Linkage and next generation sequencing (NGS) data in six large Danish families with dyslexia

Hans Eiberg (h.r.l.eiberg@gmail.com)
University of Copenhagen
Lars Hansen

Article

Keywords: Reading Disability, Genome Wide Linkage Analysis, chromosome 13q12.3, 15q23-q24.1, 18q11.21, 19p13.3 and 21q22.3

Posted Date: March 24th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2711629/v1

Abstract

Dyslexia is a common learning disability exhibited as a delay in acquiring reading skills despite adequate intelligence, and reading single real words are impaired in many dyslexics. Reading disability or developmental dyslexia (DD) is a neurodevelopmental disorder affecting children globally, and the molecular mechanisms underlying it are largely underdetermined, while loci and susceptibility genes are suggested by genetic mapping in families or cohorts and by genome wide association studies (GWAS). To identify a possible genetic cause, we genotyped and performed genome wide linkage analysis employing the programs LIPED and SNP6-LINK of six multigenerational families with autosomal dominant inherited dyslexia. The linkage analyses resulted in informative haplotypes segregating with the dyslectic trait in all families and a LOD score of Z>4 at 13q12.3 and 19p13.3, and a LOD score of Z>3 at 15q23-q24.1, 18q11.21, and 21q22.3. The five mapped regions are supported by previous linkage or associations studies of dyslexia. Whole genome sequencing (WGS) of affected individuals in the six family's revealed rare regulatory variants in the mapped regions.

Introduction

Developing dyslexia (DD) including reading and language disorder is heritable, with genetic effects accounting for between 45% and 61% of the phenotypic variance [1,2]. Dyslexia is associated with functional and activation abnormalities within reading areas of the brain[3,4]. Identification of candidate genes for reading disability has involved linkage analysis, mapping quantitative-trait loci (QTLs) and genome wide association studies (GWAS)[5]. Seven loci are reported in OMIM: DYX1 on 15q21 (OMIM 127700), DYX2 on 6p22-p21 (OMIM 600202), DYX3 on 2p16-p15 (OMIM 604254), DYX5 on 3p12-q13 (OMIM 606896), DYX6 on 18p11.2 (OMIM 606616), DYX8 1p36-p34 (OMIM 608995) and DYX9 on Xq27.3 (OMIM 300509). Fine mapping of the regions has identified susceptible candidate genes as *DNAAF4*, DYX1[6], *DCDC2* and*KIAA0319*, DYX2[7,8], and *ROBO1*, DYX5[9]. The strength of gene-mapping studies has been limited that only putative functional variants affecting the genes are reported[10]. The genetic complexity of DD, where factors as incomplete penetrance, phenocopies, genetic heterogeneity and oligogenicity make identification of single genetic variants for DD difficult[11,12].

GWA studies have in the last decade identified additional loci with SNPs associated with DD. Since the first GWAS using a 100K SNP chip and pooled DNA from 5760 children[13] several other GWAS studies querying common genetic variants across the whole genome for association with DD in larger populations have been carried out [14,15,16,17,18,19]. A recent GWA study by Doust et al.[20] reported 42 loci for DD where 15 were in genes linked to cognitive ability/educational attainment, and 27 were new and potentially more specific to dyslexia. Notable none of the GWAS identified DD loci overlap with previous reported DYX loci and illustrate the genetic heterogeneity of DD and suggest the DYX1-9 loci may represent rare familiar forms. The combination of whole genome linkage (WGL) analyses in large families and WGS of the linkage regions has resulted in novel DD candidate genes. Examples are an intron variant with cis-regulatory effect on *SEMA3C*[21] and a missense variant in the gene *SPY*[22]. We have used a similar approach of WGL analyses of six large Danish DD families and analysed the linkage regions by WGS. The analyses revealed five different loci for DD (Table 1), and the WGS analyses of the linkage regions revealed a series of putative regulatory variants close to genes expressed in brain that might contribute to the DD phenotype.

Table 1

Family	Chr. band	Linkage region (hg19)	Delimiting markers	Size of linkage region	LOD score [2)]
E06	13q12.3	13:29,324,683 – 30,941,039	rs1005969-rs1472248718	1,616,356	4.42 (1.52)
A29	15q23-q24.1	15:70,362,585 – 73,666,730	rs12593849-rs59322618	3,304,145	3.01 (2.10)
A67	18p11.21	18:10,905,079 – 11,901,888	rs7241188-rs9963456	996,809	3.87 (3.57)
B41	19p13.3	19:366,412-6,755,007	rs689396-rs2305806	6,388,595	4.72 (2.64)
A09	19p13.3	19:1,364,306- 2,827,300	rs17673260-rs2159561	1,462,994	2.64 (1.7)
A22	21q22.3	21:44,828,031–48,129,895	rs857552-rs2839367 [1)]	3,301,865	3.30 (1.52)

1) Close to telomer position.

2) LOD scores for dyslexia. In parenthesis are LOD scores calculated using the first collected samples genotyped using SNP6.0 arrays. Final LOD scores are calculated between dyslexia and the haplotype (p = 0.001) for all family members.

Materials And Methods

Family material.

Six families with dyslexia were collected from the Copenhagen Family Bank[23]. The families were followed over a period of mere than 40 years and the DD phenotypes were self-reported by interview of parents and sibs. All participants were orally informed and provided written consent.

Genotyping and whole genome linkage analysis

DNA was extracted using EDTA (ethylenediamine tetra-acetic acid) blood by standard phenol/chloroform extraction protocols and genotyped by SNP arrays. Affymetrix SNP6.0 array was used in family A09, A22, A67, B71, E06 and CytoScan™ HD SNP array (ThermoFisher Scientific) was used for family A29. More than 800.000 markers were included in the WGL analysis, and the Birdseed text files were converted to a single input file employing the SNP6-LINK program package and analyzed by LIPED[24]. Initial calculation of LOD scores using the SNP-array data was carried out with a frequency of 0.02 for the dyslexic trait and an allele frequency of 0.05 for the SNPs. The LOD scores were sorted by chromosome and position and plotted graphicly and continuous regions with positive LOD scores > 2.0 were mapped. Additional family members in four families were sampled and genotyped for selected SNPs by Sanger sequencing (SI, Table S1), and a final LOD score was calculated including all individuals and SNP markers with an allele frequency of 0.05 and a risk haplotype frequency of 0.001 and a penetrans of 0.05.

WGS data analyses

WGS (BGI Europe, Copenhagen Denmark) was done by standard methods. Briefly a ≤ 800bp insert normal library was created for selected individuals and reads were aligned to human reference sequence hg19, GRCh37 using the BWA (0.7.15) aligner[25]. Variant calling was done employing GATK (4.0.11.0)[26], and variant annotation and filtration was done using VarSeq (Golden Helix, USA) and a minimum coverage of 20 reads was obtained. The WGS data was filtered for heterozygous SNVs and indels with minor allele frequency (MAF) values < 0.02. Variants in repeated regions were excluded after employing the RepeatMasker track in the UCSC browser[27]. For families with two individuals sequenced, shared variants were selected for the analyses. The variants were analyzed for position in coding regions, in non-coding RNA genes, intron location, location in regulatory regions and elements and intergenic positions. The following track in the UCSC browser[27] was employed for the filtration of variants: ENCODE regulation for histone methylation markers, DNaseI Hypersensitivity Clusters and Transcription Factor ChIP-seq Clusters, GeneHancer, and JASPAR Transcription Factor Binding Site Database. Variant Effect Predictor (VEP, Ensembl)[28] was employed for identification of variants affecting regulatory regions, and gene expression data analyses was obtained from the GTEx Portal[29] and the Human Protein Atlas[30].

Results

Six large unrelated families were recruited from Copenhagen Family Bank (CFB)[23] all with DD segregating in autosomal dominant mode (Figure 1A-F). The families were genotyped using SNP-array technology and subsequently WGL analysis revealed one single continuous region of positive LOD scores for family E06, B41, A29, A67 and A22 and two regions for family A09 (Figure 2). A risk haplotype was constructed for each family and by including additional family members (SI, Table S1) and fine mapping of the regions by Sanger sequencing of selected SNP markers excluded one of the linkage regions in family A09. A final maximum LOD score could be calculated for all six families and the five DD loci were mapped to 13q12.3 (family E06), 15q23-q24.1 (family A29), 18p11.21 (family A67), 19p13.3 (families B41 and A09) and 21q22.3 (family A22) (Table 1). The linkage region for family A09 was embedded in the linkage region for family B41.

One or two affected individuals in each family were selected for WGS and analyzed for putative DD associated variants. Heterozygote variants with MAF<0.01 in the European population were selected and variants in repeated region were excluded as these regions are highly variable. Finally, variants in regulatory regions were selected and neighboring genes were analyzed expression in the brain. The result for the WGS analysis is shown in Table 2.

Characterization of the six DD families

Family E06 represent four generations and include 20 individuals whereof 9 were reported with DD (Figure 1A). Eleven individuals were genotyped using SNP-arrays and additional 8 individuals were genotyped for selected SNPs (SI, Table S1). The WGL analyses revealed a continuous region with maximum positive LOD score of Z=1.5 on chromosome 13q12.3 (Figure 2A). Including the additional 8 individuals in the genotype analysis a final LOD score of Z=4.42 was obtained setting the haplotype frequency p=0.001. The linkage region covered 1.6Mbp delimited by the markers rs1005969 and rs1472248718 due to recombination in individuals IV:3 and IV:7.

Family A29 represent three generations with 14 individuals whereof seven were reported with dyslexia (Figure 1B). A total of 11 individuals were genotyped by SNP-arrays, and additional informative markers were analyzed by Sanger sequencing in three individuals (SI, Table S1). The WGL analysis identified a continuous region of positive LOD scores on chromosome 15q23-q24.1 with Z=2.1 (Figure 4B) and by including genotype data for additional family members, a final positive LOD score of Z=3.01 was obtained. The linkage region was 3.3Mbp delimited by the markers rs12593849 and rs59322618 due to recombination in II:1 and II:3.

Family A67 represent five generations and 21 individuals where 14 were reported with dyslexia (Figure 1C). All 21 members were genotyped by SNP-array and the following WGL analysis resulted in a maximum LOD score of Z=3.57at 18p11.21 (Figure 2C) and a LOD score of Z=3.87 was obtained by setting the haplotype frequency p=0.001. The mapped region represented 1Mbp flanked by the markers rs7241188 and rs9963456 due to recombination in individuals III:3 and III:9.

Families B41 and A09. Both families map to 19p13.3 and the 1.5Mbp region found in A09 is embedded in the 6.4Mbp region for family B41. Family B41 represent tree generations and 31 individuals whereof 9 were reported with dyslexia (Figure 2B). LOD score calculation using SNP-array data for 12 persons resulted in a continuous region with a LOD score of Z=2.64 (Figure 4D). Including additional seven family members tested for selected SNPs and a STS marker (D19S209) (SI, Table S1) resulted in a final LOD score Z=4.72 with a haplotype frequency p=0.001. The region represents 6,4Mbp delimited by the markers rs689396 and rs2305806 due to recombination in individuals II:1 and II:2.

Family A09 comprises 16 individuals in three generations with five members reported with dyslexia (Figure 3A). Genotyping of 8 individuals resulted in two continuous regions with a positive LOD score of Z=1.7 (Figure 4E). Genotyping of additional four individuals (SI, Table S1) for selected SNP markers excluded one of the linkage regions leaving the region at 19p13.3 with a positive LOD score of maximum Z=2.64. The region represented 1.5Mbp delimited by the markers rs17673260 and rs2159561 due to recombination in II:3 and III:3.

Family A22 represent a three generations family with 21 individuals whereof seven were reported with dyslexia (Figure 3B). 18 members were genotyped by SNP-arrays and a linkage region at 21q22.3 with a maximum LOD score Z=3.30 was determined (Figure 4F. Incomplete penetrance was observed for individual III:8. The region covered 3,3 Mbp and was delimited by the markers rs857552 and rs2839367 (telomeric) due to a recombination in individual III:7.

The WGS analyses

In the 13q12.3 locus, two DNA variants, rs573197999 and rs117556116, of interest in family E06 were found in II:2 and IV:3. Both SNPs were in regulatory regions in the vicinity of *MTUS2*; rs117556116 in a CTCF binding site upstream for *MTUS2* and rs573197999 in a promotor region in intron 1 with a cluster of transcription factors (Table 2). *MTSU2* is expressed in heart and is regional enhanced in the cerebral cortex and single-cell RNA data shows expression in neurons (HPA). In the 15q23-q24.1 locus, two SNPs were found in II:1 and II:5 in family A29. rs35304292 is approximately 2kbp downstream for the lncRNA gene *LINC02205*, and rs1196366533 in intron 14 of *LRRC49* in a region with several TFBS. None of genes are brain specific. In the 18p11.21 locus (family A67, II:2) one SNP, rs143669678, was found in a regulatory enhancer and CTCF-binding region in intron 2 of *PIEZO2*, a gene expressed in brain (Table 2).

In locus 19p13.3 a total of 15 SNPs identified, eight in family B41 and seven in family A09 (Table 2). Three SNPs in family B41 were close to the gene *PWWP3A* (rs144512862, rs116900972, and rs147204443), all three SNPs were assigned to regulatory regions. *PWWP3A* has low tissue specificity but is expressed in the brain. Three SNPs in family B41 and two in family A09 were in the vicinity of the gene *APC2*, all in regulatory regions. rs201353187 was further annotated as a splice site acceptor variant for the gene *C19orf25* (function unknown). *APC2* is expressed specially in the brain whereas *C19orf25* is ubiquitously expressed. One SNP, rs115178429, was found in a regulatory region in intron 1 of *ABHD17A*, a gene ubiquitously expressed including the brain, and two SNPs were in or close to *BTBD2.* rs149364482 represent a missense variant (p.Ser373Ile; NP_060267.2) predicted tolerated by PolyPhen-2 and rs146449301 in a regulatory region in an intron of *BTBD2*. *BTBD2* is ubiquitously expressed with low tissue specificity. The remaining three SNPs were in regulatory regulator regions close to *TCF3* and *GNG7*, where *GNG7* is highly expressed in the basal ganglia (HPA). Two SNPs were found outside the shared region, both in regulatory regions in intron 2 of *TLE5* (rs117195808) and in intron 2 of *PIP5K1C* (rs1364917700), respectively (Table 2).

In family A22, locus 19p13.3, three SNPs of interest in regulatory regions were found in individuals I:2 and II:1. All three SNPs are in regulatory regions; rs539002811 in intron 9 of *PDXK* in a cluster of transcription factors, rs572129208 in intron 2 of *ADARB1* and rs560135812 upstream for the genes *LINC01694*, *SLC19A1* and *PCBP3* (Table 2). *PDXK* and *PCBP3* have enhanced expression in brain, *ADARB1* and *SLC19A1* have low tissue specificity but are expressed in the brain.

Discussion

The genetic etiology of dyslexia could be complex as demonstrated by many studies (for reviews see[14,17,31,32]). A small number of genes are characterized as DD susceptibility genes, but single deleterious mutations are still missing. DD is supposed to be a neurodevelopmental disability characterized as polygenetic with a strong genetic component and heritability estimated to 40-60%[17], and a potential molecular mechanism is supposed to be linked to neuronal migration[33]. Comorbidity between dyslexia and other neurodevelopmental disorders has been reported for overlapping loci for DD, autism, and ADHD[34,35]. Other studies show families with autosomal dominant inheritance of DD and cases of incomplete penetrance[6,7,8,9,36,37].

In the present study, we have analyzed six families with autosomal dominant DD and mapped five DD loci that all have been reported linked or associated to DD (SI, Table S2). The 13p12 locus in family E06 has been reported by Igo et al.[36] as the strongest signal to the markers D13S1304-ATA5A09 for single word reading in a genome wide scan of 108 DD families with a LOD score of Z=2.94. Luciano et al.[14] and Truong et al.[37] reported three SNPs, rs9508555, rs2892463 and rs7997649, in the linkage region associated with non-word repetition and DD. None of the three studies suggested candidate gene for the 13p12 locus.

The 15q23 locus in family A29 is supported by a CNV found in a study of more than 1300 DD cases[38]. The CNV deletion suggested *MYO9A* as a candidate, but the WGS analysis failed to identify variants in the gene. The 18p11.2 locus in family A67 mapped to in the DYX6 locus (OMIM 606616). Studies found linkages for DD to D18S464 (p=0.00004) and D18S53 (p=0.0002), both close to the A67 linkage region, but a candidate gene is not suggested for the DYX6 locus[39,40]. The SNPs rs7507114 and rs1846090 have been reported for association with DD and *PTPN2* has been suggested as a candidate gene [41,42] but the WGS analysis of A67 failed to identify variants in the gene.

The locus at the telomeric region of 19p11.2 found in families B41 and A09 have been reported in three different studies. A CNV deletion of 46 kbp[38] in the linkage region for family B41 but distal to the A06 region include the genes *BSG* and intron 1 of *HCN2*, but the WGS analyses failed to find variants in these genes. Luciano et al.[14] identified several SNPs in the shared region for B41 and A09 and suggests *DAZAP1* as a DD candidate gene, but again the WGS analyses in both families failed to find variants in the gene. Finally, a microdeletion has been reported in the shared region with dyslexia as part of the phenotype[43]. Six SNPs from the WGS analyses are in the microdeletion, five in or close to the gene *APC2* gene (rs557485888, rs529177770, rs201353187, rs12974027 and rs147084328) and one (rs118087435) close to *TCF3* (Table 2). Three SNPs reported by Luciano et al.[14] distal to the microdeletion are close to three SNPs found in family B47 (rs144512862, rs116900972 and rs147204443). The gene *APC2* is of interest that the gene shows tissue specific expression in the brain and suggested to be involved in nervous system development. Further has *APC2* has been associated to intellectual developmental (MRT74, OMIM 617169) and involved in cortical dysplasia (CDCBM10, OMIM 618677) both for autosomal recessive inheritance.

Finally, the locus at 21q22.3 telomeric in family A22 is reported in five other studies (SI, Table S2). In two families with DD, translocation breakpoints have been reported[44] with a phenotypically spectrum that includes language impairment and developmental coordination. A 175 kbp deletion that includes *PCNT*, *DIP2A* and *S100B* is reported for a family with dyslexia as part of the phenotype[45]. Other studies have shown association to *DIP2A*[46] or *S100B*[47], and finally, a GWA study by Gialluisi et al.[16] found association to rs73234886 in the *TSPEAR* gene and close to *KRTAP10-12*. The WGS analysis in family A22 did not find any variants in the aforementioned genes, and none of the three candidate SNPs in A22 were close to the reported loci (Table 2).

In conclusion we have mapped five different loci for DD by linkage analyses of six families. For all five loci we can find support in the literature either as linkage analysis, association studies or CNV mapping. The combination of linkage analyses of large families and the use of NGS of the mapped regions is relatively new[21,22]. The lack of deleterious mutation in coding genes causing DD suggest mutations for DD must involve regulation of genes involved in neurological processes. The large number of GWA studies done in the last decade by questioning school children with DD suggest that many different genes are involved in DD[16,17,20,35]. It is notable that the five DD loci found in this study are not found by the recent GWA studies. An explanation can be that GWAS only find regions where there is a relative high frequency of a founder mutation in the population.

These studies based on large populations are not able to detect DD loci that segregate in single families as a monogenic trait. Few family studies combining linkage analysis and NGS technology have resulted in variants in putative regulatory regions and found in families with AD inheritance of DD. In the present study, we have proposed candidate variants for regulatory regions for genes expressed in the brain, but additional functional analysis is needed for a more and deep knowledge of the DD etiology.

A better understanding of a molecular mechanistic genetic model of DD is needed for analysis of regulatory variants in possible candidate genes. An approach used by Price et al.[35] with a hypothesis-driven model combined with a GWAS where SNPs near or in genes involved in neuronal migration/axon guidance or implicated in autism spectrum might be useful for analyses of candidate variants in mapped linkage regions as done in this study. Future genetic analyses, either family studies or GWAS studies are needed to clarify the genetic components in DD and the overlapping comorbidities to other neurogenetic disorders.

Declarations

Acknowledgments

We gratefully acknowledge the families who have made all these studies possible.

The authors would like to thank the families for their participation in the project and Annemette Friis Mikkelsen is thanked for her excellent laboratory assistance.

Author contribution

Both authors have contributed to and approved the final manuscript. HE has designed the study, done the linkage and WGS data analyses, LH has done data analyses, drafted and written the manuscript together with HE.

Funding

This work was supported by research grants from The Augustinus Foundation, The Aase and Ejnar Danielsen's Foundation and The Lundbeck Foundation. Denmark, j.nr. R44-2009-4224.

Ethical Approval

The study protocols adhered to the tenets of the Declaration of Helsinki and the ARVO statement on human subjects and is approved by the Danish National Committee on Health Research Ethics in 2019 (H-19019167).

Competing interest

The authors declare that there is no conflict of interest, and there are not any competing financial interests in relation to the work described.

Data Availability Statement

Data are available from the corresponding author on reasonable request

Web sources

SNP6-LINK package is available from the corresponding author

UCSC (reference sequence GRCh37/hg19) (https://genome.ucsc.edu/)

VEP, Variant Effect Predictor/Ensembl (http://www.ensembl.org/index.html)

GTEx Portal (Release V8) The data used for the RNA analyses described in this manuscript were obtained from the GTEx Portal on 01/01/23 (https://gtexportal.org/home/)

HPA, The Human Protein Atlas (Version: 22.0**)**: https://www.proteinatlas.org/

PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/)

Supplementary information

Supplementary information is available at MP's website

References

1. Bishop DV, Laws G, Adams, C, Norbury, CF. High heritability of speech and language impairments in 6-year-old twins demonstrated using parent and teacher report. *Behav Genet* 2006; **36:** 173–184. https://doi.org/10.1007/s10519-005-9020-0
2. Hawke, JL, Wadsworth SJ, DeFries JC. Genetic influences on reading difficulties in boys and girls: the Colorado twin study. *Dyslexia* 2006; **12:** 21–29. https://doi.org/10.1002/dys.301
3. Aylward E, Richards T, Berninger V, Nagy W, Field K, Grimme A, et al. Instructional treatment associated with changes in brain activation in children with dyslexia. *Neurology* 2003; **61:** 212–219. https://doi.org/ 10.1212/01.wnl.0000068363.05974.64
4. Odegard TN, Ring J, Smith S, Biggan J, Black J. Differentiating the neural response to intervention in children with developmental dyslexia. *Ann Dyslexia* 2008; **58:** 1-14. https:/doi.org/10.1007/s11881-008-0014-5
5. Erbeli F, Rice M, Paracchini S. Insights into Dyslexia Genetics Research from the Last Two Decades. *Brain Sci* 2021; **12:** 27. https://doi.org/10.3390/brainsci12010027

6. Taipale M, Kaminen N, Nopola-Hemmi J, Haltia T, Myllyluoma B. Lyytinen H et al. A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proc Natl Acad Sci U S A* 2003; **30:** 100:11553-8. https://doi.org/10.1073/pnas.1833911100.
7. Meng H, Smith SD, Hager K, Held M, Liu J, Olson RK, et al. DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proc Nat Acad Sci USA* 2005; **102:** 17053–17058. https://doi.org/10.1073/pnas.0508591102
8. Cope N, Harold D, Hill G, Moskvina V, Stevenson J, Holmans P, et al. Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *Am JHum Genet* 2005; **76:** 581–591. https://doi.org/10.1086/429131
9. Hannula-Jouppi K, Kaminen-Ahola N, Taipale M, Eklund R, Nopola-Hemmi J, Kaariainen H, et al. The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS Genet* 2005; **1:** 467–474. https://doi.org/10.1371/journal.pgen.0010050
10. Mascheretti S, De Luca A, Trezzi V, Peruzzo D, A Nordio A , Marino C, et al. Neurogenetics of developmental dyslexia: from genes to behavior through brain neuroimaging and cognitive and sensorial mechanisms. *Transl Psychiatry* 2017; **7:** e987. https://doi.org/10.1038/tp.2016.240
11. Fisher SE, Franks C, Marlow AJ, MacPhie IL, Newbury DF, Cardon LR, et al. Independent genome-wide scans identify a chromosome 18 quantitative trait locus influencing dyslexia. *Nat Genet* 2002; **30:** 86–91. https://doi.org/ 10.1038/ng792
12. Svensson I, Nilsson S, Wahlström J, Jernås M, Carlsson LM, Hjelmquist E. Familial dyslexia in a large Swedish family: a whole genome linkage scan *Behav Genet* 2011; **41:** 43-9. https://doi.org/10.1007/s10519-010-9395-4
13. Meaburn EL, Harlaar N, Craig IW, Schalkwyk LC, Plomin R. Quantitative trait locus association scan of early reading disability and ability using pooled DNA and 100K SNP microarrays in a sample of 5760 children. *Mol Psychiatry* 2008; **13:** 729–740. https://doi.org/10.1038/sj.mp.4002063
14. Luciano M, Evans DM, Hansell NK, Medland SE, Montgomery GW, Martin NG, et al. A genome-wide association study for reading and language abilities in two population cohorts. Genes, *Brain Behav* 2013; **12:** 645–52. https://doi.org/10.1111/gbb.12053
15. Gialluisi A, Newbury DF, Wilcutt EG, Olson RK, DeFries JC, Brandler WM, et al. Genome-wide screening for DNA variants associated with reading and language traits. Genes *Brain Behav* 2014; **13:** 686–701. https://doi.org/10.1111/gbb.12158
16. Gialluisi A, Andlauer TFM, Mirza-Schreiber N, Moll K, Becker J, Hoffmann P, et al. Genome-wide association scan identifies new variants associated with a cognitive predictor of dyslexia. *Transl Psychiatry* 2019; **9:** 77. https://doi.org/10.1038/s41398-019-0402-0
17. Gialluisi A, Andlauer TFM, Mirza-Schreiber N, Moll K, Becker J, Hoffmann P, et al. Genome-wide association study reveals new insights into the heritability and genetic correlates of developmental dyslexia. *Mol Psychiatry* 2021; **26:** 3004–3017. https://doi.org/10.1038/s41380-020-00898-x
18. Truong DT, Adams AK, Paniagua S, Frijters JC, Boada R, Hill DE, et al. Multivariate genome-wide association study of rapid automatised naming and rapid alternating stimulus in Hispanic American and African–American youth. *J Med Genet* 2019; **56:** 557–566. https://doi.org/10.1136/jmedgenet-2018-105874
19. Eising E, Mirza-Schreiber N, de Zeeuw EL, Wang CA, Truong DT, Allegrini AG, et al. Genome-wide association analyses of individual differences in quantitatively assessed reading- and language-related skills in up to 34,000 people. *Proc Nat Acad Sci USA* 2022; **119:** e2202764119. https://doi.org/10.1073/pnas.2202764119
20. Doust C, Fontanillas P, Eising E, Gordon SD, Wang Z, Alagöz G, et al. Discovery of 42 genome-wide significant loci associated with dyslexia, *Nature Genetics* 2022; 54: 1621–1629. https://doi.org/10.1038/s41588-022-01192-y
21. Carrion-Castillo A, Estruch SB, Maassen B, Franke B, Francks C, Fisher SE. Whole-genome sequencing identifies functional noncoding variation in SEMA3C that cosegregates with dyslexia in a multigenerational family. *Hum Genet* 2021; **140:** 1183-1200. https://doi.org/10.1007/s00439-021-02289-w
22. Grimm T, Garshasbi M, Puettmann L, Chen W, Ullmann R, Müller-Myhsok B, et al. A Novel Locus and Candidate Gene for Familial Developmental Dyslexia on Chromosome 4q. Z Kinder *Jugendpsychiatr Psychother* 2020; **48:** 478-489. https://doi.org/10.1024/1422-4917/a000758
23. Eiberg H, Nielsen LS, Klausen J, Dahlén M, Kristensen M, Bisgaard ML, et al. Linkage between serum cholinesterase 2 (CHE2) and gamma-crystalline gene cluster (CRYG): assignment to chromosome 2. *Clin Genet* 1989; **35:** 313-321. https://doi.org/10.1111/j.1399-0004.1989.tb02951.x
24. Ott JA. computer program for linkage analysis of general human pedigrees. Am J Hum Genet. 1976; 28:528-529. PMC168511425.
25. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25:** 1754-60. https://doi.org/10.1093/bioinformatics/btp324
26. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; **43:** 491-8. https://doi.org/10.1038/ng.806

27. Nassar LR, Barber GP, Benet-Pagès A, Casper J, Clawson H, Mark Diekhans M et al. The UCSC Genome Browser database: 2023 update. *Nucleic Acids Research* 2023; **51:** Issue D1, D1188–D1195, https://doi.org/10.1093/nar/gkac1072
28. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biology* 2016; **17:** 122. https://doi.org/10.1186/s13059-016-0974-4
29. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 01/01/23
30. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome.*Science* 2015; **347:** 1260419. https://doi.org//10.1126/science.1260419
31. DeFries J.C., Fulker DW, LaBuda MC. Evidence for a genetic aetiology in reading disability of twins. *Nature* 1987; **329:** 537-539. https://doi.org/10.1038/329537a0
32. Carrion-Castillo A, Franke B, Fisher SE. Molecular genetics of dyslexia: an overview. *Dyslexia* 2013; **19:** 214–240. https://doi.org/10.1002/dys.1464
33. Guidi LG, Velayos-Baeza A, Martinez-Garay I, Monaco AP, Paracchini S, Bishop DVM, et al. The neuronal migration hypothesis of dyslexia: A critical evaluation 30 years on. *Eur J Neurosci* 2018; **48:** 3212–33. https://doi.org/10.1111/ejn.14149
34. Price KM, Wigg KG, Feng Y, Blokland K, Wilkinson M, He G et al. Genome-wide association study of word reading: Overlap with risk genes for neurodevelopmental disorders. *Genes Brain Behav* 2020; **19:** e12648. https://doi.org/10.1111/gbb.12648
35. Price KM, Wigg KG, Eising E, Feng Y, Blokland K, Wilkinson M, et al. Hypothesis-driven genome-wide association studies provide novel insights into genetics of reading disabilities. *Transl Psychiatry* 2022; **12:** 495. doi: 10.1038/s41398-022-02250-z
36. 36 Igo RP, Chapman NH, Berninger VW, Matsushita M, Brkanac Z, et al. Genomewide scan for real-word reading subphenotypes of dyslexia: novel chromosome 13 locus and genetic complexity. *Am J Med Genet B Neuropsychiatr Genet* 2006; **141B:** 15-27. https://doi.org/10.1002/ajmg.b.30245
37. Truong DT, Shriberg LD, Smith SD, Chapman KL, Scheer-Cohen AR, DeMille MM, et al. Multipoint genome-wide linkage scan for nonword repetition in a multigenerational family further supports chromosome 13q as a locus for verbal trait disorders. *Hum Genet* 2016; **135:** 1329-1341. https://doi.org/ 10.1007/s00439-016-1717-z
38. Chapman J. The Molecular Genetics of Developmental Dyslexia. Thesis 2011. Cardiff University (United Kingdom) ProQuest Dissertations Publishing, 2011. U585465
39. Bates TC, Luciano M, Castles A, Coltheart M, Wright MJ, Martin NG. Replication of reported linkages for dyslexia and spelling and suggestive evidence for novel regions on chromosomes 4 and 17. *Eur J Hum Genet* 2007; **15:** 194–203. https://doi.org/10.1038/sj.ejhg.5201739
40. Fisher, SE, Francks, C., Marlow, AJ, MacPhie IL, Newbury DF, Cardon LR, et al. Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. *Nature Genet* 2002; **30:** 86-91. https://doi.org/10.1038/ng792
41. Seshadri S, DeStefano AL, Au R, Massaro JM, Beiser AS, Kelly-Hayes M, et al. . Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham Study. *BMC Med Genet* 2007; **8 Suppl 1:** S15. https:/doi.org/10.1186/1471-2350-8-S1-S15
42. Scerri TS, Paracchini S,Morris A, MacPhie IL, Talcott J, Stein J, et al. Identification of candidate genes for dyslexia susceptibility on chromosome 18. *PLoS One* 2010; **5:** e13712. https:/doi.org/10.1371/journal.pone.0013712
43. Palumbo P, Palumbo O, Leone MP, Stallone R, Palladino T, Zelante L, et al. Clinical and molecular characterization of a de novo 19p13.3 microdeletion. *Mol Cytogenet* 2016; **9:** 40. http://doi.org/10.1186/s13039-016-0252-x.
44. Coton J, Labalme A, Till M, Bussy G, Krifi Papoz S, Lesca, et al. Characterization of two familial cases presenting with a syndromic specific learning disorder and carrying (17q;21q) unbalanced translocations. *Clin Case Rep* 2018; **6:** 827-834. https://doi.org/10.1002/ccr3.1450
45. Poelmans G, Engelen JJ, Van Lent-Albrechts J, Smeets HJ, Schoenmakers E, Franke B, et al. Identification of novel dyslexia candidate genes through the analysis of a chromosomal deletion. *Am J Med Genet B Neuropsychiatr Genet* 2009; **150B:** 140-7. https://doi.org/10.1002/ajmg.b.30787
46. Kong R, Shao S, Wang J, Zhang X, Guo S, Zou L, et al. Genetic variant in DIP2A gene is associated with developmental dyslexia in Chinese population. *Am J Med Genet B Neuropsychiatr Genet* 2016; **171B:** 203-8. https://doi.org/10.1002/ajmg.b.32392
47. Matsson H, Huss M, Persson H, Einarsdottir E, Tiraboschi E, Nopola-Hemmi J, et al. Polymorphisms in DCDC2 and S100B associate with developmental dyslexia. *J Hum Genet* 2015; **60:** 399-401. https://doi.org/10.1038/jhg.2015.37

Tables

Table 1.

Family	Chr. band	Linkage region (hg19)	Delimiting markers	Size of linkage region	LOD score [2)]
E06	13q12.3	13:29,324,683-30,941,039	rs1005969-rs1472248718	1,616,356	4.42 (1.52)
A29	15q23-q24.1	15:70,362,585-73,666,730	rs12593849-rs59322618	3,304,145	3.01 (2.10)
A67	18p11.21	18:10,905,079-11,901,888	rs7241188-rs9963456	996,809	3.87 (3.57)
B41	19p13.3	19:366,412-6,755,007	rs689396-rs2305806	6,388,595	4.72 (2.64)
A09	19p13.3	19:1,364,306- 2,827,300	rs17673260-rs2159561	1,462,994	2.64 (1.7)
A22	21q22.3	21:44,828,031-48,129,895	rs857552-rs2839367 [1)]	3,301,865	3.30 (1.52)

1. Close to telomer position.
2. LOD scores for dyslexia. In parenthesis are LOD scores calculated using the first collected samples genotyped using SNP6.0 arrays. Final LOD scores are calculated between dyslexia and the haplotype (p=0.001) for all family members.

Table 2

Candidate SNPs found in the mapped linkage regions

Family	SNP	Location hg19	Minor allele	MAF	Consequence	Gene	Biotype
E06	rs117556116	13:29371767	A	0.00281	regulatory_region_variant	*MTUS2*	CTCF_binding site
	rs573197999	13:30003980	T	0.00238	intron variant	*MTUS2*	Promoter, CTCF_binding_site
A29	rs35304292	15:70794485	T	0.00581	downstream_gene_variant	*LINC02205*	lncRNA
	rs1196366533	15:71306799-71306800	delG	0.00000	intron_variant, many TFBS	*LRRC49*	protein_coding
A67	rs143669678	18:11026113	A	0.00728	intron_variant	*PIEZO2*	protein_coding
					regulatory_region_variant	-	CTCF_binding_site
					regulatory_region_variant	-	enhancer
B41	rs144512862	19:1377234	T	0.01182	3_prime_UTR_variant	*MUM1/ PWWP3A*	protein_coding
					regulatory_region_variant	-	promoter_flanking_region
	rs116900972	19:1378011	G	0.007999	3_prime_UTR_variant	*MUM1/ PWWP3A*	protein_coding
					regulatory_region_variant	-	CTCF_binding_site
					regulatory_region_variant	-	promoter_flanking_region
	rs147204443	19:1380013	T	0.01221	downstream_gene_variant	*MUM1/ PWWP3A*	protein_coding
					regulatory_region_variant	-	promoter_flanking_region
	rs557485888	19:1445779	T	0.003428	upstream_gene_variant	*APC2*	protein_coding
					regulatory_region_variant	-	promoter
					regulatory_region_variant	-	CTCF_binding_site
					TF_binding_site_variant	-	-
	rs529177770	19:1472092	GGGGG	0.000015	3_prime_UTR_variant	*APC2*	protein_coding
	rs201353187	19:1475258	G	0.006159	downstream_gene_variant	*APC2*	protein_coding
					splice_acceptor_variant	*C19orf25*	protein_coding
					regulatory_region_variant	-	CTCF_binding_site
	rs115178429	19:1882327	A	0.000647	intron_variant	*ABHD17A*	protein_coding
					regulatory_region_variant	-	promoter
	rs149364482	19:1987562	A	0.006151	missense_variant	*BTBD2*	protein_coding
					regulatory_region_variant	-	CTCF_binding_site
	rs117195808	19:3060851	A	0.003428	intron_variant	*TLE5*	protein_coding
					regulatory_region_variant	-	promoter
	rs1364917700	19:3666035	T	0.000015	intron_variant	*PIP5K1C*	protein_coding
					regulatory_region_variant	-	CTCF_binding_site
					TF_binding_site_variant	-	-
A09	rs12974027	19:1457509	C	0.01625	intron_variant	*APC2*	protein_coding
					regulatory_region_variant	-	CTCF_binding_site
	rs147084328	19:1476199	T	0.00805	downstream_gene_variant	*APC2*	protein_coding

					intron_variant	*C19orf25*	protein_coding
	rs118087435	19:1613257	T	0.00056	intron_variant	*TCF3*	protein_coding
	rs193271498	19:1874187	G	0.00112	regulatory_region_variant	-	promoter
	rs146449301	19:2013973	C	0.00190	intron_variant	*BTBD2*	protein_coding
					regulatory_region_variant	-	promoter
	rs148452202	19:2527577	A	0.01625	intron_variant	*GNG7*	protein_coding
					regulatory_region_variant	-	promoter_flanking_region
	rs147662493	19:2528978	T	0.00805	intron_variant	*GNG7*	protein_coding
A22	rs539002811	21:45175002	A	0.00266	intron_variant	*PDXK*	protein_coding
					regulatory_region_variant	-	TF_binding_site
	rs572129208	21:46584395	T	0.00351	intron_variant	*ADARB1*	protein_coding
					regulatory_region_variant	-	promoter_flanking_region
	rs560135812	21:46994131	T	0.00539	regulatory_region_variant	-	promoter_flanking_region

Figures

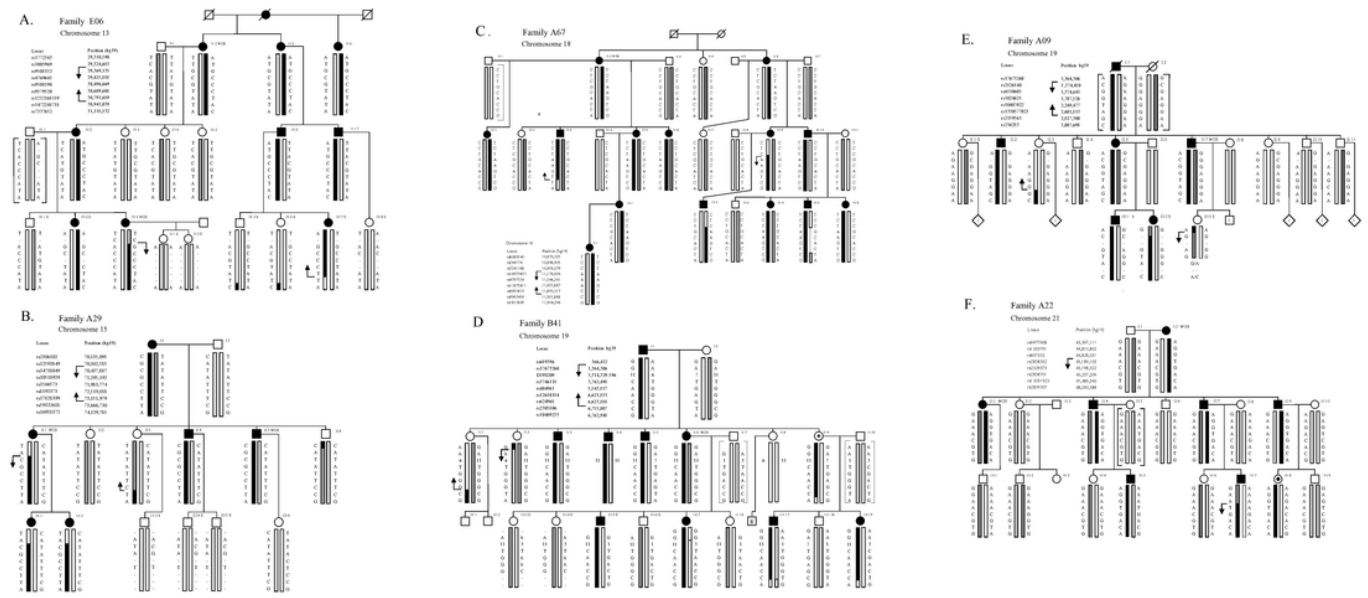

Figure 1

Pedigrees for six Danish families with DD segregation in an autosomal dominant order. **A.** Pedigree of family mapped to locus 13q12.3. **B.** Pedigree of family A29 mapped to

locus 15q23-q24.1. **C.** Pedigree of family A67 mapped to 18p11.21.**D.** Pedigree of family B41 mapped to locus 19p13.3. **E.** Pedigree of family A09 at locus 19p13.3. **F**. Pedigree of family A22 at locus 21q22.3. Females are shown as circles, males as squares, healthy individuals have open symbols, affected individuals have filled black symbols. Symbols with a black dot indicated healthy carriers of the disease trait. WGS denotes individual's whole genome sequenced and S denotes individuals genotyped for selected SNPs. Parenthesis denotes inferred haplotypes.

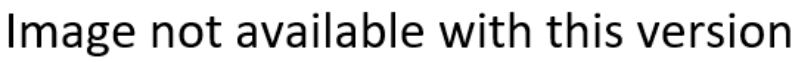

Figure 2

Graphic presentation of the LOD scores from the initial WGL analysis of the six families.

Figure 3

These images are not available with this version.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarydata.docx