheenen et al., 2019). In particular, language disorders share genetic variance with dyslexia, with twin studies demonstrating shared genetic aetiology of reading impairments with non-word repetition in children with SLI (Bishop, 2001). ADHD is also genetically correlated with dyslexia, with genetic factors accounting for 60% of the comorbidity of reading difficulties with inattention symptoms in a twin design (Wadsworth et al., 2015). Genetic factors account for significantly less of the comorbidity between reading difficulties and hyperactivity/impulsivity symptoms, however. These overlaps complicate both identification of developmental disorders and disentangling genetic aetiologies. To determine whether there are dyslexia-specific genes and which genetic variants are shared with other developmental disorders, genetic studies should contrast dyslexia only-samples with samples of dyslexia comorbid with other disorders. This would reveal which genetic factors are unique to dyslexia, and which overlap with SLI and/or ADHD and/or others. Genomic SEM is another approach by which genetic relationships between developmental disorders can be further elucidated (Grotzinger et al., 2019), which will be expanded on further in the discussion. Genomic SEM analyses genetic correlations and SNP heritability from

GWAS summary statistics from different complex traits whose overlap in genetic architecture is unknown (Grotzinger et al., 2019).

Measurements of reading ability may differ between studies, although a multivariate twin study demonstrated that reading tests scores are genetically correlated (Byrne et al., 2013), and genomic structural equation modelling (SEM) has demonstrated the shared genetic architecture of reading skills (Eising et al., 2022). More problematically, as previously discussed, criteria for identifying dyslexia are often based on an arbitrary threshold on reading tests which can vary between different studies in different places. Studies conducted in different decades, or survey questions asking participants to self-report a past diagnosis, may differ vastly in their diagnostic criteria. Historically, dyslexia was diagnosed when reading skill was discrepant from IQ, however the current diagnostic net for dyslexia is much broader. Thus, criteria for dyslexia identification are inconsistent both longitudinally and spatially (between different research groups in different countries).

Thesis aims

The aim of this thesis is to discover novel genetic variants associated with dyslexia and variation in reading skill in order to better understand the aetiology of reading difficulties, which in turn, may inform prediction, identification and intervention strategies in the future. I will address this aim firstly by performing the largest GWAS of dyslexia to date, and secondly, by investigating methods to increase statistical power to detect genetic variants associated with reading ability, for which studies have historically been underpowered. It is important to uncover the genetic mechanisms underlying variation in specific reading skills to better understand the causes of reading difficulties. This will also help to address the question of whether dyslexia is a distinct phenotype or whether it represents the tail end of a continuum in reading ability.

Through the GWAS, which is well powered to detect variants of small effect size, I hope to identify novel genetic variants and biological pathways associated with a self-reported dyslexia diagnosis. I will also perform hypothesis-driven association analyses of candidate genes and biological pathways previously associated with dyslexia. Previous genetic studies of dyslexia and reading ability have been limited in power to detect variants of small effect size which are now known to comprise the polygenic architecture. We must re-evaluate previously proposed candidate genes and biological pathways for dyslexia/reading ability in the light of well-powered GWAS. I will also calculate the genetic correlations between dyslexia and a wide range of cognitive, psychiatric, physical and socioeconomic measures to increase our understanding of co-occurring traits and disorders.

Previously, genetic studies of reading ability have been underpowered to detect the variants of small effect size which contribute to variation in the population. To increase statistical power to detect genetic variants, larger sample sizes are required. Most genetic studies of reading ability, and especially dyslexia, have been conducted in affected children and adolescents. Such samples tend to be modest in size because it harder to recruit and genotype large numbers of children. In comparison, adult participants from the general population are easier to recruit and genotype in greater numbers. There is some support that the normal range of reading ability in older adults shares the same genetic aetiology as in childhood and adolescence (Luciano et al., 2018). Therefore, it could be investigated further whether adult cohorts are a valid approach for gene discovery of reading ability generalizable to childhood.

Currently, large genotyped samples are available (e.g., UK Biobank, Sudlow et al. (2015), typically in adults, however few have quantitative measures of reading ability, which take considerable time to collect. Meta-analysis of multiple smaller cohorts with measures of reading ability is one approach to obtaining larger samples which has proven fruitful (Eising et al., 2022; Gialluisi et al., 2019). However, in recent examples, power was still limited by sample size. GWAS sample sizes could be increased adding sample of adults who have been measured on validated reading tests.

There may also be power gains by using large cohorts that have no quantitative measures of reading ability, but have collected data which have been shown to correlate with reading ability. For example, reading books is indicative of higher literacy skill, along with a diversity of reading practices (books, newspapers, magazines and documents) (M. C. Smith, 1996) and some studies have collected information on reading practices as part of lifestyle questionnaires. Another example is health literacy, which can be predicted by reading ability (Kripalani et al., 2006) and is assessed in

some studies. Such measures could be used as proxy phenotypes for reading ability, enabling the use of valuable genotyped samples for discovery of genetic factors associated with reading ability, which could be incorporated in GWAS meta-analyses.

The aims of this thesis can be divided into the following objectives:

1. To identify novel genetic markers associated with dyslexia.
2. To investigate whether any of these occur in candidate genes and/or pathways for reading ability and/or dyslexia.
3. To investigate which traits are genetically correlated with dyslexia.
4. To investigate whether the following methods are valid for increasing power to detect variants associated with quantitative reading skill:
    a) Unselected adult cohorts
    b) Proxy measures of reading abilities

Together, these will further our understanding of the genetics of reading ability and dyslexia. In the next chapter, I will perform a GWAS of dyslexia.

CHAPTER 2: DISCOVERING GENETIC FACTORS ASSOCIATED WITH DYSLEXIA

Introduction

Dyslexia is a developmental disorder which impacts accurate word reading and spelling in approximately one in 10 individuals (Katusic et al., 2001; Shaywitz et al., 1990). Dyslexia is highly heritable (40-80%) (DeFries et al., 1987; Olson et al., 1989) but the underlying genetic mechanisms are poorly understood. Previous genetic studies of dyslexia have been underpowered to detect the genetic variants of small effect size which comprise the polygenic architecture. These have tended to be linkage analyses, candidate gene and biological pathway-based tests, and small genome-wide association analyses in samples of affected children, adolescents and young adults. Whilst numerous loci and candidate genes and biological pathways have been proposed, attempts to replicate findings have been inconsistent. A recent relatively well-powered GWAS meta-analysis of over 30,000 individuals identified a genome-wide significant association of rs11208009 with word reading (Eising et al., 2022). However, greater sample sizes are required to have sufficient statistical power for discovering further genetic factors. The overall aim of this thesis is discovery of genetic variants associated with dyslexia and variation in reading skill. In this chapter, I seek to address Objectives 1, 2 and 3, that are described at end of Chapter 1. To address Objective 1, I will carry out a GWAS of self-reported dyslexia diagnosis in 51,800 individuals plus one million controls and discuss any novel variants associated with dyslexia. To address Objective 2, I will carry out gene-based association tests of candidate genes and biological pathways for dyslexia, to see if I can replicate the results of previous studies. I will also carry out non-hypothesis driven gene-based tests and compare the results. Finally, to address Objective 3, I will analyse the genetic

correlation between self-reported dyslexia and 98 other traits in order to determine which traits are most highly genetically correlated with dyslexia. The work for this chapter was published in the article which follows. My contribution was the analyses downstream of the GWAS, including the functional annotations, the partitioned heritability and the genetic correlation

nature genetics

Article https://doi.org/10.1038/s41588-022-01192-y

Discovery of 42 genome-wide significant loci associated with dyslexia

Received: 28 August 2021

Accepted: 23 August 2022

Published online: 20 October 2022

Check for updates

Catherine Doust[1], Pierre Fontanillas[2], Else Eising[3], Scott D. Gordon[4], Zhengjun Wang[5], Gökberk Alagöz[3], Barbara Molz[3], 23andMe Research Team*, Quantitative Trait Working Group of the GenLang Consortium*, Beate St Pourcain[3,6,7], Clyde Francks[3,6], Riccardo E. Marioni[8], Jingjing Zhao[5], Silvia Paracchini[9], Joel B. Talcott[10], Anthony P. Monaco[11], John F. Stein[12], Jeffrey R. Gruen[13], Richard K. Olson[14,15], Erik G. Willcutt[14,15], John C. DeFries[14,15], Bruce F. Pennington[16], Shelley D. Smith[17], Margaret J. Wright[18], Nicholas G. Martin[4], Adam Auton, Timothy C. Bates[1], Simon E. Fisher[3,6] and Michelle Luciano[1] ✉

Reading and writing are crucial life skills but roughly one in ten children are affected by dyslexia, which can persist into adulthood. Family studies of dyslexia suggest heritability up to 70%, yet few convincing genetic markers have been found. Here we performed a genome-wide association study of 51,800 adults self-reporting a dyslexia diagnosis and 1,087,070 controls and identified 42 independent genome-wide significant loci: 15 in genes linked to cognitive ability/educational attainment, and 27 new and potentially more specific to dyslexia. We validated 23 loci (13 new) in independent cohorts of Chinese and European ancestry. Genetic etiology of dyslexia was similar between sexes, and genetic covariance with many traits was found, including ambidexterity, but not neuroanatomical measures of language-related circuitry. Dyslexia polygenic scores explained up to 6% of variance in reading traits, and might in future contribute to earlier identification and remediation of dyslexia.

The ability to read is crucial for success at school and access to employment, information and health and social services, and is related to attained socioeconomic status[1]. Dyslexia is a neurodevelopmental disorder characterized by severe reading difficulties, present in 5–17.5% of the population, depending on diagnostic criteria[2,3]. It often involves impaired phonological processing (the decoding of sound units, or phonemes, within words) and frequently co-occurs with psychiatric and other developmental disorders[4], especially attention-deficit hyperactivity disorder

[1]Department of Psychology, University of Edinburgh, Edinburgh, UK. [2]23andMe, Inc., Sunnyvale, CA, USA. [3]Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands. [4]Genetic Epidemiology Laboratory, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [5]School of Psychology, Shaanxi Normal University and Shaanxi Key Research Center of Child Mental and Behavioral Health, Xi'an, China. [6]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands. [7]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [8]Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK. [9]School of Medicine, University of St Andrews, St Andrews, UK. [10]Institute of Health and Neurodevelopment, Aston University, Birmingham, UK. [11]Office of the President, Tufts University, Medford, MA, USA. [12]Department of Physiology, Anatomy and Genetics, Oxford University, Oxford, UK. [13]Departments of Pediatrics and Genetics, Yale Medical School, New Haven, CT, USA. [14]Department of Psychology and Neuroscience, University of Colorado, Boulder, CO, USA. [15]Institute for Behavioral Genetics, University of Colorado, Boulder, CO, USA. [16]Department of Psychology, University of Denver, Denver, CO, USA. [17]Department of Neurological Sciences, College of Medicine, University of Nebraska Medical Center, Omaha, NE, USA. [18]Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: michelle.luciano@ed.ac.uk

(ADHD)[5,6] and speech and language disorders[7,8]. Dyslexia may represent the low extreme of a continuum of reading ability, a complex multifactorial trait with heritability estimates ranging from 40% to 80%[9,10]. Identifying genetic risk factors not only aids increased understanding of the biological mechanisms, but may also expand diagnostic capabilities, facilitating earlier identification of individuals prone to dyslexia and co-occurring disorders for specific support.

Previous genome-wide investigations of dyslexia have been limited to linkage analyses of affected families[11] or modest ($n < 2,300$ cases) association studies of diagnosed children and adolescents[12]. Candidate genes from linkage studies show inconsistent replication, and genome-wide association studies (GWAS) have not found significant associations, although *LOC388780* and *VEPH1* were supported in gene-based tests[12]. Larger cohorts are vital for increasing sensitivity to detect new genetic associations of small effect. Here, we present the largest dyslexia GWAS to date, with 51,800 adults self-reporting a dyslexia diagnosis and 1,087,070 controls, all of whom are research participants with the personal genetics company 23andMe, Inc. We validate our association discoveries in independent cohorts, provide functional annotations of significant variants (mainly single-nucleotide polymorphisms (SNPs)) and potential causal genes, and estimates of SNP-based heritability. Lastly, we investigate genetic correlations with reading and related skills, health, socioeconomic, and psychiatric measures, and evaluate the evidence for previously implicated dyslexia candidate genes in our well-powered results.

Results

Genome-wide associations

The full dataset included 51,800 (21,513 males, 30,287 females) participants responding 'yes' to the question 'Have you been diagnosed with dyslexia?' (cases) and 1,087,070 (446,054 males, 641,016 females) participants responding 'no' (controls). Participants were aged 18 years or over (mean ages of cases and controls were 49.6 years (s.d. 16.2) and 51.7 years (s.d. 16.6), respectively). We identified 42 independent genome-wide significant associated loci ($P < 5 \times 10^{-8}$) and 64 loci with suggestive significance ($P < 1 \times 10^{-6}$) (Fig. 1 and Supplementary Table 1). Genomic inflation was moderate ($\lambda_{GC} = 1.18$) and consistent with polygenicity (see Q–Q plot, Extended Data Fig. 1). We also performed sex-specific GWAS and age-specific GWAS (younger or older than 55 years) because dyslexia prevalence was higher in our younger (5.34% in 20- to 30-year-olds) than older (3.23% in 80- to 90-year-olds) participants. These subsample analyses showed high consistency with the main GWAS (of the full sample). Genetic correlation estimated by linkage disequilibrium (LD) score regression (LDSC) was 0.91 (95% confidence intervals (CI): 0.86–0.96; $P = 8.26 \times 10^{-253}$) in males and females, and 0.97 (95% CI: 0.91–1.02; $P = 2.32 \times 10^{-268}$) between younger and older adults.

Of the 17 genome-wide significant variants in the female GWAS (Extended Data Fig. 2), all but four (rs61190714, rs4387605, rs12031924 and rs57892111) were significant in the main GWAS and, of these four, three were in LD with an SNP that approached significance ($P < 3.3 \times 10^{-7}$ or smaller) in the main analysis. Intergenic SNP rs57892111 (located between *TFAP2B* and *PKHD1* on chromosome 6p) was not among the significant or suggestive SNPs of the main analysis, and so may represent a female-specific variant. There is no evidence from existing GWAS that this SNP is associated with any other human trait. Of the six genome-wide significant variants in the male GWAS (Extended Data Fig. 3), all were significant in the main GWAS.

In the main GWAS, all significant variants were autosomal, except rs5904158 at Xq27.3 (for regional association plots, see Supplementary Fig. 1). A total of 17 index variants were in high LD with published (genome-wide significant) associated SNPs in the NHGRI GWAS Catalog[13] (15 were associated with cognitive/educational traits; Supplementary Tables 1 and 2). Thus, a total of 27 associated loci showed no evidence of published genome-wide associations with traits expected to overlap with dyslexia (for example, educational attainment, cognitive ability) and were considered new (Table 1).

Of 38 associated loci (the 4 remaining were tagged by indels unavailable in validation cohorts), 3 (rs13082684, rs34349354 and rs11393101) were significant at a Bonferroni-corrected level ($P < 0.05/38$) in the GenLang consortium GWAS meta-analysis of reading ($n = 33,959$) and spelling ($n = 18,514$) ability[14]. At $P < 0.05$, 18 were associated in GenLang, 3 in the NeuroDys case-control GWAS[12] ($n = 2,274$ cases), and 5 in the Chinese Reading Study (CRS) of reading accuracy and fluency ($n = 2,270$; Supplementary Note) (Table 1 and Supplementary Tables 3–6).

Gene-based tests identified 173 significantly associated genes (Supplementary Table 7) but no significantly enriched biological pathways (Supplementary Table 8). We estimated the LDSC liability-scale SNP-based heritability of dyslexia to be $h^2_{SNP} = 0.152$ (standard error = 0.006) using the 23andMe sample prevalence of 5%, and $h^2_{SNP} = 0.189$ (standard error = 0.008) using a 10% prevalence of dyslexia, which is more typical of the general population[2,3].

Fine-mapping and functional annotations

Within the credible variant set (Supplementary Table 1), missense variants were the most common (55%) of the coding variants; Extended Data Figure 4 summarizes all predicted variant effects. Predicted deleterious variants by SIFT (Sorting Intolerant From Tolerant) score were identified in *R3HCC1L*, *SH2B3*, *CCDC171*, *C1orf87*, *LOXL4*, *DLAT*, *ALG9* and *SORT1*. Within the credible variant set, no genes were especially intolerant to functional variation (smallest LoFtool (Loss-of-Function) percentile was 0.39). For the 42 associated loci, the most probable gene targets of each were estimated by the Overall V2G (Variant-to-Gene) score from OpenTargets (Supplementary Table 9). Two index variants (missense variant rs12737449 (*C1orf87*) and rs3735260 (*AUTS2*)) could be causal because they had combined annotation dependent depletion (CADD) scores suggestive of deleteriousness to gene function according to Kircher et al.[15] (Supplementary Table 10). The *AUTS2* variant RegulomeDB rank of 2b indicated a regulatory role; its chromatin state supported location at an active transcription start site[16,17].

Of the 173 significant genes from genome-wide gene-based tests in MAGMA (see Supplementary Table 11 for their functions), 129 could be functionally annotated (Supplementary Table 12). Protein-coding and noncoding sequences are actively conserved in approximately three-quarters of these genes, 63% are more intolerant to variation than average and 33% are intolerant to loss-of-function mutations. Gene property analysis for general tissues and 13 brain tissues confirmed the importance of the brain and specific brain regions (Supplementary Tables 13 and 14). Levels of brain expression for 125 of the 173 significant genes from gene-based tests could be mapped in FUMA and are shown in Supplementary Table 15. A total of 20 genes showed high general brain expression levels and, of these, 3 (*PPP1R1B*, *NPM1* and *WASF3*) were located near significant SNP associations. Of the 12 brain regions assessed, gene expression was generally highest in the cerebellar hemisphere, cerebellum, and cerebral cortex, consistent with the results of gene property analysis.

Partitioned heritability

SNP-based heritability of dyslexia partitioned by functional annotation showed significant enrichment for conserved regions and H3K4me1 clusters (Supplementary Table 16 and Extended Data Fig. 5). There was enrichment in genes expressed in the frontal cortex, cortex and anterior cingulate cortex ($P < 4.17 \times 10^{-3}$) (Supplementary Table 17 and Extended Data Fig. 6), but not for brain cell type (Supplementary Table 18 and Extended Data Fig. 7). Enrichment was seen in enhancer and promoter regions, identified by the presence of H3K4me1 and H3K4me3 chromatin marks, respectively, in multiple central nervous system (CNS) tissues (Supplementary Tables 19 and 20 and Extended Data Figs. 8 and 9). Reading, an offshoot of spoken language, is a uniquely human

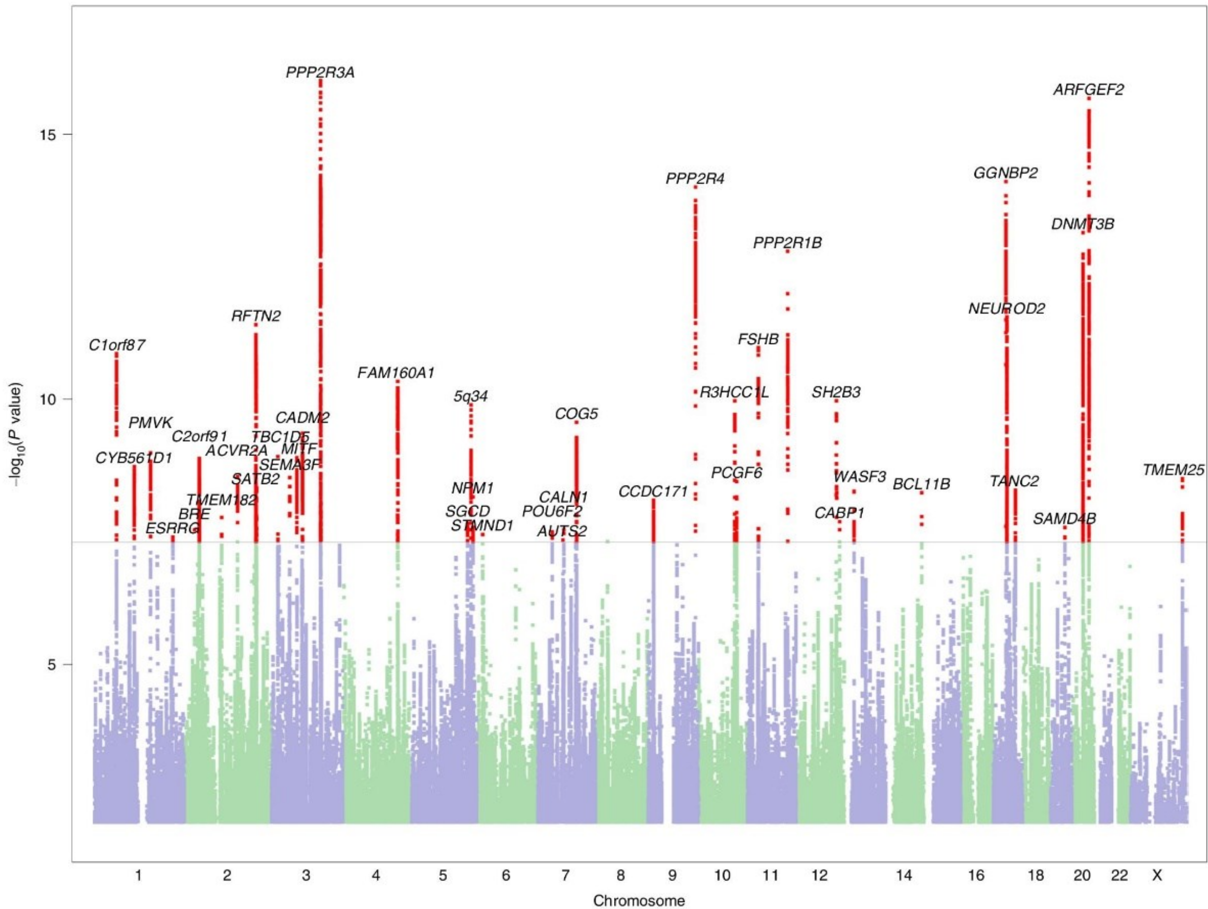


Fig. 1 | Manhattan plot of the genome-wide association analysis of dyslexia. The *y* axis represents the $-\log_{10} P$ value for association of SNPs with self-reported dyslexia diagnosis from 51,800 individuals and 1,087,070 controls. The threshold for genome-wide significance ($P < 5 \times 10^{-8}$) is represented by a horizontal grey line. Genome-wide significant variants in the 42 genome-wide significant loci are red. Variants located within a distance of <250 kb of each other are considered as one locus.

trait, but there was no enrichment for a range of annotations related to human evolution spanning the last 30 million to 50,000 years[18] (Supplementary Table 21).

Genetic correlations and LDSC

Genetic correlations were estimated for 98 traits (Fig. 2 and Supplementary Table 22), including reading and spelling measures, from GenLang (Fig. 3), and brain subcortical structure volumes, total cortical surface area and thickness from the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) consortium. A total of 63 traits showed genetic correlations with dyslexia at the Bonferroni-corrected significance threshold ($P < 0.05/98$; Fig. 2). Genetic correlations (r_g) with quantitative reading and spelling measures ranged from −0.70 to −0.75 (lowest 95% CI of −0.60, highest 95% CI of −0.86), and were −0.62 (95% CI: −0.50, −0.74) and −0.45 (95% CI: −0.26, −0.64) with phoneme awareness and nonword repetition measures, respectively. The childhood/adolescent performance (nonverbal) intelligence quotient (IQ) r_g was lower (−0.19; 95% CI: −0.08, −0.30) than that for adult verbal-numerical reasoning[19] (−0.50; 95% CI: −0.45, −0.55) but similar to that for childhood IQ[20] (−0.32; 95% CIs: −0.21, −0.43) and educational attainment[21] (−0.22; 95% CI: −0.15, −0.29). Traits showing positive r_g included jobs involving heavy manual work[21] (0.40; (95% CI: 0.34, 0.45)), work-related/vocational qualifications[21] (0.50; 95% CI: 0.41, 0.59), ADHD[22] (0.53; 95% CI: 0.29, 0.77), equal use of right and left hands[21] (0.38; 95% CI: 0.19, 0.57) and pain measures[21] (average = 0.31; 95% CI: 0.21, 0.41). Of the 11 ENIGMA measures tested, only intracranial volume was significantly correlated with dyslexia ($r_g = -0.14$; 95% CI: −0.06, −0.22). Targeted investigation of 80 structural neuroimaging measures from UK Biobank, including surface-based morphometry and diffusion-weighted imaging for brain circuitry linked to language, were nonsignificant at a Bonferroni-corrected significance level for number of independent traits. Phenotype independence was estimated by spectral decomposition of the phenotypic correlation matrix implied by the bivariate LDSC intercept from GWAS summary statistics of these traits, using the PhenoSpD toolkit[23] (Supplementary Table 23).

Polygenic score analyses

Dyslexia polygenic scores (PGS) based on the 23andMe dyslexia GWAS were computed in four independent cohorts and, overall, higher PGS were associated with lower reading and spelling accuracy (Supplementary Table 24). In two Australian population-based samples (1,647 adolescents, 1,163 adults), the dyslexia PGS explained up to 3.6% of variance in the reading and spelling measures, being most predictive of lower performance on tests of nonword reading, an index of phonological decoding. Dyslexia PGS did not correlate with scores on tests of nonword repetition (considered a marker of phonological short-term

Table 1 | New SNP associations with dyslexia, including gene-based results, eQTL status, expression in brain and validation in three independent cohorts (GenLang Consortium, CRS and NeuroDys)

Cytoband	SNP	Effect allele	Frequency	Odds Ratio	GWAS *P*	Gene(s)	Most probable gene	Validation cohort (*P* uncorrected for multiple testing)
chr1q21.3	rs4845687	A	0.56	1.044	1.1×10^{-9}	*KCNN3, PMVK*	*PMVK*[ab]	GenLang (0.02)
chr2q22.3	rs497418	A	0.38	1.043	3.0×10^{-9}	*ACVR2A*	AC062032.1[c]	GenLang (0.009)
chr2q33.1	rs72916919	G	0.51	1.049	4.1×10^{-12}	***RFTN2***	*MARS2*[a]	NeuroDys (0.02), GenLang (0.02)
chr3p12.1	rs10511073	A	0.37	1.046	4.6×10^{-10}	***CADM2***	***CADM2***[a]	GenLang (0.02)
chr3q22.3	rs13082684	A	0.24	1.069	1.0×10^{-16}	***PPP2R3A***	***PPP2R3A*** (intron)[a]	GenLang (0.0004); not in CRS
chr6p22.3	rs2876430	T	0.34	1.041	3.7×10^{-8}	*ATXN1, STMND1*	*STMND1*	GenLang (0.04)
chr7p14.1	rs62453457	G	0.48	1.039	3.3×10^{-8}	***POU6F2***	***POU6F2***	CRS (0.04)
chr7q11.22	rs3735260	G	0.08	1.075	4.7×10^{-8}	***AUTS2***	***AUTS2***	GenLang (0.02)
chr7q11.22	rs77059784	G	0.97	1.123	3.0×10^{-8}	***CALN1***	***CALN1***	GenLang (0.02); not in CRS
chr9q34.11	rs9696811	C	0.69	1.069	1.1×10^{-16}	***PPP2R3A***	AL158151.4[abc]	GenLang (0.03)
chr11q23.1	rs138127836	A	0.65	1.056	1.7×10^{-13}	***PPP2R1B***	***PPP2R1B*** (intron)[ab]	GenLang (0.02)
chr17q23.3	rs72841395[c]	C	0.77	1.049	5.4×10^{-9}	***TANC2***	***TANC2***[a]	GenLang (0.005)
chrXq27.3	rs5904158	GTA	0.65	1.037	3.3×10^{-8}	*TMEM257, CXorf51B*[b]	AL109653.3[c]	GenLang (0.02); not in NeuroDys/CRS
chr2q12.1	rs367982014	CAAT	0.29	1.045	1.8×10^{-8}	***TMEM182***	***MFSD9***[a]	Not available
chr3p24.3	rs373178590	G	0.51	1.046	1.3×10^{-9}	***TBC1D5***	***TBC1D5*** (intron)[a]	Not available
chr10q24.33	rs34732054	C	0.57	1.045	3.7×10^{-9}	***PCGF6***	***USMG5***[a]	Not available
chr13q12.13	rs375018025	CA	0.57	1.044	5.6×10^{-9}	*CDK8*, ***WASF3***	***WASF3***	Not available
chr1p32.1	rs12737449	G	0.85	1.070	1.4×10^{-11}	***C1orf87***	***C1orf87*** (missense)[a]	Not significant
chr2p23.2	rs1969131	T	0.17	1.053	3.0×10^{-8}	***BABAM2***	***BABAM2***	Not significant
chr3q26.33	rs7625418	C	0.21	1.056	4.3×10^{-9}	*PEX5L, TTC14*	*TTC14*[a]	Not significant
chr3p13	rs13097431	G	0.58	1.044	1.3×10^{-9}	***MITF***	***MITF***[a]	Not significant
chr5q33.3	rs867009	G	0.36	1.041	2.3×10^{-9}	***SGCD***	***SGCD***[a]	Not significant
chr9p22.3	rs3122702	T	0.5	1.041	8.3×10^{-9}	***CCDC171***	***CCDC171***[ab]	Not significant
chr10q24.2	rs10786387	C	0.68	1.049	1.1×10^{-10}	***CRTAC1, R3HCC1L***	***R3HCC1L***[a]	Not significant
chr11p14.1	rs676217	G	0.37	1.050	1.1×10^{-11}	*KCNA4*, ***FSHB***	*ARL14EP*[ab]	Not significant
chr19q13.2	rs60963584	A	0.89	1.065	2.7×10^{-8}	***GMFG, SAMD4B***	***SAMD4B***[a]	Not significant
chr20q11.21	rs4911257	C	0.39	1.055	7.5×10^{-14}	***DNMT3B***	***DNMT3B*** (intron)[ab]	Not significant

Statistics for each variant are from the 23andMe GWAS (see Supplementary Table 1 for all 42 significant variants). Genes that are significant in gene-based tests are set in bold. Multi-allelic effect alleles represent insertions. The most probable gene is that most likely to be causal based on genetic and functional genomic data tied to the tag SNP (https://platform.opentargets.org/). [a]eQTL. [b]eQTL linked to brain expression. [c]Not available in gene-based results.

memory). In developmental cohorts enriched for reading difficulties, the dyslexia PGS explained 3.7% (UKdys; $n = 930$) and 5.6% (CLDRC; $n = 717$) of variance in word recognition tests.

Analyses of dyslexia associations from the literature

Of 75 previously reported dyslexia associations, none showed genome-wide significance in our analyses (Supplementary Table 25). Of these targeted variants, 19 (in *ATP2C2, CMIP, CNTNAP2, DCDC2, DIP2A, DYX1C1, FOXP2, KIAA0319L* and *PCNT*) showed association surviving Bonferroni correction that accounted for LD ($P < 0.05/68.7$). In gene-based tests of 14 candidate genes from the literature[24,25], association at a Bonferroni level ($P < 0.05/14$) was seen for *KIAA0319L* ($P = 1.84 \times 10^{-4}$) and *ROBO1* ($P = 1.53 \times 10^{-3}$) (Supplementary Table 26). The *CNTNAP2* association approached corrected replication-level significance ($P = 0.004$). Targeted gene set analysis of three pathways previously implicated in dyslexia (Supplementary Table 27) showed replication-level support ($P = 2.00 \times 10^{-3}$) for the axon guidance pathway (comprising 216 genes).

Discussion

In the largest GWAS of dyslexia to date (>50,000 self-reported diagnoses), we identified 42 significant independent loci. Of these,

 https://doi.org/10.1038/s41588-022-01192-y

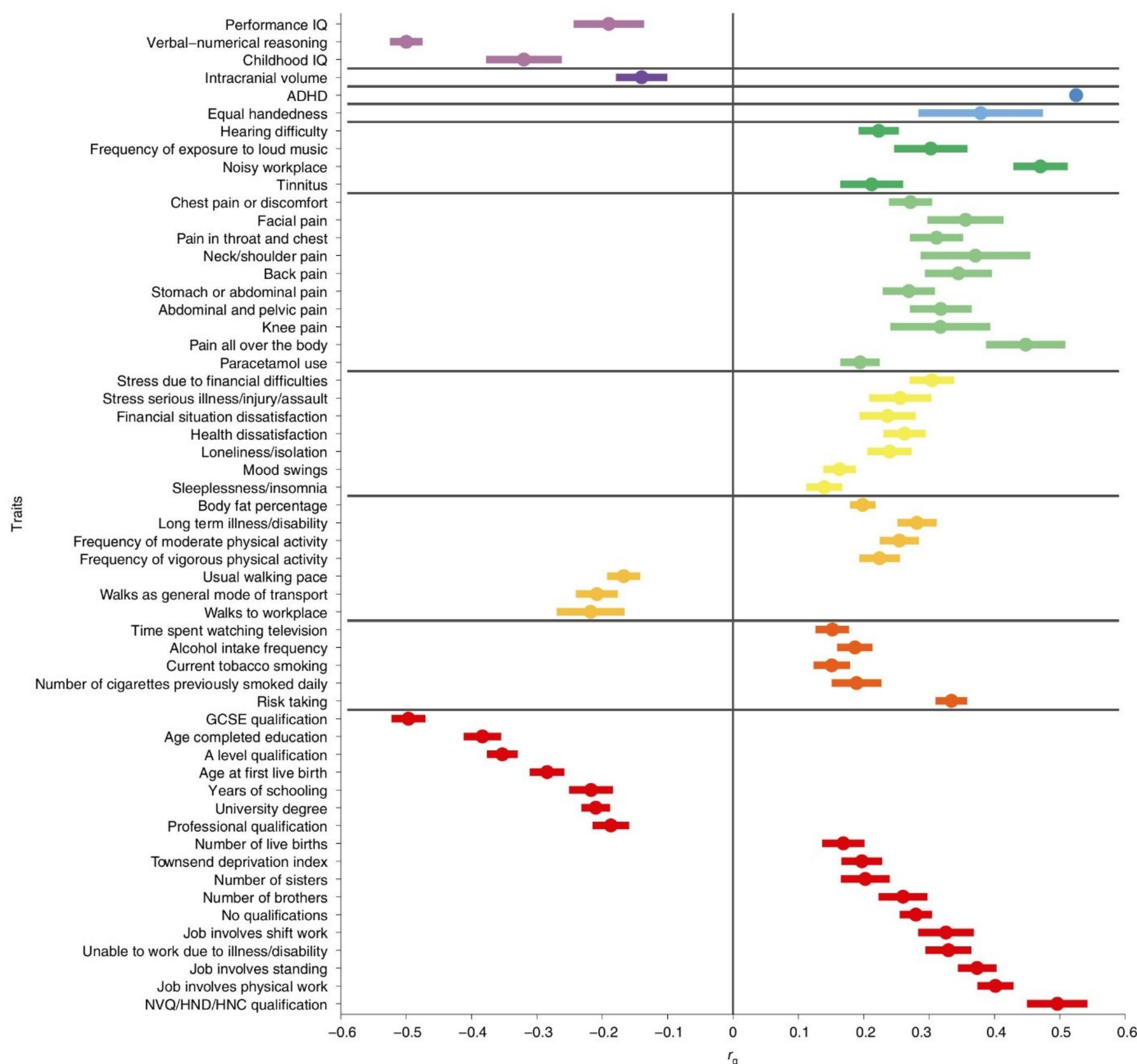


Fig. 2 | Genetic correlations of dyslexia with other phenotypes. Significant ($P < 5 \times 10^{-4}$) genetic correlations (r_g) between self-reported dyslexia diagnosis from 23andMe and other phenotypes from the LD Hub database and Enhancing Neuro Imaging Genetics Through Meta-Analysis (ENIGMA). We tested 98 traits but present only those that were significant after Bonferroni correction. Center points represent genetic correlations, and error bars represent standard errors around the estimate; exact values can be found in Supplementary Table 22. The vertical line indicates a genetic correlation of zero, and the horizontal lines divide groups of related traits. GCSE, General Certificate of Secondary Education; HNC, Higher National Certificate; HND, Higher National Diploma; NVQ, National Vocational Qualification.

27 represent new associations that have not been uncovered in GWAS of related cognitive traits; 12 of the new associations were validated in the GenLang consortium GWAS meta-analysis of reading/spelling in English and other European languages[14], and 1 in a Chinese language cohort. Of the significant SNPs, 36% overlapped with variants from general cognitive ability GWAS, consistent with twin studies that find that genetic variation in reading disability is explained by general and reading-specific cognitive ability[10]. Similar to other complex traits, and consistent with high polygenicity, each significant locus showed small effects (odds ratios (ORs) ranging from 1.04 to 1.12). Our estimated SNP-based heritability of 19% (assuming a 10% dyslexia population prevalence) was equal to that reported in a smaller GWAS[12], but lower than heritability estimates from twin studies (40–80%)[26,27]. This difference may be due partly to effects of rare and structural variants[28], which have been implicated in reading and related traits[29,30].

Whereas *AUTS2* has been implicated in autism[31], intellectual disability[32] and dyslexia[33], the variant we uncovered (rs3735260) represents the strongest *AUTS2* SNP association with a neurodevelopmental trait to date. Amongst our findings were other known neurodevelopmental genes, such as *TANC2* (implicated in language delay and intellectual disability[34,35]) and, especially, *GGNBP2* (linked to neurodevelopmental delay[36] and autism[37]) with variant rs34349354 supported in all our validation cohorts. However, rs34349354 is also associated with cognitive performance[38], and based on expression quantitative trait loci (eQTL) evidence is more likely linked to *ZNHIT3*, colocalizing with molecular QTLs (opentargets.org). Notably, none of the more established

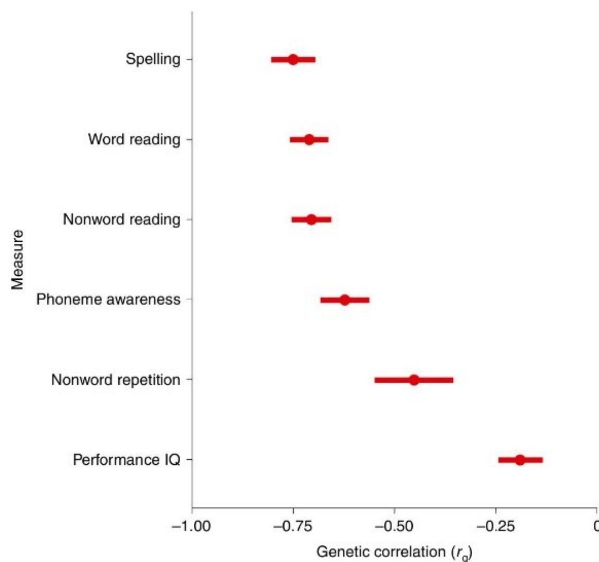


Fig. 3 | Genetic correlations between dyslexia and measures of reading, language and nonverbal IQ. Genetic correlations (r_g) between self-reported dyslexia diagnosis from 23andMe and measures of reading, language and performance (nonverbal) IQ in the GenLang consortium. Center points represent genetic correlations estimated in LDSC, and error bars represent standard errors around the estimate; exact values can be found in Supplementary Table 22.

candidate genes for dyslexia approached genome-wide significance in our results.

Like other human complex traits, partitioning of SNP-based heritability revealed enrichment in conserved regions[39]. We further observed enrichment in the histone mark H3K4me1 (which has also been reported for ASD[40]), and at H3K4me1 and H3K4me3 clusters in the CNS (marking enhancers and promoters, respectively). Since reading/writing systems are built on our capacities for spoken language, it is plausible that evolutionary changes on the human lineage helped shape the underlying genetic architecture[41]. However, we did not find enrichment of significant associations for curated annotations spanning different periods of hominin prehistory.

Our self-reported dyslexia diagnosis binary trait showed strong negative genetic correlations with quantitative reading and spelling measures, supporting the validity of this measure in the 23andMe cohort, and suggesting that reading skills and disorder are not qualitatively distinct. The positive genetic correlation between hearing difficulties and dyslexia is consistent with genetic correlations reported for childhood reading skill[42], suggesting that hearing problems at an early age could affect acquisition of phonological processing skills.

Dyslexia showed moderately negative genetic correlations with adult verbal-numerical reasoning, but there was a lack of a strong genetic correlation of dyslexia with (nonverbal) performance IQ. This would be consistent with phenotypic observations that individuals with dyslexia are disadvantaged on verbal IQ tests[43]. Educational attainment correlations were also not strong, which might reflect school adjustments and other support that counteract disadvantage in academic learning.

There was little evidence of common genetic variation in dyslexia being related to interindividual differences in subcortical volumes, or structural connectivity and morphometry for brain regions implicated in language processing in adults. Thus, the phenotypic correlations previously reported between dyslexia and aspects of neuroanatomy may in large part reflect environmental shaping of the brain, perhaps through the process of reading itself[44]. Left-handedness and ambidexterity show small genetic overlap with each other[45] yet are both phenotypically linked to neurodevelopmental disorders/cognitive abilities[46,47]. We report a significant genetic correlation between dyslexia and self-reported equal hand use, but not left-handedness, supporting theories linking ambidexterity and dyslexia[48].

Dyslexia and ADHD[5,6] often co-occur (24% reporting ADHD in our cases versus 9% in controls), and we show a moderate genetic correlation between the two, potentially reflecting shared endophenotypes like deficits in working memory and attention[49]. Although we did not find significant genetic correlations between dyslexia and ASD, the GWAS for the latter encompassed diverse neurodevelopmental phenotypes, including subgroups with varying educational attainment and IQ[40]. Genetic correlations with pain-related traits suggest that individuals with dyslexia may have a lower threshold for pain perception. Links between pain and other neurodevelopmental disorders have been reported[50,51].

Dyslexia polygenic scores were correlated with lower achievement on reading and spelling tests in population-based and reading-disorder enriched samples, especially for nonword reading, a measure of phonological decoding that is typically impaired in dyslexia. Polygenic scores could become a valuable tool to help identify children with a propensity for dyslexia, enabling learning support before development of reading skills. However, a limitation of our study is the potential for collider bias arising from sample selection (that is, people without dyslexia and from higher socioeconomic positions), which we were unable to quantify; thus, care should be taken in future research when using polygenic scores based on many variants[52].

In summary, we report 42 new independent genome-wide significant loci associated with dyslexia, 27 of which have not been associated with cognitive-educational traits and should be prioritized for follow up as dyslexia candidates. Functional annotation of the variants highlights the importance of conserved and enhancer regions of the genome for this trait. Dyslexia shows positive genetic correlations with ADHD, vocational qualifications, physical occupations, ambidexterity and pain perception, and negative correlations with academic qualifications and cognitive ability; family-based methods are needed to dissociate pleiotropic and causal effects.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-022-01192-y.

References

1. Ritchie, S. J. & Bates, T. C. Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychol. Sci.* **24**, 1301–1308 (2013).
2. Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M. & Escobar, M. D. Prevalence of reading disability in boys and girls: results of the Connecticut Longitudinal Study. *JAMA* **264**, 998–1002 (1990).
3. Katusic, S. K., Colligan, R. C., Barbaresi, W. J., Schaid, D. J. & Jacobsen, S. J. Incidence of reading disability in a population-based birth cohort, 1976–1982, Rochester, Minn. *Mayo Clin. Proc.* **76**, 1081–1092 (2001).
4. Carroll, J. M., Maughan, B., Goodman, R. & Meltzer, H. Literacy difficulties and psychiatric disorders: evidence for comorbidity. *J. Child Psychol. Psychiatry* **46**, 524–532 (2005).
5. Margari, L. et al. Neuropsychopathological comorbidities in learning disorders. *BMC Neurol.* **13**, 198 (2013).
6. Willcutt, E. G., Pennington, B. F. & DeFries, J. C. Twin study of the etiology of comorbidity between reading disability and attention-deficit/hyperactivity disorder. *Am. J. Med. Genet.* **96**, 293–301 (2000).

7. McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M. & Mengler, E. D. On the 'specifics' of specific reading disability and specific language impairment. *J. Child Psychol. Psychiatry* **41**, 869–874 (2000).
8. Catts, H. W., Fey, M. E., Tomblin, J. B. & Zhang, X. A longitudinal investigation of reading outcomes in children with language impairments. *J. Speech Lang. Hear. Res.* **45**, 1142–1157 (2002).
9. Bates, T. C. et al. Genetic and environmental bases of reading and spelling: a unified genetic dual route model. *Read. Writ.* **20**, 147–171 (2007).
10. Haworth, C. M. A. et al. Generalist genes and learning disabilities: a multivariate genetic analysis of low performance in reading, mathematics, language and general cognitive ability in a sample of 8000 12-year-old twins. *J. Child Psychol. Psychiatry* **50**, 1318–1325 (2009).
11. Fisher, S. E. & DeFries, J. C. Developmental dyslexia: genetic dissection of a complex cognitive trait. *Nat. Rev. Neurosci.* **3**, 767–780 (2002).
12. Gialluisi, A. et al. Genome-wide association study reveals new insights into the heritability and genetic correlates of developmental dyslexia. *Mol. Psychiatry* **26**, 3004–3017 (2021).
13. Buniello, A. et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2018).
14. Eising, E. et al. Genome-wide analyses of individual differences in quantitatively assessed reading- and language-related skills in up to 34,000 people. *Proc. Natl Acad. Sci. USA* **119**, e2202764119 (2022).
15. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
16. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
17. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
18. Tilot, A. K. et al. The evolutionary history of common genetic variants influencing human cortical surface area. *Cerebral Cortex* **31**, 1873–1887 (2020).
19. Sniekers, S. et al. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat. Genet.* **49**, 1107–1112 (2017).
20. Benyamin, B. et al. Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Mol. Psychiatry* **19**, 253–258 (2014).
21. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
22. Middeldorp, C. M. et al. A genome-wide association meta-analysis of attention-deficit/hyperactivity disorder symptoms in population-based pediatric cohorts. *J. Am. Acad. Child Adolesc. Psychiatry* **55**, 896–905.e6 (2016).
23. Zheng, J. et al. PhenoSpD: an integrated toolkit for phenotypic correlation estimation and multiple testing correction using GWAS summary statistics. *Gigascience* **7**, giy090 (2018).
24. Luciano, M., Gow, A. J., Pattie, A., Bates, T. C. & Deary, I. J. The influence of dyslexia candidate genes on reading skill in old age. *Behav. Genet.* **48**, 351–360 (2018).
25. Doust, C. et al. The association of dyslexia and developmental speech and language disorder candidate genes with reading and language abilities in adults. *Twin Res. Hum. Genet.* **23**, 23–32 (2020).
26. Davis, C. J., Knopik, V. S., Olson, R. K., Wadsworth, S. J. & DeFries, J. C. Genetics and environmental influences on rapid naming and reading ability. *Ann. Dyslexia* **51**, 231–247 (2001).
27. Gayán, J. & Olson, R. K. Genetic and environmental influences on orthographic and phonological skills in children with reading disabilities. *Dev. Neuropsychol.* **20**, 483–507 (2001).
28. Hannula-Jouppi, K. et al. The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS Genet.* **1**, e50 (2005).
29. Ganna, A. et al. Ultra-rare disruptive and damaging mutations influence educational attainment in the general population. *Nat. Neurosci.* **19**, 1563–1565 (2016).
30. Gialluisi, A. et al. Investigating the effects of copy number variants on reading and language performance. *J. Neurodev. Disord.* **8**, 17–17 (2016).
31. Oksenberg, N., Stevison, L., Wall, J. D. & Ahituv, N. Function and regulation of AUTS2, a gene implicated in autism and human evolution. *PLoS Genet.* **9**, e1003221 (2013).
32. Beunders, G. et al. Two male adults with pathogenic AUTS2 variants, including a two-base pair deletion, further delineate the AUTS2 syndrome. *Eur. J. Human Genet.* **23**, 803–807 (2015).
33. Girirajan, S. et al. Relative burden of large CNVs on a range of neurodevelopmental phenotypes. *PLoS Genet.* **7**, e1002334 (2011).
34. Wessel, K. et al. 17q23.2q23.3 de novo duplication in association with speech and language disorder, learning difficulties, incoordination, motor skill impairment, and behavioral disturbances: a case report. *BMC Med. Genet.* **18**, 119 (2017).
35. Guo, H. et al. Disruptive mutations in TANC2 define a neurodevelopmental syndrome associated with psychiatric disorders. *Nat. Commun.* **10**, 4679 (2019).
36. Pasmant, E. et al. Characterization of a 7.6-Mb germline deletion encompassing the NF1 locus and about a hundred genes in an NF1 contiguous gene syndrome patient. *Eur. J. Hum. Genet.* **16**, 1459–1466 (2008).
37. Takata, A. et al. Integrative analyses of de novo mutations provide deeper biological insights into autism spectrum disorder. *Cell Reports* **22**, 734–747 (2018).
38. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
39. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
40. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
41. Mozzi, A. et al. The evolutionary history of genes involved in spoken and written language: beyond FOXP2. *Sci. Rep.* **6**, 22157 (2016).
42. Schmitz, J., Abbondanza, F. & Paracchini, S. Genome-wide association study and polygenic risk score analysis for hearing measures in children. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **186**, 318–328 (2021).
43. Vellutino, F. Alternative conceptualizations of dyslexia: evidence in support of a verbal-deficit hypothesis. *Harvard Educ. Rev.* **47**, 334–354 (2012).
44. Dehaene, S., Cohen, L., Morais, J. & Kolinsky, R. Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nat. Rev. Neurosci.* **16**, 234–244 (2015).
45. Cuellar-Partida, G. et al. Genome-wide association study identifies 48 common genetic variants associated with handedness. *Nat. Hum. Behav.* **5**, 59–70 (2021).
46. Papadatou-Pastou, M. et al. Human handedness: a meta-analysis. *Psychol. Bull.* **146**, 481–524 (2020).
47. Peters, M., Reimers, S. & Manning, J. T. Hand preference for writing and associations with selected demographic and behavioral variables in 255,100 subjects: the BBC internet study. *Brain Cogn.* **62**, 177–189 (2006).

48. Brandler, W. M. & Paracchini, S. The genetic relationship between handedness and neurodevelopmental disorders. *Trends Mol. Med.* **20**, 83–90 (2014).
49. Willcutt, E. G., Pennington, B. F., Olson, R. K., Chhabildas, N. & Hulslander, J. Neuropsychological analyses of comorbidity between reading disability and attention deficit hyperactivity disorder: in search of the common deficit. *Dev. Neuropsychol.* **27**, 35–78 (2005).
50. Gu, X. et al. Heightened brain response to pain anticipation in high-functioning adults with autism spectrum disorder. *Eur. J. Neurosci.* **47**, 592–601 (2018).
51. Whitney, D. G. & Shapiro, D. N. National prevalence of pain among children and adolescents with autism spectrum disorders. *JAMA Pediatr.* **173**, 1203–1205 (2019).
52. Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M. & Davey Smith, G. Collider scope: when selection bias can substantially influence observed associations. *Int. J. Epidemiol.* **47**, 226–235 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022, corrected publication 2023

23andMe Research Team

Stella Aslibekyan[2], Adam Auton[2], Elizabeth Babalola[2], Robert K. Bell[2], Jessica Bielenberg[2], Katarzyna Bryc[2], Emily Bullis[2], Daniella Coker[2], Gabriel Cuellar Partida[2], Devika Dhamija[2], Sayantan Das[2], Sarah L. Elson[2], Teresa Filshtein[2], Kipper Fletez-Brant[2], Pierre Fontanillas[2], Will Freyman[2], Pooja M. Gandhi[2], Karl Heilbron[2], Barry Hicks[2], David A. Hinds[2], Ethan M. Jewett[2], Yunxuan Jiang[2], Katelyn Kukar[2], Keng-Han Lin[2], Maya Lowe[2], Jey McCreight[2], Matthew H. McIntyre[2], Steven J. Micheletti[2], Meghan E. Moreno[2], Joanna L. Mountain[2], Priyanka Nandakumar[2], Elizabeth S. Noblin[2], Jared O'Connell[2], Aaron A. Petrakovitz[2], G. David Poznik[2], Morgan Schumacher[2], Anjali J. Shastri[2], Janie F. Shelton[2], Jingchunzi Shi[2], Suyash Shringarpure[2], Vinh Tran[2], Joyce Y. Tung[2], Xin Wang[2], Wei Wang[2], Catherine H. Weldon[2], Peter Wilton[2], Alejandro Hernandez[2], Corinna Wong[2] and Christophe Toukam Tchakouté[2]

Quantitative Trait Working Group of the GenLang Consortium

Filippo Abbondanza[9], Andrea G. Allegrini[19], Till F. M. Andlauer[20,21], Cathy L. Barr[22,23,24], Timothy C. Bates[1], Manon Bernard[25], Kirsten Blokland[23], Milene Bonte[26], Dorret I. Boomsma[27,28,29], Thomas Bourgeron[30,31], Daniel Brandeis[32,33,34,35], Manuel Carreiras[36,37,38], Fabiola Ceroni[39,40], Valéria Csépe[41,42], Philip S. Dale[43], John C. DeFries[14,15], Peter F. de Jong[44], Jean Francois Démonet[45], Eveline L. de Zeeuw[28], Else Eising[3], Yu Feng[46], Simon E. Fisher[3,6], Marie-Christine J. Franken[47], Clyde Francks[3,6], Margot Gerritse[3], Alessandro Gialluisi[20,48], Scott D. Gordon[4], Jeffrey R. Gruen[13], Sharon L. Guger[49], Marianna E. Hayiou-Thomas[50], Juan Hernández-Cabrera[51], Jouke-Jan Hottenga[28], Charles Hulme[52], Philip R. Jansen[53,54,55], Juha Kere[56,57], Elizabeth N. Kerr[49,58,59], Tanner Koomar[60], Karin Landerl[61,62], Gabriel T. Leonard[63], Zhijie Liao[64], Maureen W. Lovett[23,59], Michelle Luciano[1], Heikki Lyytinen[65], Nicholas G. Martin[4], Angela Martinelli[9], Urs Maurer[66], Jacob J. Michaelson[60], Nazanin Mirza-Schreiber[67], Kristina Moll[68], Anthony P. Monaco[11], Angela T. Morgan[69,70,71], Bertram Müller-Myhsok[20,72], Dianne F. Newbury[40], Markus M. Nöthen[73], Richard K. Olson[14,15], Silvia Paracchini[9], Tomas Paus[74,75,76], Zdenka Pausova[25,77], Craig E. Pennell[78,79,80], Bruce F. Pennington[16], Robert J. Plomin[19], Kaitlyn M. Price[23,24,46], Franck Ramus[81], Sheena Reilly[69,82], Louis Richer[83], Kaili Rimfeld[19], Gerd Schulte-Körne[68], Chin Yang Shapland[7,84], Nuala H. Simpson[85], Shelley D. Smith[17], Margaret J. Snowling[85,86], Beate St Pourcain[3,6,7], John F. Stein[87], Lisa J. Strug[88,89], Joel B. Talcott[10], Henning Tiemeier[53,90], J. Bruce Tomblin[91], Dongnhu T. Truong[13], Elsje van Bergen[27,28,92], Marc P. van der Schroeff[93,94], Marjolein Van Donkelaar[3], Ellen Verhoef[3], Carol A. Wang[78,79], Kate E. Watkins[85], Andrew J. O. Whitehouse[95], Karen G. Wigg[46], Erik G. Willcutt[14,15], Margaret Wilkinson[23], Margaret J. Wright[18] and Gu Zhu[4]

[19]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. [20]Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany. [21]Department of Neurology, Klinikum rechts der Isar, School of Medicine,Technical University of Munich, Munich, Germany. [22]Division of Experimental and Translational Neuroscience, Krembil Research Institute, University Health Network, Toronto, Ontario, Canada. [23]Program in Neuroscience and Mental Health, Hospital for Sick Children, Toronto, Ontario, Canada. [24]Department of Physiology, University of Toronto, Toronto, Ontario, Canada. [25]Departments of Physiology and Nutritional Sciences, Hospital for Sick Children, Toronto, Ontario, Canada. [26]Department of Cognitive Neuroscience and Maastricht Brain Imaging Center, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, the Netherlands. [27]Netherlands Twin Register, Amsterdam, the Netherlands. [28]Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [29]Amsterdam Reproduction and Development (AR&D) Research Institute, Amsterdam, the Netherlands. [30]Human Genetics and Cognitive Functions Unit, Institut Pasteur, Paris, France. [31]CNRS UMR 3571, Université de Paris, Paris, France. [32]Department of Child and Adolescent Psychiatry and Psychotherapy, Psychiatric Hospital, University of Zurich, Zurich, Switzerland. [33]Zurich Center for Integrative Human Physiology (ZIHP), University of Zurich and ETH Zurich, Zurich, Switzerland. [34]Neuroscience Center Zurich, University of Zurich and ETH Zurich, Zurich, Switzerland. [35]Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [36]Basque Center on Cognition, Brain and Language (BCBL), Donostia-San Sebastian, Spain. [37]Ikerbasque, Basque Foundation for Science, Bilbao, Spain. [38]Lengua Vasca y Comunicación, University of the Basque Country (UPV/EHU),

Bilbao, Spain. [39]Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy. [40]Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, UK. [41]Brain Imaging Centre, Research Centre for Natural Sciences, Budapest, Hungary. [42]Multilingualism Doctoral School, Faculty of Modern Philology and Social Sciences, University of Pannonia, Veszprém, Hungary. [43]Department of Speech and Hearing Sciences, University of New Mexico, Albuquerque, NM, USA. [44]Department of Child Development and Education, University of Amsterdam, Amsterdam, the Netherlands. [45]Leenaards Memory Centre, Department of Clinical Neurosciences Lausanne University Hospital (CHUV), University of Lausanne, Lausanne, Switzerland. [46]Genetics and Development Division, Krembil Research Institute, University Health Network, Toronto, Ontario, Canada. [47]Department of Otorhinolaryngology, Erasmus University Medical Centre, Rotterdam, the Netherlands. [48]Department of Epidemiology and Prevention, IRCCS Istituto Neurologico Mediterraneo Neuromed, Pozzilli, Italy. [49]Department of Psychology, Hospital for Sick Children, Toronto, Ontario, Canada. [50]Department of Psychology, University of York, York, UK. [51]Departamento de Psicología Clínica Psicobiología y Metodología, Universidad de La Laguna, Santa Cruz de Tenerife, Spain. [52]Department of Education, University of Oxford, Oxford, UK. [53]Department of Child and Adolescent Psychiatry/Psychology, Erasmus University Medical Center, Rotterdam, the Netherlands. [54]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University, Amsterdam, the Netherlands. [55]Department of Human Genetics, VU Medical Center, Amsterdam UMC, Amsterdam, the Netherlands. [56]Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden. [57]Stem Cells and Metabolism Research Program, University of Helsinki, and Folkhälsan Research Center, Helsinki, Finland. [58]Department of Neurology, Hospital for Sick Children, Toronto, Ontario, Canada. [59]Department of Paediatrics, The University of Toronto, Toronto, Ontario, Canada. [60]Department of Psychiatry, University of Iowa, Iowa City, IA, USA. [61]Institute of Psychology, University of Graz, Graz, Austria. [62]BioTechMed-Graz, Graz, Austria. [63]Cognitive Neuroscience Neurology and Neurosurgery, Montreal, Quebec, Canada. [64]Department of Psychology, University of Toronto, Toronto, Ontario, Canada. [65]Department of Psychology, University of Jyväskylä, Jyväskylä, Finland. [66]Department of Psychology, The Chinese University of Hong Kong, Hong Kong, China. [67]Institute of Neurogenomics, Helmholtz Zentrum München, Munich, Germany. [68]Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, LMU University Hospital Munich, Munich, Germany. [69]Speech and Language, Murdoch Children's Research Institute, Melbourne, Victoria, Australia. [70]Department of Audiology and Speech Pathology, University of Melbourne, Melbourne, Victoria, Australia. [71]Speech Pathology Department, Royal Children's Hospital, Melbourne, Victoria, Australia. [72]Department of Health Science, University of Liverpool, Liverpool, UK. [73]Institute of Human Genetics, University Hospital of Bonn, Bonn, Germany. [74]Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada. [75]Departments of Psychiatry and Neuroscience and Centre Hospitalier Universitaire Sainte Justine, University of Montreal, Montreal, Quebec, Canada. [76]Department of Psychology, University of Toronto, Toronto, Ontario, Canada. [77]Hospital for Sick Children, Toronto, Ontario, Canada. [78]School of Medicine and Public Health, College of Health, Medicine and Wellbeing, University of Newcastle, Newcastle, New South Wales, Australia. [79]Mothers and Babies Research Centre, Hunter Medical Research Institute, Newcastle, New South Wales, Australia. [80]Maternity and Gynaecology, John Hunter Hospital, Newcastle, New South Wales, Australia. [81]Laboratoire de Sciences Cognitives et Psycholinguistique, Ecole Normale Supérieure, PSL University, EHESS, CNRS, Paris, France. [82]Menzies Health Institute Queensland, Griffith University, Gold Coast, Queensland, Australia. [83]Department of Health Sciences, Université du Québec à Chicoutimi, Chicoutimi, Quebec, Canada. [84]Population Health Sciences, University of Bristol, Bristol, UK. [85]Department of Experimental Psychology, University of Oxford, Oxford, UK. [86]St John's College, University of Oxford, Oxford, UK. [87]Department of Physiology, Anatomy and Genetics, Oxford University, Oxford, UK. [88]Departments of Statistical Sciences and Computer Science and Division of Biostatistics, University of Toronto, Toronto, Ontario, Canada. [89]Program in Genetics and Genome Biology and The Centre for Applied Genomics, Hospital For Sick Children, Toronto, Ontario, Canada. [90]Harvard T.H. Chan School of Public Health, Boston, MA, USA. [91]Communication Sciences and Disorders, University of Iowa, Iowa City, IA, USA. [92]Research Institute LEARN!, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [93]Department of Otolaryngology, Head and Neck Surgery, Erasmus MC, Rotterdam, the Netherlands. [94]Generation R Study Group, Erasmus MC, Rotterdam, the Netherlands. [95]Telethon Kids Institute, The University of Western Australia, Perth, Western Australia, Australia.

Methods

GWAS participants

Participants were drawn from the customer base of 23andMe, Inc., a consumer genetics company. Participants provided informed consent and participated in the research online, under a protocol approved by the external AAHRPP-accredited IRB, Ethical and Independent Review Services (www.eandireview.com). They included 51,800 (21,513 male, 30,287 female) participants who responded 'yes' to the question 'Have you been diagnosed with dyslexia?' (cases) and 1,087,070 (446,054 male, 641,016 female) participants who responded 'no' (controls). Age ranged from 18 to 110 years, with the prevalence of dyslexia higher for younger participants (5.34% in those aged 20–30 years) than older participants (3.23% in those aged 80–90 years). The negative linear relationship between dyslexia prevalence and participant age was expected given that screening for specific learning difficulties has only become commonplace in more recent decades. Moreover, this aligns with findings from the subsample (4.3%) of participants who reported age of diagnosis: younger participants were diagnosed at an earlier age (for example, 9.7 years (±4.7) for 20- to 30-year-olds) than older participants (for example, 22.4 years (±17.8) for 80- to 90-year-olds). The prevalence of dyslexia in our sample was similar for women (4.51%) and men (4.6%), although the slightly higher prevalence in males in this very large sample was statistically significant ($P < 8.7 \times 10^{-6}$). Such a prevalence lies at the lower end of the range typically reported in the US population[3] and might represent the more severe cases of dyslexia given that a formal diagnosis was required; additionally, people with dyslexia might opt out of survey research that requires reading, further restricting the sample range.

Genotyping and imputation

DNA was extracted from saliva samples and genotyped on one of five genotyping platforms by the National Genetics Institute (NGI). In the present analysis, only participants with European ancestry were included. Details about the genotyping arrays, quality control of samples and ancestry derivation can be found in Fontanillas et al.[53] and the Supplementary Note. Phased genotypes were imputed to a combined reference panel of the 1000 Genomes Phase 3 haplotypes (May 2015) and the UK10K imputation reference panel using Minimac3 (see Das et al.[54]).

Association analysis

Association analysis was performed on genotyped and imputed SNP dosage data using logistic regression and assuming an additive model of allelic effects. For X-chromosome analysis, male genotypes were treated as homozygous diploid. Covariates included age, age squared, gender, the first five ancestry principal components and genotype platform. SNP significance was evaluated by a likelihood ratio test, and genome-wide significance was determined as $P < 5 \times 10^{-8}$ (suggestive significance level as $P < 1 \times 10^{-6}$). Only reliably imputed SNPs ($r^2 > 0.80$) and those with minor allele frequency (MAF) > 0.01 are presented (n = 7,995,923). We define associated regions by first identifying all variants with $P < 5 \times 10^{-8}$, then grouping these variants into regions separated by gaps of at least 250 kb. Index variants are the variants with smallest P value within each associated region. We use the same approach for regions with suggestive associations, but by first identifying all variants with $P < 10^{-5}$. Subsidiary genome-wide association analysis of separate male (n = 21,513 cases, 446,054 controls) and female (n = 30,287 cases, 641,016 controls) groups, and younger (below 55 years; n = 30,763 cases, 582,276 controls) and older (55 and above; n = 21,037 cases, 504,794 controls) groups was performed. The latter was to check whether reliability of diagnosis (assumed to be higher in the younger sample whose recall of diagnosis should be better and who would have been exposed to greater levels of dyslexia screening) affected the GWAS signal.

We also looked to independently validate our genome-wide significant variants within (1) a published GWAS meta-analysis of 2,274 dyslexia cases from nine European countries representing six different languages (NeuroDys) by Gialluisi et al.[55]; (2) a population sample (Chinese Reading Study; CRS) of children measured on quantitative traits of reading accuracy and reading fluency (n = 2,270; described in the Supplementary Note), and; (3) within the GenLang quantitative trait GWAS meta-analysis of word reading (up to n = 33,959) and spelling (up to n = 18,514) skills measured in cohorts of children and adolescents from Europe, the United States and Australia, and representing seven European languages, of which English was the most common[14].

Genomic control

Top SNPs are reported from the more conservative GWAS results adjusted for genomic control (Fig. 1, Extended Data Figs. 1–4, and Supplementary Tables 1, 2, 9 and 10), whereas downstream analyses (including gene-set analysis, enrichment and heritability partitioning, genetic correlations, polygenic prediction, candidate gene replication) are based on GWAS results without genomic control.

Gene-based analyses

The GWAS results were used to calculate gene-based P values for association with dyslexia by performing the gene analysis in MAGMA v.1.08 (ref. [56]) through the FUMA interface[57] using standard settings. In total, 19,039 genes were tested, and P values were judged based on a Bonferroni-corrected significance threshold of $P < 2.63 \times 10^{-6}$. We also performed gene set analyses for association of biological pathways (all available gene ontology (GO) terms and curated gene sets from the Molecular Signatures Database (MsigDB)[58,59]) with dyslexia in MAGMA through the FUMA interface. The total number of pathways tested was 15,486, and P values were judged based on a Bonferroni-corrected significance threshold of $P < 3.23 \times 10^{-6}$.

Biological annotations

Genome-wide significant variants and nearby gene(s) were annotated using external reference data and evaluated for functional or regulatory impact. A 99% credible set of potentially causal variants for SNPs in significant regions was based on approximate Bayes factor (ABFs)[60] assuming a prior variance of 0.1, and using the method of Maller et al.[61] to define these sets. Variant effect prediction of these was done in ENSEMBL (release 104)[62]. For genome-wide significant variants, we considered: gene context (whether a variant is intergenic or located within a specific functional region within a gene locus); deleteriousness (Combined Annotation Dependent Depletion (CADD) score); functionality (RegulomeDB (RDB) category); chromatin state (minimum and common 15-core chromatin state); and SNP-trait associations reported in the NHGRI GWAS Catalog[13].

For each variant, the most probable gene target was identified using the Open Target Genetics portal[63], which draws on evidence from QTL and chromatin interaction experiments, functional predictions and distance from a gene's transcription start site. For genome-wide significant genes, we considered: loss-of-function intolerance (probability of loss-of-function Intolerance (pLI) score); variation intolerance (residual variation intolerance score, RVIS); variation intolerance in noncoding regions (noncoding RVIS, ncRVIS); evolutionary constraint of noncoding regions (noncoding genomic evolutionary rate profiling (ncGERP) score); evolutionary constraint of protein-coding regions (protein-coding genomic evolutionary rate profiling (pcGERP) score); deleteriousness across noncoding regions (noncoding CADD (ncCADD) score); combined functionality of variants in noncoding regions (noncoding genome-wide annotation of variants (ncGWAVA) score); and expression in 12 brain tissues (amygdala, anterior cingulate cortex, caudate basal ganglia, cerebellar hemisphere, cerebellum, cortex, frontal cortex, hippocampus, hypothalamus, nucleus accumbens basal ganglia, putamen basal ganglia and substantia nigra). All annotations were obtained through FUMA[57] except RVIS, ncGERP, pcGERP, ncCADD and ncGWAVA, which were taken from

Petrovski et al.[64]. Details of each annotation including original sources are in the Supplementary Note.

Partitioned heritability

We partitioned SNP heritability of dyslexia using stratified LDSC, as described by Finucane et al.[39], to determine whether SNPs that share the greatest proportion of the heritability are also clustered in specific functional categories in the genome. Overall, we performed 266 different tests, which would give a very conservative Bonferroni-corrected significance level of 1.88×10^{-4}, but because there will be overlap among annotation groups, we also report corrections to significance within different classes of annotation, each of which we now describe. Partitioning was performed for the 24 main functional annotations defined by Finucane et al.[39]. LD scores, regression weights and allele frequencies are from European ancestry samples and were retrieved from https://alkesgroup.broadinstitute.org/LDSCORE. Heritability estimates were considered statistically significant if the P value surpassed an α level of 2.08×10^{-3}, derived by Bonferroni correction based on 24 tests.

We also estimated the enrichment for heritability of dyslexia for tissue-specific annotations, while controlling for the annotations in the baseline model, including gene expression in three brain cell types, gene expression in 12 brain regions, and chromatin marks H3K4me1 and H3K4me3 in multiple tissues (108 and 114, respectively) since these marks are enriched at enhancers[65] and promoters[66], respectively. Enrichment is the proportion of SNP heritability divided by the proportion of SNPs. For the brain cell types, we estimated enrichment for heritability of dyslexia for genes expressed in neurons, astrocytes, and oligodendrocytes using data from Cahoy et al.[67]. Enrichments were considered statistically significant if the P value surpassed an α level of 0.017, derived by Bonferroni correction based on three tests. The gene expression data used to estimate the enrichment of heritability in genes expressed in certain brain regions was from the GTEx database[68], and the Bonferroni-derived α level for enrichment was 4.17×10^{-3} (based on 12 tests). Chromatin annotations include data from the Roadmap Epigenomics consortium[17] and EN-TEx[69,70]. For H3K4me1, the Bonferroni-derived α level for enrichment was 4.63×10^{-4} (based on 108 tests) and, for H3K4me3, the Bonferroni-derived α level for enrichment was 4.39×10^{-4} (based on 114 tests).

Evolutionary annotations. Although reading and writing is a human cultural invention, it builds on fundamental pathways involved in language processing. Therefore, we investigated whether annotations related to human evolution were significantly enriched for heritability of dyslexia by applying an evolutionary analysis pipeline adapted from Tilot et al.[18]. These analyses capture a range of periods in an evolutionary timeframe on the lineage that led to humans, from approximately 30 million years ago to 50,000 years ago.

Enrichment of heritability was estimated in adult brain human gained enhancers (HGEs)[71], fetal brain HGEs[72], ancient selective sweep regions[73], Neanderthal-introgressed SNPs[74] and Neanderthal-depleted regions[75] (see Supplementary Note for a description of each annotation); and controlled for using the baselineLD v.2 model from Gazal et al.[76]. Heritability enrichment in human adult and fetal HGEs were additionally controlled for adult and fetal brain active regulatory elements from the Roadmap Epigenomics resource[17]. Active regulatory elements were defined using chromHMM[16]. Enrichment P values were judged by an α level of 10^{-2}, derived by Bonferroni correction based on five tests.

Genetic correlations

Genetic correlations within the 23andMe GWAS of dyslexia. Genetic correlation between self-reported dyslexia diagnosis in males and females, and between younger (<55 years old) and older (≥55 years old) adults was calculated using LDSC[77,78].

Genetic correlations of dyslexia with other traits. We present the pairwise genetic correlation of dyslexia with 98 traits. Summary statistics for most of these traits are publicly available through LD Hub[77–79]—a centralized database and web interface that automates the LDSC regression analysis pipeline. A selection of brain magnetic resonance imaging measures obtained from the ENIGMA-3 consortium[80–83], and measures of reading and spelling accuracy, and performance IQ from the GenLang Consortium[14] were analyzed locally using LDSC. Word reading accuracy in GenLang was measured by the number of correct words read aloud from a list in a time restricted or unrestricted fashion. Examples of tools that include this measure are Test of Word Reading Efficiency (TOWRE), the British Ability Scales (BAS) and the Wide Range Achievement Test (WRAT). Spelling accuracy in GenLang was measured by the number of words correctly spelled orally or in writing. The words were dictated as single words or in a sentence. Examples of tools that include this measure are the BAS, WRAT and Wechsler Objective Reading Dimensions (WORD). Performance IQ in GenLang was based on subtests of IQ tests that did not depend on verbal cues, as included for example in the BAS and Wechsler Intelligence Scale for Children (WISC). Trait descriptions and summary statistic sources are in Supplementary Table 22. Bonferroni correction for multiple testing derived an adjusted critical P value of 5.1×10^{-4} from 98 independent tests.

Genetic correlations were further estimated in a targeted analysis of structural brain magnetic resonance imaging measures from UK Biobank, which were more comprehensive than those currently available from ENIGMA, along with further advantages such as hemisphere-specific data and greater homogeneity in cohort and scanning procedures. GWAS summary statistics from brain imaging-derived phenotypes for 33,000 participants were downloaded from the Oxford Brain Imaging Genetics Server[84]. Structural brain imaging traits encompassed both diffusion tensor imaging and surface-based morphometric phenotypes[85] where selected tracts or regions of interest had a known link to language. For diffusion tensor imaging, fractional anisotropy values derived from both tract-based-spatial statistics and probabilistic tractography were used for available tracts spanning the extended language network[86]. For surface-based morphometric (cortical volume, surface area and thickness) GWAS, summary statistics for regions of interest derived from the Desikan-Killiany atlas (white surface) were used, again selected for their relevance in language processing, based on previous literature[87–90]. To correct for multiple testing, phenotypic correlations between the UK Biobank imaging indices were derived and analyzed by PhenoSpD[23] to obtain the number of independent variables (36.08) to use for Bonferroni correction (adjusted critical P value of 1.39×10^{-3}).

Polygenic score analyses

Dyslexia polygenic scores were based on increasingly larger numbers of SNPs corresponding to their association P values from the 23andMe GWAS ($P < 5 \times 10^{-8}$, $P < 1 \times 10^{-5}$, $P < 0.001$, $P < 0.01$, $P < 0.05$, $P < 0.1$, $P < 0.5$, 1). They were calculated in four independent cohorts. Two were general population cohorts from Australia: $n = 1,640$ (772 families) adolescents/young adults (Brisbane adolescents)[91]; $n = 1,165$ (966 families) older adults (Brisbane adults)[25]. The other two were family-based samples selected for dyslexia: one from the United Kingdom (UKdys), $n = 930$ (595 families); the other from the United States (Colorado Learning Disabilities Research Center, CLDRC), $n = 717$ (336 families)[92]. In the Australian samples, polygenic scores were calculated on 1000 Genomes Phase 3 (v.20101123) imputed genetic data using PLINK[93]. Only reliably imputed SNPs ($R^2 > 0.80$) and those with a minor allele frequency >0.01 were included, and the default clumping procedure was used where index SNPs formed a clump with other SNPs in LD ($R^2 > 0.1$) and within a 250 kb distance. In the UKdys and CLDRC samples, polygenic scores were calculated on Haplotype Reference Consortium imputed genetic data using PRSice[94], with the same imputation quality and MAF exclusions for the base (23andMe GWAS) sample, and clumping parameters.

Polygenic scores were then used as predictors in linear models of quantitative trait outcomes (Australia: word, nonword (phonetic), irregular word (lexical) reading and spelling tests from an extended version of the Components of Reading Examination[95], and two nonword repetition tests which are sensitive to developmental language disorders—Dollaghan and Campbell[96], Gathercole and Baddeley[97]; UKdys and CLDRC: word recognition). All quantitative traits were pre-adjusted for sex, age and ancestry principal components (10 principal components in UKdys and CLDR; 20 principal components in Australian samples). Further adjustments were made for imputation run (separate runs for different genotyping arrays) in the Australian samples, and for nonverbal IQ in all samples (except for the Australian adults), and for hearing difficulties in the Australian older adults. Because the cohorts included related family members (twins or siblings), linear mixed models (lme) were specified in RStudio[98], with family membership modeled as a random effect and the dyslexia polygenic score as a fixed effect. Where monozygotic twins were present, their trait scores were averaged and they were used as a single case.

Evaluation of candidates from previous literature

We used the results of the 23andMe dyslexia GWAS to assess variants, genes and biological pathways previously associated with or implicated in dyslexia and/or variation in reading and spelling ability in past association studies, linkage analyses and other studies.

Previously reported variants. We assessed 75 previously reported variants within our summary statistics, adopting a replication/validation significance threshold of $P < 7.28 \times 10^{-4}$, derived by Bonferroni correction based on 68.7 independent tests derived through matrix spectral decomposition, taking into account LD (see Doust et al.[25] for details on how these variants were selected). The sources for each variant are provided in Supplementary Table 26.

Dyslexia candidate genes. We evaluated gene-based results from MAGMA v.1.08 (ref. [56]) for overrepresentation of genome-wide significant variants from the 23andMe dyslexia GWAS within the loci of 14 candidate genes from earlier literature: *CMIP*, *CNTNAP2*, *CYP19A1*, *DCDC2*, *DIP2A*, *DYX1C1*, *GCFC2*, *KIAA0319*, *KIAA0319L*, *MRPL19*, *PCNT*, *PRMT2*, *S100B* and *ROBO1*. The rationale for this selection is detailed by Luciano et al.[24] and Doust et al.[5]. The critical *P* value, based on Bonferroni correction for 14 tests, was 3.57×10^{-3}.

Candidate dyslexia gene sets. We performed a gene set analysis in MAGMA to test for overrepresentation of genome-wide significant variants within (1) a set of transcriptional targets of *FOXP2*, a highly conserved transcription factor linked to speech and language impairment[99]; and (2) two biological pathways previously suggested to play a role in dyslexia susceptibility[100,101]—axon guidance (GO:0007411: 'chemotaxis process that directs the migration of an axon growth cone to a specific target site'; 216 genes) and neuron migration (GO:0001764: 'movement of an immature neuron from germinal zones to specific positions where they will reside as they mature'; 145 genes). An adjusted critical *P* value of 0.017 was derived using Bonferroni correction based on three independent tests.

Ethical standards

Participants provided informed consent and participated in the research online, under a protocol approved by the external AAHRPP-accredited IRB, Ethical and Independent Review Services. Participants were included in the analysis on the basis of consent status as checked at the time data analyses were initiated.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The full summary statistics for each dyslexia GWAS presented in this paper will be made available through 23andMe website (https://research.23andme.com/dataset-access/) to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. The top 10,000 associated SNPs from the main GWAS can be downloaded from https://doi.org/10.7488/ds/3465.

References

53. Fontanillas, P. et al. Disease risk scores for skin cancers. *Nat. Commun.* **12**, 160 (2021).
54. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
55. Gialluisi, A. et al. Genome-wide association scan identifies new variants associated with a cognitive predictor of dyslexia. *Transl. Psychiatry* **9**, 77 (2019).
56. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
57. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
58. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
59. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
60. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Human Genet.* **81**, 208–227 (2007).
61. Maller, J. B. et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012).
62. Howe, K. L. et al. Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891 (2020).
63. Carvalho-Silva, D. et al. Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res.* **47**, D1056–D1065 (2018).
64. Petrovski, S. et al. The intolerance of regulatory sequence to genetic variation predicts gene dosage sensitivity. *PLoS Genet.* **11**, e1005492 (2015).
65. Rada-Iglesias, A. Is H3K4me1 at enhancers correlative or causative? *Nat. Genet.* **50**, 4–5 (2018).
66. Heintzman, N. D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).
67. Cahoy, J. D. et al. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264 (2008).
68. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648 (2015).
69. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
70. Dunham, I. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
71. Vermunt, M. W. et al. Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nat. Neurosci.* **19**, 494–503 (2016).
72. Reilly, S. K. et al. Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* **347**, 1155–1159 (2015).

73. Peyrégne, S., Boyle, M. J., Dannemann, M. & Prüfer, K. Detecting ancient positive selection in humans using extended lineage sorting. *Genome Res.* **27**, 1563–1572 (2017).
74. Simonti, C. N. et al. The phenotypic legacy of admixture between modern humans and Neandertals. *Science* **351**, 737–741 (2016).
75. Vernot, B. et al. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science* **352**, 235–239 (2016).
76. Gazal, S. et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49**, 1421–1427 (2017).
77. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
78. Bulik-Sullivan, B. K. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
79. Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2016).
80. Grasby, K. L. et al. The genetic architecture of the human cerebral cortex. *Science* **367**, eaay6690 (2020).
81. Satizabal, C. L. et al. Genetic architecture of subcortical brain structures in 38,851 individuals. *Nat. Genet.* **51**, 1624–1636 (2019).
82. Hibar, D. P. et al. Novel genetic loci associated with hippocampal volume. *Nat. Commun.* **8**, 13624 (2017).
83. Adams, H. H. et al. Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat. Neurosci.* **19**, 1569–1582 (2016).
84. Smith, S. M. et al. An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat. Neurosci.* **24**, 737–745 (2021).
85. Alfaro-Almagro, F. et al. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage* **166**, 400–424 (2018).
86. Forkel, S. J. & Catani, M. *The Oxford Handbook of Neurolinguistics: Diffusion Imaging Methods in Language Sciences* (Oxford Univ. Press, Oxford, 2019).
87. Price, C. J. The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann. N. Y. Acad. Sci.* **1191**, 62–88 (2010).
88. Richardson, F. M. & Price, C. J. Structural MRI studies of language function in the undamaged brain. *Brain Struct. Funct.* **213**, 511–523 (2009).
89. Perdue, M. V., Mednick, J., Pugh, K. R. & Landi, N. Gray matter structure is associated with reading skill in typically developing young readers. *Cereb. Cortex* **30**, 5449–5459 (2020).
90. Roehrich-Gascon, D., Small, S. L. & Tremblay, P. Structural correlates of spoken language abilities: a surface-based region-of interest morphometry study. *Brain Lang.* **149**, 46–54 (2015).
91. Luciano, M. et al. A genome-wide association study for reading and language abilities in two population cohorts. *Genes Brain Behav.* **12**, 645–652 (2013).
92. Gialluisi, A. et al. Genome-wide screening for DNA variants associated with reading and language traits. *Genes Brain Behav.* **13**, 686–701 (2014).
93. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Human Genet.* **81**, 559–575 (2007).
94. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: polygenic risk score software. *Bioinformatics* **31**, 1466–1468 (2015).
95. Bates, T. C. et al. Behaviour genetic analyses of reading and spelling: a component processes approach. *Aust. J. Psychol.* **56**, 115–126 (2004).
96. Dollaghan, C. & Campbell, T. F. Nonword repetition and child language impairment. *J. Speech Lang. Hear. Res.* **41**, 1136–1146 (1998).
97. Gathercole, S. E., Willis, C. S., Baddeley, A. D. & Emslie, H. The Children's Test of Nonword Repetition: a test of phonological working memory. *Memory* **2**, 103–127 (1994).
98. RStudio Team. RStudio: Integrated Development for R. (Boston, MA, 2020).
99. Ayub, Q. et al. FOXP2 Targets show evidence of positive selection in European populations. *Am. J. Human Genet.* **92**, 696–706 (2013).
100. Poelmans, G., Buitelaar, J. K., Pauls, D. L. & Franke, B. A theoretical molecular network for dyslexia: integrating available genetic findings. *Mol. Psychiatry* **16**, 365–382 (2011).
101. Guidi, L. G. et al. The neuronal migration hypothesis of dyslexia: a critical evaluation 30 years on. *Eur. J. Neurosci.* **48**, 3212–3233 (2018).

Acknowledgements

We thank the research participants and employees of 23andMe Inc, the GenLang Consortium, the Brisbane Adults Reading Study, and the CRS. E.E., G.A., B.M., B.S.P., C.F. and S.E.F. are supported by the Max Planck Society (Germany). The CRS was supported by grants from the National Natural Science Foundation of China (Grant No. 61807023), Funds for Humanities and Social Sciences Research of the Ministry of Education (Grant No. 19YJC190023 and 17XJC190010) and General Project of Shaanxi Natural Science Basic Research Program (2018JQ8015) (Grant No. 2018JQ8015 and 2021JQ-309). S.P. is funded by the Royal Society. Acknowledgements for the GenLang Consortium appear in the Supplementary Note.

Author contributions

M.L., S.E.F., T.C.B. and N.G.M. conceived the study, with M.L. overseeing general analysis and A.A. overseeing 23andMe analysis. C.D., P.F., E.E., G.A., S.D.G., Z.W., B.M. and M.L. performed statistical and/or downstream annotation analysis. R.E.M. advised C.D. on some analysis. C.D. drafted the manuscript, with sections contributed by P.F., E.E., G.A., Z.W. and M.L. B.S.P., C.F. and S.E.F. supervised the GenLang GWAS. J.Z. managed the Chinese Reading Study. S.P., J.B.T., A.P.M. and J.F.S. managed the UKDys study. J.R.G., R.K.O., E.G.W., J.C.D., B.F.P. and S.D.S. managed the CLDRC study. M.J.W., T.C.B. and N.G.M. managed the Australian adolescent twin studies. M.L., T.C.B., S.E.F. and N.G.M. managed the Australian adult reading study. All authors critically reviewed the manuscript.

Competing interests

P.F., A.A. and the 23andMe Research Team are employed by and hold stock or stock options in 23andMe, Inc. The remaining authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-022-01192-y.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-022-01192-y.

Correspondence and requests for materials should be addressed to Michelle Luciano.

Peer review information *Nature Genetics* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Supplementary Methods

23andMe Genotyping and Imputation

Samples were genotyped on one of five genotyping platforms. The V1 and V2 platforms were variants of the Illumina HumanHap550 + BeadChip, including about 25,000 custom SNPs selected by 23andMe, with a total of about 560,000 SNPs. The V3 platform was based on the Illumina OmniExpress + BeadChip, with custom content to improve the overlap with our V2 array, with a total of ~950,000 SNPs. The V4 platform is a fully custom array, including a lower redundancy subset of V2 and V3 SNPs with additional coverage of lower-frequency coding variation, and ~570,000 SNPs. The v5 platform, in current use, is an Illumina Infinium Global Screening Array (~640,000 SNPs) supplemented with ~50,000 SNPs of custom content. Samples that failed to reach 98.5% call rate were excluded from the study.

Individuals were only included if they had > 97% European ancestry, as determined through an analysis of local ancestry (see 1 for further details on the methodology used). Briefly, this analysis first partitions phased genomic data into short windows of ~100 SNPs. Within each window, a support vector machine is used to classify individual haplotypes into one of 31 reference populations. The support vector machine classifications are then fed into a Hidden Markov Model (HMM) that accounts for switch errors and incorrect assignments and gives probabilities for each reference population in each window. Finally, simulated admixed individuals are used to recalibrate the HMM probabilities so that the reported assignments are consistent with the simulated admixture proportions. The reference population data are derived from public data sets (the Human Genome Diversity Project, HapMap and 1000 Genomes) and from 23andMe research participants who have reported having four grandparents from the same country.

A maximal set of unrelated individuals was chosen for each analysis using a segmental identity-by-descent (IBD) estimation algorithm[2]. Individuals were defined as related if they shared more than 700 cM IBD, including regions where the two individuals share either one or both genomic segments identical-by-descent. This level of relatedness (roughly 20% of the genome) corresponds 6 approximately to the minimal expected sharing between first cousins in an outbred population. For the purposes of GWAS, if a case was found to be related to a control, the case was preferentially kept in the sample.

Participant genotype data were imputed against a single unified imputation reference panel, combining the May 2015 release of the 1000 Genomes Phase 3 haplotypes and the UK10K imputation reference panel. Data for each genotyping platform were phased and imputed separately. Variants that were only genotyped on the ‘V1’ platform were flagged due to small sample size, and variants on chrM or chrY, because many of these are not currently called reliably. Using trio data, variants that failed a test for parent–offspring transmission were also flagged; specifically, the child’s allele count was regressed against the mean parental allele count and variants with fitted β < 0.6 and p < 10-20 for a test of β< 10-20 in Europeans, or a call rate of < 90%, were also flagged. Genotyped variants were also tested for batch effects and variants with p < 10-50 by analysis of variance of genotypes against a factor dividing genotyping date into 20 roughly equal-sized buckets were flagged. For imputed GWAS results, variants with average r 2 < 0.5 or minimum r 2 < 0.3 in any imputation batch were flagged, as well as SNPs that had strong evidence of an imputation batch effect, using an analysis of variance of the imputed dosages against a factor representing imputation batch; results with p < 10-50 were flagged. Each variant flagged by QC on genotyped or imputation data were excluded from the GWAS analysis.

Chinese Reading Study Sample

Participants

3,127 Grade 3 to Grade 6 primary students aged nine to 14 years were recruited from three cities and four districts in China (Xi'an-YT, Xi'an-CB, Qingyang, and Baotou). In total, 2,476 participants were eligible for subsequent genotyping and association analysis. Ethical approval was obtained for each cohort at the local level and written informed consent was obtained from all the participants' parents.

Phenotypic Measures

Reading accuracy: A Chinese character recognition test was employed to measure each child's reading accuracy[3-5]. The test consisted of 150 single Chinese characters selected from China's Elementary School Textbooks (1996). The average frequency of the characters was 182 per million (ranging from 0 to 2,282), and the reliability of this test was 0.95[3]. Each child was individually tested and was required to read aloud each character at a time. Reading fluency: A word list reading task 3 was used to measure each child's reading fluency. In this task, children were asked to name a list of 180 two-character words as rapidly and accurately as possible. All these words were from primary school textbooks and have been learned before Grade 3, such as "我们 (we)" and "太阳 (sun)". The mean frequency of these words was 212.77 per million[6]. Since words included in this task were all simple, this task was administrated to test children's reading fluency. The total time for naming the whole word list was recorded as the measurement of reading fluency.

Genotype Quality Control, Imputation, and Analysis

DNA was extracted from saliva samples, and individuals were genotyped using the Illumina Asian screening array (650K) by Beijing Compass Biotechnology. Quality

control was performed using standard quality control metrics. Eight samples were excluded as they had sex discrepancies between the records and the genetically inferred data[7,8]. Next, we removed 53 samples who had unexpected duplicates or probable relatives (PI-HAT > 0.20). Then, SNPs were filtered out if they showed a variant call rate < 0.95, a minor allele frequency (MAF) < 0.01, a missing genotype data (mind) < 0.90, or a Hardy-Weinberg Equilibrium (HWE) $p < 10^{-5}$ within each dataset. For imputation, autosomal variants were aligned to the 1000G genomes phase 1v3 reference panel. Imputation was performed using the Michigan imputation Server 4.0 in 5Mb chunks with 500kb buffers, filtering out variants that were monomorphic in the Genome Asia Pilot (GAsP). Chunks with 51% genotyped variants or concordance rate < 0.92 were fused with neighbouring chunks and re-imputed. Finally, imputed variants were filtered out for r2 < 0.60, MAF < 0.02, mind < 0.1, HWE $p < 10^{-8}$ [5] using Plink (v1.90). After quality control procedures had been performed, 2,415 children with 4,261,603 SNPs were included in the final analysis. Association analyses were performed using PLINK, fitting an additive model to the linear regression model with adjustment for sex, age, and the first two principal components[8].

Biological Annotations

Genome-wide significant variants and the closest gene(s) were annotated using external reference data through FUMA v1.3.6a[9] (unless otherwise specified) and evaluated for functional or regulatory impact. Specifically, we considered the following annotations of SNPs reaching genome-wide significance ($p < 5 \times 10^{-8}$) (Supplementary Table 10):

- **Gene context:**

- **Distance**: The distance of the variant to the nearest gene in kb. Variants within the gene body or 1 kb up- or downstream of the transcription start site (TSS) or transcription end site (TES) have a value of zero.
- **Function**: Whether a variant is intergenic or the functional region in which the variant is located within a gene or RNA locus (e.g., 5' UTR).

- **Combined Annotation Dependent Depletion (CADD) score:** A score of the deleteriousness of variants computed from 63 integrated annotations[10]. The higher the score, the more deleterious a variant is: 12.37 is the threshold indicated by the study of potentially actionable exonic pathogenic single-nucleotide variants in European- and African ancestry patients[11].
- **RegulomeDB category (RDB):** A variant classification system in which variants are grouped according to evidence of having a functional consequence from Category 6 (minimal evidence) to Category 1a (likely to affect binding and linked to expression of a gene target)[12].
- **Chromatin state:** The minimum and the most common 15-core chromatin state across 127 tissue/cell types predicted by ChromHMM[13] from 15 (quiescent/low) to 1 (active TSS).
- **GWAS Catalog:** SNP-trait associations reported in the NHGRI-EBI Catalog of human GWAS[14], including for each variant: the trait(s), the effect allele(s), the PubMed ID(s), the study title(s) and the study sample size(s) (Supplementary Table 2).

And the following annotations of genes which were significant in genome-wide gene-based tests (Supplementary Table 12):

- **Probability of Loss-of-function Intolerance (pLI) score:** A score of intolerance to functional mutation from the ExAC database[15] ranging from zero

to one. The closer the score is to one, the more intolerant the gene is to loss-of-function mutations. The threshold suggested by Lek, et al.[15] for likely disease-causing variants is ≥ 0.9.

- **Non-coding Residual Variation Intolerance Score (ncRVIS):** A score of intolerance to mutation to non-coding variants[16]. Where ncRVIS is zero, the gene has the average number of noncoding variants given its total mutational burden; when ncRVIS is greater than zero, the gene has less non-coding variation than expected; when ncRVIS is less than zero, it has more. The ncRVIS percentile reflects the rank of the gene amongst all genes. The more negative the ncRVIS, or the lower the percentile, the more intolerant to non-coding variation the gene is.
- **Residual Variation Intolerance Score (ncRVIS) percentile: As** for ncRVIS score but the percentile of the average RVIS score for the whole gene sequence.
- **Non-coding Genomic Evolutionary Rate Profiling (ncGERP) score:** Identifies constraint in non-coding regions by quantifying deficits in substitutions[16]. It is calculated by taking the average GERP++ score (see Davydov, et al.[17]) across the non-coding sequence. The higher the ncGERP score, the fewer substitutions are present than what would be expected as a result of a neutral rate of evolution, and thus the more conserved are the non-coding regions of the gene. The ncGERP percentile reflects the rank of the gene amongst all genes.
- **Protein-coding Genomic Evolutionary Rate Profiling (pcGERP) percentile:** As for ncGERP score but the percentile of the average GERP score for protein-coding sequence[16] .

- **Non-coding Combined Annotation Dependent Depletion (ncCADD) score:** As for CADD score but the average variant score across the non-coding sequence of the gene[16].
- **Non-coding Genome-Wide Annotation of Variants (ncGWAVA) score:** Predicts the combined functionality of non-coding variants across non-coding sequence[16]. It is the average GWAVA score (see Ritchie, et al.[18]) of variants in the non-coding sequence, ranging from zero to one. The closer ncGWAVA is to one, the more likely the variants in non-coding regions of the gene are functional.
- **Expression in the brain:** Average log2 expression in transcripts per million (TPM) per tissue type per gene from the GTEx v8 dataset[19] for 12 brain tissues: Amygdala, Anterior Cingulate Cortex, Caudate Basal Ganglia, Cerebellar Hemisphere, Cerebellum, Cortex, Frontal Cortex, Hippocampus, Hypothalamus, Nucleus Accumbens Basal Ganglia, Putamen Basal Ganglia, and Substantia Nigra (Supplementary Table 15).

Partitioned Heritability

Evolutionary Analysis

Enrichment of heritability was estimated for the following evolutionary annotations (as described in Tilot, et al.[20]):

- **Human Gained Enhancers and Promoters:** These regulatory regions were identified based on differential H3K27ac and H3K4me2 patterns in the adult and foetal brain tissues of humans, macaques and mice [19, 20], and shown to be present to a significantly lesser degree in macaques and mice. Thus, these regulatory elements were gained in the last 30 million years of human evolution and may be involved in the emergence of human-specific traits[21,22].

- **Ancient selective sweep regions:** These consist of unusually long genomic regions that reached fixation in human populations possibly due to adaptive advantages in the last 250 650 thousand years[23].
- **Neanderthal-introgressed SNPs:** The genomic variants introduced into the human genome by the admixture of Homo sapiens and Neanderthal populations around 50-60,000 years ago[24].
- **Neanderthal Depleted Regions:** Large regions in the human genome that are depleted for Neanderthal ancestry, possibly due to the deleterious effect of the archaic sequences in hybrid individuals[25].

Genlang Quantitative Trait Consortium Acknowledgements

General Acknowledgements

BM, BMM, BSP, CF, EE, EV, GA, MvD and SEF are supported by the Max Planck Society. AG and TFMA were supported by the Munich Cluster for Systems Neurology (SyNergy), and AG was supported by Fondazione Umberto Veronesi. ATM is supported by the National Health and Medical Research Council of Australia (NHMRC) Fellowships (1105008; 1195955) and Centre of Research Excellence grant (1116976). AJOW is supported by an Investigator Grant from the NHMRC (1173896). BSP and MvD are supported by the Simons Foundation Autism Research Initiative grant (514787 to BSP). CYS works in the Medical Research Council Integrative Epidemiology Unit at the University of Bristol (MC_UU_00011/3). DIB acknowledges the Royal Netherlands Academy of Science Professor Award (PAH/6635). EE is supported by a National Institutes of Health (NIH) R01 grant (R01DC016977). EGW is supported by NICHD P50 HD 27802. FR is supported by Agence Nationale de la Recherche (ANR06-NEURO-019-01, ANR-17-EURE-0017 IEC, ANR-10-IDEX-0001-02 PSL, ANR-11-BSV4-014-01), European Commission (LSHM-CT-2005-018696). HT is supported by a VICI grant of the Netherlands Organization for Scientific Research (NWO) and The Netherlands Organisation for Health Research and Development (ZonMW) (016.VICI.170.200). JCDF was supported by the National Institute of Child Health and Human Development (NICHD) (P50 HD 27802). JJM, JBTo, and TK were supported by the National Institutes of Health (NIH) (R01 DC014489). KP was supported by the Hospital for Sick Children Research Training Program (Restracomp). KR is supported by a Sir Henry Wellcome Postdoctoral Fellowship. MJS is supported by a Wellcome Trust Programme grant (WT082032MA). SP and FA are supported by the Royal Society (UF150663; RGF\EA\180141). TB is

supported by Institut Pasteur, the Bettencourt-Schueller Foundation, Université de Paris.

We are extremely grateful to all the children, twins, families and participants who took part and are taking part in the 22 cohorts whose data contributed to these GWAS meta-analyses. We also thank the staff working on the different cohorts, including volunteers, study coordinators, interviewers, teachers, nurses, research scientists, general practitioners, midwives, psychologists, psychometrists, 20 computer and laboratory technicians and colleagues who assisted in the quality control and preparation of the imputed GWAS data and the pharmacies and hospitals that were involved.

Adolescent Brain Cognitive Development^{SM} Study

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive DevelopmentSM (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the NIH and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators

designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data used in this report are summarized in this NDA study (https://doi.org/10.15154/1522876).

Aston Cohort

Data collection for the Aston Cohort has been supported by funding from the European Union Horizon 2020 Programme (641652), The Waterloo Foundation (797/17290), and with the invaluable support of the participants and their families. Data analyses were supported by the St Andrews Bioinformatics Unit which is funded by Wellcome Trust ISSF awards (105621/Z/14/Z and 204821/Z/16/Z).

Avon Longitudinal Study of Parents and their Children (ALSPAC)

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory 21 technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. GWAS data were generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. A comprehensive list of grants funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf). The publication is the work of the authors and they will serve as guarantor for the ALSPAC contribution to this article.

The Basque Center on Cognition, Brain and Language (BCBL)

Data collection was supported by the Basque Government through the BERC program, and by the Agencia Estatal de Investigación through BCBL Severo Ochoa excellence accreditation.

Brisbane Adolescent Twins Study (BATS)

The research was supported by the Australian Research Council (A7960034, A79906588, A79801419, DP0212016 and DP0343921), with genotyping funded by the NHMRC (Medical Bioinformatics Genomics Proteomics Program, 389891).

Colorado Learning Disabilities Research Center (CLDRC)

The CLDRC was supported by a NICHD grant P50 HD 27802.

The Early Language in Victoria Study (ELVS)

Funding by the NHMRC grant number 436958 is gratefully appreciated.

Familial Influences on Literacy Abilities (FIOLA)

The FIOLA cohort (PIs EvB and PFdJ) is supported by the University of Amsterdam, the MPI Nijmegen, and by fellowships awarded to EvB (NWO's Rubicon 446-12-005 and VENI 451-15-017). We are grateful to the NEMO Science Museum.

Generation R

The Generation R Study is conducted by the Erasmus Medical Center in close collaboration with School of Law and Faculty of Social Sciences of the Erasmus University Rotterdam, the Municipal Health 22 Service Rotterdam area, the Rotterdam Homecare Foundation, and the Stichting Trombosedienst & Artsenlaboratorium Rijnmond (STAR[1]MDC), Rotterdam.

The Genes, Reading and Dyslexia (GRaD) Study

The GRaD Study was funded by the generous support of the Manton Foundation, the NIH (P50- HD027802; K99-HD094902), and the Lambert Family.

Neurodys

We would like to acknowledge our project partners Catherine Billard, Caroline Bogliotti, Vanessa Bongiovanni, Laure Bricout, Camille Chabernaud, Yves Chaix, Isabelle Comte-Gervais, Florence DelteilPinton, Jean-François Démonet, Fabien Fauchereau, Florence George, Christophe-Loïc Gérard, Ferenc Honbolygó, Guillaume Huguet, Stéphanie Iannuzzi, Marie Lageat, Marie-France Leheuzey, MarieThérèse Lenormand, Marion Liébert, Emilie Longeras, Emilie Racaud, Isabelle Soares-Boucaud, Sylviane Valdois, Nadège Villiermet, and Johannes Ziegler. This project was funded by the EU FP6 grant to the Neurodys consortium, the SNSF (32-108130) and Austrian Science Fund (18351-B02).

Netherlands Twin Register (NTR)

Funding from the NWO and ZonMW is gratefully acknowledged: Twin family database for behavior genomics studies (NWO 480-04-004); Longitudinal data collection from teachers of Dutch twins and their siblings (481-08-011), Twin-family-study of individual differences in school achievement (NWOFES, 056-32-010), Genotype-phenotype database for behavior genetic and genetic epidemiological studies (ZonMw Middelgroot 911-09-032); Genetic influences on stability and change in psychopathology from childhood to young adulthood (ZonMW 912-10-020), Why some children thrive (OCW_Gravity program –NWO-024.001.003), Netherlands Twin Registry Repository: researching the interplay between genome and environment (NWO-Groot 480-15-001/674); BBMRI –NL (184.021.007 and 184.033.111):

Biobanking and Biomolecular Resources Research Infrastructure; Spinozapremie (NWO- 56-464-14192); Amsterdam Public Health (APH) and Amsterdam Reproduction and Development (AR&D) Research Institutes; European Science Council Genetics of Mental Illness (ERC Advanced, 230374); Aggression in Children: Unraveling gene-environment interplay to inform 23 Treatment and InterventiON strategies (ACTION) project (European Union Seventh Framework Program (FP7/2007-2013) no 602768); Rutgers University Cell and DNA Repository cooperative agreement (NIMH U24 MH068457-06); Developmental Study of Attention Problems in Young Twins (NIMH, RO1 MH58799-03); Grand Opportunity grant Developmental trajectories of psychopathology (NIMH 1RC2 MH089995) and the Avera Institute for Human Genetics.

Pediatric Imaging, Neurocognition and Genetics Study (PING)

Data collection and sharing for this project was funded by the Pediatric Imaging, Neurocognition and Genetics Study (PING) (NIH grant RC2DA029475). PING is funded by the National Institute on Drug Abuse and the Eunice Kennedy Shriver NICHD. PING data are disseminated by the PING Coordinating Center at the Center for Human Development, University of California, San Diego.

The Philadelphia Neurodevelopmental Cohort (PNC)

Support for the collection of the data sets was provided by grant RC2MH089983 awarded to Raquel Gur and RC2MH089924 awarded to Hakon Hakonarson. All subjects were recruited through the Center for Applied Genomics at The Children's Hospital in Philadelphia. Reading performance was assessed as part of the cognitive test battery and collected by Gur et al. (2012). Genotyping was funded by an Institutional Development Award to the Center for Applied Genomics from The

Children’s Hospital of Philadelphia and a donation from Adele and Daniel Kubert. We would also like to thank the NIH data repository. The dataset analysed was accessed via dbGaP number 19848 and 12043 (Neurodevelopmental Genomics: Trajectories of Complex Phenotypes (phs000607.v3.p2)).

The Raine Study

The Raine Study was supported by the NHMRC (572613, 403981 and 1059711), the Canadian Institutes of Health Research (MOP-82893), and WA Health, Government of Western Australia (Future Health WA G06302). Funding was also generously provided by Safe Work Australia. The authors gratefully acknowledge the NHMRC for their long-term funding to the study over the last 30 years and also the following institutes for providing funding for Core Management of the Raine Study: The University of Western Australia (UWA), Curtin University, Women and Infants Research Foundation, Telethon Kids 24 Institute, Edith Cowan University, Murdoch University, The University of Notre Dame Australia and The Raine Medical Research Foundation. This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and Government of Western Australia.

Saguenay Youth Study (SYS)

The Canadian Institutes of Health Research (ZP, TP), Heart and Stroke Foundation of Quebec (ZP), and the Canadian Foundation for Innovation (ZP) support the Saguenay Youth Study.

SLI Consortium (SLIC)

SLI Consortium members are as follows: Wellcome Trust Centre for Human Genetics, Oxford: D. F. Newbury, N. H. Simpson, F. Ceroni, A. P. Monaco; Max Planck Institute

for Psycholinguistics, Nijmegen: S. E. Fisher, C. Francks; Newcomen Centre, Evelina Children’s Hospital, St Thomas’ Hospital, London: G. Baird, V. Slonims; Child and Adolescent Psychiatry Department and Medical Research Council Centre for Social, Developmental, and Genetic Psychiatry, Institute of Psychiatry, London: P. F. Bolton; Medical Research Council Centre for Social, Developmental, and Genetic Psychiatry Institute of Psychiatry, London: E. Simonoff; Salvesen Mindroom Centre, Child Life & Health, School of Clinical Sciences, University of Edinburgh: A. O’Hare; Cell Biology & Genetics Research Centre, St. George’s University of London: J. Nasir; Queen’s Medical Research Institute, University of Edinburgh: J. Seckl; Department of Speech and Language Therapy, Royal Hospital for Sick Children, Edinburgh: H. Cowie; Speech and Hearing Sciences, Queen Margaret University: A. Clark, J. Watson; Department of Educational and Professional Studies, University of Strathclyde: W. Cohen; Department of Child Health, the University of Aberdeen: A. Everitt, E. R. Hennessy, D. Shaw, P. J. Helms; Audiology and Deafness, School of Psychological Sciences, University of Manchester: Z. Simkin, G. Conti-Ramsden; Department of Experimental Psychology, University of Oxford: D. V. M. Bishop; Biostatistics Department, Institute of Psychiatry, London: A. Pickles. The collection and genetic characterisation of SLIC samples was funded by the Wellcome Trust (076566) and the UK Medical Research Council (G1000569).

The Twins Early Development Study (TEDS)

TEDS is supported by a programme grant to RP from the UK Medical Research Council (MR/V012878/1 and previously MR/M021475/1), with additional support from the NIH (AG046938). The research leading to these results has also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013/ grant agreement number 602768).

Toronto

Support for this project was provided by grants from the Canadian Institutes of Health Research (MOP-133440).

UK Dyslexia

Cohort recruitment and data collection was supported by Wellcome Trust (076566/Z/05/Z and 075491/Z/04) and The Waterloo Foundation (grants to JBTa and SP; 797–1720). Genotype data were generated and analyses funded by the EU (Neurodys, 018696) and the Royal Society (UF100463 grant to SP).

York

The York cohort was funded by Wellcome Trust Programme Grant 082036/B/07/Z.

Supplementary References

1. Durand, E.Y. et al. A scalable pipeline for local ancestry inference using tens of thousands of reference haplotypes. bioRxiv, 2021.01.19.427308 (2021).

2. Henn, B.M. et al. Cryptic Distant Relatives Are Common in Both Isolated and Cosmopolitan Genetic Samples. PLOS ONE 7, e34267 (2012).

3. Pan, J. & Shu, H. Rapid Automatized Naming and Its Unique Contribution to Reading: Evidence from Chinese Dyslexia. Reading development and difficulties in monolingual and bilingual Chinese children, 125-138 (2014).

4. Lei, L. et al. Developmental trajectories of reading development and impairment from ages 3 to 8 years in Chinese children. Journal of Child Psychology and Psychiatry 52, 212-220 (2011).

5. Song, S., Zhang, Y., Shu, H., Su, M. & McBride, C. Universal and Specific Predictors of Chinese Children with Dyslexia - Exploring the Cognitive Deficits and Subtypes. Frontiers in psychology 10, 2904-2904 (2020).

6. Wang, H., Chang, B. R., Li, Y. S., Lin, L. H., Liu, J., Sun, Y. L., et al. . Modern Chinese Frequency Dictionary, (Beijing Language Colleague Press, Beijing, 1986).

7. Anderson, C.A. et al. Data quality control in genetic case-control association studies. Nat Protoc 5, 1564-73 (2010).

8. Chang, C.C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015).

9. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun 8, 1826 (2017). 13

10. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics 46, 310-315 (2014).

11. Amendola, L.M. et al. Actionable exomic incidental findings in 6503 participants: challenges of variant classification. Genome research 25, 305-315 (2015).

12. Boyle, A.P. et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 22, 1790-7 (2012).

13. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. Nature Methods 9, 215-216 (2012).

14. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 47, D1005-d1012 (2019).

15. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285-291 (2016).

16. Petrovski, S. et al. The Intolerance of Regulatory Sequence to Genetic Variation Predicts Gene Dosage Sensitivity. PLOS Genetics 11, e1005492 (2015).

17. Davydov, E.V. et al. Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. PLOS Computational Biology 6, e1001025 (2010).

18. Ritchie, G.R.S., Dunham, I., Zeggini, E. & Flicek, P. Functional annotation of noncoding sequence variants. Nature methods 11, 294-296 (2014).

19. Aguet, F. et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. bioRxiv, 787903 (2019).

20. Tilot, A.K. et al. The Evolutionary History of Common Genetic Variants Influencing Human Cortical Surface Area. Cerebral Cortex 31, 1873-1887 (2020).

21. Vermunt, M.W. et al. Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. Nat Neurosci 19, 494-503 (2016).

22. Reilly, S.K. et al. Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. Science 347, 1155-9 (2015).

23. Peyrégne, S., Boyle, M.J., Dannemann, M. & Prüfer, K. Detecting ancient positive selection in humans using extended lineage sorting. Genome Res 27, 1563-1572 (2017).

24. Vernot, B. & Akey, J.M. Resurrecting surviving Neandertal lineages from modern human genomes. Science 343, 1017-21 (2014).

25. Vernot, B. et al. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. Science 352, 235-239 (2016).

26. Bates, T.C. et al. Dyslexia and DYX1C1: deficits in reading and spelling associated with a missense mutation. Molecular Psychiatry 15, 1190-6 (2010).

27. Brkanac, Z. et al. Evaluation of candidate genes for DYX1 and DYX2 in families with dyslexia. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics 144b, 556-60 (2007).

28. Carrion-Castillo, A. et al. Association analysis of dyslexia candidate genes in a Dutch longitudinal sample. European Journal of Human Genetics 25, 452-460 (2017).

29. Chen, Y., Zhao, H., Zhang, Y.X. & Zuo, P.X. DCDC2 gene polymorphisms are associated with developmental dyslexia in Chinese Uyghur children. Neural Regeneration Research 12, 259- 266 (2017).

30. Cope, N.A. et al. No support for association between Dyslexia Susceptibility 1 Candidate 1 and developmental dyslexia. Molecular Psychiatry 10, 237-238 (2005).

31. Couto, J.M. et al. The KIAA0319-like (KIAA0319L) gene on chromosome 1p34 as a candidate for reading disabilities. Journal of Neurogenetics 22, 295-313 (2008).

32. Couto, J.M. et al. Association of reading disabilities with regions marked by acetylated H3 histones in KIAA0319. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics 153b, 447-462 (2010).

33. Dahdouh, F. et al. Further evidence for DYX1C1 as a susceptibility factor for dyslexia. Psychiatric Genetics 19, 59-63 (2009).

34. Francks, C. et al. A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. American Journal of Human Genetics 75, 1046-58 (2004).

35. Harold, D. et al. Further evidence that the KIAA0319 gene confers susceptibility to developmental dyslexia. Molecular Psychiatry 11, 1085-1091 (2006).

36. Kong, R. et al. Genetic variant in DIP2A gene is associated with developmental dyslexia in Chinese population. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics 171, 203-208 (2016).

37. Lind, P.A. et al. Dyslexia and DCDC2: normal variation in reading and spelling is associated with DCDC2 polymorphisms in an Australian population sample. European Journal of Human Genetics 18, 668-73 (2010).

38. Mascheretti, S. et al. KIAA0319 and ROBO1: evidence on association with reading and pleiotropic effects on language and mathematics abilities in developmental dyslexia. Journal of Human Genetics 59, 189-197 (2014).

39. Matsson, H. et al. Polymorphisms in DCDC2 and S100B associate with developmental dyslexia. Journal of Human Genetics 60, 399-401 (2015).

40. Mozzi, A. et al. A common genetic variant in FOXP2 is associated with language-based learning (dis)abilities: Evidence from two Italian independent samples. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics 174B, 578–586 (2017).

41. Müller, B. et al. ATP2C2 and DYX1C1 are putative modulators of dyslexia-related MMR. Brain and Behavior 7, e00851-e00851 (2017).

42. Müller, B. et al. Association, characterisation and meta-analysis of SNPs linked to general reading ability in a German dyslexia case-control cohort. Scientific Reports 6, 27901 (2016).

43. Newbury, D.F. et al. Investigation of dyslexia and SLI risk variants in reading- and language impaired subjects. Behavior Genetics 41, 90-104 (2011).

44. Newbury, D.F. et al. CMIP and ATP2C2 modulate phonological short-term memory in language impairment. American Journal of Human Genetics 85, 264-272 (2009).

45. Paracchini, S. et al. Analysis of dyslexia candidate genes in the Raine cohort representing the general Australian population. Genes, Brain and Behavior 10, 158-65 (2011).

46. Peter, B. et al. 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, 39-49 (2011).

47. Poelmans, G. et al. Identification of novel dyslexia candidate genes through the analysis of a chromosomal deletion. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics 150B, 140-7 (2009).

48. Scerri, T.S. et al. Putative functional alleles of DYX1C1 are not associated with dyslexia susceptibility in a large sample of sibling pairs from the UK. Journal of Medical Genetics 41, 853-7 (2004).

49. Scerri, T.S. et al. DCDC2, KIAA0319 and CMIP are associated with reading-related traits. Biological Psychiatry 70, 237-45 (2011).

50. Schumacher, J. et al. Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. American Journal of Human Genetics 78, 52-62 (2006).

51. Shao, S. et al. The Roles of Genes in the Neuronal Migration and Neurite Outgrowth Network in Developmental Dyslexia: Single- and Multiple-Risk Genetic Variants. Molecular Neurobiology 53, 3967-3975 (2016).

52. Sun, X. et al. ROBO1 polymorphisms, callosal connectivity, and reading skills. Human Brain Mapping 38, 2616-2626 (2017).

53. Taipale, M. et al. A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. Proceedings of the National Academy of Sciences of the United States of America 100, 11553-8 (2003).

54. Tolosa, A. et al. FOXP2 gene and language impairment in schizophrenia: association and epigenetic studies. BMC Medical Genetics 11, 114-114 (2010).

55. Tran, C. et al. Association of the ROBO1 gene with reading disabilities in a family-based analysis. Genes, Brain and Behavior 13, 430-438 (2014).

56. Venkatesh, S.K., Siddaiah, A., Padakannaya, P. & Ramachandra, N.B. Lack of association between genetic polymorphisms in ROBO1, MRPL19/C2ORF3 and THEM2 with developmental dyslexia. Gene 529, 215-9 (2013).

57. Vernes, S.C. et al. A functional genetic link between distinct developmental language disorders. New England Journal of Medicine 359, 2337-2345 (2008).

58. Whitehouse, A.J.O., Bishop, D.V.M., Ang, Q.W., Pennell, C.E. & Fisher, S.E. CNTNAP2 variants affect early language development in the general population. Genes, Brain and Behavior 10, 451-456 (2011).

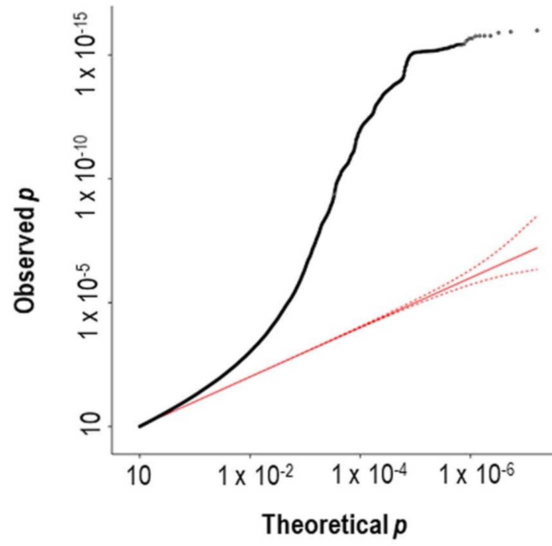
59. Wigg, K.G. et al. Support for EKN1 as the susceptibility locus for dyslexia on 15q21. Molecular Psychiatry 9, 1111-21 (2004).

60. Gialluisi, A. et al. Genome-wide association scan identifies new variants associated with a cognitive predictor of dyslexia. Translational Psychiatry 9, 77 (2019).

Extended Data

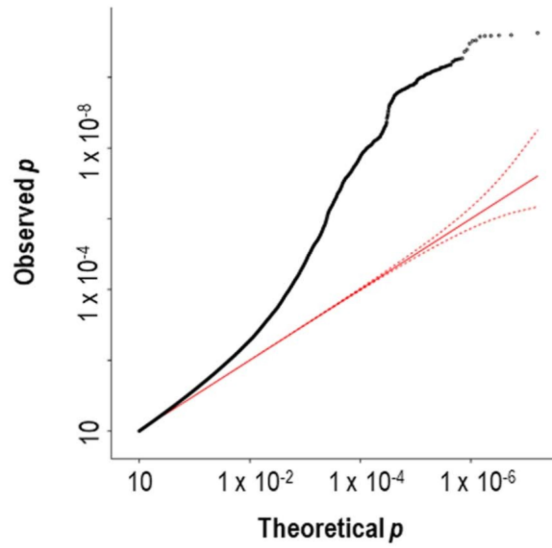
All

a



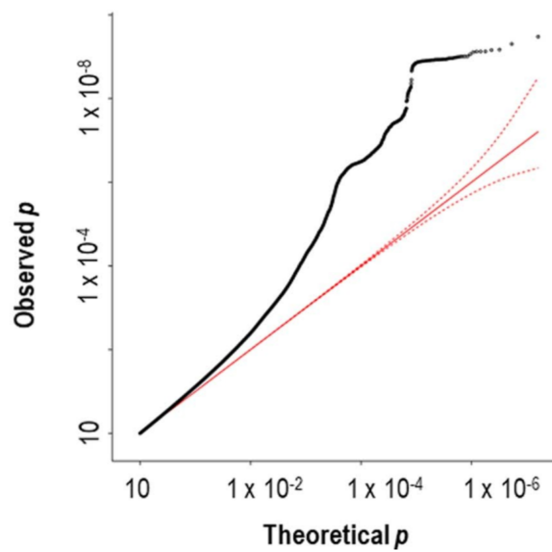
Female

b



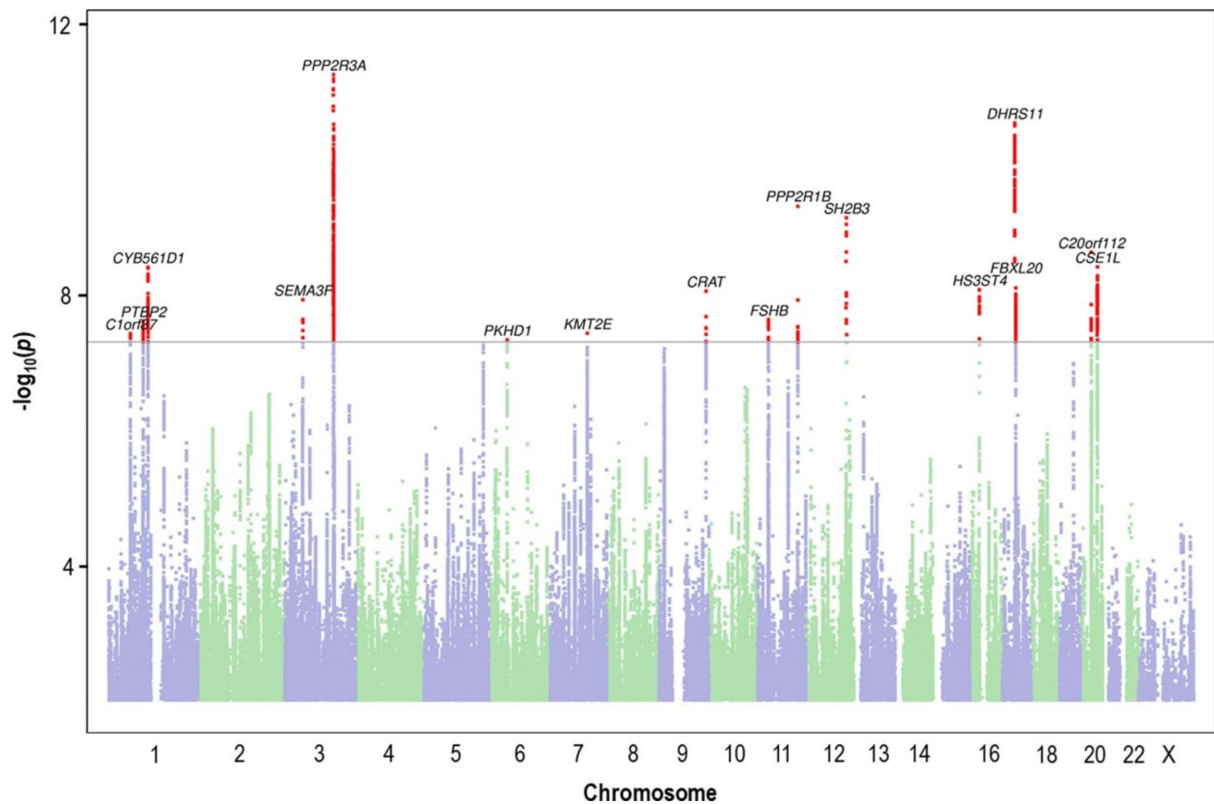
Male

c



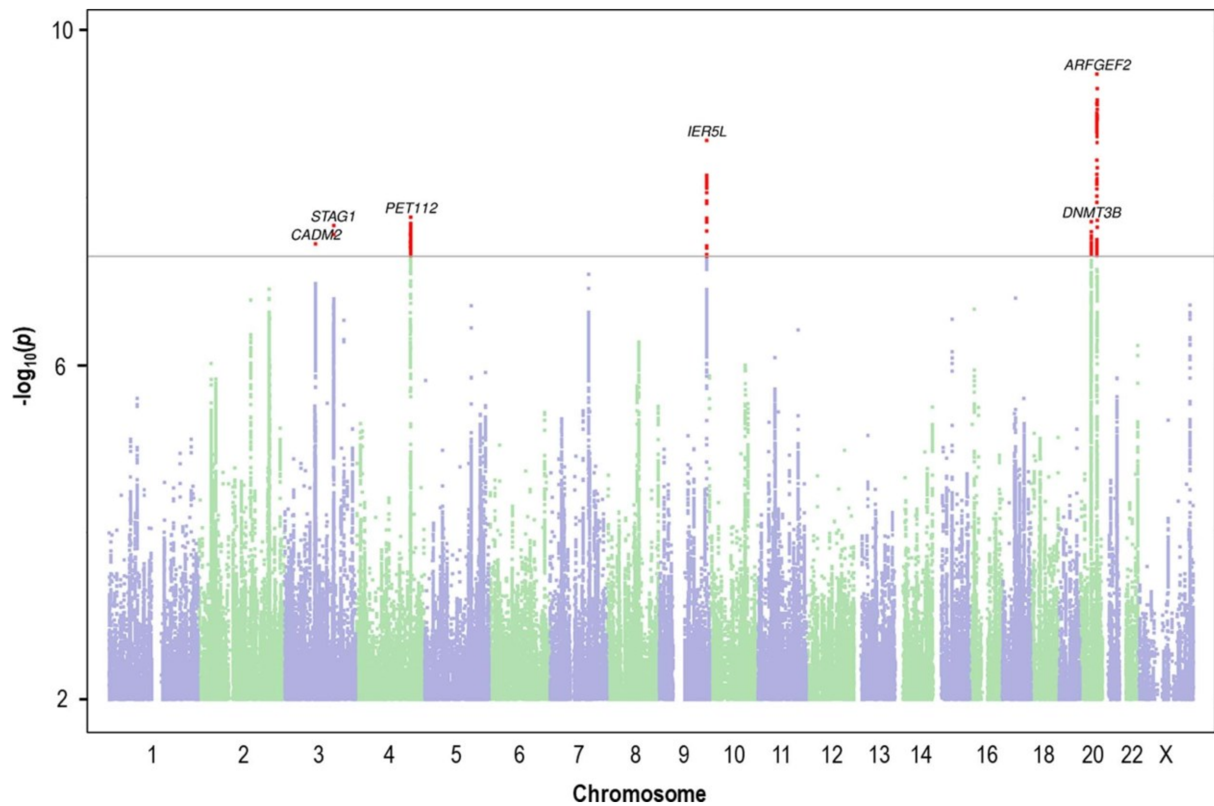
Extended Data Figure 1. **QQ plot of dyslexia GWAS results**

a-**c**, Quantile-quantile (Q-Q) plots of observed versus expected *P* values for associations of single nucleotide polymorphisms with self-reported dyslexia diagnosis in a genome-wide association analysis for all participants (*n* = 51,800 cases, 1,087,070 controls) (**a**), female participants (*n* = 30,287 cases, 641,016 controls) (**b**), and male participants (*n* = 21,513 cases, 446,054 controls) (**c**). The solid red line represents the distribution of *P* values under the null hypothesis, and the dashed red line represent 95% confidence intervals. The black circles represent the observed distribution of *P* values.



Extended Data Figure 2. **Manhattan plot of dyslexia GWAS results for females**

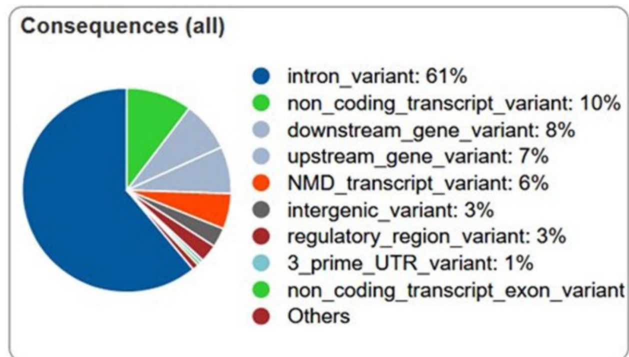
The *y*-axis represents the $-\log_{10} P$ value for association of single nucleotide polymorphisms with self-reported dyslexia diagnosis from 30,287 female individuals and 641,016 female controls. The threshold for genome-wide significance ($P < 5 \times 10^{-8}$) is represented by a horizontal grey line. Genome-wide significant variants in the 17 genome-wide significant loci are red. Variants located within a distance of 250 kb of each other are considered as one locus.

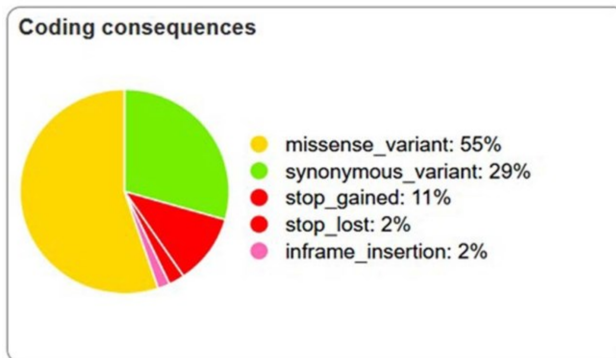


Extended Data Figure 3. **Manhattan plot of dyslexia GWAS results for males**

The *y*-axis represents the $-\log_{10} P$ value for association of single nucleotide polymorphisms with self-reported dyslexia diagnosis from 21,513 male individuals and 446,054 male controls. The threshold for genome-wide significance ($P < 5 \times 10^{-8}$) is represented by a horizontal grey line. Genome-wide significant variants in the 6 genome-wide significant loci are red. Variants located within a distance of 250 kb of each other are considered as one locus.

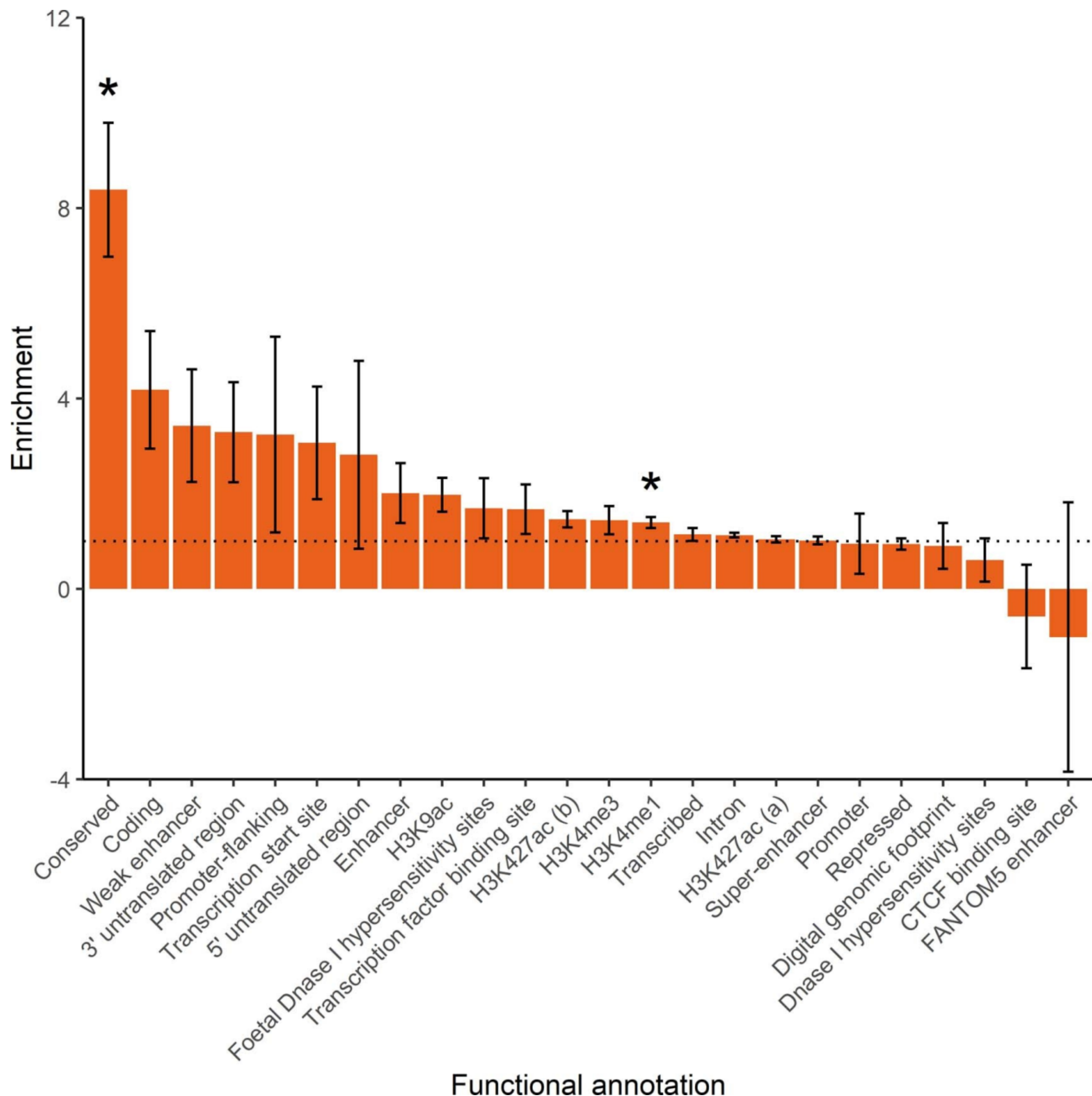
Category	Count
Variants processed	6210
Variants filtered out	0
Novel / existing variants	0 (0.0) / 6210 (100.0)
Overlapped genes	238
Overlapped transcripts	1176
Overlapped regulatory features	569





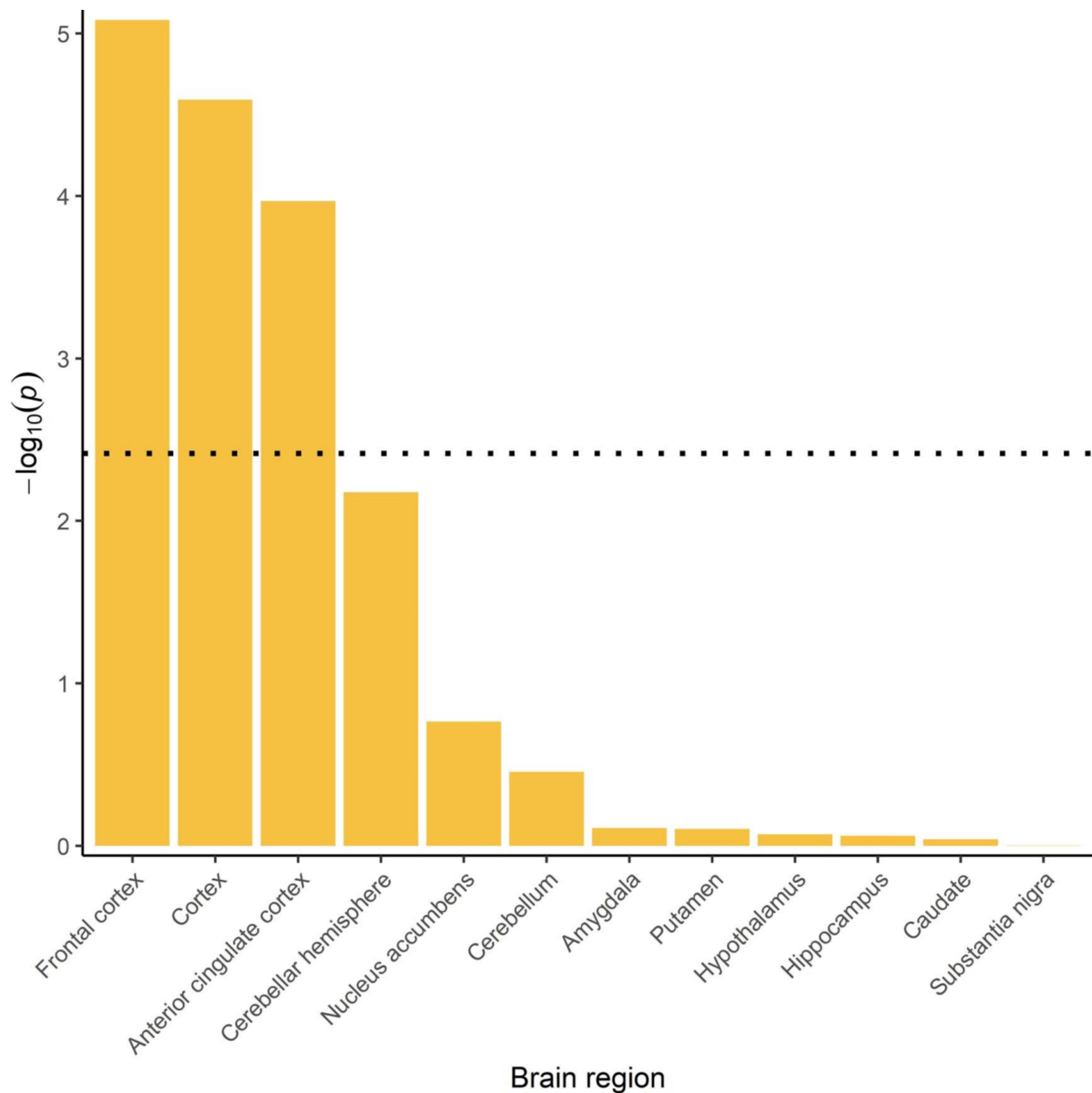
Extended Data Figure 4. **Variant effect predictor summary for the credible set of variants significantly associated with dyslexia**

Summary information is output from the online variant effect predictor in ENSEMBL (release 104). All our variants were present in the 1000 Genomes reference panel so are considered existing, and no pre-filtering (for example, on MAF; consequence type) was done.



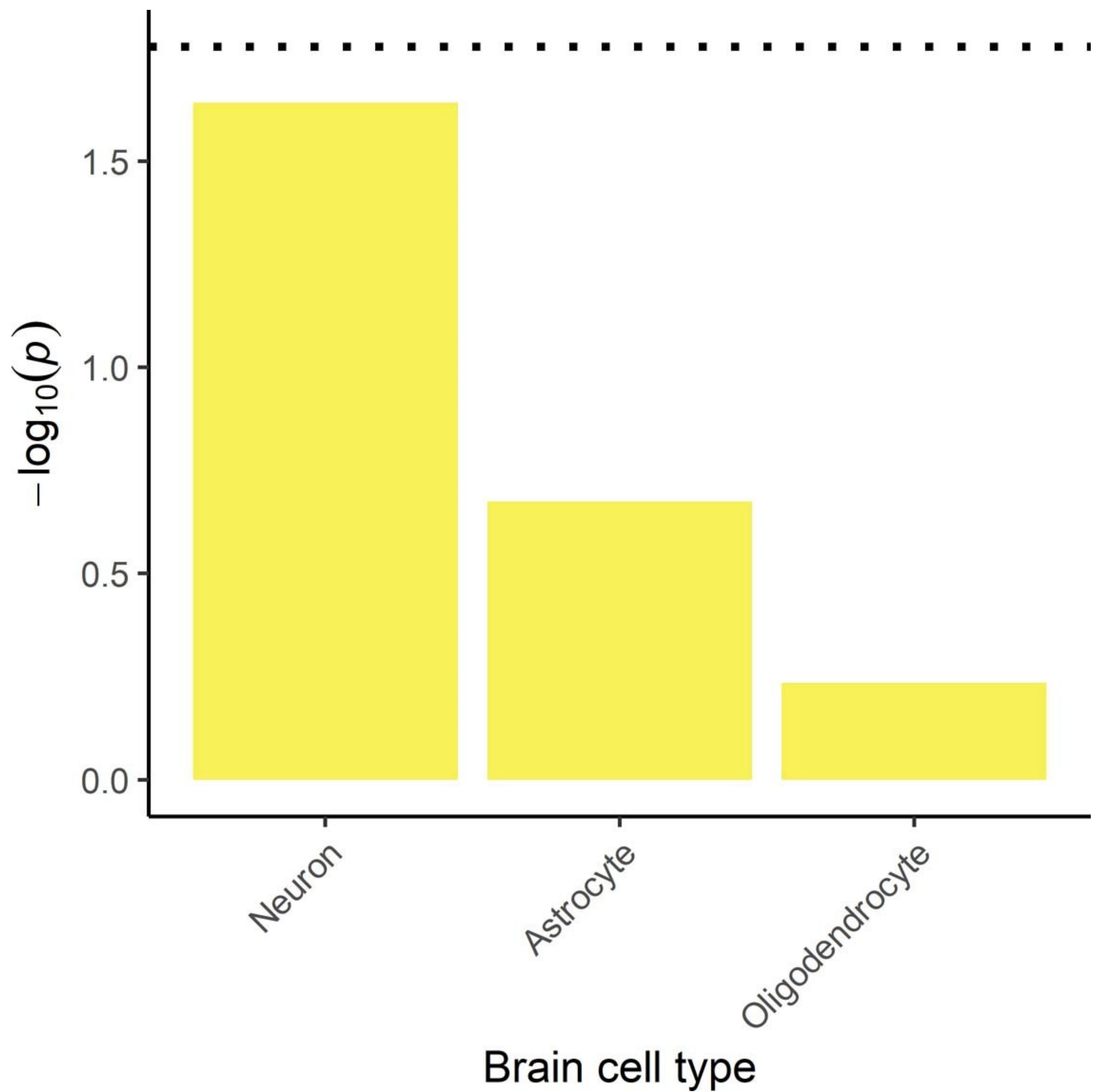
Extended Data Figure 5. **Enrichment estimates for major functional annotations**

The 24 major functional annotations were defined by Finucane et al.39. Enrichment is the proportion of h^2/proportion of SNPs. The horizontal dotted line indicates no enrichment (where proportion of h^2/proportion of SNPs = 1). Error bars represent standard errors of the enrichment estimates. Asterisks indicate enrichment estimates are significant based on a Bonferroni-derived P value of $< 2.08 \times 10^{-3}$ (for 24 tests). Exact values of enrichment statistic, standard error, and P value can be found in Supplementary Table 16.



Extended Data Figure 6. **Heritability of dyslexia partitioned by brain tissue gene expression**

The $-\log_{10} P$ value of the enrichment estimates for heritability of dyslexia for genes expressed in 12 brain regions. The horizontal dotted line indicates significance after Bonferroni correction for 12 tests ($P < 4.17 \times 10^{-3}$).

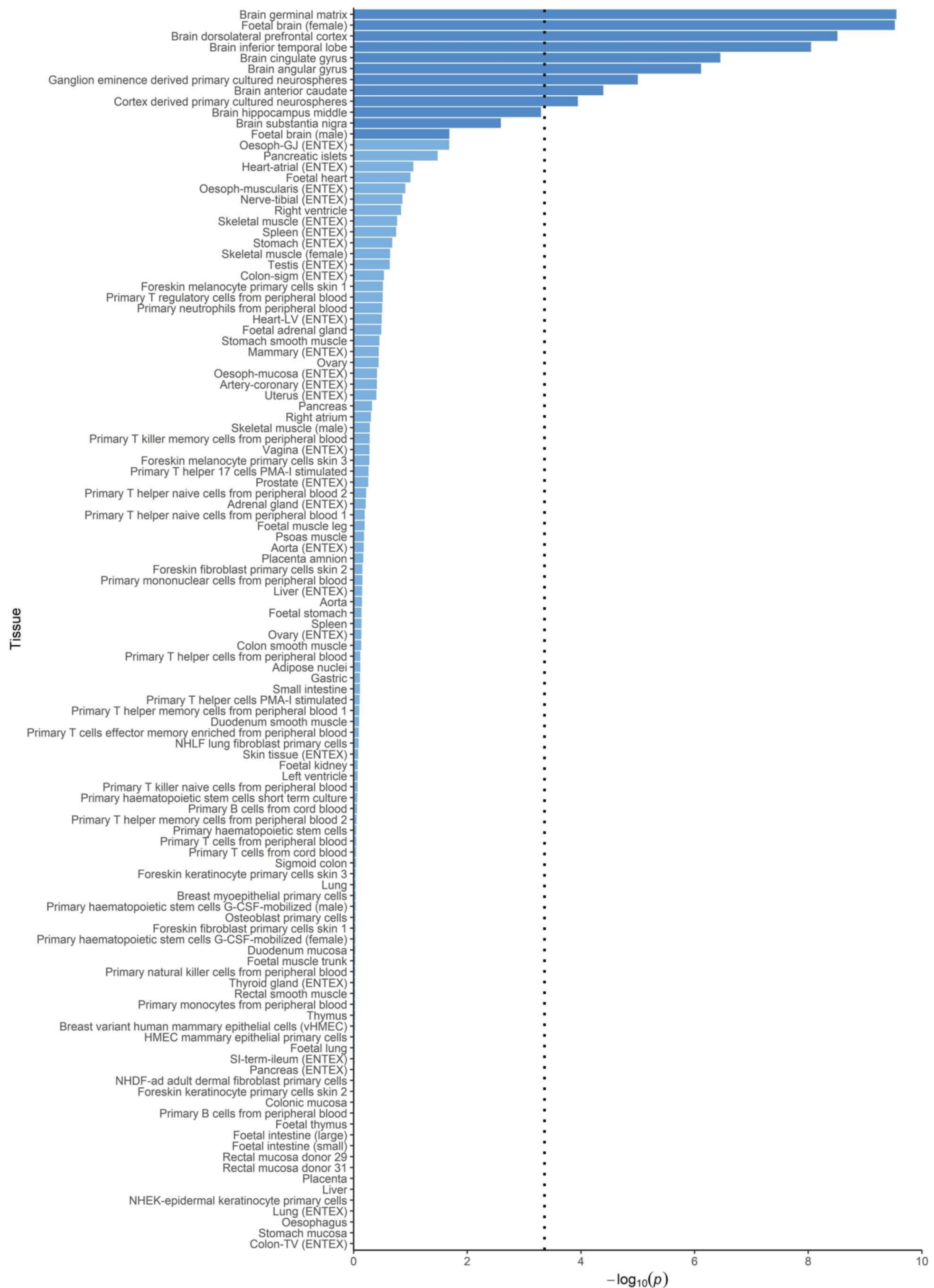


Extended Data Figure 7. **Heritability of dyslexia partitioned by brain cell type**

The -$\log_{10}$ *P* value of the enrichment estimates for heritability of dyslexia for brain cell types. The horizontal dotted line indicates significance after Bonferroni correction for three tests ($P < 1.67 \times 10^{-2}$).

Extended Data Figure 8. **Heritability of dyslexia partitioned by cell-type specific H3K4me1**

The $-\log_{10} P$ value of the enrichment estimates for heritability of dyslexia for variants located within H3K4me1 peaks of different tissues. Central nervous systems tissues are represented in dark green and other tissues are represented in light green. The vertical dotted line indicates significance after Bonferroni correction for 114 tests ($P < 4.39 \times 10^{-4}$).



Extended Data Figure 9. **Heritability of dyslexia partitioned by cell-type specific H3K4me3**

The $-\log_{10} P$ value of the enrichment estimates for heritability of dyslexia for variants located within H3K4me3 peaks of different tissues. Central nervous systems tissues are represented in dark blue and other tissues are represented in light blue. The vertical dotted line indicates significance after Bonferroni correction for 114 tests ($P < 4.39 \times 10^{-4}$).

Further Supplementary Material available at: https://www.nature.com/articles/s41588-022-01192-y - Sec31 and https://www.nature.com/articles/s41588-022-01192-y - Sec31.

Conclusion

In this chapter, I presented the 42 genome-wide significant loci associated with dyslexia, many of which could be replicated in recent well-sized GWAS meta-analyses (Eising et al., 2022; Gialluisi et al., 2020; Wang et al., 2023). SNP heritability was similar to other GWAS (Gialluisi et al., 2020), but smaller than estimates from twin studies (DeFries et al., 1987; Olson et al., 1989). Whilst this GWAS was substantially larger than all previous genetic studies of dyslexia, even larger sample sizes will be required to uncover more of the genetic factors associated with dyslexia. Future studies should also seek to be sufficiently well-powered to conduct GWAS in non-European ancestries, to ensure our understanding of the genetics of dyslexia is inclusive. Previous candidate genes and biological pathways did not reach genome-wide significance, suggesting their importance should be re-evaluated. However, a number of genome-wide significant SNPs were located in other neurodevelopmental genes (*AUTS2*, *TANC2*, *GGNBP2*). Further, no candidate gene sets/biological pathways reached genome-wide significance (although the axon guidance pathway received replication-level support), whilst more than half of the most significant biological pathways were novel pathways involved in neuronal processes. Therefore, future candidate genes and biological pathways should be designated through the unbiased, systematic approach applied here. Lastly, self-reported dyslexia was shown to be highly genetically correlated with quantitative measures of reading and spelling, suggesting dyslexia is not a distinct trait to normal variation in reading ability. Additionally, we showed dyslexia and ADHD were genetically correlated, in line with the multiple deficit model of developmental disorders.

In the next chapter, I will investigate the use of unselected adult cohorts for the discovery of genetic factors associated with reading ability.

CHAPTER 3: INVESTIGATING THE USE OF UNSELECTED ADULT SAMPLES FOR DISCOVERY OF GENETIC FACTORS ASSOCIATED WITH READING ABILITY

Introduction

In Chapter 2, I discovered novel genetic factors associated with dyslexia and demonstrated that dyslexia is highly genetically correlated with quantitative reading skills. To better understand the biological mechanisms that cause dyslexia, it is important to also uncover the genetics of normal variation in reading ability. Unfortunately, few very large genotyped samples, such as the one I used in Chapter 2, are currently available. Most samples with quantitative measures of reading skills are insufficiently powered for gene discovery. Part of the aim of my thesis, therefore, is to investigate methods to increase power to detect the genetic variants of small effect size that are associated with variation in reading ability. This chapter addresses Objective 3: to investigate whether unselected adult samples can boost power for GWAS of reading ability. Previous genetic studies of reading ability have largely been conducted using small samples of affected children and adolescents. Recently, Luciano et al., (2018) suggested adult samples may be valid for genetic studies of developmental disorders, using a proxy reading measure. Here, I will perform GWAS with validated reading and spelling measures in a new cohort of Australian adults. Since the sample is modest in size (n = 1,505), I will aim to replicate candidate genes and biological pathways to determine whether adult samples are of value for future, larger GWAS. The work for this chapter was published in the article which follows. My contribution included the phenotypic analyses, the statistical analyses and the gene-based tests.

Twin Research and Human Genetics (2020), **23**, 23–32
doi:10.1017/thg.2020.7

CAMBRIDGE
UNIVERSITY PRESS

Article

The Association of Dyslexia and Developmental Speech and Language Disorder Candidate Genes with Reading and Language Abilities in Adults

Catherine Doust[1], Scott D. Gordon[2], Natalie Garden[2], Simon E. Fisher[3,4], Nicholas G. Martin[2], Timothy C. Bates[1] and Michelle Luciano[1]

[1]Department of Psychology, The University of Edinburgh, Edinburgh, Scotland, UK, [2]Genetic Epidemiology Laboratory, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia, [3]Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands and [4]Donders Institute for Brain, Cognition and Behaviour, Nijmegen, the Netherlands

Abstract

Reading and language abilities are critical for educational achievement and success in adulthood. Variation in these traits is highly heritable, but the underlying genetic architecture is largely undiscovered. Genetic studies of reading and language skills traditionally focus on children with developmental disorders; however, much larger unselected adult samples are available, increasing power to identify associations with specific genetic variants of small effect size. We introduce an Australian adult population cohort (41.7–73.2 years of age, $N = 1505$) in which we obtained data using validated measures of several aspects of reading and language abilities. We performed genetic association analysis for a reading and spelling composite score, nonword reading (assessing phonological processing: a core component in learning to read), phonetic spelling, self-reported reading impairment and nonword repetition (a marker of language ability). Given the limited power in a sample of this size (~80% power to find a minimum effect size of 0.005), we focused on analyzing candidate genes that have been associated with dyslexia and developmental speech and language disorders in prior studies. In gene-based tests, *FOXP2*, a gene implicated in speech/language disorders, was associated with nonword repetition ($p < .001$), phonetic spelling ($p = .002$) and the reading and spelling composite score ($p < .001$). Gene-set analyses of candidate dyslexia and speech/language disorder genes were not significant. These findings contribute to the assessment of genetic associations in reading and language disorders, crucial for understanding their etiology and informing intervention strategies, and validate the approach of using unselected adult samples for gene discovery in language and reading.

Keywords: Dyslexia; developmental speech and language disorder; DLD; reading ability; language ability; *FOXP2*

(Received 13 November 2019; accepted 12 February 2020; First Published online 6 April 2020)

Reading ability is critical for achievement in school, which in turn impacts on success in adulthood (Ritchie & Bates, 2013). Both impairments and normal variability in reading and language capabilities are highly heritable (Bates et al., 2004; Harlaar et al., 2005), but little is known about the genetic architecture underlying these complex traits. Identifying the key genetic factors that contribute is important for understanding the etiology of reading and language disorders and therefore informing intervention strategies. A key to progress in molecular understanding is increased sample size of study cohorts. To date, most data in language disorder have come from affected samples, often of school age. These samples are modest in size, limiting power. By contrast, large genotyped samples of thousands of unselected adults are now being accumulated (e.g. UK Biobank), although collection of data on reading/language-related skills has seldom been prioritized. Here, we report the results of phenotyping a range of reading, spelling and language measures in an unselected adult sample of >1000 people, followed by testing for replication of prior associations, to validate this approach for future, large-scale studies of language-related traits and associated disorders.

A number of candidate genes for dyslexia and developmental language disorder (DLD; previously known as specific language impairment or SLD) have been identified through linkage mapping and targeted association (e.g. Francks et al., 2004; Meng et al., 2005; Nopola-Hemmi et al., 2001) and replicated in genetic association studies of children, adolescents and young adults (e.g. Bates et al., 2007; Bates et al., 2010; Scerri et al., 2011). However, many hundreds of quantitative trait loci (QTL) of small effect size (<1%) are likely to contribute to these complex, heterogeneous disorders, and much of the relevant genetic variance still remains unaccounted for (Bishop, 2015; Carrion-Castillo et al., 2013; Deriziotis & Fisher, 2017). Higher powered genome-wide association (GWA) studies derived from larger cohorts are needed to provide further validations of known candidates and to increase sensitivity for identifying new QTL.

Unselected adult cohorts are often orders of magnitude larger than even the largest case–control studies of children. Because

Authors for correspondence: Catherine Doust, Email: s1571212@sms.ed.ac.uk and Michelle Luciano, Email: michelle.luciano@ed.ac.uk

Cite this article: Doust C, Gordon SD, Garden N, Fisher SE, Martin NG, Bates TC, and Luciano M. (2020) The Association of Dyslexia and Developmental Speech and Language Disorder Candidate Genes with Reading and Language Abilities in Adults. *Twin Research and Human Genetics* **23**: 23–32, https://doi.org/10.1017/thg.2020.7

© The Author(s) 2020. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/thg.2020.7 Published online by Cambridge University Press

specific reading and language impairments are theoretically viewed as the extreme end of a continuum of normally varying ability (Leonard, 1991; Rodgers, 1983), selecting samples from the general population should remain sensitive for detecting relevant genetic factors — an expectation borne out in research on normal adolescents for both dyslexia and poor reading skill (e.g. Lind et al., 2010). Cognitive abilities have been shown to remain stable throughout life (Deary et al., 2000), and reading comprehension measured in adolescents explains ~80% of the variance in adult reading comprehension (Smith, 1993). Measures of reading ability taken even in adulthood may be as informative as adolescent measures. As maximal reading skill is not reached until the mid-20s, perhaps involving similar mechanisms to those that underlie increasing heritability of intelligence after childhood (McArdle et al., 2002), adult cohorts may provide even more sensitive tests of genetic (as opposed to environmental) variation than do child cohorts. It is currently not known whether general reading ability in adults is underpinned by the same genetic factors as in children with dyslexia.

To probe the utility of unselected adult cohorts, Luciano et al. (2018) tested a set of 14 dyslexia candidate genes originally associated with reading disability in children in a meta-analysis of two cohorts of older adults (mean age = 79 years). They found that the gene set was significantly associated with a reading index ($p = .016$) and that individual single nucleotide polymorphism (SNP) associations, although not significant, had allelic effects in the same direction as earlier studies. These results suggest that the same genetic factors underlying reading disability in children may contribute to variation in the normal range of reading ability in later life. However, the measures used to create the reading index in the unselected adults were not ideal. Specifically, Luciano et al. (2018) employed only two word recognition tests: the National Adult Reading Test (Nelson & Willison, 1991) and the Wechsler Test of Adult Reading (Holdnack, 2001), which both require pronunciation of irregular words. Performance on such tests is strongly influenced by vocabulary size, and since the latter is correlated with intelligence quotient (IQ), with these tasks it is hard to disentangle reading skill from general cognitive ability (Dykiert & Deary, 2013). Here, we report an association analysis of the same set of candidate genes in an unselected Australian adult sample using validated reading and language measures, including nonword reading to assess phonological processing, a core component of reading skill.

Our strategy for the present study was to identify adults who had already been genotyped across the genome in the context of earlier genome-wide association studies (GWAS) and to perform targeted phenotyping with reading, spelling and language measures. There were three main aims for the research: (1) to demonstrate the reliability and validity of the reading, spelling and language measures (see Table 1) in adults since such studies are uncommon; (2) to confirm in a middle-aged sample (mean = 58.7 years) that while skill may vary with age such variation is not a significant issue for gene finding and (3) to demonstrate the validity of using unselected adults to identify genetic factors associated with reading and language abilities.

Our long-term goal is to contribute to large-scale GWAS meta-analyses of speech, language and reading skills, given that genomic studies of these phenotypes are lagging behind those of other genetically complex traits (Deriziotis & Fisher, 2017). Since the cohort described here by itself lacks power for fully genome-wide investigations, for the current study, we focused attention on the most prominent genetic associations from the prior literature. Specifically, we analyzed a set of 14 genes that have been reported to show associations with dyslexia — *CMIP*, *CNTNAP2*, *CYP19A1*, *DCDC2*, *DIP2A*, *DXY1C1*, *GCFC2* (or *C2orf3*), *KIAA0319*, *KIAA0319L*, *MRPL19*, *ROBO1*, *PCNT*, *PRMT2* and *S100B* (see Luciano et al., 2018, for rationale). We also analyzed a set of five genes previously associated with language disorders of various kinds — *ATP2C2*, *CMIP*, *CNTNAP2*, *FOXP2* and *TM4SF20*. The rationale for selecting these five genes is as follows. Studies of nonword repetition in a DLD cohort collected by the UK SLI Consortium identified associations with SNPs in *ATP2C2* and *CMIP* (Newbury et al., 2009) as well as in *CNTNAP2* (Vernes et al., 2008). Nonword repetition was chosen for those studies (and also the present work) since it is a measure of phonological short-term memory that is often impaired in DLD (Gathercole et al., 1994; Newbury et al., 2005). Mutations in the *FOXP2* gene have been reported to segregate with severe speech and language disorders, mainly characterized by childhood apraxia of speech, in a large family pedigree (Lai et al., 2001), and additional *FOXP2* mutations have been found in independent cases with similar impairments (Morgan et al., 1993). *TM4SF20* was associated with early language delay in Southeast Asian families (Wiszniewski et al., 2013). In addition to analyses of the gene sets as a whole, we examined individual SNPs from within the relevant candidate genes that were previously reported to be associated with reading/language ability or impairment.

Finally, we included as a target the axon guidance pathway (GO:0007411: 'chemotaxis process that directs the migration of an axon growth cone to a specific target site'; 216 genes) and the neuron migration pathway (GO:0001764: 'movement of an immature neuron from germinal zones to specific positions where they will reside as they mature'; 214 genes) which have both been suggested to be implicated in dyslexia (Poelmans et al., 2011), although see Guidi et al. (2018) for a critical review.

Materials and Methods

Participants

In 2017, we recruited participants from earlier twin studies at the QIMR Berghofer Medical Research Institute in Australia. The final cohort consisted of 1550 participants (78.06% female), 1505 of whom had previously been genome-wide genotyped using SNP arrays and were living in Australia. Ages ranged from 41.7 to 73.2 years (mean = 58.7, $SD = 7.8$). Self-report data on dyslexia, DLD and related traits were collected in all 1505 participants (including 227 sibling pairs, 76 of whom were monozygotic (MZ) twins). Reading and language test data were collected in 1112 participants (including 197 sibling pairs and 70 MZ twins). All participants were free from neurological conditions and major psychiatric illness at the time of testing.

Genotyping

Participants had been genotyped on standard Illumina SNP arrays, the chip model varying, and merged after quality control (QC; including Mendelian checks, as data are typically family based). Within each batch, and across batches, sample errors or failures were identified using sex and relatedness tests, and either corrected or removed as appropriate. Samples were also removed if they were below a 97% call rate or (at a later stage post-merging) of non-European ancestry as judged from nonclustering with known European populations in a principal component analysis (PCA). Markers in a batch were dropped due to Illumina-recommended QC filters (e.g. GenTrain score), as well as: (1) there were issues

https://doi.org/10.1017/thg.2020.7 Published online by Cambridge University Press

Table 1. Full assessment battery administered to the Brisbane adult cohort

Measure	Test details/response options	Purpose
CC2A	Reading and correct pronunciation of 55 regular and 55 irregular words and 55 nonwords	Regular word reading assesses general reading skill Irregular word reading assesses lexical reading only Nonword reading assesses phonological processing only
Spelling	Spell 36 regular and irregular words	Regular word spelling assesses general spelling skill Irregular word spelling specifically assesses lexical spelling skill
Phonetic spelling	Spell 18 regular and irregular words 'as it sounds'	Assesses non-lexical spelling skill
Gathercole and Baddeley Nonword Repetition Test	Repeat 40 nonwords of 2, 3, 4 or 5 syllables back to the interviewer	Assesses the phonological loop, a component of learning language
Self-reported book reading	Frequency of reading books (not newspapers or magazines): daily, weekly, monthly, yearly or never	Correlates with reading ability (Smith, 1996)
Self-reported reading or language difficulties	Yes/no If yes, was it self-evaluated or noticed by others, and what is the formal diagnosis (if any)?	
Child with reading or language difficulties	Yes/no If yes, was it self-evaluated or noticed by others, and what is the formal diagnosis (if any)?	
Age	Years	Covariates
Sex	Female/male	
School years	Highest year of school completed (year 6 and below to year 12)	
Higher education	Yes/no	
Qualifications	Any obtained	
Self-estimated IQ	Do not know/well below average/low average/average/high average/superior/very superior	
Hearing difficulties	Yes/no	
Stutter	Yes/no	May share genetic etiology with reading and language abilities
Developmental delay	Yes/no	
Dyspraxia	Yes/no	
ASD	Yes/no	
Attention deficit hyperactivity disorder	Yes/no	
Tourette syndrome	Yes/no	
Eating disorder	Yes/no	
Obsessive–compulsive disorder	Yes/no	
Depression	Yes/no	
Bipolar disorder	Yes/no	
Social anxiety	Yes/no	
Personality disorder	Yes/no	

CC2A = Castles and Coltheart Test 2 Adults, IQ = intelligence quotient, ASD = Autism Spectrum Disorder.

with map placement or strand alignment in a Basic Local Alignment Search Tool search of primers; (2) call rate was <95%; (3) $p < 1 \times 10^{-6}$ in Hardy–Weinberg equilibrium tests; (4) minor allele frequency (MAF) was <1%; (5) (for chromosome X) male heterozygosity <%; (6) for older chips, there was a low mean GenCall score, <0.7 (Duffy et al., 2018; Medland et al., 2009).

Data were imputed to the Haplotype Reference Consortium reference panel version r1.1 (Haplotype Reference Consortium et al., 2016) and SNPs with a MAF of <.05 and an imputation accuracy of <.8 were excluded. Imputed genotypes were taken from three imputation runs (each using Eagle for phasing and minimac3 (autosomes) or minimac4 (chromosome X) for imputation, on the University of Michigan Imputation Server). Each run used individuals genotyped in a specific chip family, one of (1) the oldest HapMap-based Illumina chips; (2) GSA chips; (3) Omni and Core+Exome/PsychArray chips; and observed markers passing QC for all corresponding batches of genotyping. The three imputation runs were then merged by taking (for each individual) preferentially (1), (2) or (3) in that order (as this generally corresponds to the best-quality imputation).

https://doi.org/10.1017/thg.2020.7 Published online by Cambridge University Press

The breakdown of chip models is (1) HapMap-based: 610 K-quad ($n = 427$), 660 K/670 K-quad ($n = 213$), CNV370 ($n = 399$), 317 K ($n = 63$); (2) GSA: GSA Avera ($n = 2$); (3) Core+Exome ($n = 152$), PsychArray ($n = 65$), Omni2.5 ($n = 34$), OmniExpress ($n = 4$).

Measures and Procedure

An approach email was sent to participants with a link to the detailed information sheet and online consent form. They were then directed to a brief self-report questionnaire, which included education, how frequently they read books (excluding magazines and newspapers), their estimated IQ, whether or not they or their child has a reading or language disorder and whether or not they have a range of other behavioral or psychiatric conditions (Table 1). Within two weeks of completion of this survey, eligible participants who provided informed online consent were contacted for a telephone interview. At the time of the interview, participants were then emailed an online link to access the tests.

Three tests were administered: the Castles and Coltheart Test 2 Adults (CC2A) reading test (Castles & Coltheart, 1993), the Gathercole and Baddeley Nonword Repetition Test (Gathercole et al., 1994) and a spelling test, including phonetic spelling (Table 1). CC2A requires the reading aloud and correct pronunciation of 55 each of regular words, irregular words and nonwords. Irregular word reading assesses the lexical route of reading while nonword reading specifically assesses phonological processing. Similarly, our spelling test includes 22 regular and 14 irregular words plus a phonetic spelling task to spell 18 irregular words 'as they sound' to assess phonological processing. Gathercole et al.'s (1994) task of nonword repetition measures language ability as related to phonological encoding and memory. Data for one individual with 10 missing items were excluded from the nonword repetition task.

Statistical Analyses

Multiple regression was used to predict each of the reading, spelling and language outcome measures from age, sex and hearing difficulties. Hearing difficulties included any respondent who reported hearing difficulties or use of a hearing aid, or any respondent identified as having hearing difficulties by the interviewer. We used residual scores in further analyses. A unitary reading and spelling ability measure was created based on the scores on the first principal component (PC) of a PCA of the regular and irregular word and nonword reading, and regular, irregular and phonetic word spelling measures.

GWA results were generated for each of five variables (the reading and spelling PC, nonword reading, phonetic spelling, nonword repetition and self-reported reading impairment). This was undertaken using the Genome-wide Complex Trait Analysis software (Yang et al., 2011), which can account for family relatedness. Where more than one member of an MZ twin pair had been assessed, only one member was selected at random for the analyses, with final genetic association samples of 1425 for self-reported reading impairment, 1290 for the reading and spelling PC, 1293 for nonword reading and 1292 for phonetic spelling and nonword repetition.

Using the available summary statistics from the above, gene-set analysis was performed for four gene sets: dyslexia candidate genes ($N = 14$), speech/language disorder candidate genes ($N = 5$), the axon guidance pathway (gene ontology (GO) term GO:0007411; 216 genes) and the neuron migration pathway (GO:0001764; 145 genes). Individual gene-based analysis was also performed for genes within the dyslexia and speech/language disorder candidate sets. Bonferroni correction derived a critical p value of .003. Analyses were performed using MAGMA (de Leeuw et al., 2015) to test for overrepresentation of significantly associated SNPs within each set and within each candidate gene. Bonferroni correction for multiple testing was too conservative because the candidate gene sets wholly overlapped with the biological pathway gene sets, so an effective number of independent tests of 2 were used to derive an adjusted critical p value of .025.

Within the candidate gene sets, 77 SNPs had previously been associated with reading or language ability or disability, or were variants identified through fluorescence *in situ* hybridization and SNP microarray analysis of a small deletion at 21q22.3 segregating with dyslexia in a family (see Supplementary Material). A total of 68.7 independent tests were derived through matrix spectral decomposition, taking into account linkage disequilibrium (Nyholt, 2004). Bonferroni correction gave an adjusted α level of 7.28×10^{-4}.

Results

Phenotypic Analyses

The distributions of the raw reading, spelling and language test scores were slightly negatively skewed. We used multiple regression to predict each of the outcome measures from age, sex and hearing difficulties (Supplementary Table S1) with the resulting residual scores (used for genetic association analysis) normally distributed. The multiple regression results indicate that age-squared was only a significant predictor for nonword repetition ($\beta = -0.00$, $p = .042$). Females scored higher than males for regular word reading ($\beta = -0.45$, $p = .006$), nonword repetition ($\beta = -0.96$, $p = .004$), regular word spelling ($\beta = -0.03$, $p < .001$) and irregular word spelling ($\beta = -0.61$, $p < .001$). Hearing difficulties were associated with worse irregular word reading ($\beta = -1.36$, $p = .029$), nonword reading ($\beta = -3.23$, $p < .001$), nonword repetition ($\beta = -0.96$, $p = .004$) and regular word spelling ($\beta = -0.04$, $p = .023$) performance. Outliers were set to a trimmed minimum of negative four standard deviations.

Table 2 shows a correlation matrix of raw reading, spelling and language scores and covariates. Minimum and maximum values, means and standard deviations for each of the variables are in Table 3, while Table 4 gives the frequencies of discrete variables. Frequent book reading correlates with higher scores on reading and spelling tasks but not with nonword repetition. More years at school is correlated with higher scores in all reading, spelling and language tasks. Self-report of a reading impairment is associated with lower scores in reading and spelling tasks but bears no relationship with nonword repetition. Self-reported language impairments do not correlate with any task, including nonword repetition.

In the PCA of reading and spelling scores, a scree plot of the eigenvalues shows the first PC is sufficient to explain the majority of variation (63.1%) in reading and spelling skills (Supplementary Figure S1).

Genetic Association Results

Quantile–quantile plots of the expected distribution of p values across SNPs within the dyslexia and speech/language disorder candidate gene sets (Supplementary Figure S2) demonstrate a slight positive deviation from the null distribution, indicative of genetic

https://doi.org/10.1017/thg.2020.7 Published online by Cambridge University Press

Table 2. Intercorrelations (Pearson's *r*) between reading, spelling and language measures and covariates

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Age (*n* = 1550)	–													
2. Sex (*n* = 1550)	.10*	–												
3. School years (*n* = 1550)	−.07*	.05*	–											
4. Book reading (*n* = 1550)	.08*	−.14*	.11*	–										
5. Reading impairment (*n* = 1520)	−.04	.04	−.06*	−.14*	–									
6. Language impairment (*n* = 1470)	−.02	.00	−.02	.01	.04	–								
7. Hearing difficulties (*n* = 1532)	.09*	.06*	−.06*	−.01	.02	.00	–							
8. Regular reading (*n* = 1386)	.06*	−.07*	.15*	.25*	−.28*	−.09*	−.03	–						
9. Irregular reading (*n* = 1386)	.10*	−.03	.20*	.26*	−.23*	−.07*	−.05	.61*	–					
10. Nonword reading (*n* = 1386)	.02	−.04	.16*	.20*	−.30*	−.09*	−.09*	.69*	.61*	–				
11. Regular spelling (*n* = 1384)	.14*	−.10*	.15*	.26*	−.26*	−.06*	−.06*	.56*	.53*	.57*	–			
12. Irregular spelling (*n* = 1376)	.07*	−.13*	.13*	.25*	−.34*	−.04	−.02	.59*	.56*	.59*	.64*	–		
13. Phonetic spelling (*n* = 1385)	.04	−.02	.15*	.11*	−.17*	−.06*	−.04	.50*	.40*	.62*	.45*	.41*	–	
14. Reading and spelling PC (*n* = 1345)	.00	.00	.20*	.26*	−.33*	−.09*	.00	.83*	.78*	.86*	.77*	.79*	.70*	–
15. Nonword repetition (*n* = 1385)	−.15*	−.10*	.11*	.08*	.00	−.02	−.28*	.19*	.22*	.26*	.21*	.14*	.22*	.25*

Note: Correlations are not adjusted for case nonindependence.
PC = principal component.
*$p \leq .05$; sex (males), impairment and hearing difficulties are coded positively.

Table 3. Summary statistics of the main continuous variables including reading, spelling and language measures

Variable	Minimum	Mean (*SD*)	Maximum
Regular word reading	70.91	95.19 (4.55)	100.00
Irregular word reading	50.91	85.31 (7.75)	100.00
Nonword reading	24.07	83.78 (12.49)	100.00
Regular word spelling	28.57	90.09 (10.57)	100.00
Irregular word spelling	21.43	89.90 (12.89)	100.00
Phonetic spelling	0.00	71.08 (20.13)	100.00
Nonword repetition	0.00	75.57 (13.04)	100.00
Age	41.69	58.70 (7.79)	73.19
School years	0.00	11.18 (1.81)	12.00

Note: For reading, spelling and language scores, summary statistics are calculated from the percentage of correct items.

signal, for phonetic spelling for dyslexia candidate gene subset of SNPs (Supplementary Figure S2(b)) and for nonword repetition for both dyslexia and speech/language disorder candidate gene SNP subsets (Supplementary Figure S2(e) and (f)). In gene-based analyses (Table 5), *FOXP2* was associated with nonword repetition ($p < .001$), phonetic spelling ($p = .002$) and the reading and spelling composite score ($p < .001$), withstanding a corrected α level of .003. For nonword repetition, *FOXP2* was in the top three most significant genes.

Gene-set analysis of the neuron migration pathway revealed a nominal association with the reading and spelling composite score ($p = .037$; Table 6), which did not survive correction for multiple testing, and gene-set analyses of 14 candidate dyslexia genes, five candidate speech/language disorder genes and the axon guidance pathway were also not significant.

Of the SNPs within the dyslexia and speech/language disorder candidate gene sets, 77 have previously been reported to be associated with reading or language ability or impairment (Supplementary Tables S2–S6). None were close to the corrected significance level of $p < 7.28 \times 10^{-4}$.

Discussion

In this study, we introduced a new population sample of previously genotyped adults for whom we have recently obtained reading and language measures. Our aim was to determine the validity of using unselected adults to identify genetic factors associated with reading and language abilities. We demonstrate the suitability of the reading and language measures to determine ability among unimpaired adults, and we confirm that age is not a confounder. Notably, there was no association between age and the most sensitive index of reading skill, namely phonological decoding (assessed through nonword reading). In our adult population, we observed associations at the gene-based level for candidate genes that have previously been implicated in dyslexia or speech/language disorders in children and adolescents; for example, finding that variation in *FOXP2* (a gene implicated in a monogenic form of speech apraxia) was associated with nonword repetition. Further, in gene pathway analyses, we find some support for associations of genes involved in neuronal migration with reading skill, albeit at a nominal level of significance that does not survive multiple-testing adjustment.

Establishing sensitive measures of adult reading and language abilities is crucial because individuals with an impairment may develop coping strategies over the life course. We demonstrated that the CC2A reading task and our spelling task, which included reading nonwords and phonetic spelling, correlate with how often individuals read books. Reading books, more so than other forms of print, is associated with higher literacy proficiency (Smith, 1996). Further, performance on the reading, spelling and language measures correlated with the number of school years individuals completed, supporting the known association between educational achievement and reading and language abilities (Garnier et al.,

https://doi.org/10.1017/thg.2020.7 Published online by Cambridge University Press

Table 4. Percentage frequencies of major discrete variables

Variable	Frequency (%)
Sex	
Female	78.06
Male	21.94
Higher education	84.38
Frequency of book reading	
Daily	29.29
Weekly	15.42
Monthly	13.55
Yearly	29.55
Never	12.19
Hearing difficulties	1.50
Reading impairment	5.26
Reading impairment evaluation	
Self-evaluated	40.00
Noticed by others (e.g. teachers)	53.75
Clinically diagnosed	6.25
Child with reading impairment	5.89
Stutter	
No	97.21
Sometimes	2.65
Yes	0.14
Other language impairment	1.90
Language impairment evaluation	
Self-evaluated	39.29
Noticed by others (e.g. teachers)	39.29
Clinically diagnosed	21.43
Child with language impairment	5.00

1997; Snowling et al., 2001). We also found that the reading and spelling scores in our cohort correlated with whether individuals self-reported a reading impairment but not with self-report of a language impairment. Unexpectedly, nonword repetition scores showed no relationship to self-report of a language impairment, even though individuals with DLD are less able to acquire phonological forms of new words (Gathercole, 2006; Newbury et al., 2005). We may be statistically underpowered to detect a relationship due to the low frequency of reports of language impairments (1.9%) in our modestly sized cohort, and as such this could be a type II error. Alternatively, this result may reflect an ascertainment bias in addition to the unreliability of self-reported measures for accurately measuring true frequencies of learning disabilities, particularly considering historical context: the youngest members of this cohort were born in 1975, six years before a standard set of criteria for diagnosing DLD (formerly known as SLI) existed (Stark & Tallal, 1981). The population frequency of language deficits not attributable to hearing impairment, low nonverbal intelligence or neurological damage is estimated to be closer to 7% (Leonard, 2014).

In gene-based analyses of prior candidate genes from the dyslexia and speech/language disorder literature, we identified associations with several reading, spelling and language measures in our cohort of largely unimpaired adults. A discussion of the individual SNP results can be found in the Supplementary Material. Variation in *FOXP2* was associated with nonword repetition as well as phonetic spelling, and a reading/spelling composite score. *FOXP2* (Forkhead Box P2) encodes a transcription factor involved in the development of the brain (among other tissues) and acts through regulating hundreds of genes (Fisher & Scharff, 2009). The gene was first identified through positional cloning studies of a severe speech and language disorders involving childhood apraxia of speech in a large multigenerational pedigree (Fisher et al., 1998). All affected cases in this family were found to carry a missense mutation in the DNA-binding domain of the encoded protein, and a translocation disrupting *FOXP2* was discovered in an unrelated individual with a similar disorder (Lai et al., 2001). Subsequently, additional rare protein-coding changes (including both missense and nonsense mutations) have been identified as causes of developmental speech and language disorders in multiple independent families and cases (MacDermot et al., 2005; Morgan et al., 1993; Reuter et al., 2017). Despite robust evidence implicating rare disruptions of *FOXP2* in severe speech and language deficits across independent studies, the contributions of common variation in this gene to language-related phenotypes remain open to debate (see Uddén et al., 2019). For example, in one of the largest prior studies to assess this issue, Mueller et al. (2016) tested for a relationship of 13 SNPs in *FOXP2* and language ability in a modestly sized population cohort of children ($N = 812$) and found no significant associations. Given our contradictory findings in the present study, further investigations using robust measures in larger samples of adults and children are warranted to resolve this long-standing question. Of note, in a recent meta-GWAS of >20k individuals diagnosed with attention deficit/hyperactivity disorder (ADHD) compared to >35k controls, SNPs in *FOXP2* were among the top genome-wide significant hits, which is intriguing in light of the known overlaps between ADHD and reading disabilities (Demontis et al., 2019).

ATP2C2 (ATPase secretory pathway Ca^{2+} transporting 2) catalyzes ATP hydrolysis coupled with calcium transportation. The gene was identified as a candidate for involvement in DLD susceptibility by the SLI Consortium (2002) following an early linkage study of families with DLD probands that included nonword repetition as a quantitative measure. In targeted analyses of the linkage region, SNPs in *ATP2C2* were found to be associated with both nonword repetition and reading measures in language-impaired individuals, but not in an unselected cohort (Newbury et al., 2011; Newbury et al., 2009). In the present study, we detected association of *ATP2C2* with nonword repetition, phonetic spelling, nonword reading and the reading/spelling composite score, although the significance levels were not robust to multiple-testing adjustment.

We also detected nominally significant associations of the dyslexia candidate genes *MRPL19* (with phonetic spelling) and *S100B* (with nonword repetition). *MRPL19* (mitochondrial ribosomal protein L19) encodes a ribosomal subunit and is involved in protein synthesis. A risk haplotype in a locus containing *MRPL19* and *C2ORF3* was associated with dyslexia in Finnish families and replicated in a German sample (Anthoni et al., 2007). Heterozygous carriers of the risk haplotype had reduced expression of both genes. *MRPL19* expression correlates with that of dyslexia candidate genes *DCDC2*, *DYX1C1*, *KIAA0319* and *ROBO1*; however, the NeuroDys study of 900 individuals with dyslexia across eight countries failed to replicate the effects of *MRPL19* (Becker et al., 2014).

https://doi.org/10.1017/thg.2020.7 Published online by Cambridge University Press

Table 5. Gene-based analysis of dyslexia and speech/language disorder candidate genes for association with measures of reading, spelling and language

				Nonword reading		Phonetic spelling		Reading and spelling PC		Self-reported reading impairment		Nonword repetition	
Gene set	Gene	Chr	SNPs (n)	*z*	*p*	*z*	*p*	*z*	*p*	*z*	*p*	*z*	*p*
Dyslexia	*KIAA0319L*	1	11	0.08	.47	0.31	.38	−0.34	.63	−0.98	.84	−0.87	.81
	GCFC2	2	81	0.37	.36	1.62	.053	0.33	.37	0.013	.49	0.76	.22
	MRPL19	2	38	0.54	.29	1.86	.031	0.75	.23	0.15	.44	1.23	.11
	ROBO1	3	1642	0.76	.22	0.83	.20	−1.26	.90	0.14	.44	0.29	.39
	DCDC2	6	467	0.54	.29	−0.43	.66	1.19	.12	−1.55	.94	1.38	.08
	KIAA0319	6	210	−0.59	.72	1.25	.11	−0.77	.78	0.22	.41	−1.03	.85
	CYP19A1	15	217	−0.54	.71	−0.35	.64	0.71	.24	−0.62	.73	−0.23	.59
	DYX1C1	15	204	0.61	.27	1.01	.16	0.31	.38	0.40	.35	−0.092	.54
	DIP2A	21	318	−0.44	.67	0.17	.43	−0.81	.79	0.85	.20	−0.13	.55
	PCNT	21	330	−1.14	.87	−0.58	.72	−1.79	.96	0.66	.25	−0.98	.84
	PRMT2	21	4	−0.21	.58	0.093	.46	0.87	.19	0.40	.35	0.69	.25
	S100B	21	25	−0.90	.82	1.03	.15	1.51	.065	0.89	.19	1.69	.046
Dyslexia and speech/ language disorders	*CNTNAP2*	7	5389	0.028	.49	−0.41	.66	−1.21	.89	−0.52	.70	−1.93	.97
	CMIP	16	736	−0.23	.59	−0.27	.61	−2.45	.99	0.39	.35	−0.95	.83
Speech/language disorders	*TM4SF20*	2	49	−1.13	.87	−0.35	.64	−1.13	.87	−0.74	.77	−1.09	.86
	FOXP2	7	547	0.44	.33	2.94	**.002**	3.78	**<.001**	−0.46	.68	3.67	**<.001**
	ATP2C2	16	494	2.23	.013	1.67	.048	1.74	.041	−0.20	.58	2.19	.014

Note: Bold type indicates nominal significance.
Chr = chromosome.

Table 6. Gene-set analysis of dyslexia and speech/language disorder candidate gene sets and neuron migration and axon guidance pathways for association with measures of reading, spelling and language

Gene set	Variable	*B*	β	*SE*	*p*
Dyslexia candidates (*N* = 14)	Nonword reading	−0.14	−0.00	0.23	.73
	Nonword repetition	0.08	0.00	0.24	.37
	Phonetic spelling	0.14	0.00	0.23	.28
	Reading and spelling PC	−0.26	−0.01	0.24	.86
	Self-reported reading impairment	0.09	0.00	0.23	.35
DLD candidates (*N* = 5)	Nonword repetition	−0.01	−0.00	0.36	.51
Neuron migration pathway (*N* = 145)	Nonword reading	0.06	0.01	0.08	.21
	Nonword repetition	−0.05	−0.00	0.08	.72
	Phonetic spelling	0.04	0.00	0.08	.31
	Reading and spelling PC	0.14	0.01	0.08	**.037**
	Self-reported reading impairment	0.08	0.01	0.08	.13
Axon guidance pathway (*N* = 216)	Nonword reading	−0.04	−0.00	0.06	.73
	Nonword repetition	0.00	0.00	0.06	.47
	Phonetic spelling	−0.05	−0.10	0.06	.82
	Reading and spelling PC	0.04	0.00	0.06	.26
	Self-reported reading impairment	0.01	0.00	0.06	.45

Note: Bold type indicates nominal significance.
PC = principal component, DLD = developmental language disorder.

S100B (S100 Calcium Binding Protein B) is involved in neurite outgrowth and neuronal migration (Huttunen et al., 2000; Poelmans et al., 2011) and was identified as one of four genes in a deleted region co-segregating with dyslexia in a family (Poelmans et al., 2009). A noncoding variant was later associated with spelling in German families (Matsson et al., 2015), but no other studies have identified the gene in association with language ability or impairment.

https://doi.org/10.1017/thg.2020.7 Published online by Cambridge University Press

An association between genes in the neuron migration pathway and the reading and spelling composite score supports proposals from Galaburda et al. (2006), Paracchini et al. (2007) and Poelmans et al. (2011), who hypothesized that dyslexia candidate genes are part of a molecular network that regulates neuronal migration and neurite outgrowth. A more recent review from Guidi et al. (2018) critically evaluated this hypothesis and suggested that there is a lack of robust evidence supporting the theory. We did not find an association of the neuron migration pathway with any measure other than our composite score, nor was the axon guidance pathway significant in our study. However, the GO terms defining these pathways are incompletely annotated and continue to expand. At the time of publishing a previous paper (Luciano et al., 2018), the neuron migration pathway contained 103 genes and the axon guidance pathway contained 203 genes, in comparison to 145 and 216 genes, respectively, at present. In the previous paper, no significant associations were found for either pathway, but here we detected an association with a reading and spelling score, albeit not robust to correction for multiple testing, highlighting the potential value of continuing to probe these pathways in their possible link to dyslexia as they are annotated with increasing resolution.

Our failure to replicate previous genetic associations may be due to a lack of statistical power to detect genetic variants of small effect size, may represent true null associations and, further, findings from prior studies could be false positives. Here, we had 78.43% power to find an effect size minimum of 0.005 (calculated using the Genetic Power Calculator; Purcell et al., 2003). Further, our participants were recruited through a twin registry, which may be subject to a sampling bias: frequencies of self-reported reading (5.26%) and language (1.90%) impairments were below the estimated population frequencies (10% and 7%, respectively). Variants may have stronger effects at the tail end of ability or in individuals with an impairment, and hence greater statistical power is required to replicate them in unselected populations compared with case–control studies. Future meta-analyses and larger GWA studies of both selected and unselected cohorts of children and adults will enable stronger conclusions to be drawn about the genetic influences on reading acquisition and continuity of reading skill over the life course and their relationship to reading disorder. In this study, we were unable to disentangle general cognitive ability from reading and language skills, which are highly correlated traits. The inclusion of an IQ test as a covariate in future studies would enable better isolation of specific reading and language abilities.

This study introduces an important new population cohort of genotyped adults with validated measures of reading and language abilities. We also measured a self-reported binary status on a range of comorbidities of dyslexia and DLD, including stutters, autism spectrum disorder (ASD) and ADHD. We have shown that at least some candidate genes associated with dyslexia and speech/language disorders in children and adolescents may show effects in unselected adult populations, demonstrating the potential of such resources (when suitably scaled-up) for the discovery of novel genetic variants associated with reading and language traits. Future studies should aim to conduct large-scale GWA analyses and meta-analyses of unselected adults to identify genetic variants that are associated with measures of reading and language abilities, accounting for general cognitive ability where possible. Analyses of relevant continuous traits in unselected populations generalize to other learning impairments and neurological traits. Ultimately, uncovering the genetic etiology of developmental disorders will enable early diagnosis and appropriate intervention.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/thg.2020.7.

Acknowledgments. We would like to thank the research interviewers and the study participants. We are grateful to Marie-Christine Opitz for data cleaning assistance.

Financial support. S.E.F. is supported by the Max Planck Society. Data collection was funded in part by a Centre for Cognitive Ageing and Cognitive Epidemiology Pilot grant.

Conflict of interest. None.

Ethical standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

References

Anthoni, H., Zucchelli, M., Matsson, H., Muller-Myhsok, B., Fransson, I., Schumacher, J., . . . Peyrard-Janvid, M. (2007). A locus on 2p12 containing the co-regulated MRPL19 and C2ORF3 genes is associated to dyslexia. *Human Molecular Genetics*, *16*, 667–677.

Bates, T. C., Castles, A., Coltheart, M., Gillespie, N., Wright, M., & Martin, N. G. (2004). Behaviour genetic analyses of reading and spelling: A component processes approach. *Australian Journal of Psychology*, *56*, 115–126.

Bates, T. C., Castles, A., Luciano, M., Wright, M., Coltheart, M., & Martin, N. G. (2007). Genetic and environmental bases of reading and spelling: A unified genetic dual route model. *Reading and Writing*, *20*, 147–171.

Bates, T. C., Lind, P. A., Luciano, M., Montgomery, G. W., Martin, N. G., & Wright, M. J. (2010). Dyslexia and DYX1C1: deficits in reading and spelling associated with a missense mutation. *Molecular Psychiatry*, *15*, 1190–1196.

Becker, J., Czamara, D., Scerri, T. S., Ramus, F., Csepe, V., Talcott, J. B., . . . Schumacher, J. (2014). Genetic analysis of dyslexia candidate genes in the European cross-linguistic NeuroDys cohort. *European Journal of Human Genetics*, *22*, 675–680.

Bishop, D. V. (2015). The interface between genetics and psychology: lessons from developmental dyslexia. *Proceedings of the Royal Society B*, *282*, 20143139.

Carrion-Castillo, A., Franke, B., & Fisher, S. E. (2013). Molecular genetics of dyslexia: an overview. *Dyslexia*, *19*, 214–240.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, *47*, 149–180.

de Leeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: Generalized gene-set analysis of GWAS data. *PLOS Computational Biology*, *11*, e1004219.

Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R., & Starr, J. M. (2000). The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish Mental Survey. *Intelligence*, *28*, 49–55.

Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., . . . Neale, B. M. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genetics*, *51*, 63–75.

Deriziotis, P., & Fisher, S. E. (2017). Speech and language: Translating the genome. *Trends in Genetics*, *33*, 642–656.

Duffy, D. L., Zhu, G., Li, X., Sanna, M., Iles, M. M., Jacobs, L. C., . . . Martin, N. G. (2018). Novel pleiotropic risk loci for melanoma and nevus density implicate multiple biological pathways. *Nature Communications*, *9*, 4774.

Dykiert, D., & Deary, I. J. (2013). Retrospective validation of WTAR and NART scores as estimators of prior cognitive ability using the Lothian Birth Cohort 1936. *Psychological Assessment*, *25*, 1361–1366.

https://doi.org/10.1017/thg.2020.7 Published online by Cambridge University Press

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

Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P., & Pembrey, M. E. (1998). Localisation of a gene implicated in a severe speech and language disorder. *Nature Genetics*, *18*, 168–170.

Francks, C., Paracchini, S., Smith, S. D., Richardson, A. J., Scerri, T. S., Cardon, L. R., … Monaco, A. P. (2004). A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. *American Journal of Human Genetics*, *75*, 1046–1058.

Galaburda, A. M., LoTurco, J., Ramus, F., Fitch, R. H., & Rosen, G. D. (2006). From genes to behavior in developmental dyslexia. *Nature Neuroscience*, *9*, 1213–1217.

Garnier, H. E., Stein, J. A., & Jacobs, J. K. (1997). The process of dropping out of high school: A 19-year perspective. *American Educational Research Journal*, *34*, 395–419.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*, 513–543.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory*, *2*, 103–127.

Guidi, L. G., Velayos-Baeza, A., Martinez-Garay, I., Monaco, A. P., Paracchini, S., Bishop, D. V. M., & Molnar, Z. (2018). The neuronal migration hypothesis of dyslexia: A critical evaluation 30 years on. *European Journal of Neuroscience*, *48*, 3212–3233.

Haplotype Reference Consortium, McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., … Marchini, J. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, *48*, 1279.

Harlaar, N., Spinath, F. M., Dale, P. S., & Plomin, R. (2005). Genetic influences on early word recognition abilities and disabilities: A study of 7-year-old twins. *Journal of Child Psychology and Psychiatry*, *46*, 373–384.

Holdnack, H. A. (2001). *Wechsler Test of Adult Reading: WTAR*. San Antonio, TX: The Psychological Corporation.

Huttunen, H. J., Kuja-Panula, J., Sorci, G., Agneletti, A. L., Donato, R., & Rauvala, H. (2000). Coregulation of neurite outgrowth and cell survival by amphoterin and S100 proteins through receptor for advanced glycation end products (RAGE) activation. *Journal of Biological Chemistry*, *275*, 40096–40105.

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

Leonard, L. B. (1991). Specific language impairment as a clinical category. *Language, Speech, and Hearing Services in Schools*, *22*, 66–68.

Leonard, L. B. (2014). *Children with specific language impairment*. Cambridge: MIT Press.

Lind, P. A., Luciano, M., Wright, M. J., Montgomery, G. W., Martin, N. G., & Bates, T. C. (2010). Dyslexia and DCDC2: normal variation in reading and spelling is associated with DCDC2 polymorphisms in an Australian population sample. *European Journal of Human Genetics*, *18*, 668–673.

Luciano, M., Gow, A. J., Pattie, A., Bates, T. C., & Deary, I. J. (2018). The influence of dyslexia candidate genes on reading skill in old age. *Behavior Genetics*, *48*, 351–360.

MacDermot, K. D., Bonora, E., Sykes, N., Coupe, A. M., Lai, C. S., Vernes, S. C., … Fisher, S. E. (2005). Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *American Journal of Human Genetics*, *76*, 1074–1080.

Matsson, H., Huss, M., Persson, H., Einarsdottir, E., Tiraboschi, E., Nopola-Hemmi, J., … Kere, J. (2015). Polymorphisms in DCDC2 and S100B associate with developmental dyslexia. *Journal of Human Genetics*, *60*, 399–401.

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, *38*, 115–142.

Medland, S. E., Nyholt, D. R., Painter, J. N., McEvoy, B. P., McRae, A. F., Zhu, G., … Martin, N. G. (2009). Common variants in the trichohyalin gene are associated with straight hair in Europeans. *American Journal of Human Genetics*, *85*, 750–755.

Meng, H., Smith, S. D., Hager, K., Held, M., Liu, J., Olson, R. K., … Gruen, J. R. (2005). 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*, 17053–17058.

Morgan, A., Fisher, S. E., Scheffer, I., & Hildebrand, M. (1993). FOXP2-related speech and language disorders. In M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens, & A. Amemiya (Eds.), *Gene reviews [Internet]*. Seattle, WA: University of Washington.

Mueller, K. L., Murray, J. C., Michaelson, J. J., Christiansen, M. H., Reilly, S., & Tomblin, J. B. (2016). Common genetic variants in FOXP2 are not associated with individual differences in language development. *PLoS One*, *11*, e0152576–e0152576.

Nelson, H. E., & Willison, J. (1991). *National Adult Reading Test (NART) test manual*. Windsor, UK: NFER-Nelson Publishing Company.

Newbury, D. F., Bishop, D. V., & Monaco, A. P. (2005). Genetic influences on language impairment and phonological short-term memory. *Trends in Cognitive Sciences*, *9*, 528–534.

Newbury, D. F., Paracchini, S., Scerri, T. S., Winchester, L., Addis, L., Richardson, A. J., … Monaco, A. P. (2011). Investigation of dyslexia and SLI risk variants in reading- and language-impaired subjects. *Behavior Genetics*, *41*, 90–104.

Newbury, D. F., Winchester, L., Addis, L., Paracchini, S., Buckingham, L. L., Clark, A., … Monaco, A. P. (2009). CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *American Journal of Human Genetics*, *85*, 264–272.

Nopola-Hemmi, J., Myllyluoma, B., Haltia, T., Taipale, M., Ollikainen, V., Ahonen, T., … Widen, E. (2001). A dominant gene for developmental dyslexia on chromosome 3. *Journal of Medical Genetics*, *38*, 658–664.

Nyholt, D. R. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *American Journal of Human Genetics*, *74*, 765–769.

Paracchini, S., Scerri, T., & Monaco, A. P. (2007). The genetic lexicon of dyslexia. *Annual Review of Genomics and Human Genetics*, *8*, 57–79.

Poelmans, G., Buitelaar, J. K., Pauls, D. L., & Franke, B. (2011). A theoretical molecular network for dyslexia: integrating available genetic findings. *Molecular Psychiatry*, *16*, 365–382.

Poelmans, G., Engelen, J. J., Van Lent-Albrechts, J., Smeets, H. J., Schoenmakers, E., Franke, B., … Schrander-Stumpel, C. (2009). Identification of novel dyslexia candidate genes through the analysis of a chromosomal deletion. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *150B*, 140–147.

Purcell, S., Cherny, S. S., & Sham, P. C. (2003). Genetic power calculator: Design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*, *19*, 149–150.

Reuter, M. S., Riess, A., Moog, U., Briggs, T. A., Chandler, K. E., Rauch, A., … Zweier, C. (2017). FOXP2 variants in 14 individuals with developmental speech and language disorders broaden the mutational and clinical spectrum. *Journal of Medical Genetics*, *54*, 64–72.

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science*, *24*, 1301–1308.

Rodgers, B. (1983). The identification and prevalence of specific reading retardation. *British Journal of Educational Psychology*, *53*, 369–373.

Scerri, T. S., Morris, A. P., Buckingham, L. L., Newbury, D. F., Miller, L. L., Monaco, A. P., … Paracchini, S. (2011). DCDC2, KIAA0319 and CMIP are associated with reading-related traits. *Biological Psychiatry*, *70*, 237–245.

SLI Consortium. (2002). A genomewide scan identifies two novel loci involved in specific language impairment. *American Journal of Human Genetics*, *70*, 384–398.

Smith, M. C. (1993). Change in reading ability and attitudes from childhood to adulthood: A life span perspective. In S. R. Yussen (Ed.), *Reading across the life span* (pp. 273–291). New York: Springer.

Smith, M. C. (1996). Differences in adults' reading practices and literacy proficiencies. *Reading Research Quarterly*, *31*, 196–219.

Snowling, M. J., Adams, J. W., Bishop, D. V., & Stothard, S. E. (2001). Educational attainments of school leavers with a preschool history of

https://doi.org/10.1017/thg.2020.7 Published online by Cambridge University Press

speech-language impairments. *International Journal of Language & Communication Disorders*, *36*, 173–183.
Stark, R. E., & Tallal, P. (1981). Selection of children with specific language deficits. *Journal of Speech and Hearing Disorders*, *46*, 114–122.
Uddén, J., Hulten, A., Bendtz, K., Mineroff, Z., Kucera, K. S., Vino, A., … **Fisher, S. E.** (2019). Toward robust functional neuroimaging genetics of cognition. *Journal of Neuroscience*, *39*, 8778–8787.
Vernes, S. C., Newbury, D. F., Abrahams, B. S., Winchester, L., Nicod, J., Groszer, M., … **Fisher, S. E.** (2008). A functional genetic link between distinct developmental language disorders. *New England Journal of Medicine*, *359*, 2337–2345.
Wiszniewski, W., Hunter, J. V., Hanchard, N. A., Willer, J. R., Shaw, C., Tian, Q., … **Lalani, S. R.** (2013). TM4SF20 ancestral deletion and susceptibility to a pediatric disorder of early language delay and cerebral white matter hyperintensities. *American Journal of Human Genetics*, *93*, 197–210.
Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*, 76–82.

https://doi.org/10.1017/thg.2020.7 Published online by Cambridge University Press

Conclusion

In this chapter, I introduced a new cohort of unselected adults with quantitative measures of reading and language skill, with the aim of investigating whether such cohorts are valid for discovering genetic factors associated with reading ability. Whilst the sample was underpowered for GWA analysis (n = 1,505), I was able to perform gene-based association tests of 14 candidate genes previously associated with dyslexia, demonstrating a significant association of *FOXP2* with a composite reading and spelling score, phonetic spelling and nonword repetition. This gene has previously been implicated in both speech/language disorders and dyslexia (Fisher & Scharff, 2009; Peter et al., 2011; Wilcke et al., 2012), and the two disorders often co-occur (Snowling et al., 2020). No other gene candidate reached significance, but this may be because their effect size is lower than previously thought (see Chapter 5 for a more in-depth discussion of candidate gene studies in the light of GWAS). Regardless, the *FOXP2* result still suggests adult population cohorts may be valid for gene discovery of reading and language ability. Since adult population cohorts are easier to recruit, genotype, and measure reading skill in large numbers, they are presently a means to boost statistical power for genetic studies of reading. For example, by including adult cohorts, such as the one presented in this chapter, in GWAS meta-analysis, or by collecting reading measures in existing well-powered adult cohorts such as the UK Biobank (Sudlow et al., 2015). Given large GWAS of dyslexia and reading ability have become available since this work was published (Doust et al., 2022; Eising et al., 2022), PGS analysis can be used as a more reliable indicator in future studies instead of candidate gene analysis.

In the following chapter, I will investigate the use of a proxy reading phenotype for discovery of genetic factors associated with reading ability.

CHAPTER 4: INVESTIGATING THE USE OF A PROXY PHENOTYPE FOR DISCOVERY OF GENETIC FACTORS ASSOCIATED WITH READING ABILITY

Introduction

Reading is a fundamental skill in modern society required for learning, work, and everyday life. Reading ability varies between individuals and can be measured by standardised tests that assess reading of regular, irregular and nonwords, accuracy and fluency of reading a text aloud, and comprehension. Performance on such tests is distributed along a continuum and is largely distinct from general cognitive ability. Variation in reading ability is partly a result of individuals' environment, but genetics studies of twins and population samples show that genetic factors may explain between 30% and 70% of the variation in reading ability (Bates et al., 2007; DeFries et al., 1987; Harlaar et al., 2005). This genetic variation is partly reading specific but does also overlap with general cognitive ability (Haworth et al., 2009).

Like other quantitative traits, reading ability is a complex trait whose variation is likely a result of many additive genetic variants of small effect size. To detect such effects through GWA analyses, large sample sizes with sufficient statistical power are required. In the past, most data used in studies on the genetics of reading have come from samples of children with reading or language disability, which tend to be small and therefore insufficiently powered for GWA analyses. Unselected adult samples tend to be collected in larger sizes, however previously, it had not been directly demonstrated whether such samples were a valid means by which to discover genetic factors associated with reading ability. Evidence suggests they are, because cognitive abilities are stable throughout life (Deary et al., 2000). In fact, reading skill does not peak until individuals are in their mid-20s (McArdle et al., 2002), therefore adult

samples may be even more sensitive tests of genetic variation than samples of children.

In Chapter 3, I demonstrated that unselected adults are indeed a valid means by which to discover genetic factors associated with reading ability (Doust et al., 2020). Whilst the study was underpowered to detect individual genetic variants associated with quantitative reading measures, in gene-based tests, candidate genes previously associated with dyslexia and/or speech/language disorders were associated with quantitative measures of reading and/or language skill. This demonstrated the potential of using unselected adult samples for discovering genetic influences of reading ability.

Many large samples of genotyped unselected adults have been collected (e.g., UK Biobank, Sudlow et al., 2015), however these rarely include quantitative reading ability phenotypes, which have only been collected in small samples with insufficient power to detect variants of small effect size. The power of large biobanks to discover variants associated with continuous reading ability could be harnessed by making use of proxy reading phenotypes. Biobanks often include forms of proxy literacy data for example survey completion time, digital literacy, health literacy, or data on hobbies that correlate with reading ability including reading books. Reading books is indicative of higher literacy skill after accounting for age, as is having a diversity of reading practices (reading books, newspapers, magazines, and personal and work documents) (M. C. Smith, 1996). Indeed, in Chapter 3, we demonstrated that reading and spelling tests were correlated with how often individuals read books.

The Health & Retirement Study (HRS) is an unselected adult sample which includes potential proxy reading variables. A longitudinal study of 20,000 older adults in the

United States of America, the HRS includes a number of measures indicative of general cognitive ability (such as numeracy skill and verbal memory) and data on how much time participants spend reading books and newspapers and/or magazines (Sonnega et al., 2014). The first aim of this chapter is to generate a proxy reading measure (that controls for general cognitive ability) which can be used in subsequent genetic analyses.

The second aim of this study is to probe whether a dyslexia polygenic score (PGS) can explain variation in this proxy measure of reading ability in the HRS unselected adult sample. In Chapter 2, we performed a GWA analysis of 51,800 adults diagnosed with dyslexia (self-reported) and 1,087,070 controls who were participants of the personal genetics company 23andMe, Inc (Doust et al., 2022). We identified 42 independent genome-wide significant loci, including 15 in genes previously linked to cognitive ability and/or educational attainment. We estimated genetic correlations between dyslexia and 98 traits including quantitative reading and spelling measures from the GenLang consortium (Eising et al., 2022), finding strong negative correlations including an r_g of -0.75 for spelling, -0.70 for nonword reading, and -0.71 for word reading. This demonstrated the validity of the self-reported dyslexia diagnosis measure, and confirms shared genetic aetiology between dyslexia and quantitative reading skill.

In Chapter 2, the 23andMe dyslexia PGS could explain up to 3.6% of variance in nonword reading (an index of phonological decoding) in two Australian population-based samples (Brisbane adolescents/young adults, n = 1,640, 772 families; Brisbane older adults, n = 1,165, 966 families); and 5.6% of variance in word recognition in developmental cohorts enriched for reading difficulties (UKdys, n = 930, 595 families; Colorado Learning Disabilities Research Centre, CLDRC; n = 717, 336 families). I was

therefore interested in whether a dyslexia PGS generated from the largest genome-wide association study (GWAS) of dyslexia to date could explain variance in the HRS proxy reading phenotype, and thus support its use in meta-analysis GWAS of quantitative reading ability.

Materials and Methods

Participants

The sample comprised participants in the Health and Retirement Study (HRS), an ongoing longitudinal study of ageing in the United States (Health and Retirement Study (Public Survey Data; Genetic Data, restricted dataset). The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan. The HRS was initiated in 1992 (n = 12,652) and is repeated every two years with recruitment of additional participants. A sub-sample of participants (initial n = 5,000) were recruited to complete the Consumption and Activities Mail Survey (CAMS) in 2001, and then every two years following, with further participants recruited in subsequent waves. This study includes 10,365 individuals who participated in both the core battery and the CAMS between 2002 and 2018. Participants were aged from 51 to 100 years (mean = 67.8 years) and sex was 58% female, 42% male. The HRS study design includes ancestries representative of the modern North American population, including two ancestry-related oversamples to increase the number of Black and Hispanic respondents from proportionate allocations of 10.0% and 5.0% respectively to 18.6% and 8.6% respectively (Heeringa & Connor, 1995).

Measures

The HRS study includes a core battery of cognitive tests administered in person or over the phone. The following measures were selected from the HRS core battery (Table 1):

Immediate and delayed word recall

Two linked tasks were used to assess episodic memory through immediate free recall and delayed free recall (Spreen & Strauss, 1998). Interviewers read a list of 20 nouns (specifically developed for the HRS) and respondents were immediately asked to recall as many of the nouns as possible, in any order. After approximately five minutes of answering of questions, the respondent was asked again to recall as many of the nouns as possible. The recall test is negatively associated with age (Schaie, 2005; Spreen & Strauss, 1998). For the purpose of my analysis, the mean score of immediate and delayed word recall was taken.

Word meaning

A vocabulary measure was used to test verbal comprehension (Spreen & Strauss, 1998). Respondents were asked to define five words from one of two randomly assigned sets (Set 1: repair, fabric, domestic, remorse, plagiarize; Set 2: conceal, enormous, perimeter, compassion, audacious). The HRS adapted this measure from the vocabulary component of the verbal comprehension subtest of the Wechsler Adult Intelligence Scale (WAIS) IV (Holdnack, 2001).

Everyday numeracy

Three numeracy questions adapted were used to assess mental arithmetic. These were read to the respondent as follows:

1. “If the chance of getting a disease is 10 percent, how many people out of 1,000 would be expected to get the disease?”
2. “If five people all have the winning numbers in the lottery and the prize is two million dollars, how much will each of them get?"
3. “Let’s say you have $200 in a savings account. The account earns ten percent interest per year. How much would you have in the account at the end of two years?”

The first question was adapted from Lipkus et al.(2001) whilst the other two were from the ELSA survey (Banks et al., 2021).

Subtraction number series

A number series module abbreviated from the Woodcock-Johnson III Cognitive Ability Battery was used to test arithmetic and working memory (Schrank, 2010). Respondents were asked to subtract seven five times, starting from 100.

Additionally, the following three variables were selected from the CAMS:

Book reading

Respondents were asked to write an answer to the question, ‘How many hours did you actually spend last week reading books?’. Respondents were requested to check a ‘0 hours’ box if they did not do this activity last week. Book reading is correlated with greater literacy skill (M. C. Smith, 1996).

Newspaper and magazine reading

Respondents were asked to write an answer to the question, ‘How many hours did you actually spend last week reading newspapers or magazines?’. Respondents were requested to check a ‘0 hours’ box if they did not do this activity last week. Smith

(1996) found reading newspaper and/or magazines specifically was unrelated to reading achievement, however they found that engaging in a variety of different reading practices (including books, documents, newspapers, and magazines) was associated with greater literacy skill.

Years in education

The number of years an individual spends in education is correlated with performance on reading tests (Garnier et al., 1997; Snowling et al., 2001) so its inclusion may help to better isolate a reading factor.

Table 1

Variables Included from the Health and Retirement Study

Variable	HRS reference	Description
Age	HRS X067	Age when participated
Sex	HRS X060	Male or female
Word recall score (immediate)	HRS D106	Participants recall as many words as possible from a list of 10
Word recall score (delayed)	HRS D148	Participants recall as many words as possible from a list of 10 after a delay (answering 12 other questions in the survey)
Word meaning	HRS D161-169	Participants define five words (e.g. perimeter) from the Wechsler Adult Intelligence Scale (Holdnack, 2001)

Subtract seven number series	HRS D142-146	Participant subtracts seven from 100, then subtracts seven from the answer a further four times
Everyday maths	HRS D178-D180	Participants asked three questions to assess everyday numeracy
Years in education	HRS B014	Highest grade of school or year of college completed by participant (up to post-graduate, 17 or more years)
Hours reading newspapers/magazines	CAMS A2	Hours spent in the previous week reading newspapers or magazines
Hours reading books	CAMS A3	Hours spent in the previous week reading books

Note. HRS = Health and Retirement Study; CAMS = Consumption and Activities Mail Survey

Phenotypic analyses

All phenotypic analysis was carried out on the complete sample prior to ancestry regression in order to use as large a dataset as possible for more robust results. The following variables were selected for use in a principal components analysis (PCA) to generate a unitary proxy reading measure: (1) word recall score, (2) subtract seven number series score, (3) word meaning score, (4) everyday maths scores, (5) years in education, (6) hours spent reading newspapers or magazines per week, and (7) hours spent reading books per week. These measures were chosen because they are have been demonstrated to be related to general cognitive ability (1-5) or reading ability (6 and 7) (see the section called Measures for further detail).

First, descriptive statistics (sample size, mean, minimum, maximum, standard deviation and frequency) were generated for each variable in R v4.1.2 (R Core Team, 2021). Second, a correlation matrix was generated to explore the relationship of each variable with each other variable. Then, in order to adjust for the effect of age, the standardised residuals from a linear regression (variable ~ age) were obtained for each variable. This was important because many tests were completed at different testing waves over a 16-year period.

Finally, to generate a unitary proxy reading measure which represents variation in reading ability specifically, rather than general cognitive ability, a PCA (with varimax rotation) was conducted using the Psych package in R (Revelle, 2023). PCA is used to transform multi-dimensional datasets into multiple dimensions; to obtain the maximal separation of reading specific variance an orthogonal rotation was used. The first PC represents the maximum amount of variance, the second PC represents the second most amount of variance, and so on. Here, it allowed for a reading ability component to be separated. Initially, multiple PCAs were performed with book reading

and/or newspaper reading included along with the five other variables. However, the final PCA included only six of the seven variables, excluding the newspaper/magazine reading variable, because variance in reading ability was better explained through a PCA which included book reading only. The reading ability component (PC2) was used in all subsequent analyses.

Genotyping

Genotyping was carried out by the HRS and the data was accessed through the NCBI Database of Genotypes and Phenotypes (dbGaP) (Mailman et al., 2007) under the approved project #20453: Genome-wide analysis of a reading skill phenotype. Salivary DNA was collected using a mouthwash collection method (prior to and in 2006) or the Oragene DNA collection kit (OGR-250) (after 2006). Genotyping was conducted using Illumina HumanOmni2.5-4v1 and HumanOmni2.5-8v1 arrays by the NIH Center for Inherited Disease Research, covering 2.5 million SNPs. See Weir (2012) for further details.

Genetic analyses

Quality control

The HRS supplied genotype data had been subjected to basic quality control measures (Weir, 2012). The following describes how the data was checked for standard measures of quality plus additional quality control steps that are necessary for PGS calculation (including the removal of ambiguous SNPs). First, the sample was checked for duplicate SNPs (n = 0), which must be removed if present. Next, the sample was checked for ambiguous SNPs (n = 0) which cannot be paired-up and therefore must be removed if present. Following this, a number of further quality control measures were performed using PLINK v2.0. In the first of these, SNPs with a

minor allele frequency of < 0.01 were removed (n = 483,723) because genotyping errors disproportionately effect SNPs with a very low MAF. Then, SNPs with a Hardy-Weinberg Equilibrium Fisher's exact test (founders only) result of $p < 1 \times 10^{-6}$ were removed (n = 546,072) because these variants are disproportionately affected by genotyping erorrs. Deviations from HWE can also be caused by natural selection, non-random mating and genetic drift. Additionally, the sample was checked for SNPs which are missing in a high proportion of individuals (0.01), which can be removed if present (n = 0). In the last of these steps, the sample was checked for individuals with a high rate (0.02) of missing genotype data (n = 0), which can be removed if present, because these data may have been subject to a DNA collection or processing error. Following these quality control steps, the dataset comprised 1,048,420 SNPs and 15,567 individuals (9,114 females, 6,453 males).

Next, samples with extremely low or extremely high rates of heterozygosity were removed because these may be present due to inbreeding or DNA contamination. Firstly, the 'indep-pairwise' function in PLINK v2.0 was used to perform pruning of variants with LD $r^2 > 0.25$ (n = 746718), using a 200 variant window and a 50 variant step size. Then, the 'het' function in PLINK v1.0 (Purcell et al., 2007) was used to calculate the rate of heterozygosity (n = 0). Specifically, the function calculates F coefficient estimates for heterozygosity. Any individuals with F coefficient > 3 standard deviations from the population mean would have been excluded if present.

Then, mismatching alleles between the two datasets were identified and resolved by recoding the HRS data as the complementary allele (i.e. strand flipping) based on an R script on GitHub (Choi et al., 2020).

Lastly, individuals with a first or second degree relative (π >0.125) in the same sample were removed (n = 2,009) (one is excluded from each pair at random) because this can lead to overfitted results which are less generalisable to the population. This quality control measure was performed using the 'rel-cutoff' function in PLINK v2.0.

Population stratification analysis

An ancestry PCA was performed in PLINK v2.0 (Chang et al., 2015) through which only individuals of European ancestry were retained to be consistent with the 23andMe dyslexia GWAS sample. Specifically, the –pca flag was used, and adjusted to extract the top ten principal components (Price et al., 2006; Weir, 2012) to be included as covariates in the PGS analysis (see below).This is standard practice when individuals of European ancestry are the majority in a sample because it avoids confounding, however, it is recognised that a lack diversity in genetic studies exacerbates inequality and efforts must be made to increase samples of different ancestries (Sirugo et al., 2019).

Polygenic score (PGS) analysis

PGS analysis is used to estimate individuals' likelihood of a disease or correlation with a continuous trait using GWA data from a separate sample (Choi et al., 2020). Polygenic score analysis was performed using PLINK v.20 and R v4.1.2. The PLINK v2.0 'score' function was used to calculate a dyslexia polygenic score from 23andMe dyslexia GWA analysis SNPs (*n* = 108,088) and p-values. A *p* value threshold of 0.05 was selected because this threshold was the most predictive in a PGS of quantitative reading measures in Chapter 2.

In the standard approach to PGS analysis of continuous traits (Choi et al., 2020), association between the polygenic score and target trait is tested by a linear

regression which is adjusted for covariates (usually age, sex, and ancestry components). In this analysis, the phenotype has already been adjusted for age, therefore sex and the first ten ancestry components were included in a linear regression of the target trait: the HRS proxy reading phenotype. The standardised residuals from this analysis were used as the dependent variable in a linear regression with the dyslexia polygenic score as the predictor.

Results

Phenotypic analyses

Descriptive statistics

Table 2 gives the sample size, minimum and maximum values, means, and standard deviations for each of these variables. Figure 1 gives the frequencies of the cognitive variables, years in education, reading-related variables, and participant age. Age and time spent reading books or newspapers/magazines was negatively skewed. Years in education was positively skewed.

Table 2

Descriptive Statistics of the Variables

Variable	n	Mean	SD	Min	Max
Age when participated in HRS core survey	8034	65.77	9.58	54	100
Age when participated in CAMS	8706	65.80	9.52	54	99
Word recall score	10365	9.71	4.06	0	20
Number series minus seven score	8363	3.37	1.85	0	5
Word meaning score	7234	5.53	2.08	0	10
Everyday maths score	6862	1.61	0.66	0	3
Hours reading papers or magazines per week	10210	4.90	6.01	0	100
Hours reading books per week	10174	3.45	6.15	0	100
Highest year in education	6189	12.39	3.15	1	17

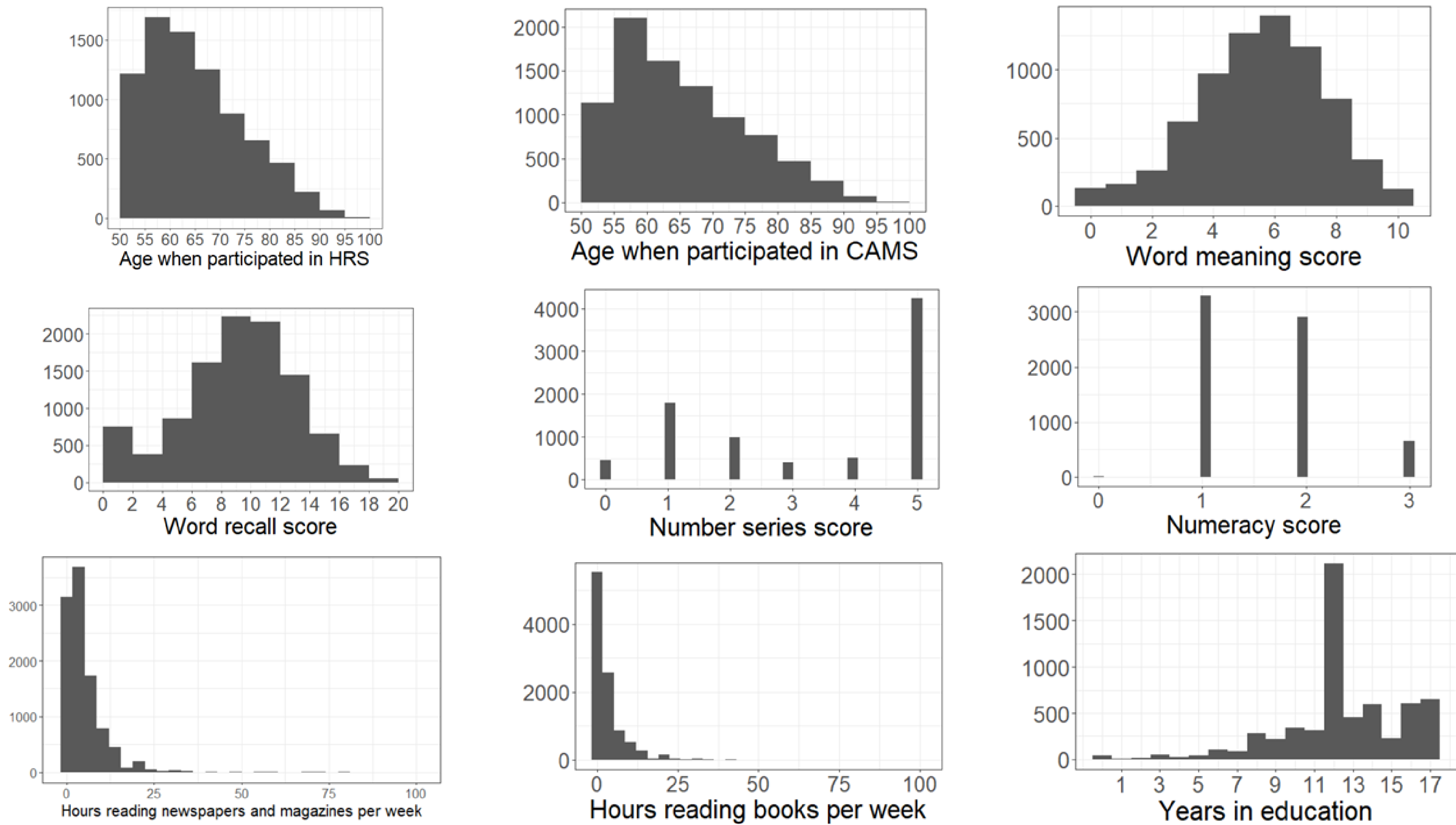


Figure 1. Frequency distribution of the variables.

Correlation between variables

Table 3 shows a correlation matrix of the cognitive measures, time spent reading books or newspapers/magazines and years in education. Time spent reading books was most highly correlated with time spent reading newspapers/magazines (r = 0.26). Of the other variables, time spent reading books and time spent reading newspapers/magazines was most highly correlated with word meaning (r = 0.11 and r = 0.12 respectively) followed by years in education (r = 0.14 and r = 0.17 respectively). The cognitive measures (word recall, word meaning, subtraction and everyday numeracy) and years in education, were more highly correlated with each other than with time spent reading books or with time spent reading newspapers/magazines (ranging from r = 0.19 to r = 0.36).

Table 3

Correlation Between Variables (Pearson's r)

	Word recall (n = 10,365)	Number series (n = 8,363)	Word meaning (n = 7,234)	Everyday maths (n = 6,862)	Newspapers/ magazines (n = 10,210)	Books (n = 10,174)	Years in education (n = 6,189)
Word recall	-						
Number series	0.19*	-					
Word meaning	0.29*	0.22*	-				
Everyday maths	0.20*	0.27*	0.25*	-			
Newspapers/magazines	-0.01	0.06*	0.11*	0.03*	-		
Books	0.07*	0.03*	0.12*	0.03*	0.26*	-	
Years in education	0.29*	0.21*	0.36*	0.29*	0.17*	0.14*	-

* $p < .05$

Age-adjusted linear regressions

In linear regressions performed for each dependent variable with age as a predictor, the results demonstrate that for all variables, significant correlations existed with variance explained ranging from .004% (vocabulary) to 6% (word recall) (Table 4). Therefore, in this sample, age had a very low effect on most cognitive measures, but did have an effect on memory (word recall score, R^2 = 0.06) and time spent reading newspapers and magazines (R^2 = 0.05). Given the significant age effects, the subsequent standardised residuals were used in all further analyses.

Table 4

Linear Regression Parameters for Age-Adjusted Variables (n = 5,106)

	R^2	**Regression coefficient**	**SE**	***p***
Number series	0.0001	0.0004	0.002	< 0.001
Numeracy	0.008	-0.006	0.00009	< 0.001
Word recall score	0.06	-0.11	0.005	< 0.001
Word meaning score	4.76×10^{-5}	0.003	0.003	< 0.001
Reading books	0.003	0.04	0.007	< 0.001
Reading newspapers/magazines	0.05	0.12	0.007	< 0.001
Years in education	0.014	-0.04	0.005	< 0.001

Principal components analysis

The final PCA excluded the reading newspapers/magazines variable (see Materials and Methods for an explanation). A scree plot of the eigenvalues shows the first and

second PCs together explained a large proportion of the variation (29% and 19% respectively) (Figure 2). The cognitive variables and years in education were moderately to highly loaded onto the first PC (ranging from 0.37 to 0.75, Figure 3 and Table 5). Time spent reading was highly loaded onto the second PC (0.93 for book reading) and word meaning score (0.28) and years in education (0.34) were also moderately loaded onto this PC (Figure 3 and Table 5). The second PC was thus used in subsequent genetic analyses as a unitary proxy measure of reading ability.

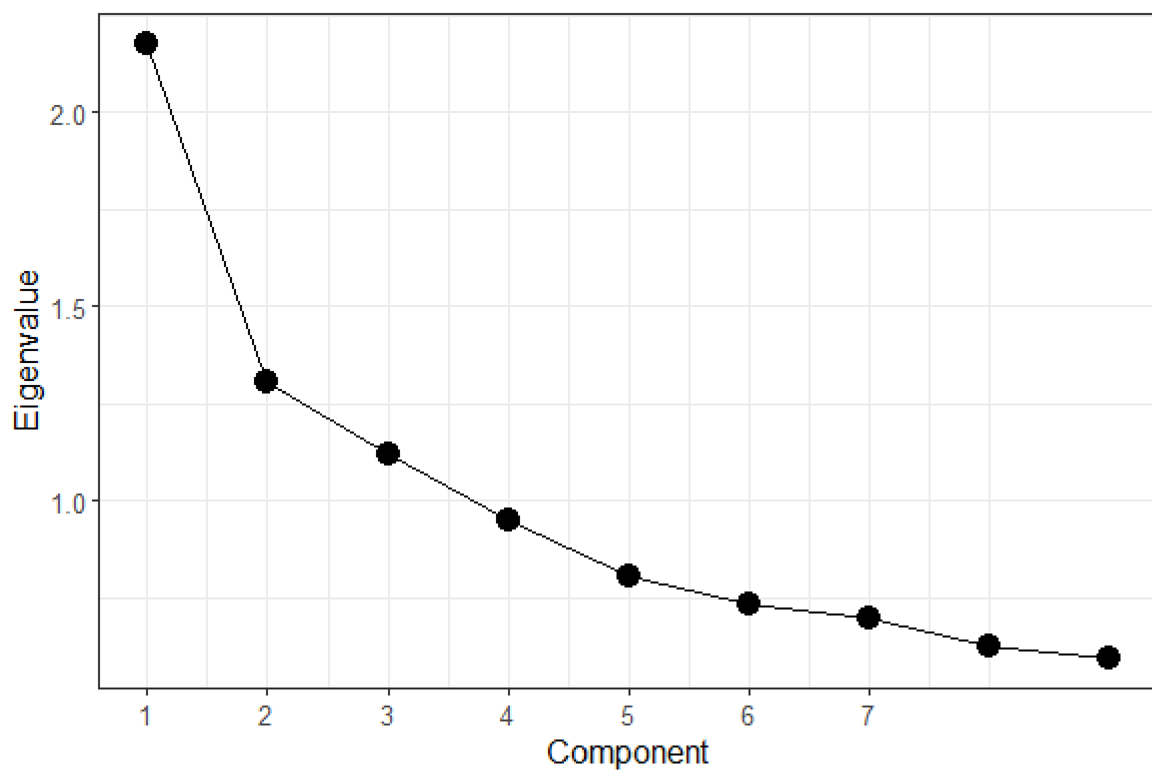


Figure 2. Scree plot of eigenvalues of the components from a principal components analysis.

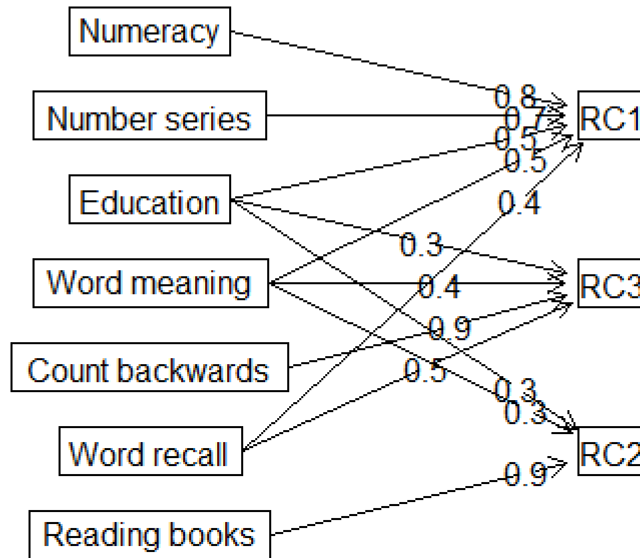


Figure 3. Component loadings (>.20) of the first two rotated components from the PCA

Table 5

Standardised Loadings of the Variables from the Principal Components Analysis (n = 5106)

	RC1	RC2	RC3	h2	u2	com
Word recall score	0.37	0.19	0.49	0.42	0.58	2.2
Number series score	0.68	-0.13	0.07	0.49	0.51	1.1
Word meaning score	0.47	0.28	0.42	0.47	0.53	2.6
Numeracy score	0.75	0.02	-0.05	0.57	0.43	1.0
Years in education	0.52	0.34	0.32	0.49	0.51	2.5
Hours reading books per week	-0.05	0.93	0.00	0.86	0.14	1.0

Note. PC = principal component; *h2* = communality; *u2* = uniqueness; and *com* = complexity

Genetic analyses

The dyslexia polygenic score constructed from the 23andMe self-reported dyslexia GWAS was not associated with the proxy reading measure (regression coefficient = -1.20, SE =0.05, *p* = 0.26 Table 6) as visualised in a scatter plot (Figure 4). It was adjusted for sex and ten ancestry principal components. Sex did not have a substantial effect on the proxy reading measure (regression coefficient = -0.04, SE = 0.04). The variables were adjusted for age prior to genetic analysis. Individuals were excluded (*n* = 3,284) due to unavailability of genetic data or failing quality control (see Materials and Methods).

Table 6

Dyslexia Polygenic Score Prediction of Proxy Reading Measure in 1822 Unselected Adults

	***p* < .05**		
Regression coefficient	**SE**	***p***	**R^2**
-1.20	0.05	0.26	6.98×10^{-4}

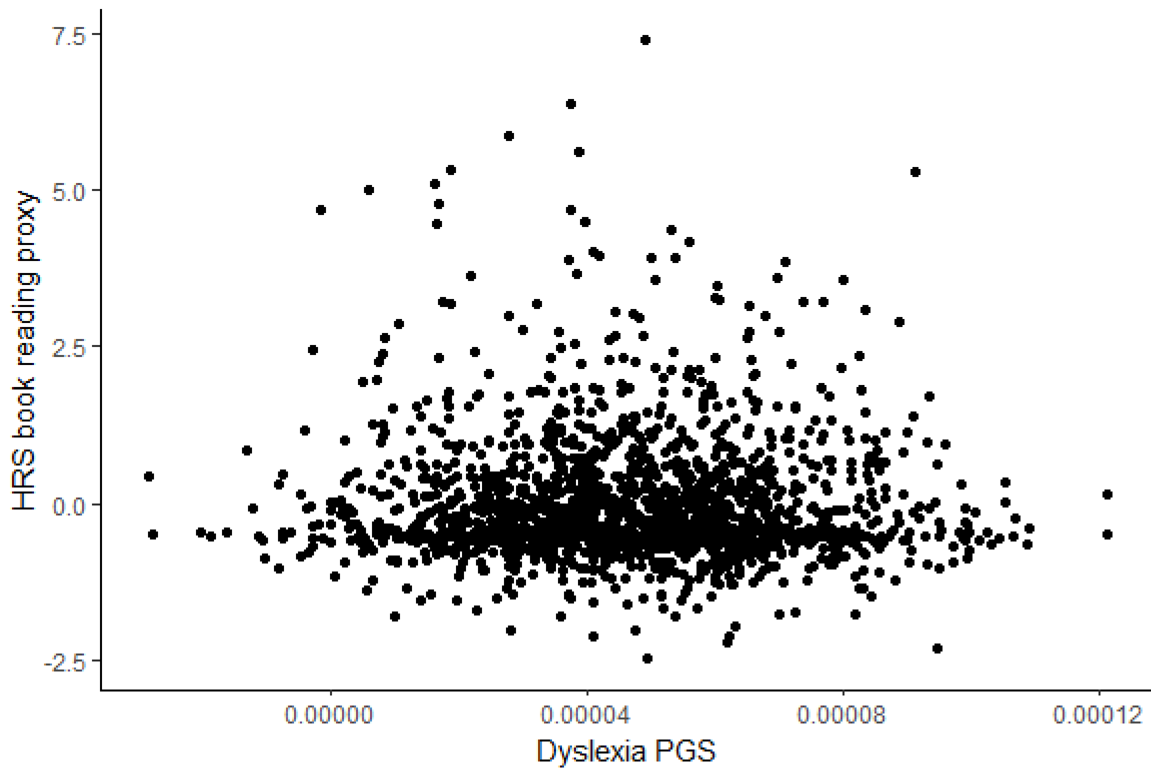


Figure 4. Scatter plot of the self-reported dyslexia PGS against the HRS proxy reading ability measure.

Discussion

In this chapter, I aimed to construct a proxy measure of reading ability from variables previously shown to correlate with reading ability. Using a proxy enabled me to make use of a large sample of unselected adults for whom standardised measures of reading ability were unavailable. I demonstrated the potential of such samples for discovering genetic factors associated with reading skill in Chapter 3 (Doust et al., 2020). Here, I probed whether a dyslexia PGS from the GWAS of dyslexia (Doust et al., 2022) performed in Chapter 4 can explain variation in this proxy measure of reading ability, to confirm its validity as a reading ability measure that could be used to boost meta-analysis GWAS of reading ability.

The main finding was that there was no association between a dyslexia PGS and this proxy reading measure. This is at odds with our finding (part of Chapter 4) that the dyslexia PGS was able to explain variation in a quantitative measure of reading ability, thus suggesting that the proxy score is not a good indicator of reading ability, which will be discussed further on.

Firstly, I found that age had a negligible effect on vocabulary and time spent reading books, despite the older age of participants, but did affect memory, consistent with previous findings (Deary et al., 2000; Salthouse, 2010). This demonstrates the validity of older adult samples for investigating verbal abilities, assuming individuals with cognitive disease, such as dementia, are accounted for.

The correlations between the variables were modest at best ($r < 0.36$), although the cognitive variables were more correlated with each other than with time spent reading, and time spent reading books was most highly correlated with time spent reading newspapers/magazines. The PCA generated two distinct components onto which cognitive and reading measures were distinctly loaded. The two components only explain 48% of the total variance, therefore a large proportion of the variance is unexplained. A cognitive factor explained 29% of the total variance and a reading factor explained a further 19% of the total variance. Alongside time spent reading books, word meaning score and years in education were moderately loaded onto the second component. Spending more time reading is associated with a greater vocabulary (Pfost et al., 2013; Stanovich et al., 1995). Higher achievement in education has also previously been associated with reading ability (Garnier et al., 1997), and further, genetic correlations have been demonstrated between reading ability and educational attainment (Eising et al., 2022).

The reading component used as a proxy measure of reading ability may have been weak because it was largely based on time spent reading books. Whilst I demonstrated that book reading correlated with reading and spelling test performance in Chapter 3 (in corroboration with Smith (1996), who found that reading books is associated with higher literacy proficiency), the correlation was moderate at best (up to $r = 0.26$). Variation in time spent reading is likely to be influenced by many other factors in addition to reading ability. For example, participants may have impaired eyesight which limits how much time they can spend reading, if at all; participants may have obligations which reduce the time they have available for reading (e.g., caring for a partner or relative); participants may lack access to books due to barriers such as poor mobility or poverty; or participants may simply enjoy other hobbies which use their time instead of reading. Therefore, the proxy reading ability score may be influenced by socio-economic and health factors, for example.

The PGS analysis which investigated whether the dyslexia PGS could explain variance in the proxy reading ability phenotype produced a null result. This may be because the target proxy reading measure may not be capturing true variance in reading ability, as discussed above. Whilst a common problem with PGS is that they are often poor proxies of actual genetic liability, and they can therefore only explain a small portion of phenotypic variance (Choi et al., 2020), in this case, the dyslexia PGS could predict quantitative reading scores in an independent sample (Chapter 2), indicating the null result is not due to a poor PGS.

Future studies where a proxy reading phenotype is used due to a lack of quantitative measures of reading ability should ideally generate a unitary proxy score from multiple reading ability related variables that are as highly correlated with quantitative reading ability as possible. These might include existing data on frequency of book reading or

other types of reading, number of books possessed, health literacy, legal literacy or digital literacy. Additionally, a single self-report item, in which participants are asked how good a reader or speller they consider themselves against a Likert scale, could reasonably be included in future surveys for large biobanks. For example, self-reported developmental language and literacy problems have previously been used to study phonological processing deficits in autism (Bishop et al., 2004). The correlation of a self-report measure with quantitative reading/spelling tests could first be investigated in smaller existing cohorts with measured reading ability such as the Brisbane adults introduced in Chapter 3.

Conclusion

In summary, I generated a proxy phenotype for reading ability in a moderately-sized unselected adult sample using a measure of time spent reading books. However, a dyslexia PGS was unable to explain variance in the proxy reading ability phenotype in this case, despite previously explaining variance in quantitative reading ability in a separate adult sample. This may be due to time spent reading books being too weak a proxy for quantitative reading ability. Future studies should look to refine the use of proxy reading measures for the discovery of genetic factors associated with reading ability in order to harness the statistical power which large biobanks of unselected adults (which often lack quantitative measures of reading abilities) can offer.

CHAPTER 5: DISCUSSION

Dyslexia affects roughly one in ten people, yet we understand little of its biological basis. The aim of this thesis was to discover genetic factors associated with dyslexia and variation in reading ability in order to better understand the aetiology of reading difficulties. This understanding is vital to inform prediction, identification and intervention strategies. In this thesis, I firstly analysed the largest GWAS of dyslexia to date, identifying 42 novel genome-wide significant loci and demonstrating genetic correlation with many traits including quantitative reading skills and ADHD. Secondly, I sought to investigate methods to increase power of GWAS of reading ability, which are historically underpowered to detect genetic variants of small effect size which underlie this polygenic trait. I first demonstrated that unselected adult samples are valid for GWAS of reading ability, and therefore adult cohorts, which tend to be larger than those of children, can increase power for gene discovery. I then tested whether proxy measures of reading ability may enable use of large cohorts for which quantitative measures of reading ability are unavailable, concluding that the use of proxy measures requires further investigation. In this discussion, I will address each of the four objectives set out in my thesis aims in Chapter 1, before discussing how my findings contribute to answering major questions about the genetics of dyslexia and reading ability.

Objective 1: To identify novel genetic markers associated with dyslexia

In Chapter 2, through performing the largest GWAS of dyslexia to date, 42 genome-wide significant independent loci were associated with a self-reported dyslexia diagnosis. More than half (27) were novel associations that have never before been reported in GWAS of dyslexia or related cognitive traits. Of these, 12 were validated

in a moderately well-powered GWAS meta-analysis of quantitative reading and spelling measures in the GenLang consortium (Eising et al., 2022), and 1 in the Chinese Reading Study (Wang et al., 2023). Approximately one third of the 42 significant SNPs were also identified in a GWAS of general cognitive ability, aligning with the finding that genetic variation in reading disability overlaps with general cognitive ability (Haworth et al., 2009). This is because constituent processes of general cognitive ability such as working memory contribute to reading acquisition (Ne'eman & Shaul, 2021). Further, general cognitive ability may be advantaged by high reading ability, allowing more information to be obtained from text and enhanced comprehension.

The effect size of every significant SNP was low (odds ratios ranged from 1.04 to 1.12), which is consistent with the high polygenicity seen for other complex traits (e.g., Demontis et al., 2017). SNP heritability was 19%, which is lower than estimates from twin studies of 40-80% (DeFries et al., 1987; Olson et al., 1989), but similar to GWAS (Eising et al., 2022; Gialluisi et al., 2020). This may be in part because rare variants may contribute to heritability. For example, it is thought that a specific rare haplotype of the *ROBO1* gene, which prevents or attenuates its expression, may contribute to dyslexia in specific families (Hannula-Jouppi et al., 2005). Such variants are likely to contribute to the heritability estimates of dyslexia in family studies.

Whilst this GWAS was the largest to date and discovered substantially more genome-wide significant loci than any prior study, it is likely that there are many thousands of other genetic loci which contribute to dyslexia (Erbeli et al., 2022), based on examples of GWAS of other complex traits. For example, the largest known GWAS to date, which was performed on height of 5.4 million individuals, identified 12,000 genetic variants associated with variation in height (Yengo et al., 2022). Future studies should aim for

equivalent, if not larger, sample sizes in order to uncover more of the genetic architecture.

Additionally, our dyslexia GWAS only included individuals with European ancestry because population stratification can produce spurious associations (Price et al., 2010). However, this standard practice means individuals of all non-European ancestries are under-represented in genetic studies (Sirugo et al., 2019) and future studies should aim to be well powered to analyse non-European ancestries, so that our understanding of the genetics of dyslexia is inclusive of all ethnicities. Identifying individual SNPs which are associated with dyslexia is important because they could contribute to a predictive genetic profile of dyslexia in the future (Lewis & Vassos, 2020), and importantly, because they indicate which genes, and by extension, which biological pathways, may be functionally relevant in dyslexia, which in turn informs our understanding of the neurobiology of dyslexia.

Objective 2: To investigate whether any of the novel genetic markers for dyslexia occur in candidate genes and/or pathways for reading ability and/or dyslexia

In Chapter 2, gene and gene set association analyses were conducted to test whether previously implicated candidate genes and biological pathways for dyslexia could be replicated. Hypothesis-free gene and gene set based tests were also performed to identify novel candidate genes and biological pathways. None of the most prominent candidate dyslexia genes reached genome-wide significance. Considering the power of this analysis compared to all those prior, this null finding prompts a re-evaluation of the previously implicated dyslexia genes of whether they are in fact important contributors in the general population, or rather Type I errors (Button et al., 2013), or

rare variants found in specific families (e.g., (Hannula-Jouppi et al., 2005; Taipale et al., 2003) (see the implications section below for a more in-depth discussion of these points).

Significant SNPs within several other genes related to neurodevelopment were observed through a discovery approach to gene-based tests. For example, rs3735260 in the Autism Susceptibility Candidate 2 (*AUTS2*) gene was identified. Not only did this SNP reach genome-wide significance, but it was the strongest association with a neurodevelopmental trait to date. The *AUTS2* gene has previously been associated autism (Oksenberg et al., 2013), intellectual disability (Beunders et al., 2015) and dyslexia (Girirajan et al., 2011). The protein acts in the Polycomb Repressive Complex 1 (PRC1) to activate transcription. Disruption of the gene in mice leads to developmental defects in the central nervous system (CNS) (Gao et al., 2014). Through functional annotations, we noted that the rs3735260 variant had a combined annotation dependent depletion (CADD) score of 17.5 which suggested deleteriousness to gene function (Kircher et al., 2014). Further, the chromatin state indicated location at an active transcription start site (Ernst & Kellis, 2012). Altogether, evidence suggests rs3735260 may impair *AUTS2* transcription and thereby impact CNS development with potentially broad effects on cognition.

Also, rs72841395 in the *TANC2* (Tetratricopeptide Repeat, Ankyrin Repeat And Coiled-Coil Containing 2) gene was identified and has previously been implicated in language delay (Wessel et al., 2017), intellectual disability, delayed motor development, autism, and psychiatric disorders (Guo et al., 2019). The TANC2 protein is involved in dendritic spine and excitatory synapse formation during embryonic development (Han et al., 2010; Stucchi et al., 2018). The fact that multiple neurodevelopmental disorders have been associated with disruption in this gene is

consistent with the comorbidity of, and genetic correlation between, these disorders (Gidziela et al., 2023).

A further association with rs34349354 in the Gametogenetin Binding Protein 2 (*GGNBP2*) gene was validated in the GenLang Consortium GWAS meta-analysis of quantitative reading measures (Eising et al., 2022), the NeuroDys GWAS meta-analysis of dyslexia cases (Gialluisi et al., 2020), and the Chinese Ready Study (CRS) GWAS meta-analyses of reading accuracy and fluency (Wang et al., 2023). This gene has also previously been associated with neurodevelopmental delay (Pasmant et al., 2008) and autism (Takata et al., 2018) however understanding of its function is limited. Takata et al. showed *GGNBP2* is co-expressed with well-established autism genes which function in pathways implicated in autism (including synaptic signalling). This promising evidence and the fact that this variant was validated in three independent samples warrants further investigation of the function of *GGNBP2* and its potential role within neurodevelopmental disorders.

A targeted gene set analysis of two candidate biological pathways for dyslexia (axon guidance and neuronal migration, based on support from Poelmans et al., (2011) and a set of transcriptional targets of *FOXP2* (Ayub et al., 2013)) was conducted. This hypothesis-driven approach is commonly used in an attempt to replicate prior findings, and is judged upon a lower significance threshold than discovery-based analyses because the smaller number of tests reduces the degree of Bonferroni correction. Here, replication-level significant enrichments were observed only for the axon guidance pathway. The axon guidance pathway is a set of 216 genes whose function is defined as a ‘chemotaxis process that directs the migration of an axon growth cone to a specific target site’ (GO:0007411). The neuronal migration hypothesis originated when post-mortem examinations of brains identified abnormal neuronal migration in

brains of individuals with dyslexia (Galaburda, 1993). Early genetic studies then identified candidate genes which molecular genetic studies showed to be involved in neuronal migration and axon guidance, such as *DCDC2* (Meng et al., 2005) and *ROBO1* (Hannula-Jouppi et al., 2005). Based on the understanding of the genetics at the time, Poelmans et al. (2011) supported a theoretical molecular network for dyslexia implicating the neuronal migration and axon guidance pathways. However, with increased understanding of the polygenicity of neurodevelopmental disorders and advancing technology in genome-wide sequence, Guidi et al., (2018) critically evaluated the hypothesis, proposing a more likely scenario that many different genes and therefore many different neurodevelopmental pathways are likely to be involved in dyslexia. In line with this, 11 of the top 20 most significant biological pathways (e.g., GO:0022008 neurogenesis and GO:0045664 regulation of neuron differentiation) associated with our dyslexia phenotype in a discovery-based approach were involved in nervous system development. The strength of this approach is that we included all GO terms (Ashburner et al., 2000; The Gene Ontology Consortium et al., 2023) and curated gene sets from the Molecular Signatures Database (MsigDB) (Subramanian et al., 2005), which enabled an unbiased perspective on potential neurobiological bases of dyslexia.

Objective 3: To investigate which traits are genetically correlated with dyslexia

In Chapter 2, to determine which traits are genetically correlated with dyslexia, we analysed the pairwise genetic correlation of 98 traits with our self-reported dyslexia diagnoses, identifying 63 which were significantly genetically correlated. The strong negative genetic correlation between self-reported dyslexia diagnosis and quantitative measures of reading and spelling ability (ranging from -0.7 to -0.75) aligns with the theory that normal variation in reading ability and dyslexia exist on a continuum as

opposed to being qualitatively distinct phenotypes (Rodgers, 1983; Shaywitz et al., 1992). It could be argued that if the same genetic factors determine reading ability and dyslexia, then the genetic correlation should be closer to one. However, the difference may be explained by error in measurement of reading skill and by noise in the self-reported dyslexia diagnosis. This dyslexia phenotype is imperfect because it is self-reported, and because individuals will have been diagnosed in different decades and different places, therefore the diagnostic criteria will be variable.

Dyslexia was genetically correlated with ADHD (0.53) but not autism (0.02). The former is in line with the high co-occurrence of ADHD and dyslexia (August & Garfinkel, 1990; Willcutt & Pennington, 2000) and their genetic correlation as estimated from twin studies (Wadsworth et al., 2015). This supports the multiple deficit model of developmental disorders that they share aetiological risk from a combination of factors, including genetic mechanisms (Pennington, 2006). The lack of genetic correlation with autism is unexpected given reports that reading difficulties are more common in children with autism than in the general population (Brimo et al., 2021) and that genetic correlations that have been demonstrated between neurodevelopmental disorders in general, including between ADHD and autism (Gidziela et al., 2023). The GWAS of autism that we used was a meta-analysis including subgroups with varying general cognitive ability. Therefore, it may be that the genetics underlying some subgroups of autism overlap more with dyslexia than other subgroups and further more refined analysis of genetic correlations between the two disorders should be conducted.

Objective 4a: To investigate whether unselected adult cohorts are valid for increasing power to detect variants associated with quantitative reading skill

To better understand dyslexia, is it also important to understanding the genetics underlying normal variation in reading ability. However, very large genotyped datasets, such as 23andMe Inc, rarely have quantitative measures of reading ability. Most genetic studies of reading ability have been conducted in small cohorts of affected children and adolescents, which are underpowered to detect variants of small effect size. Meta-analysis are one means by which to boost power. A recent large meta-analysis of reading ability identified a novel genome-wide significant SNP associated with word reading (Eising et al., 2022), however this study was still limited in power, given the thousands of SNPs that are likely to contribute to reading ability that remain undetected. Therefore, a major aim of this thesis was to investigate further methods to increase power of GWAS of reading ability.

In Chapter 3, we introduced a new sample of unselected genotyped adults from whom quantitative measures of reading and language ability were collected (Doust et al., 2020). The aim was to determine whether such samples are a valid means to discover genetic markers of reading ability. If so, unselected adult samples can be used to boost power of GWAS, especially by including existing adult samples in meta-analyses. Further, adult population samples are easier to recruit, genotype, and obtain quantitative measures of reading ability from, compared to affected children.

Firstly, we demonstrated that age did not affect phonological decoding scores assessed by nonword reading, even in older adults. This is important because it validates measuring reading ability in older adult samples. Further, through candidate gene analysis, we identified association of the *FOXP2* gene, which has previously been linked to dyslexia (Peter et al., 2011; Wilcke et al., 2012) and speech/language disorders (Fisher & Scharff, 2009), with nonword repetition, phonetic spelling, and a reading and spelling composite score. This study was underpowered however, to

replicate individual SNPs, because of its modest size (n = 1,505). Further, this study was published prior to the release of large GWAS of dyslexia and reading ability (Doust et al., 2022; Eising et al., 2022), thus PGS analysis was not a reasonable approach at the time. Regardless, the findings suggest that genetic factors linked to reading and speech/language disability in children may also affect normal variation in adult ability. Again, this provides evidence in support of a continuous model of reading ability and dyslexia (Rodgers, 1983; Shaywitz et al., 1992), with shared genetic aetiology. Further, in Chapter 2, genetic correlation between younger and older adults with the self-reported dyslexia diagnosis was .97. This test was conducted to check whether the phenotype was noisier in older participants as a result of past approaches to dyslexia screening, when many individuals with dyslexia may not have been diagnosed and therefore treated as controls in our analyses, reducing power to detect genetic variants. The extremely high genetic correlation indicates the older participant phenotype is not a concern in this case. Overall, unselected adult samples should be considered for future GWAS of reading ability, as a means to obtain larger sample sizes and boost statistical power to detect small genetic effects.

Objective 4b: To investigate whether proxy measures of reading ability are valid for increasing power to detect variants associated with quantitative reading skill

Another potential way of boosting statistical power for discovery of genetic markers of reading ability is to use proxy measures of reading ability that correlate with quantitative measures of reading ability (e.g., Luciano et al., 2018). Proxies may already be available in large cohorts or easily obtained by adding a single question (e.g., “How good are you at reading?”), thus enabling much larger GWAS of reading ability than before. In Chapter 4, I investigated the use of a proxy reading phenotype

in an unselected older adult sample generated from ‘hours spent reading books per week’ and ‘hours spent reading newspapers and magazines per week’ and controlled for cognitive measures. A dyslexia PGS constructed from the 23andMe GWAS (Chapter 2) was not able to explain any variation in the proxy phenotype. This was unexpected since Luciano et al., (2018) previously generated a proxy reading ability measure from word recognition tests and a book reading frequency question, controlled for general cognitive ability, with promising results. The finding here may be because this particular proxy would be too weak a correlate with reading ability, as a result of the influence of other factors on the amount of time that participants spent reading, such as eyesight, time commitments and other preferences, which were not controlled for. However, Luciano et al., (2018) did not control for such factors in a similar study, and were able to extract a viable reading proxy. The advantage of their study was a more appropriate range of cognitive tests, including an irregular word reading test that is a marker of premorbid IQ but from which they could extract reading specific variance when used alongside book reading. Here, I was limited by the cognitive tests available in the HRS, which did not include any form of quantitative reading test. Despite the null finding, since certain variables, including reading habits, are highly correlated with reading ability, the use of proxies warrants further investigation. Indeed, in Chapter 3, we demonstrated that book reading correlated with performance on reading and spelling tests (Doust et al., 2020), and a diversity of reading practices is correlated with higher literacy skill (M. C. Smith, 1996). Additionally, health literacy is dependent on reading ability (Lee, 1999), and therefore is another possible proxy. It will be important for future studies to control for potential confounders, where possible, to generate a proxy which is as strongly correlated with

reading ability as possible. Further, studies could compare the use of individual proxies with a combined multi-proxy of reading ability.

Implications of the findings for our understanding of the genetics of dyslexia and reading ability

What does GWAS tell us about candidate gene findings?

The findings from recent large-scale GWAS suggest that prior genetic studies of reading and dyslexia should be re-evaluated in the light of this robust, systematic approach. Evidence comes from two moderately-sized GWAS meta-analysis (Eising et al., 2022; Gialluisi et al., 2020) and the GWAS we conducted in Chapter 2, which was the largest to date (Doust et al., 2022). All three studies found limited evidence for association of previously implicated candidate genes and biological pathways for dyslexia. The findings from the linkage and association analyses which first proposed gene candidates might result from three possible scenarios: 1) the association found is true, but is a rare variant found only in specific families; 2) the effects are context specific and may interact with other SNPs common in the particular population; 3) the association found is a Type I error. Evidence for the first scenario comes from reports of structural variants such as a translocation in *DYX1C1* associated with dyslexia (Taipale et al., 2003). The third scenario can result because most genetic studies of dyslexia have been insufficiently powered to detect small effects. The result of small sample sizes is an overinflation of effect size and statistically significant results that are unlikely to reflect a true effect (Button et al., 2013). The inconsistent reproducibility of results is evidence for this (Chapter 1, Table 1). However, we demonstrated in Chapter 2 that some previously-reported SNPs (Supplementary Table 25) and candidate genes (Supplementary Table 26)/biological pathways (Supplementary Table 27) were significant at a replication level in our GWAS of dyslexia (including

KIAA0319L, ROBO1 and *CNTNAP2* genes and the axon guidance pathway). It may be that such effects are valid, but much smaller than previously suggested. Overall, we should designate new candidate genes for dyslexia based on a genome-wide discovery-based approach conducted in sufficiently powered samples. From these, we can start afresh in seeking an unbiased understanding of the biological mechanisms of dyslexia.

What do the genetic findings tell us about the relationship between reading ability and dyslexia?

Whilst some have proposed dyslexia represents the lower end of a continuum of population variation in reading ability (Rodgers, 1983; Shaywitz et al., 1992), others have argued it is a qualitatively distinct phenotype (Hoover & Gough, 1990; Rutter & Yule, 1975). The high genetic correlation of our self-reported dyslexia phenotype with multiple validated quantitative measures of reading ability in the GenLang Consortium (Chapter 2, Doust et al., 2022) indicates a strongly shared genetic aetiology and therefore supports the theory that dyslexia is on a continuum with reading ability. Additionally, in Chapter 3, candidate gene tests demonstrated that a candidate gene for dyslexia was also associated with normal variation in reading ability, again indicating shared genetic mechanisms, albeit in a modestly-sized sample and not at genome-wide replication level. However, overall, our findings align with the liability threshold model, in which multiple variables which contribute to continuous variation in a trait are summed to produce a liability score, enabling a categorical outcome of case or control determined by whether an individual score exceeds a threshold (Hujoel et al., 2020). The liability threshold model is underpinned by a normal distribution, and further, in Fisher's infinitesimal theory, normally distributed traits are polygenic, and caused by a large number of genes of a small effect size, in addition to environmental

factors (Fisher, 1919). However, the correlation we reported between dyslexia and reading ability did not exceed r_g = -0.75. One reason for this may be that the dyslexia phenotype is imperfect, being self-reported and from participants ranging in age, therefore diagnostic criteria would have varied. Also, the dyslexia results do not take into account comorbidity of other neurodevelopmental disorders, which might skew the phenotype slightly away from purely reading difficulties.

What more do we need to know to determine the relationship between reading ability and dyslexia?

To better understand the relationship between normal variation in reading ability and dyslexia it will be important to fully elucidate the genetic architecture through well-powered discovery-based quantitative genetic studies of both reading ability and dyslexia. We should continue to seek to uncover the specific genetic profiles of different skills within reading (e.g., Gialluisi et al., 2019; Wang et al., 2023), including phonological awareness. This will uncover whether genetic variants underlying dyslexia are the same as those underlying specific aspects of reading (e.g., comprehension or decoding) or whether some more closely overlap with dyslexia than others. Also, this will reveal whether there are any genetic factors that are not involved in reading ability. As discussed in the introduction to this thesis, definitions of dyslexia can be broad and include non-reading related features, for example, in the widely-accepted Rose definition, individuals with dyslexia may have difficulties with language, mental mathematics and motor coordination (Rose, 2009). Indeed, there are reports of traits common in individuals with dyslexia which are not obviously linked to reading ability, for example, motor coordination (Habib, 2021). These may be the presence of other (undiagnosed) neurodevelopmental disorders (see further discussion on this below) or perhaps they are dyslexia-specific traits which are separate from reading

skill. Determining whether or not there are non-reading related genetic factors involved would help to clarify a definition of dyslexia. If non-reading related factors for dyslexia exist, they may be dyslexia specific, or more likely, they may represent an overlap with other developmental disorders such as speech and language disorders, dyscalculia, and dyspraxia, in line with the multiple deficit model of developmental disorders (Pennington, 2006). GWAS by subtraction would help to disentangle the genetic factors involved (e.g., Demange et al., 2021). Future studies should look to perform GWAS of dyslexia in which reading ability is removed. Any existing remainder should be analysed for correlation with other traits, particularly neurodevelopmental disorders. This understanding would enable us to determine the neurological processes that enable us to read, how these might be impaired in dyslexia, and further, whether or not this impairment impacts non-cognitive functions. The presence of non-reading related dyslexia-specific traits could be circumstantial, that is arising in response to reading difficulties as a coping strategy, they could be innate, or they might not exist at all reflecting misdiagnosis. Fully understanding the genetic mechanisms at work would help to elucidate this.

A criticism of the continuum model is that it takes a deficit-based view of dyslexia. Whilst it is highly important to understand the aetiology of reading difficulties to ensure we can best support all children and adults to achieve functional literacy, some argue that we should also investigate proposed strengths of individuals with dyslexia, such as creativity, and avoid terming dyslexia as a disorder (Taylor & Vestergaard, 2022). Investigating non-reading related traits of dyslexia, both those perceived as strengths and those perceived as deficits, whilst controlling for overlap with other neurodevelopmental disorders, would increase our understanding of the relationship between variation in reading ability and dyslexia. Further, since individuals with

reading difficulties are subject to discrimination (Schumacher et al., 2007) and increased internalising problems such as anxiety and depression (Francis et al., 2019), finding evidence of strengths associated with dyslexia could have a positive impact on individuals with dyslexia.

What do the genetic findings tell us about the relationships between dyslexia and other developmental disorders?

There are high rates of co-occurrence between developmental disorders, including dyslexia and SLI (Snowling et al., 2020) and dyslexia and ADHD (August & Garfinkel, 1990; Willcutt & Pennington, 2000). The multiple deficit model of developmental disorders proposes that these disorders share aetiological risk from multiple factors, rather than each being caused by a single factor (Pennington, 2006). In Chapter 2, I demonstrated that dyslexia was highly genetically correlated with ADHD (r_g = .53), in line with the multiple deficit model. We did not test the genetic correlation between ADHD and SLI because this data was unavailable. A similarly high genetic correlation between dyslexia and ADHD has been found (Wadsworth et al., 2015). Further, in another, albeit smaller, GWAS of dyslexia, Gialluisi et al., (2020) demonstrated that dyslexia PGS could explain a proportion of variation in ADHD. A genetic twin study of the aetiology of the co-occurrence of dyslexia and ADHD indicated that overlap of dyslexia with the inattention dimension of ADHD is largely a result of genetic factors, whereas genetic factors contribute substantially less to the overlap between dyslexia with the hyperactivity/impulsivity dimension of ADHD (Wadsworth et al., 2015). This suggests the genetic correlation between dyslexia and ADHD we identified in Chapter 2 may be picking up genetic signal from the inattentive dimension. It may be that inattention negatively affects concentration required to attend to decoding principles for reading acquisition and for practising reading.

Genetic correlation indicates a number of possible mechanisms: a) a shared genetic aetiology and therefore a shared functional basis, and/or b) a causal relationship, in which the features of one disorder cause features of the other, and/or c) a third unmeasured variable (e.g., perinatal factors) affects both disorders. If the first mechanism is true, one might expect to see common features in the two disorders, resulting from the same impaired biological pathways. If the second mechanism is true, one might expect one disorder to exhibit traits than could be caused by traits of the other disorder. Both mechanisms provide evidence in favour of the multiple deficit model of developmental disorders because this theory proposes that conditions such as dyslexia and ADHD are heterogeneous disorders resulting from multiple additive and interacting factors (Pennington, 2006). Between language disorders and dyslexia, there are clear overlapping features. Oral language ability is the foundation of reading and both disorders are associated with reading difficulties, language disorders more with reading comprehension, and dyslexia more with decoding (Snowling et al., 2020). However, it is less clear whether there are common features between ADHD and dyslexia. In an investigation of shared cognitive deficits of ADHD and dyslexia, impaired processing speed was the only cognitive variable which correlated with both reading ability and inattention (McGrath et al., 2011). This study was limited to performing pairwise correlations of case-control samples, therefore was unable to answer these more nuanced questions about the relationship of dyslexia to other disorders. As discussed in the previous paragraph, it could be that inattention itself is a causative factor in dyslexia.

What more do we need to know to determine the relationship between dyslexia and other developmental disorders?

Despite evidence of high comorbidity between developmental disorders, the causal mechanisms are yet to be clearly understood. Substantial genetic correlations between dyslexia and ADHD and dyslexia and SLI have been demonstrated but what underlies the unexplained remainder? Are there dyslexia-specific genetic factors, or does dyslexia result from a specific combination of additive and interactive factors which also contribute to other developmental disorders?

One important approach to help answer these questions will be to conduct (sufficiently well-powered) GWAS of 'pure' samples of dyslexia that control for ADHD, SLI, and other developmental disorders, and compare these to comorbid samples. This will uncover any non-reading related or dyslexia-specific genetic factors, as discussed above.

Another important approach is genomic structural equation modelling (SEM) of multiple developmental disorders to determine genetic correlations between many disorders, as opposed to pairwise correlations (Grotzinger et al., 2018). This creates a structural model in which disorders are clustered according to shared genetic liability and enables a wider perspective on the genetic links between them. Through genomic SEM of dyslexia, ADHD, autism, Tourette syndrome and six psychiatric disorders, five correlated genomic latent factors were generated (internalising disorders, psychotic disorders, compulsive disorders, neurodevelopmental conditions, and attention and learning difficulties (Ciulkinyte et al., 2023, unpublished). Dyslexia and ADHD were influenced by the attention and learning difficulties factor, which was moderately correlated with the internalising disorders factor, however, most of the genomic variance in dyslexia was unique. The authors also identified 49 genomic risk loci

associated with both dyslexia and ADHD, which were not previously indentified in GWAS of each individual disorder. This study demonstrates the value of genomic SEM in understanding the genetics of dyslexia and related disorders.

How can understanding the genetics of reading ability and dyslexia help refine dyslexia identification criteria?

By understanding the genetics of different aspects of reading skill we will gain a more nuanced understanding of dyslexia, for example, are the genes underlying phonological decoding skill impacted by variants in all individuals with dyslexia? Or do some groups of individuals with dyslexia have distinct genetic profiles implicating different biological pathways? For example, Castles & Coltheart (1993) distinguish surface dyslexia, in which there is greater difficulty reading irregular words, from phonological dyslexia, in which there is greater difficulty reading nonwords. If there are subtypes of dyslexia, understanding the genetic aetiology, and therefore the biological mechanisms, would help to define distinct identification criteria. Further, genetic studies can help to determine whether the discrepancy diagnostic model is valid. In Chapter 2, we demonstrated that dyslexia PGS prediction of reading ability was equivalent whether or not IQ was controlled for (Supplementary Table 24), supporting the case that dyslexia can occur with varying general cognitive ability.

Identification of dyslexia is also currently complicated by comorbidity with other developmental disorders. As discussed above, disentangling the genetic aetiology of developmental disorders will help us to understand whether dyslexia is caused by unique biological mechanisms or whether the multiple deficit model applies. An understanding of the genetic mechanisms of developmental disorders, and how they may interact, will help us to understand the biological mechanisms of dyslexia and whether they overlap with the biological mechanisms causing other disorders such as

SLI and ADHD. Currently, children tend to be diagnosed with a single disorder (whichever presents the greatest difficulty), and additional disorders can be overlooked. A better understanding of the aetiology of developmental disorders as a whole can be used to more clearly define features of each condition, including which features are likely to overlap with other disorders, thus refining identification.

With further discovery of the genetic factors underlying dyslexia, it will be feasible to construct a PGS for dyslexia which can predict whether infants are likely to develop dyslexia (Lewis & Vassos, 2020). This will enable interventions to be implemented before children even start school, as opposed to only following identification, which currently is often only in older primary children, or even later (Torppa et al., 2015). Interventions can therefore not only take the form of targeted reading instruction without delay, but they can also target early years' risk factors. Particularly, language skill in early childhood affects later phonological awareness and is a risk factor for dyslexia, thus children likely to develop dyslexia could benefit from language intervention if they were identified earlier. Additionally, early identification would enable targeted intervention against environmental risk factors such as a poor home literacy environment (e.g., providing access to resources and opportunities for parents to engage in activities which promote language and reading development with their young children).

Conclusion

To summarise, in this thesis I presented 42 novel variants associated with dyslexia and identified new genes and biological pathways that may be part of its biological basis. I discussed how large-scale, systematic, genome-wide, discovery-based approaches are prompting a re-evaluation of previous candidate genes and biological

pathways for dyslexia. I identified numerous traits genetically correlated with dyslexia including quantitative reading skills, suggesting dyslexia and reading are not distinct phenotypes; and ADHD, in line with the multiple deficit model of developmental disorders. I also investigated methods to boost power for GWAS of reading ability, since previous studies have been underpowered for gene discovery, and an understanding of the genetics of normal variation in reading skills is important for our understanding of dyslexia. Whilst the use of proxy reading ability measures was inconclusive, unselected adult cohorts with validated reading measures proved a viable means by which to boost sample sizes and thereby increase statistical power to detect genetic variants of small effect size. Future studies should seek to perform high-powered GWAS of both dyslexia and specific reading skills, and multivariate GWAS of dyslexia and reading skill, in different ancestry groups, and consider controlling for multiple developmental disorders to better understand the genetics of dyslexia against the background of comorbidity.

REFERENCES

Adlof, S. M. (2017). Understanding Word Reading Difficulties in Children With SLI. *Perspectives of the ASHA Special Interest Groups*, *2*, 71–77. https://doi.org/10.1044/persp2.SIG1.71

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed* (pp. xliv, 947). American Psychiatric Publishing, Inc. https://doi.org/10.1176/appi.books.9780890425596

APA. (2023). *APA Dictionary of Psychology*. https://dictionary.apa.org/

Arnett, A. B., Pennington, B. F., Peterson, R. L., Willcutt, E. G., DeFries, J. C., & Olson, R. K. (2017). Explaining the sex difference in dyslexia. *Journal of Child Psychology and Psychiatry*, *58*(6), 719–727. https://doi.org/10.1111/jcpp.12691

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), Article 1. https://doi.org/10.1038/75556

August, G. J., & Garfinkel, B. D. (1990). Comorbidity of ADHD and reading disability among clinic-referred children. *Journal of Abnormal Child Psychology*, *18*(1), 29–45. https://doi.org/10.1007/BF00919454

Ayub, Q., Yngvadottir, B., Chen, Y., Xue, Y., Hu, M., Vernes, S. C., Fisher, S. E., & Tyler-Smith, C. (2013). FOXP2 Targets Show Evidence of Positive Selection in European Populations. In *The American Journal of Human Genetics* (Vol. 92, Issue 5, pp. 696–706).

Banks, J., Batty, G., Breedvelt, J., Coughlin, K., Crawford, R., Marmot, M., Nazroo, J., Oldfield, Z., Steel, N., Steptoe, A., Wood, M., & Zaninotto, P. (2021). *English Longitudinal Study of Ageing: Waves 0-9, 1998-2019. [Data collection]*. UK Data Service.

Bates, T. C., Castles, A., Luciano, M., Wright, M., Coltheart, M., & G. Martin, N. (2007). Genetic and environmental bases of reading and spelling: A unified genetic dual route model. In *Reading and Writing* (Vol. 20, Issue 20, pp. 147–171).

Beunders, G., de Munnik, S. A., Van der Aa, N., Ceulemans, B., Voorhoeve, E., Groffen, A. J., Nillesen, W. M., Meijers-Heijboer, E. J., Frank Kooy, R., Yntema, H. G., & Sistermans, E. A. (2015). Two male adults with pathogenic AUTS2 variants, including a two-base pair deletion, further delineate the AUTS2 syndrome. In *European Journal of Human Genetics* (Vol. 23, Issue 6, pp. 803–807).

Bisgaard, M. L., Eiberg, H., Møller, N., Niebuhr, E., & Mohr, J. (1987). Dyslexia and chromosome 15 heteromorphism: Negative lod score in a Danish material. *Clinical Genetics*, *32*(2), 118–119. https://doi.org/10.1111/j.1399-0004.1987.tb03337.x

Bishop, D. V. (2001). Genetic influences on language impairment and literacy problems in children: Same or different? *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(2), 189–198.

Bishop, Maybery, M., Wong, D., Maley, A., Hill, W., & Hallmayer, J. (2004). Are phonological processing deficits part of the broad autism phenotype? *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *128B*(1), 54–60. https://doi.org/10.1002/ajmg.b.30039

Brimo, K., Dinkler, L., Gillberg, C., Lichtenstein, P., Lundström, S., & Åsberg Johnels, J. (2021). The co-occurrence of neurodevelopmental problems in dyslexia. *Dyslexia*, *27*(3), 277–293. https://doi.org/10.1002/dys.1681

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews. Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Byrne, B., Wadsworth, S. J., Boehme, K., Talk, A. C., Coventry, W. L., Olson, R. K., Samuelsson, S., & Corley, R. (2013). Multivariate genetic analysis of learning and early reading development. *Scientific Studies of Reading : The Official Journal of the Society for the Scientific Study of Reading*, *17*(3), 224–242. https://doi.org/10.1080/10888438.2011.654298

Cardon, L. R., Smith, S. D., Fulker, D. W., Kimberling, W. J., Pennington, B. F., & DeFries, J. C. (1994). Quantitative trait locus for reading disability on chromosome 6. *Science (New York, N.Y.)*, *266*(5183), 276–279. https://doi.org/10.1126/science.7939663

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, *47*(2), 149–180.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. In *Gigascience* (NLM; 2015/02/28 ed., Vol. 4, p. 7).

Chapman, N. H., Igo, R. P., Thomson, J. B., Matsushita, M., Brkanac, Z., Holzman, T., Berninger, V. W., Wijsman, E. M., & Raskind, W. H. (2004). Linkage analyses of four regions previously implicated in dyslexia: Confirmation of a locus on chromosome 15q. *American Journal of Medical Genetics. Part B,*

Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics, *131B*(1), 67–75. https://doi.org/10.1002/ajmg.b.30018

Chenxi Liu, Dan Wang, Chaojie Liu, Junnan Jiang, Xuemei Wang, Haihong Chen, Xin Ju, & Xinping Zhang. (2020). What is the meaning of health literacy? A systematic review and qualitative synthesis. *Family Medicine and Community Health*, *8*(2), e000351. https://doi.org/10.1136/fmch-2020-000351

Choi, S. W., Mak, T. S.-H., & O'Reilly, P. F. (2020). Tutorial: A guide to performing polygenic risk score analyses. *Nature Protocols*, *15*(9), 2759–2772. https://doi.org/10.1038/s41596-020-0353-1

Christopher, M. E., Hulslander, J., Byrne, B., Samuelsson, S., Keenan, J. M., Pennington, B., DeFries, J. C., Wadsworth, S. J., Willcutt, E., & Olson, R. K. (2013). The Genetic and Environmental Etiologies of Individual Differences in Early Reading Growth in Australia, the United States, and Scandinavia. *Journal of Experimental Child Psychology*, *115*(3), 453–467. https://doi.org/10.1016/j.jecp.2013.03.008

Ciulkinyte, A., Mountford, H. S., Fontanillas, P., 23andMe Research Team, Bates, T. C., Martin, N. G., Fisher, S. E., & Luciano, M. (2023). Genetic neurodevelopmental clustering and dyslexia. *Unpublished Manuscript*.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). *DRC: A Dual Route Cascaded model of visual word recognition and reading aloud* (Vol. 108, pp. 204–256).

Corballis, M. C. (2017). Language Evolution: A Changing Perspective. *Trends in Cognitive Sciences*, *21*(4), 229–236. https://doi.org/10.1016/j.tics.2017.01.013

Critchley, M. (1970). *The Dyslexic Child*. Heinemann Medical.

de Kovel, C. G. F., Hol, F. A., Heister, J. G. a. M., Willemen, J. J. H. T., Sandkuijl, L. A., Franke, B., & Padberg, G. W. (2004). Genomewide scan identifies susceptibility locus for dyslexia on Xq27 in an extended Dutch family. *Journal of Medical Genetics*, *41*(9), 652–657. https://doi.org/10.1136/jmg.2003.012294

Deary, I. J., Spinath, F. M., & Bates, T. C. (2006). Genetics of intelligence. *European Journal of Human Genetics: EJHG*, *14*(6), 690–700. https://doi.org/10.1038/sj.ejhg.5201588

Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R., & Starr, J. M. (2000). The Stability of Individual Differences in Mental Ability from Childhood to Old Age: Follow-up of the 1932 Scottish Mental Survey. In *Intelligence* (Vol. 28, Issue 1, pp. 49–55).

Deffenbacher, K. E., Kenyon, J. B., Hoover, D. M., Olson, R. K., Pennington, B. F., DeFries, J. C., & Smith, S. D. (2004). Refinement of the 6p21.3 quantitative trait locus influencing dyslexia: Linkage and association analyses. *Human Genetics*, *115*(2), 128–138. https://doi.org/10.1007/s00439-004-1126-6

DeFries, J. C., Fulker, D. W., & LaBuda, M. C. (1987). Evidence for a genetic aetiology in reading disability of twins. *Nature*, *329*(6139), 537–539. https://doi.org/10.1038/329537a0

Demange, P. A., Malanchini, M., Mallard, T. T., Biroli, P., Cox, S. R., Grotzinger, A. D., Tucker-Drob, E. M., Abdellaoui, A., Arseneault, L., van Bergen, E., Boomsma, D. I., Caspi, A., Corcoran, D. L., Domingue, B. W., Harris, K. M., Ip, H. F., Mitchell, C., Moffitt, T. E., Poulton, R., … Nivard, M. G. (2021). Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. *Nature Genetics*, *53*(1), Article 1. https://doi.org/10.1038/s41588-020-00754-2

Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Churchhouse, C., Dumont, A., Eriksson, N., Gandal, M., Goldstein, J., Grove, J., Hansen, C. S., Hauberg, M. E., … Neale, B. M. (2017). Discovery of the first genome-wide significant risk loci for ADHD. In *BioRxiv* (p. 145581).

Doust, C., Fontanillas, P., Eising, E., Gordon, S. D., Wang, Z., Alagöz, G., Molz, B., Aslibekyan, S., Auton, A., Babalola, E., Bell, R. K., Bielenberg, J., Bryc, K., Bullis, E., Coker, D., Partida, G. C., Dhamija, D., Das, S., Elson, S. L., … Quantitative Trait Working Group of the GenLang Consortium. (2022). Discovery of 42 genome-wide significant loci associated with dyslexia. *Nature Genetics*, *54*(11), 1621–1629. https://doi.org/10.1038/s41588-022-01192-y

Doust, C., Gordon, S. D., Garden, N., Fisher, S. E., Martin, N. G., Bates, T. C., & Luciano, M. (2020). The Association of Dyslexia and Developmental Speech and Language Disorder Candidate Genes with Reading and Language Abilities in Adults. In *Twin Research and Human Genetics* (Cambridge University Press; 2020/04/06 ed., Vol. 23, Issue 1, pp. 23–32).

Eising, E., Mirza-Schreiber, N., de Zeeuw, E. L., Wang, C. A., Truong, D. T., Allegrini, A. G., Shapland, C. Y., Zhu, G., Wigg, K. G., Gerritse, M. L., Molz, B., Alagöz, G., Gialluisi, A., Abbondanza, F., Rimfeld, K., van Donkelaar, M., Liao, Z., Jansen, P. R., Andlauer, T. F. M., … Fisher, S. E. (2022). Genome-wide analyses of individual differences in quantitatively assessed reading- and language-related skills in up to 34,000 people. *Proceedings of the National Academy of Sciences*, *119*(35), e2202764119. https://doi.org/10.1073/pnas.2202764119

Erbeli, F., Rice, M., & Paracchini, S. (2022). Insights into Dyslexia Genetics Research from the Last Two Decades. *Brain Sciences*, *12*(1), Article 1. https://doi.org/10.3390/brainsci12010027

Ernst, J., & Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. In *Nature Methods* (Vol. 9, Issue 3, pp. 215–216).

Fagerheim, T., Raeymaekers, P., Toenessen, F. E., Sandkuijl, L. A., Lubs, H. A., & Tranebjaerg, L. (2000). A genome wide search for dyslexia loci in a large Norwegian family. *American Journal of Human Genetics*, *67*(4), 310–310.

Fagerheim, T., Raeymaekers, P., Tønnessen, F. E., Pedersen, M., Tranebjaerg, L., & Lubs, H. A. (1999). A new gene (DYX3) for dyslexia is located on chromosome 2. *Journal of Medical Genetics*, *36*(9), 664–669.

Fagerheim, T., Toennessen, F. E., & Lubs, H. A. (2002). Refinement of the dyslexia 3 locus (DYX3) on chromosome 2p15-p16 and mutation screening of candidate genes. *American Journal of Human Genetics*, *71*(4), 461–461.

Fergusson, D. M., & Woodward, L. J. (1999). Maternal age and educational and psychosocial outcomes in early adulthood. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *40*(3), 479–489.

Field, L. L., & Kaplan, B. J. (1998). Absence of linkage of phonological coding dyslexia to chromosome 6p23-p21.3 in a large family data set. *American Journal of Human Genetics*, *63*(5), 1448–1456. https://doi.org/10.1086/302107

Fisher. (1919). XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Earth and Environmental Science Transactions of The Royal Society of Edinburgh*, *52*(2), 399–433. https://doi.org/10.1017/S0080456800012163

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., & Monaco, A. P. (2002). Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. *Nature Genetics*, *30*(1), 86–91. https://doi.org/10.1038/ng792

Fisher, S. E., Marlow, A. J., Lamb, J., Maestrini, E., Williams, D. F., Richardson, A. J., Weeks, D. E., Stein, J. F., & Monaco, A. P. (1999). A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. *American Journal of Human Genetics*, *64*(1), 146–156.

Fisher, S. E., & Scharff, C. (2009). FOXP2 as a molecular window into speech and language. In *Trends in Genetics* (Vol. 25, Issue 4, pp. 166–177).

Francis, D. A., Caruana, N., Hudson, J. L., & McArthur, G. M. (2019). The association between poor reading and internalising problems: A systematic review and meta-analysis. *Clinical Psychology Review*, *67*, 45–60. https://doi.org/10.1016/j.cpr.2018.09.002

Francks, C., Fisher, S. E., Olson, R. K., Pennington, B. F., Smith, S. D., DeFries, J. C., & Monaco, A. P. (2002). Fine mapping of the chromosome 2p12-16 dyslexia susceptibility locus: Quantitative association analysis and positional candidate genes SEMA4F and OTX1. *Psychiatric Genetics*, *12*(1), 35–41. https://doi.org/10.1097/00041444-200203000-00005

Friend, A., DeFries, J. C., & Olson, R. K. (2008). Parental education moderates genetic influences on reading disability. *Psychological Science*, *19*(11), 1124–1130. https://doi.org/10.1111/j.1467-9280.2008.02213.x

Fulker, D. W., Cardon, L. R., Defries, J. C., Kimberling, W. J., Pennington, B. F., & Smith, S. D. (1991). Multiple regression analysis of sib-pair data on reading to detect quantitative trait loci. *Reading and Writing*, *3*(3), 299–313. https://doi.org/10.1007/BF00354964

Galaburda, A. M. (1993). Neurology of developmental dyslexia. *Current Opinion in Neurobiology*, *3*(2), 237–242. https://doi.org/10.1016/0959-4388(93)90216-I

Galaburda, A. M., LoTurco, J., Ramus, F., Fitch, R. H., & Rosen, G. D. (2006). From genes to behavior in developmental dyslexia. In *Nature Neuroscience* (Vol. 9, Issue 10, pp. 1213–1217).

Gao, Z., Lee, P., Stafford, J. M., von Schimmelmann, M., Schaefer, A., & Reinberg, D. (2014). AUTS2 confers gene activation to Polycomb group proteins in the CNS. *Nature*, *516*(7531), 349–354. https://doi.org/10.1038/nature13921

Garnier, H. E., Stein, J. A., & Jacobs, J. K. (1997). The process of dropping out of high school: A 19-year perspective. In *American Educational Research Journal* (Vol. 34, Issue 2, pp. 395–419).

Gayán, J., Smith, S. D., Cherny, S. S., Cardon, L. R., Fulker, D. W., Brower, A. M., Olson, R. K., Pennington, B. F., & DeFries, J. C. (1999). Quantitative-trait locus for specific language and reading deficits on chromosome 6p. *American Journal of Human Genetics*, *64*(1), 157–164. https://doi.org/10.1086/302191

Gialluisi, A., Andlauer, T. F. M., Mirza-Schreiber, N., Moll, K., Becker, J., Hoffmann, P., Ludwig, K. U., Czamara, D., Pourcain, B. S., Honbolygó, F., Tóth, D., Csépe, V., Huguet, G., Chaix, Y., Iannuzzi, S., Demonet, J.-F., Morris, A. P., Hulslander, J., Willcutt, E. G., … Schulte-Körne, G. (2020). Genome-wide association study reveals new insights into the heritability and genetic correlates of developmental dyslexia. In *Molecular Psychiatry*.

https://doi.org/10.1038/s41380-020-00898-x, https://www.nature.com/articles/s41380-020-00898-x.pdf

Gialluisi, A., Andlauer, T. F. M., Mirza-Schreiber, N., Moll, K., Becker, J., Hoffmann, P., Ludwig, K. U., Czamara, D., St Pourcain, B., Brandler, W., Honbolygó, F., Tóth, D., Csépe, V., Huguet, G., Morris, A. P., Hulslander, J., Willcutt, E. G., DeFries, J. C., Olson, R. K., … Schulte-Körne, G. (2019). Genome-wide association scan identifies new variants associated with a cognitive predictor of dyslexia. In *Translational Psychiatry* (Vol. 9, Issue 1, p. 77).

Gidziela, A., Ahmadzadeh, Y. I., Michelini, G., Allegrini, A. G., Agnew-Blais, J., Lau, L. Y., Duret, M., Procopio, F., Daly, E., Ronald, A., Rimfeld, K., & Malanchini, M. (2023). A meta-analysis of genetic effects associated with neurodevelopmental disorders and co-occurring conditions. *Nature Human Behaviour*, *7*(4), Article 4. https://doi.org/10.1038/s41562-023-01530-y

Girirajan, S., Brkanac, Z., Coe, B. P., Baker, C., Vives, L., Vu, T. H., Shafer, N., Bernier, R., Ferrero, G. B., Silengo, M., Warren, S. T., Moreno, C. S., Fichera, M., Romano, C., Raskind, W. H., & Eichler, E. E. (2011). Relative Burden of Large CNVs on a Range of Neurodevelopmental Phenotypes. In *PLOS Genetics* (Vol. 7, Issue 11, p. e1002334).

Gough, P. B., & Tunmer, W. E. (1986). Decoding, Reading, and Reading Disability. *Remedial and Special Education*, *7*(1), 6–10. https://doi.org/10.1177/074193258600700104

Granocchio, E., De Salvatore, M., Bonanomi, E., & Sarti, D. (2023). Sex-related differences in reading achievement. *Journal of Neuroscience Research*, *101*(5), 668–678. https://doi.org/10.1002/jnr.24913

Grigorenko, E. L., Wood, F. B., Golovyan, L., Meyer, M., Romano, C., & Pauls, D. (2003). Continuing the search for dyslexia genes on 6p. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *118B*(1), 89–98. https://doi.org/10.1002/ajmg.b.10032

Grigorenko, E. L., Wood, F. B., Meyer, M. S., Hart, L. A., Speed, W. C., Shuster, A., & Pauls, D. L. (1997). Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. *American Journal of Human Genetics*, *60*(1), 27–39.

Grigorenko, E. L., Wood, F. B., Meyer, M. S., & Pauls, D. L. (2000). Chromosome 6p influences on different dyslexia-related cognitive processes: Further confirmation. *American Journal of Human Genetics*, *66*(2), 715–723. https://doi.org/10.1086/302755

Grigorenko, E. L., Wood, F. B., Meyer, M. S., Pauls, J. E., Hart, L. A., & Pauls, D. L. (2001). Linkage studies suggest a possible locus for developmental dyslexia on chromosome 1p. *American Journal of Medical Genetics*, *105*(1), 120–129.

Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., Ip, H. F., Marioni, R. E., McIntosh, A. M., Deary, I. J., Koellinger, P. D., Harden, K. P., Nivard, M. G., & Tucker-Drob, E. M. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. In *Nature Human Behaviour* (Vol. 3, Issue 5, pp. 513–525).

Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., Ip, H. F., McIntosh, A. M., Deary, I. J., Koellinger, P. D., Harden, K. P., Nivard, M. G., & Tucker-Drob, E. M. (2018). Genomic SEM Provides Insights into the Multivariate Genetic Architecture of Complex Traits. In *BioRxiv* (p. 305029).

Guidi, L. G., Velayos-Baeza, A., Martinez-Garay, I., Monaco, A. P., Paracchini, S., Bishop, D. V. M., & Molnar, Z. (2018). The neuronal migration hypothesis of dyslexia: A critical evaluation 30 years on. In *European Journal of Neuroscience* (NLM; 2018/09/16 ed., Vol. 48, Issue 10, pp. 3212–3233).

Guo, H., Bettella, E., Marcogliese, P. C., Zhao, R., Andrews, J. C., Nowakowski, T. J., Gillentine, M. A., Hoekzema, K., Wang, T., Wu, H., Jangam, S., Liu, C., Ni, H., Willemsen, M. H., van Bon, B. W., Rinne, T., Stevens, S. J. C., Kleefstra, T., Brunner, H. G., … University of Washington Center for Mendelian, G. (2019). Disruptive mutations in TANC2 define a neurodevelopmental syndrome associated with psychiatric disorders. In *Nature Communications* (Vol. 10, Issue 1, p. 4679).

Habib, M. (2021). The Neurological Basis of Developmental Dyslexia and Related Disorders: A Reappraisal of the Temporal Hypothesis, Twenty Years on. *Brain Sciences*, *11*(6), Article 6. https://doi.org/10.3390/brainsci11060708

Hamilton, L. G., Hayiou-Thomas, M. E., Hulme, C., & Snowling, M. J. (2016). The Home Literacy Environment as a Predictor of the Early Literacy Development of Children at Family-Risk of Dyslexia. *Scientific Studies of Reading*, *20*(5), 401–419. https://doi.org/10.1080/10888438.2016.1213266

Han, S., Nam, J., Li, Y., Kim, S., Cho, S.-H., Cho, Y. S., Choi, S.-Y., Choi, J., Han, K., Kim, Y., Na, M., Kim, H., Bae, Y. C., Choi, S.-Y., & Kim, E. (2010). Regulation of Dendritic Spines, Spatial Memory, and Embryonic Development by the TANC Family of PSD-95-Interacting Proteins. *Journal of Neuroscience*, *30*(45), 15102–15112. https://doi.org/10.1523/JNEUROSCI.3128-10.2010

Hannula-Jouppi, K., Kaminen-Ahola, N., Taipale, M., Eklund, R., Nopola-Hemmi, J., Kääriäinen, H., & Kere, J. (2005). The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS Genet*, *1*(4), e50.

Harlaar, N., Spinath, F. M., Dale, P. S., & Plomin, R. (2005). Genetic influences on early word recognition abilities and disabilities: A study of 7-year-old twins. *Journal of Child Psychology and Psychiatry*, *46*(4), 373–384.

Haworth, C. M. A., Kovas, Y., Harlaar, N., Hayiou-Thomas, M. E., Petrill, S. A., Dale, P. S., & Plomin, R. (2009). Generalist genes and learning disabilities: A multivariate genetic analysis of low performance in reading, mathematics, language and general cognitive ability in a sample of 8000 12-year-old twins. In *Journal of Child Psychology and Psychiatry* (Vol. 50, Issue 10, pp. 1318–1325).

Hayiou-Thomas, M. E., Harlaar, N., Dale, P. S., & Plomin, R. (2010). Preschool speech, language skills, and reading at 7, 9, and 10 years: Etiology of the relationship. In *J Speech Lang Hear Res* (NLM; 2010/04/03 ed., Vol. 53, Issue 2, pp. 311–332).

Hill, W. D., Davies, N. M., Ritchie, S. J., Skene, N. G., Bryois, J., Bell, S., Di Angelantonio, E., Roberts, D. J., Xueyi, S., Davies, G., Liewald, D. C. M., Porteous, D. J., Hayward, C., Butterworth, A. S., McIntosh, A. M., Gale, C. R., & Deary, I. J. (2019). Genome-wide analysis identifies molecular systems and 149 genetic loci associated with income. In *Nature Communications* (Vol. 10, Issue 1, p. 5741).

Hill, W. G. (2013). Heritability. In S. Maloy & K. Hughes (Eds.), *Brenner's Encyclopedia of Genetics (Second Edition)* (pp. 432–434). Academic Press. https://doi.org/10.1016/B978-0-12-374984-0.00695-1

Hinshelwood, J. (1895). WORD - BLINDNESS AND VISUAL MEMORY. *The Lancet*, *146*(3773), 1564–1570. https://doi.org/10.1016/S0140-6736(01)98764-1

Hinshelwood, J. (1900). CONGENITAL WORD-BLINDNESS. *The Lancet*, *155*(4004), 1506–1508. https://doi.org/10.1016/S0140-6736(01)99645-X

Hinshelwood, J. (1902). FOUR CASES OF WORD-BLINDNESS. *The Lancet*, *159*(4093), 358–363. https://doi.org/10.1016/S0140-6736(01)81505-1

Hinshelwood, J. (1907). Four Cases Of Congenital Word-Blindness Occurring In The Same Family. *The British Medical Journal*, *2*(2444), 1229–1232.

Hoff, E., & Tian, C. (2005). Socioeconomic status and cultural influences on language. *Journal of Communication Disorders*, *38*(4), 271–278. https://doi.org/10.1016/j.jcomdis.2005.02.003

Holdnack, H. A. (2001). *Wechsler Test of Adult Reading: WTAR*. The Psychological Corporation.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160. https://doi.org/10.1007/BF00401799

Hsiung, G.-Y. R., Kaplan, B. J., Petryshen, T. L., Lu, S., & Field, L. L. (2004). 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*(1), 112–119. https://doi.org/10.1002/ajmg.b.20082

Hujoel, M. L. A., Gazal, S., Loh, P.-R., Patterson, N., & Price, A. L. (2020). Liability threshold modeling of case–control status and family history of disease increases association power. *Nature Genetics*, *52*(5), Article 5. https://doi.org/10.1038/s41588-020-0613-6

Hulme, C., & Snowling, M. J. (2016). Reading disorders and dyslexia. *Current Opinion in Pediatrics*, *28*(6), 731–735. https://doi.org/10.1097/MOP.0000000000000411

Igo, R. P., Chapman, N. H., Berninger, V. W., Matsushita, M., Brkanac, Z., Rothstein, J. H., Holzman, T., Nielsen, K., Raskind, W. H., & Wijsman, E. M. (2006). Genomewide scan for real-word reading subphenotypes of dyslexia: Novel chromosome 13 locus and genetic complexity. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, *141B*(1), 15–27. https://doi.org/10.1002/ajmg.b.30245

Kaminen, N., Hannula-Jouppi, K., Kestilä, M., Lahermo, P., Muller, K., Kaaranen, M., Myllyluoma, B., Voutilainen, A., Lyytinen, H., Nopola-Hemmi, J., & Kere, J. (2003). A genome scan for developmental dyslexia confirms linkage to chromosome 2p11 and suggests a new locus on 7q32. *Journal of Medical Genetics*, *40*(5), 340–345.

Kaplan, D. E., Gayán, J., Ahn, J., Won, T.-W., Pauls, D., Olson, R. K., DeFries, J. C., Wood, F., Pennington, B. F., Page, G. P., Smith, S. D., & Gruen, J. R. (2002). Evidence for linkage and association with reading disability on 6p21.3-22. *American Journal of Human Genetics*, *70*(5), 1287–1298. https://doi.org/10.1086/340449

Katusic, S. K., Colligan, R. C., Barbaresi, W. J., Schaid, D. J., & Jacobsen, S. J. (2001). Incidence of reading disability in a population-based birth cohort, 1976-1982, Rochester, Minn. *Mayo Clin Proc*, *76*(11), 1081–1092.

Kearns, D. M., & Al Ghanem, R. (2019). The role of semantic information in children's word reading: Does meaning affect readers' ability to say polysyllabic words

aloud? *Journal of Educational Psychology*, *111*(6), 933–956. https://doi.org/10.1037/edu0000316

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, *46*(3), 310–315.

Krafnick, A. J., & Evans, T. M. (2019). Neurobiological Sex Differences in Developmental Dyslexia. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02669

Kremen, W. S., Jacobson, K. C., Xian, H., Eisen, S. A., Waterman, B., Toomey, R., Neale, M. C., Tsuang, M. T., & Lyons, M. J. (2005). Heritability of word recognition in middle-aged men varies as a function of parental education. *Behavior Genetics*, *35*(4), 417–433. https://doi.org/10.1007/s10519-004-3876-2

Lee, P. P. (1999). Why Literacy Matters: Links Between Reading Ability and Health. *Archives of Ophthalmology*, *117*(1), 100–103. https://doi.org/10.1001/archopht.117.1.100

Lewis, C. M., & Vassos, E. (2020). Polygenic risk scores: From research tools to clinical instruments. *Genome Medicine*, *12*(1), 44. https://doi.org/10.1186/s13073-020-00742-5

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, *21*(1), 37–44. https://doi.org/10.1177/0272989X0102100105

Liu, L., Wang, J., Shao, S., Luo, X., Kong, R., Zhang, X., & Song, R. (2016). Descriptive epidemiology of prenatal and perinatal risk factors in a Chinese

population with reading disorder. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep36697

Luciano, M., Gow, A. J., Pattie, A., Bates, T. C., & Deary, I. J. (2018). The Influence of Dyslexia Candidate Genes on Reading Skill in Old Age. *Behavior Genetics*, *48*(5), 351–360.

Mailman, M. D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., Hao, L., Kiang, A., Paschall, J., Phan, L., Popova, N., Pretel, S., Ziyabari, L., Lee, M., Shao, Y., Wang, Z. Y., Sirotkin, K., Ward, M., Kholodov, M., … Sherry, S. T. (2007). The NCBI dbGaP database of genotypes and phenotypes. *Nature Genetics*, *39*(10), 1181–1186. https://doi.org/10.1038/ng1007-1181

Marino, C., Giorda, R., Vanzin, L., Nobile, M., Lorusso, M., Baschirotto, C., Riva, L., Molteni, M., & Battaglia, M. (2004). A locus on 15q15-15qter influences dyslexia: Further support from a transmission/disequilibrium study in an Italian speaking population. *Journal of Medical Genetics*, *41*(1), 42–46. https://doi.org/10.1136/jmg.2003.010603

Marlow, A. J., Fisher, S. E., Francks, C., MacPhie, I. L., Cherny, S. S., Richardson, A. J., Talcott, J. B., Stein, J. F., Monaco, A. P., & Cardon, L. R. (2003). Use of multivariate linkage analysis for dissection of a complex cognitive trait. *American Journal of Human Genetics*, *72*(3), 561–570. https://doi.org/10.1086/368201

Mascheretti, S., Andreola, C., Scaini, S., & Sulpizio, S. (2018). Beyond genes: A systematic review of environmental risk factors in specific reading disorder. *Research in Developmental Disabilities*, *82*, 147–152. https://doi.org/10.1016/j.ridd.2018.03.005

Mascheretti, S., Marino, C., Simone, D., Quadrelli, E., Riva, V., Cellino, M. R., Maziade, M., Brombin, C., & Battaglia, M. (2015). Putative risk factors in developmental dyslexia: A case-control study of Italian children. *Journal of Learning Disabilities*, *48*(2), 120–129. https://doi.org/10.1177/0022219413492853

Mascheretti, S., Riva, V., Giorda, R., Beri, S., Lanzoni, L. F. E., Cellino, M. R., & Marino, C. (2014). M. In *Journal of Human Genetics* (Vol. 59, Issue 4, pp. 189–197).

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, *38*(1), 115–142.

McGrath, L. M., Pennington, B. F., Shanahan, M. A., Santerre-Lemmon, L. E., Barnard, H. D., Willcutt, E. G., DeFries, J. C., & Olson, R. K. (2011). A multiple deficit model of Reading Disability and Attention-Deficit/Hyperactivity Disorder: Searching for shared cognitive deficits. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *52*(5), 547–557. https://doi.org/10.1111/j.1469-7610.2010.02346.x

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., & Gruen, J. R. (2005). 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*(47), 17053–17058.

Moll, K., Snowling, M. J., & Hulme, C. (2020). Introduction to the Special Issue “Comorbidities between Reading Disorders and Other Developmental Disorders.” *Scientific Studies of Reading*, *24*(1), 1–6. https://doi.org/10.1080/10888438.2019.1702045

Morgan, A., Fisher, S. E., Scheffer, I., & Hildebrand, M. (1993). FOXP2-Related Speech and Language Disorders. In M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens, & A. Amemiya (Eds.), *GeneReviews*. University of Washington, Seattle.

Morris, D. W., Robinson, L., Turic, D., Duke, M., Webb, V., Milham, C., Hopkin, E., Pound, K., Fernando, S., Easton, M., Hamshere, M., Williams, N., McGuffin, P., Stevenson, J., Krawczak, M., Owen, M. J., O’Donovan, M. C., & Williams, J. (2000). Family-based association mapping provides evidence for a gene for reading disability on chromosome 15q. *Human Molecular Genetics*, *9*(5), 843–848. https://doi.org/10.1093/hmg/9.5.843

Ne’eman, A., & Shaul, S. (2021). Readiness or Impairment: Cognitive and Linguistic Differences Between Children Who Learn to Read and Those Who Exhibit Difficulties With Reading in Kindergarten Compared to Their Achievements at the End of First Grade. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.614996

NHS. (2018, August 14). *Dyslexia*. Nhs.Uk. https://www.nhs.uk/conditions/dyslexia/

NICHD Early Child Care Research Network. (2005). Pathways to reading: The role of oral language in the transition to reading. *Developmental Psychology*, *41*(2), 428–442. https://doi.org/10.1037/0012-1649.41.2.428

Nopola-Hemmi, J., Myllyluoma, B., Haltia, T., Taipale, M., Ollikainen, V., Ahonen, T., Voutilainen, A., Kere, J., & Widen, E. (2001). A dominant gene for

developmental dyslexia on chromosome 3. In *Journal of Medical Genetics* (NLM; 2001/10/05 ed., Vol. 38, Issue 10, pp. 658–664).

Norton, N., Heiervang, E., Hamshere, M., Williams, N., Stevenson, J., Owen, M. J., O'Donovan, M. C., & Williams, J. (2000). Suggestive evidence of linkage with reading disability in a large norwegian family. *American Journal of Medical Genetics - Neuropsychiatric Genetics*, *96*(4), 556.

Nöthen, M. M., Schulte-Körne, G., Grimm, T., Cichon, S., Vogt, I. R., Müller-Myhsok, B., Propping, P., & Remschmidt, H. (1999). Genetic linkage analysis with dyslexia: Evidence for linkage of spelling disability to chromosome 15. *European Child & Adolescent Psychiatry*, *8 Suppl 3*, 56–59. https://doi.org/10.1007/pl00010696

OECD. (2019). *PISA 2018 Results (Volume II): Where All Students Can Succeed*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/education/pisa-2018-results-volume-ii_b5fd1b8f-en

OECD. (2022). *OECD/INFE Toolkit for Measuring Financial Literacy and Financial Inclusion 2022*. OECD.

Oksenberg, N., Stevison, L., Wall, J. D., & Ahituv, N. (2013). Function and Regulation of AUTS2, a Gene Implicated in Autism and Human Evolution. In *PLOS Genetics* (Vol. 9, Issue 1, p. e1003221).

Olson, R., Wise, B., Conners, F., Rack, J., & Fulker, D. (1989). Specific deficits in component reading and language skills: Genetic and environmental influences. *Journal of Learning Disabilities*, *22*(6), 339–348. https://doi.org/10.1177/002221948902200604

Paracchini, S., Scerri, T., & Monaco, A. P. (2007). The genetic lexicon of dyslexia. *Annual Review of Genomics and Human Genetics*, *8*, 57–79.

Pasmant, E., de Saint-Trivier, A., Laurendeau, I., Dieux-Coeslier, A., Parfait, B., Vidaud, M., Vidaud, D., & Bièche, I. (2008). Characterization of a 7.6-Mb germline deletion encompassing the NF1 locus and about a hundred genes in an NF1 contiguous gene syndrome patient. *Eur J Hum Genet*, *16*(12), 1459–1466.

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, *101*(2), 385–413.

Peter, B., Raskind, W. H., Matsushita, M., Lisowski, M., Vu, T., Berninger, V. W., Wijsman, E. M., & Brkanac, Z. (2011). 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*(1), 39–49.

Peterson, R. L., & Pennington, B. F. (2012). Seminar: Developmental Dyslexia. *Lancet*, *379*(9830), 1997–2007. https://doi.org/10.1016/S0140-6736(12)60198-6

Petryshen, T. L., Kaplan, B. J., Fu Liu, M., de French, N. S., Tobias, R., Hughes, M. L., & Field, L. L. (2001). Evidence for a susceptibility locus on chromosome 6q influencing phonological coding dyslexia. *American Journal of Medical Genetics*, *105*(6), 507–517. https://doi.org/10.1002/ajmg.1475

Petryshen, T. L., Kaplan, B. J., Hughes, M. L., Tzenova, J., & Field, L. L. (2002). Supportive evidence for the DYX3 dyslexia susceptibility gene in Canadian families. *Journal of Medical Genetics*, *39*(2), 125–126. https://doi.org/10.1136/jmg.39.2.125

Petryshen, T. L., Kaplan, B. J., Liu, M. F., & Field, L. L. (2000). Absence of significant linkage between phonological coding dyslexia and chromosome 6p23-21.3, as

determined by use of quantitative-trait methods: Confirmation of qualitative analyses. *American Journal of Human Genetics*, *66*(2), 708–714. https://doi.org/10.1086/302764

Pfost, M., Dörfler, T., & Artelt, C. (2013). Students' extracurricular reading behavior and the development of vocabulary and reading comprehension. *Learning and Individual Differences*, *26*, 89–102. https://doi.org/10.1016/j.lindif.2013.04.008

Poelmans, G., Buitelaar, J. K., Pauls, D. L., & Franke, B. (2011). A theoretical molecular network for dyslexia: Integrating available genetic findings. In *Molecular Psychiatry* (NLM; 2010/10/20 ed., Vol. 16, Issue 4, pp. 365–382).

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*(8), Article 8. https://doi.org/10.1038/ng1847

Price, A. L., Zaitlen, N. A., Reich, D., & Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics*, *11*(7), Article 7. https://doi.org/10.1038/nrg2813

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, *81*(3), 559–575.

Quinn, J. M., & Wagner, R. K. (2015). Gender Differences in Reading Impairment and in the Identification of Impaired Readers: Results From a Large-Scale Study of At-Risk Readers. *Journal of Learning Disabilities*, *48*(4), 433–445. https://doi.org/10.1177/0022219413508323

R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rabin, M., Wen, X. L., Hepburn, M., Lubs, H. A., Feldman, E., & Duara, R. (1993). Suggestive linkage of developmental dyslexia to chromosome 1p34-p36. *The Lancet*, *342*(8864), 178. https://doi.org/10.1016/0140-6736(93)91384-X

Raskind, W. H., Igo, R. P., Chapman, N. H., Berninger, V. W., Thomson, J. B., Matsushita, M., Brkanac, Z., Holzman, T., Brown, M., & Wijsman, E. M. (2005). A genome scan in multigenerational families with dyslexia: Identification of a novel locus on chromosome 2q that contributes to phonological decoding efficiency. *Molecular Psychiatry*, *10*(7), 699–711. https://doi.org/10.1038/sj.mp.4001657

Revelle, W. (2023). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University. https://cran.r-project.org/web/packages/psych/citation.html

Rodgers, B. (1983). The identification and prevalence of specific reading retardation. *British Journal of Educational Psychology*, *53 (Pt 3)*, 369–373.

Rose, J. (2009). *Identifying and Teaching Children and Young People with Dyslexia and Literacy Difficulties: An independent report from Sir Jim Rose to the Secretary of State for Children, Schools and Families*. DCSF Publicatinos. https://webarchive.nationalarchives.gov.uk/ukgwa/20130401151715/http://www.education.gov.uk/publications/eOrderingDownload/00659-2009DOM-EN.pdf

Rosenberg, J., Pennington, B. F., Willcutt, E. G., & Olson, R. K. (2012). Gene by environment interactions influencing reading disability and the inattentive

symptom dimension of attention deficit/hyperactivity disorder. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *53*(3), 243–251. https://doi.org/10.1111/j.1469-7610.2011.02452.x

Rutter, M., Caspi, A., Fergusson, D., Horwood, L. J., Goodman, R., Maughan, B., Moffitt, T. E., Meltzer, H., & Carroll, J. (2004). Sex Differences in Developmental Reading DisabilityNew Findings From 4 Epidemiological Studies. *JAMA*, *291*(16), 2007–2012. https://doi.org/10.1001/jama.291.16.2007

Rutter, M., & Yule, W. (1975). The concept of specific reading retardation. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *16*(3), 181–197. https://doi.org/10.1111/j.1469-7610.1975.tb01269.x

Salthouse, T. (2010). *Major Issues in Cognitive Aging*. Oxford University Press, USA.

Samuelsson, S., Finnström, O., Flodmark, O., Gäddlin, P.-O., Leijon, I., & Wadsby, M. (2006). A Longitudinal Study of Reading Skills Among Very-Low-Birthweight Children: Is There a Catch-up? *Journal of Pediatric Psychology*, *31*(9), 967–977. https://doi.org/10.1093/jpepsy/jsj108

Sawyer, D., Krishnamani, R., Hannig, V. L., Farcia, M., Kim, J. K., Haines, J. L., & Phillips III, J. A. (1998). Genetic analysis of phonologic core deficit dyslexia (PCDD). *Am J Hum Genet*, *63*, A307.

Scerri, T. S., & Schulte-Korne, G. (2010). Genetics of developmental dyslexia. *Eur Child Adolesc Psychiatry*, *19*(3), 179–197.

Schaie, K. W. (2005). *Developmental influences on adult intelligence: The Seattle longitudinal study* (pp. x, 496). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195156737.001.0001

Schrank, F. (2010). Woodcock-Johnson III Tests of Cognitive Abilities. In *Handbook of Pediatric Neuropsychology* (pp. 415–434). Springer.

Schulte-Körne, G., Grimm, T., Nöthen, M. M., Müller-Myhsok, B., Cichon, S., Vogt, I. R., Propping, P., & Remschmidt, H. (1998). Evidence for linkage of spelling disability to chromosome 15. *American Journal of Human Genetics*, *63*(1), 279–282. https://doi.org/10.1086/301919

Schumacher, J., Hoffmann, P., Schmäl, C., Schulte-Körne, G., & Nöthen, M. M. (2007). Genetics of dyslexia: The evolving landscape. *Journal of Medical Genetics*, *44*(5), 289–297. https://doi.org/10.1136/jmg.2006.046516

Schumacher, J., König, I. R., Plume, E., Propping, P., Warnke, A., Manthey, M., Duell, M., Kleensang, A., Repsilber, D., Preis, M., Remschmidt, H., Ziegler, A., Nöthen, M. M., & Schulte-Körne, G. (2006). Linkage analyses of chromosomal region 18p11-q12 in dyslexia. *Journal of Neural Transmission (Vienna, Austria: 1996)*, *113*(3), 417–423. https://doi.org/10.1007/s00702-005-0336-y

Schumacher, J., König, I. R., Schröder, T., Duell, M., Plume, E., Propping, P., Warnke, A., Libertus, C., Ziegler, A., Müller-Myhsok, B., Schulte-Körne, G., & Nöthen, M. M. (2008). Further evidence for a susceptibility locus contributing to reading disability on chromosome 15q15-q21. *Psychiatric Genetics*, *18*(3), 137–142. https://doi.org/10.1097/YPG.0b013e3282fb7fc6

Scottish Government. (2010). *Literacy Action Plan An Action Plan to Improve Literacy in Scotland*. Scottish Government.

Seal, R. L., Braschi, B., Gray, K., Jones, T. E. M., Tweedie, S., Haim-Vilmovsky, L., & Bruford, E. A. (2023). Genenames.org: The HGNC resources in 2023. *Nucleic Acids Research*, *51*(D1), D1003–D1009. https://doi.org/10.1093/nar/gkac888

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of

reading ability. *The New England Journal of Medicine*, *326*(3), 145–150. https://doi.org/10.1056/NEJM199201163260301

Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., & Escobar, M. D. (1990). Prevalence of Reading Disability in Boys and Girls: Results of the Connecticut Longitudinal Study. In *JAMA* (Vol. 264, Issue 8, pp. 998–1002).

Simpson, J., & Everatt, J. (2005). Reception class predictors of literacy skills. *The British Journal of Educational Psychology*, *75*(Pt 2), 171–188. https://doi.org/10.1348/000709904X24780

Sirugo, G., Williams, S. M., & Tishkoff, S. A. (2019). The Missing Diversity in Human Genetic Studies. *Cell*, *177*(1), 26–31. https://doi.org/10.1016/j.cell.2019.02.048

Smith, M. C. (1990). Reading habits and attitudes of adults at different levels of education and occupation. *Reading Research and Instruction*, *30*(1), 50–58. https://doi.org/10.1080/19388079009558033

Smith, M. C. (1996). Differences in adults' reading practices and literacy proficiencies. In *Reading Research Quarterly* (Vol. 31, Issue 2, pp. 196–219).

Smith, S. D., Kimberling, W. J., Pennington, B. F. %J R., & Writing. (1991). *Screening for multiple genes influencing dyslexia* (Vol. 3, Issue 3, pp. 285–298).

Snowling, M. J., Adams, J. W., Bishop, D. V., & Stothard, S. E. (2001). Educational attainments of school leavers with a preschool history of speech-language impairments. *International Journal of Language & Communication Disorders*, *36*(2), 173–183.

Snowling, M. J., Hayiou-Thomas, M. E., Nash, H. M., & Hulme, C. (2020). Dyslexia and Developmental Language Disorder: Comorbid disorders with distinct effects on reading comprehension. *Journal of Child Psychology and Psychiatry*, *61*(6), 672–680. https://doi.org/10.1111/jcpp.13140

Snowling, M. J., & Hulme, C. (Eds.). (2005). *The science of reading: A handbook* (pp. xiv, 661). Blackwell Publishing.

Soleimani, F., Zaheri, F., & Abdi, F. (2014). Long-Term Neurodevelopmental Outcomes After Preterm Birth. *Iranian Red Crescent Medical Journal*, *16*(6). https://doi.org/10.5812/ircmj.17965

Sonnega, A., Faul, J. D., Ofstedal, M. B., Langa, K. M., Phillips, J. W. R., & Weir, D. R. (2014). Cohort Profile: The Health and Retirement Study (HRS). *International Journal of Epidemiology*, *43*(2), 576–585.

Spreen, O., & Strauss, E. (1998). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. Oxford University Press, USA.

Stanovich, K. E., West, R. F., & Harrison, M. R. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology*, *31*(5), 811–826. https://doi.org/10.1037/0012-1649.31.5.811

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

Steven G Heeringa, & Connor, J. (1995). *Technical Description of the Health and Retirement Study Sample Design*. Institute for Social Research, University of Michigan.

Stevenson, J., Graham, P., Fredman, G., & McLoughlin, V. (1987). A twin study of genetic influences on reading and spelling ability and disability. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *28*(2), 229–247. https://doi.org/10.1111/j.1469-7610.1987.tb00207.x

Stucchi, R., Plucińska, G., Hummel, J. J. A., Zahavi, E. E., Juan, I. G. S., Klykov, O., Scheltema, R. A., Altelaar, A. F. M., & Hoogenraad, C. C. (2018). Regulation of KIF1A-Driven Dense Core Vesicle Transport: Ca2+/CaM Controls DCV Binding

and Liprin-α/TANC2 Recruits DCVs to Postsynaptic Sites. *Cell Reports*, *24*(3), 685–700. https://doi.org/10.1016/j.celrep.2018.06.071

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. In *Proceedings of the National Academy of Sciences* (Vol. 102, Issue 43, pp. 15545–15550).

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Taipale, M., Kaminen, N., Nopola-Hemmi, J., Haltia, T., Myllyluoma, B., Lyytinen, H., Muller, K., Kaaranen, M., Lindsberg, P. J., Hannula-Jouppi, K., & Kere, J. (2003). A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(20), 11553–11558.

Takata, A., Miyake, N., Tsurusaki, Y., Fukai, R., Miyatake, S., Koshimizu, E., Kushima, I., Okada, T., Morikawa, M., Uno, Y., Ishizuka, K., Nakamura, K., Tsujii, M., Yoshikawa, T., Toyota, T., Okamoto, N., Hiraki, Y., Hashimoto, R., Yasuda, Y., … Matsumoto, N. (2018). Integrative Analyses of De Novo Mutations Provide Deeper Biological Insights into Autism Spectrum Disorder. In *Cell Reports* (Vol. 22, Issue 3, pp. 734–747).

Tallerman, M., & Gibson, K. R. (2011). *The Oxford Handbook of Language Evolution*. Oxford University Press.

Tanaka, H., Black, J. M., Hulme, C., Stanley, L. M., Kesler, S. R., Whitfield-Gabrieli, S., Reiss, A. L., Gabrieli, J. D. E., & Hoeft, F. (2011). The Brain Basis of the Phonological Deficit in Dyslexia is Independent of IQ. *Psychological Science*, *22*(11), 1442–1451. https://doi.org/10.1177/0956797611419521

Taylor, H., & Vestergaard, M. D. (2022). Developmental Dyslexia: Disorder or Specialization in Exploration? *Frontiers in Psychology*, *13*. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.889245

The Gene Ontology Consortium, Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J., Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., … Westerfield, M. (2023). The Gene Ontology knowledgebase in 2023. *Genetics*, *224*(1), iyad031. https://doi.org/10.1093/genetics/iyad031

Thompson, P. A., Hulme, C., Nash, H. M., Gooch, D., Hayiou-Thomas, E., & Snowling, M. J. (2015). Developmental dyslexia: Predicting individual risk. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *56*(9), 976–987. https://doi.org/10.1111/jcpp.12412

Tilot, A. K., Khramtsova, E. A., Liang, D., Grasby, K. L., Jahanshad, N., Painter, J., Colodro-Conde, L., Bralten, J., Hibar, D. P., Lind, P. A., Liu, S., Brotman, S. M., Thompson, P. M., Medland, S. E., Macciardi, F., Stranger, B. E., Davis, L. K., Fisher, S. E., & Stein, J. L. (2020). The Evolutionary History of Common Genetic Variants Influencing Human Cortical Surface Area. In *Cerebral Cortex* (Vol. 31, Issue 4, pp. 1873–1887).

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research: JSLHR*, *40*(6), 1245–1260. https://doi.org/10.1044/jslhr.4006.1245

Torppa, M., Eklund, K., van Bergen, E., & Lyytinen, H. (2015). Late-Emerging and Resolving Dyslexia: A Follow-Up Study from Age 3 to 14. *Journal of Abnormal Child Psychology*, *43*(7), 1389–1401. https://doi.org/10.1007/s10802-015-0003-1

Tran, C., Gagnon, F., Wigg, K. G., Feng, Y., Gomez, L., Cate-Carter, T. D., Kerr, E. N., Field, L. L., Kaplan, B. J., Lovett, M. W., & Barr, C. L. (2013). A family-based association analysis and meta-analysis of the reading disabilities candidate gene DYX1C1. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, *162B*(2), 146–156. https://doi.org/10.1002/ajmg.b.32123

Turic, D., Robinson, L., Duke, M., Morris, D. W., Webb, V., Hamshere, M., Milham, C., Hopkin, E., Pound, K., Fernando, S., Grierson, A., Easton, M., Williams, N., Van Den Bree, M., Chowdhury, R., Gruen, J., Stevenson, J., Krawczak, M., Owen, M. J., … Williams, J. (2003). Linkage disequilibrium mapping provides further evidence of a gene for reading disability on chromosome 6p21.3-22. *Molecular Psychiatry*, *8*(2), 176–185. https://doi.org/10.1038/sj.mp.4001216

Tzenova, J., Kaplan, B. J., Petryshen, T. L., & Field, L. L. (2004). 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*, *127B*(1), 117–124. https://doi.org/10.1002/ajmg.b.20139

UNESCO Institute of Statistics. (2018). *A global framework of reference on digital literacy skills for indicator 4.4.2—UNESCO Digital Library*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000265403.locale=en

UNESCO Institute of Statistics. (2020, June 22). *UNESCO Glossary: Functional literacy*. https://uis.unesco.org/en/glossary-term/functional-literacy

van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H., & Wray, N. R. (2019). Genetic correlations of polygenic disease traits: From theory to practice. *Nature Reviews Genetics*, *20*(10), Article 10. https://doi.org/10.1038/s41576-019-0137-z

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., & Fisher, S. E. (2008). A functional genetic link between distinct developmental language disorders. In *New England Journal of Medicine* (Vol. 359, Issue 22, pp. 2337–2345).

Wadsworth, S. J., Corley, R. P., Hewitt, J. K., Plomin, R., & DeFries, J. C. (2002). In *Journal of Child Psychology and Psychiatry* (Vol. 43, Issue 6, pp. 769–774).

Wadsworth, S. J., DeFries, J. C., Willcutt, E. G., Pennington, B. F., & Olson, R. K. (2015). The Colorado Longitudinal Twin Study of Reading Difficulties and ADHD: Etiologies of Comorbidity and Stability. *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies*, *18*(6), 755–761. https://doi.org/10.1017/thg.2015.66

Wang, Z., Zhao, S., Zhang, L., Yang, Q., Cheng, C., Ding, N., Zhu, Z., Shu, H., Liu, C., & Zhao, J. (2023). A genome-wide association study identifies a new variant associated with word reading fluency in Chinese children. *Genes, Brain, and Behavior*, *22*(1), e12833. https://doi.org/10.1111/gbb.12833

Weir, D. R. (2012). *Quality Control Report for Genotypic Data*. University of Washington. https://hrsonline.isr.umich.edu/sitedocs/genetics/HRS_QC_REPORT_MAR2012.pdf?_gl=1*108pdqv*_ga*MTQ0OTYzMTE2MS4xNjQzODg4Njk3*_ga_FF28MW3MW2*MTY4OTg0ODkxNC4xMS4xLjE2ODk4NDk3NjUuMC4wLjA.&_ga=2.250299270.1447185050.1689848914-1449631161.1643888697&_gl=1*108pdqv*_ga*MTQ0OTYzMTE2MS4xNjQzODg4Njk3*_ga_FF28MW3MW2*MTY4OTg0ODkxNC4xMS4xLjE2ODk4NDk3NjUuMC4wLjA.&_ga=2.250299270.1447185050.1689848914-1449631161.1643888697

Wessel, K., Suleiman, J., Khalaf, T. E., Kishore, S., Rolfs, A., & El-Hattab, A. W. (2017). 17q23.2q23.3 de novo duplication in association with speech and language disorder, learning difficulties, incoordination, motor skill impairment, and behavioral disturbances: A case report. In *BMC Med Genet* (NLM; 2017/10/27 ed., Vol. 18, Issue 1, p. 119).

Wilcke, A., Ligges, C., Burkhardt, J., Alexander, M., Wolf, C., Quente, E., Ahnert, P., Hoffmann, P., Becker, A., Müller-Myhsok, B., Cichon, S., Boltze, J., & Kirsten, H. (2012). Imaging genetics of FOXP2 in dyslexia. *European Journal of Human Genetics: EJHG*, *20*(2), 224–229. https://doi.org/10.1038/ejhg.2011.160

Willcutt, E. G., & Pennington, B. F. (2000). Comorbidity of reading disability and attention-deficit/hyperactivity disorder: Differences by gender and subtype. *Journal of Learning Disabilities*, *33*(2), 179–191. https://doi.org/10.1177/002221940003300206

Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J., Bartell, E., Sakaue, S., Graff, M., Eliasen, A. U., Jiang, Y., Raghavan, S., Miao, J., Arias, J. D., Graham, S. E.,

Mukamel, R. E., Spracklen, C. N., Yin, X., Chen, S.-H., Ferreira, T., Highland, H. H., … Hirschhorn, J. N. (2022). A saturated map of common genetic variants associated with human height. *Nature*, *610*(7933), Article 7933. https://doi.org/10.1038/s41586-022-05275-y

Yushu, G. (2010). The Sumerian Account of the Invention of Writing—A New Interpretation. *Procedia - Social and Behavioral Sciences*, *2*(5), 7446–7453. https://doi.org/10.1016/j.sbspro.2010.05.110