ISSN: 2993-5377

Research Article

Open Access Journal of Applied Science and Technology

Linkage and Next Generation Sequencing (NGS) Data in Six Large Danish Families with Dyslexia

Hans Eiberg[1*] and Lars Hansen[2,3]

[1]*RCLINK, Stormosevej 18, 2765 Smørum. Denmark*

[2]*Department of Cellular and Molecular Medicine, University of Copenhagen, Denmark*

[3]*Department of Science and Environment, Roskilde University, Roskilde, Denmark*

[*]Corresponding Author
Hans Eiberg, RCLINK, Stormosevej 18, 2765 Smørum. Denmark.

Submitted: 2023, Aug 08; **Accepted**: 2023, Aug 29: **Published**: 2023, Sep 04

Citation: Eiberg, H., Hansen, L. (2023). Linkage and Next Generation Sequencing (NGS) Data in Six Large Danish Families With Dyslexia. *OA J Applied Sci Technol, 1*(2), 113-125.

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, and the molecular mechanisms underlying 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 dyslexic 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. Whole genome sequencing of dyslexic individuals in the six family's failed to identify protein located mutations and a catalogue of possible regulatory variants are suggested as causative for dyslexia.

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

1. 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* for DYX1, *DCDC2* and *KIAA0319* for DYX2, and *ROBO1* for DYX5 [6-9]. The strength of gene-mapping studies has been limited that only putative functional variants affecting genes are reported, and the genetic complexity of DD, where factors as incomplete penetrance, phenocopies, genetic heterogeneity and oligogenic inheritance, has hampered identification of single genetic variants for DD [10-12].

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 several other GWAS studies querying common genetic variants across the whole genome for association with DD in larger populations have been reported [13-19]. A recent GWA study by Doust et al., 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 [20]. 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 of large families followed by whole genome sequencing (WGS) of the linkage regions have resulted in novel putative DD candidate genes as *SEMA3C* and *SPY* [21, 22]. We have used a similar approach of WGL analyses of six large Danish DD families and WGS of one or two family members in each family. Five loci linked with the DD trait (Table 1) that two families mapped to the same region, and one locus overlapped with the DYX6 locus. WGS analyses of the linkage regions resulted in

a catalogue of putative regulatory variants that might contribute to the DD phenotype.

Family	Chr. band	Linkage region (hg19)	Delimiting markers	Size of linkage region	LOD score [1]
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 (1.70)
A22	21q22.3	21:44,828,031-48,129,895	rs857552-rs2839367	3,301,865	3.30 (1.52)

[1] Initial LOD scores calculated from SNP6.0 arrays shown in parenthesis.

Table 1: The Mapped Linkage Regions for the Six Families with Segregating Dyslexia

2. Materials and Methods

2.1. Family Material

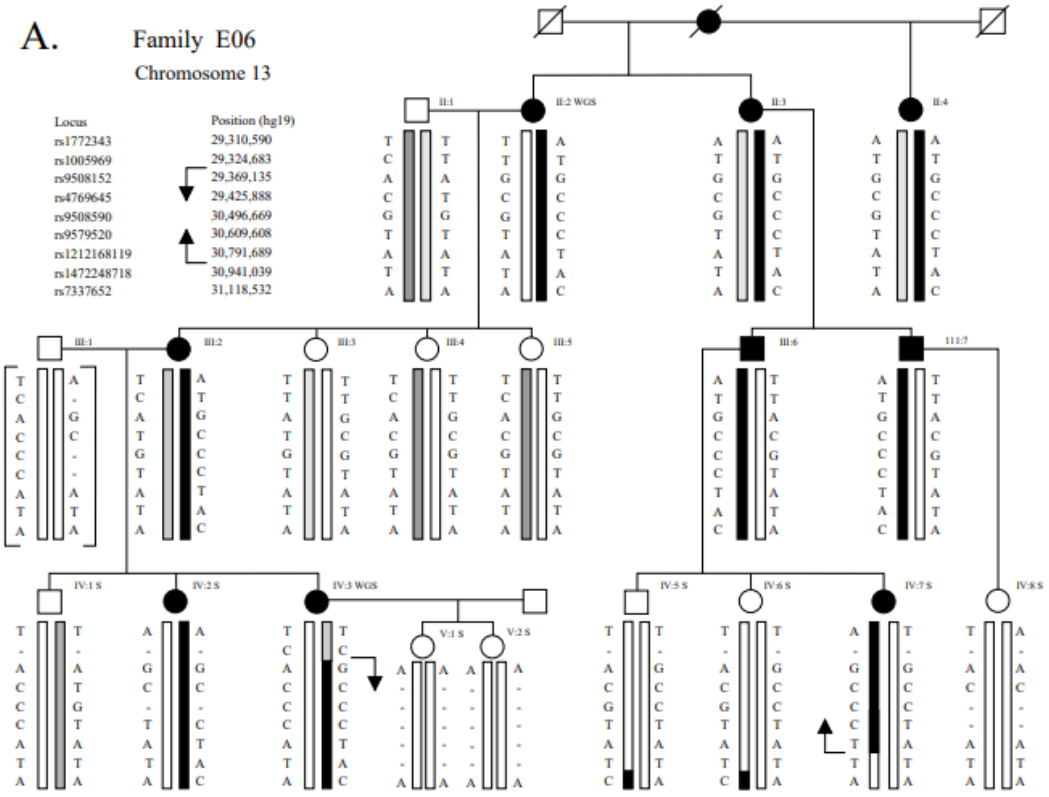
Six families with dyslexia were collected from the Copenhagen Family Bank [23]. The families were followed over a period of over 40 years with new family members registered. The DD phenotypes were self-reported by interview of parents and sibs over the entire period, and all participants were orally informed about the project and provided written consent achieved. All individuals have been detailed informed about the linkage mapping of the family by SNP array technology and WGS analyses and the consequences the data analyses of selected family members.

2.2. Genotyping and Whole Genome Linkage Analysis

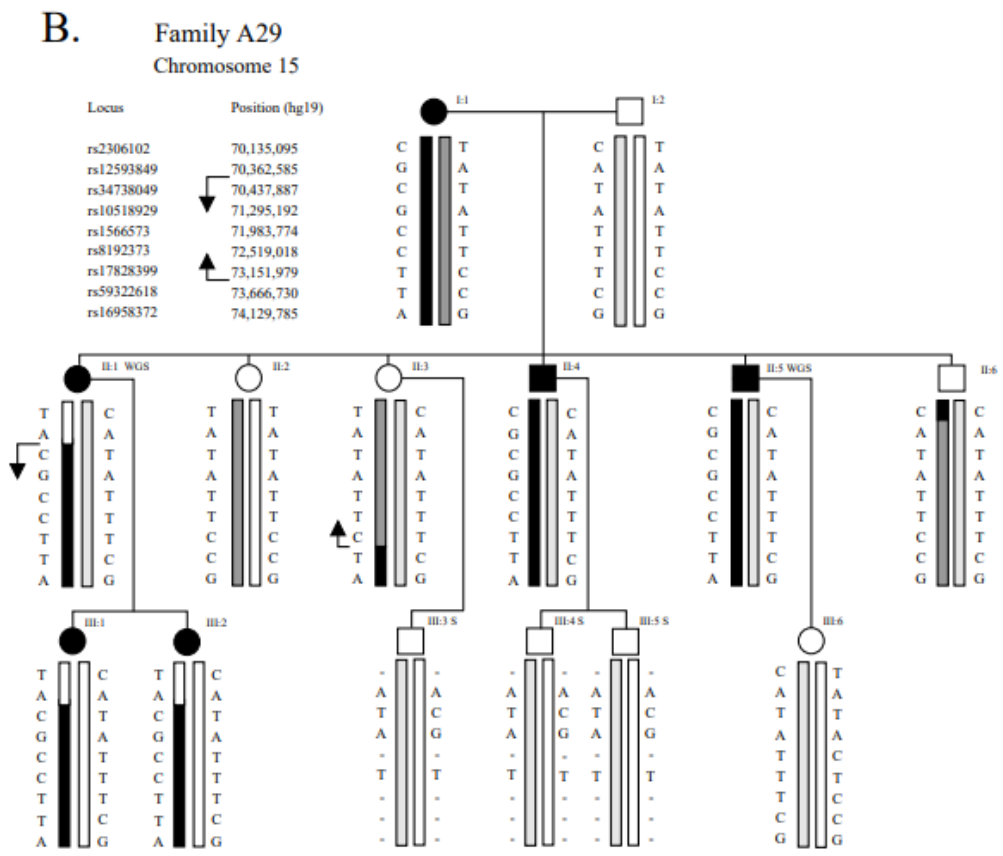
DNA was extracted using EDTA (ethylenediamine tetra-acetic acid) blood by standard phenol/chloroform extraction protocols and genotyped by SNP arrays. Family A09, A22, A67, B71, E06 were genotyped using Affymetrix SNP6.0 arrays (Affymetrix, Thermo Fisher Scientific Inc) and family A29 was genotyped applying CytoScan™ HD SNP arrays (Affymetrix, Thermo Fisher Scientific Inc). More than 900,000 high average heterozygosity markers were included in the WGL analyses. Birdseed files with genotype, chromosome, and position were converted to input files and analyzed by the SNP6-LINK program package. A two-points linkage analysis was carried employing the program LIPED, initial calculation of LOD scores for the six families was carried out with a frequency of 0.02 for the dyslexic trait and a penetrance of 100% and an allele frequency of 0.5 for the SNPs. The LOD scores were sorted by chromosome and position using the SNP6-LINK program package and plotted graphicly using Excel and continuous regions with positive LOD scores >2.0 were mapped [24]. In four families (E06, A29, B41, A09) additional members were sampled and genotyped by Sanger sequencing. The genotyping included the following individuals: family E06, SNP array: II:1-4; III:2-7; IV:3, Sanger sequencing: IV:1,2,5-8; V:1,2; family A29, SNP array: I:1,2; II:1-6; III:1-2,6, Sanger sequencing: III:3-5; family A67, SNP array: I:2-4,6,7; III:1-3,5,11; IV:1-6; V:1; Sanger sequencing: none; family B41, SNP array: I:1-2; II:1-3,5,6,9; III:7,17-19, Sanger sequencing or STS markers: II:4,8; III:3-6, 8, family A09, SNP array: II:2-5,7,9-11; Sanger sequencing II:1; III:1-3; family A22, SNP array I:1,2; II:1-2, 6-10; III:1-2, 4-9); Sanger sequencing: none. Fine mapping of recombination events included Sanger sequencing of the following SNPS in all family members, E06: rs1772343, rs1212168119, rs1472248718; A29: rs12593849 and rs8192373; B41: rs12610184, D19S209; A09: rs17673260, rs2526140, rs6510605, rs1024625, rs10407022, rs1330577023, rs2159561, rs216283. A final LOD score for each family using genotypes from SNP arrays and Sanger sequencing was calculated with LIPED with a frequency for the dyslexia trait of 0.02 and a penetrans of 0.05 and a risk haplotype frequency of 0.01 [24].

2.3. WGS-Data Analyses

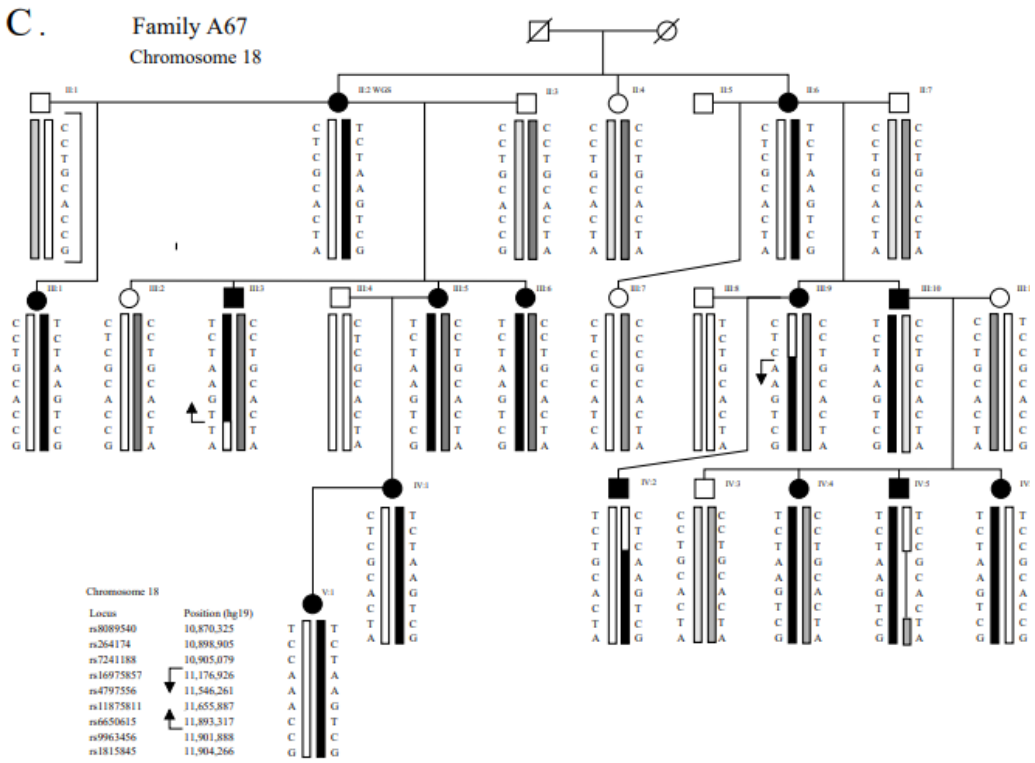
WGS was done by BGI Europe (Copenhagen, Denmark) by standard methods. Briefly a ≤800bp insert normal library was created for selected individuals and reads were aligned to human reference sequence hg19, GRCh37 using the BWA (0.7.15) aligner [25]. Variant calling was done employing GATK (4.0.11.0) and variant annotation, filtration was done using VarSeq (Golden Helix, USA), and a minimum coverage of 20 reads was obtained [26]. The WGS data was filtered for heterozygous SNVs and indels with a minor allele frequency (MAF) <0.02. Repeat Masker excluded variants in repeated regions in the UCSC browser [27]. In families with two individuals, deep sequenced, shared variants were selected for the analyses. The following individuals were selected for WGS: family E06: II:2; IV:3, family A29: II:1,5, family A67: II:2, family B41: II:6; family A09: II:7; family A22: I:2, II:1 (Figure 1A-F). The filtered variants were analyzed for positions in protein coding genes (exon/introns), in non-coding RNA genes, intergenic regions and for affecting regulatory regions employing the Variant Effect Predictor (VEP, Ensembl, assembly GRCh37/hg19), and gene expression data was obtained from the GTEx Portal and the Human Protein Atlas [28-30].



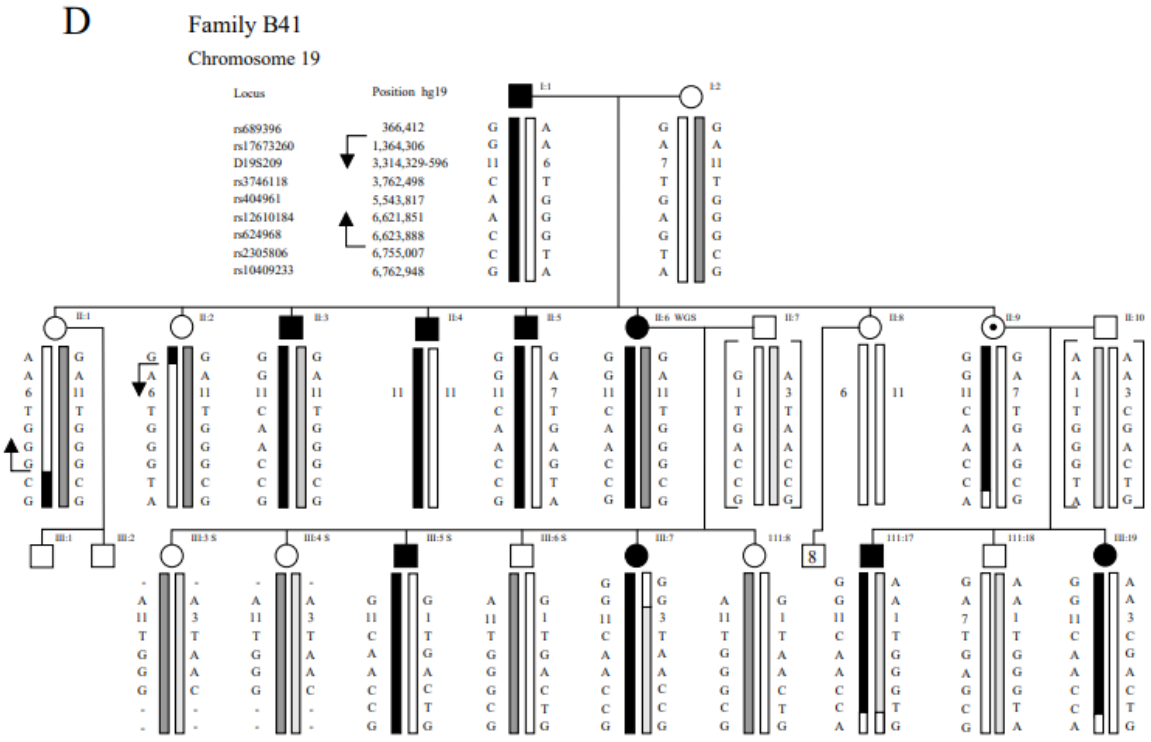
A. Pedigree of Family E06 That Map To Locus 13q12.3.



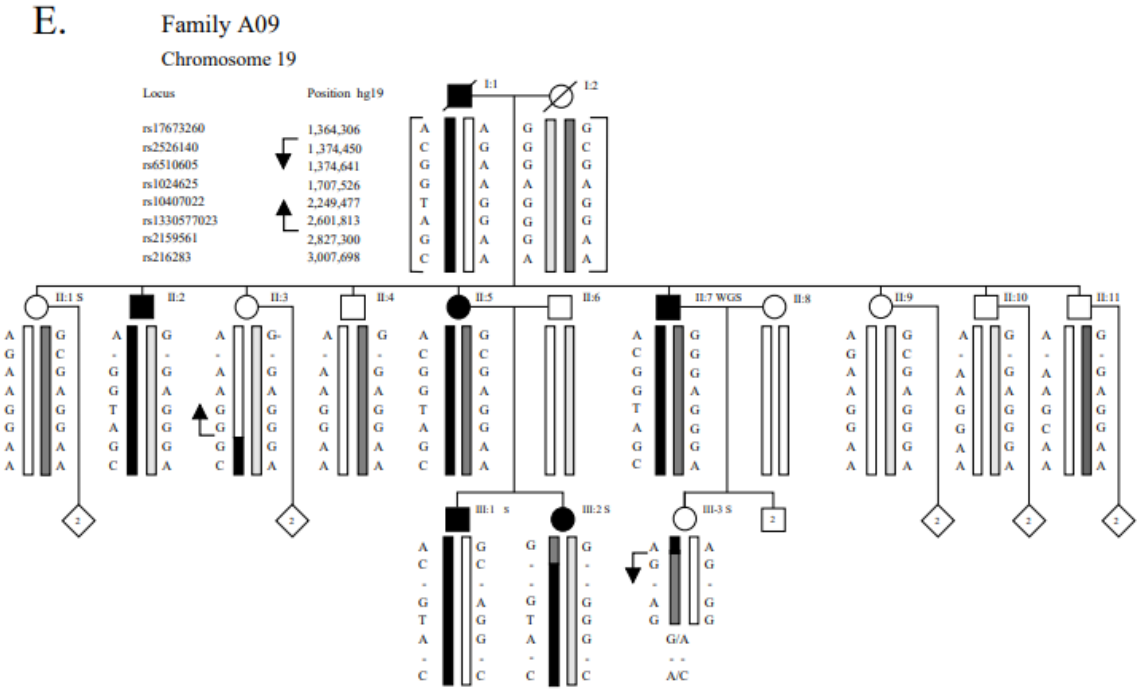
B. Pedigree of Family A29 Mapped to Locus 15q23-q24.1.



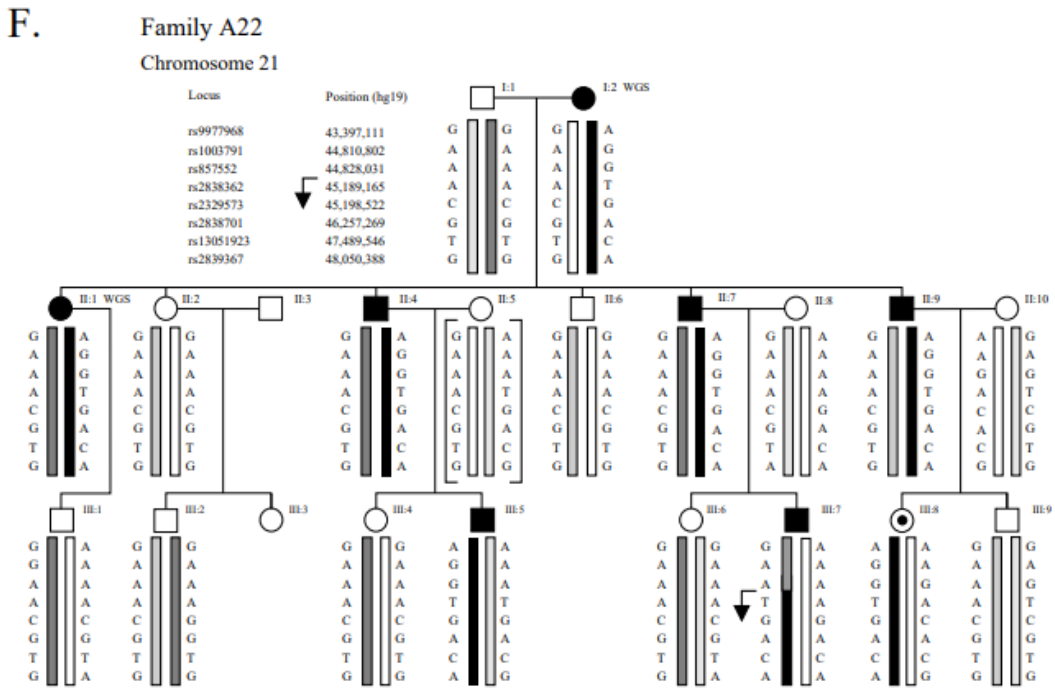
C. Pedigree of Family A67 Mapped to 18p11.21.



D. Pedigree of Family B41 Mapped to Locus 19p13.3.



E. Pedigree of Family A09 at Locus 19p13.3.



E. Pedigree of Family A22 at Locus 21q22.3.

Figure 1: Pedigrees for six Danish families with DD segregation in an autosomal dominant order. For all six pedigrees 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 individuals that have been whole genome sequenced and S denotes individuals genotyped for informative SNPs by Sanger sequencing. All other individuals were included in the SNP array analyses. Parenthesis denotes inferred haplotypes.

3. Results

3.1. Linkage Analyses

The linkage analyses of the six families applying the SNP array technology resulted in genotype data for the entire genome. Calculation of LOD scores followed by a graphic presentation revealed single regions with positive LOD scores and the remaining part of the genome could be excluded by negative LOD scores. For each family only one single region was observed except for family A09 that two regions on chromosome 19 were observed (Figure 2).

Figure 2

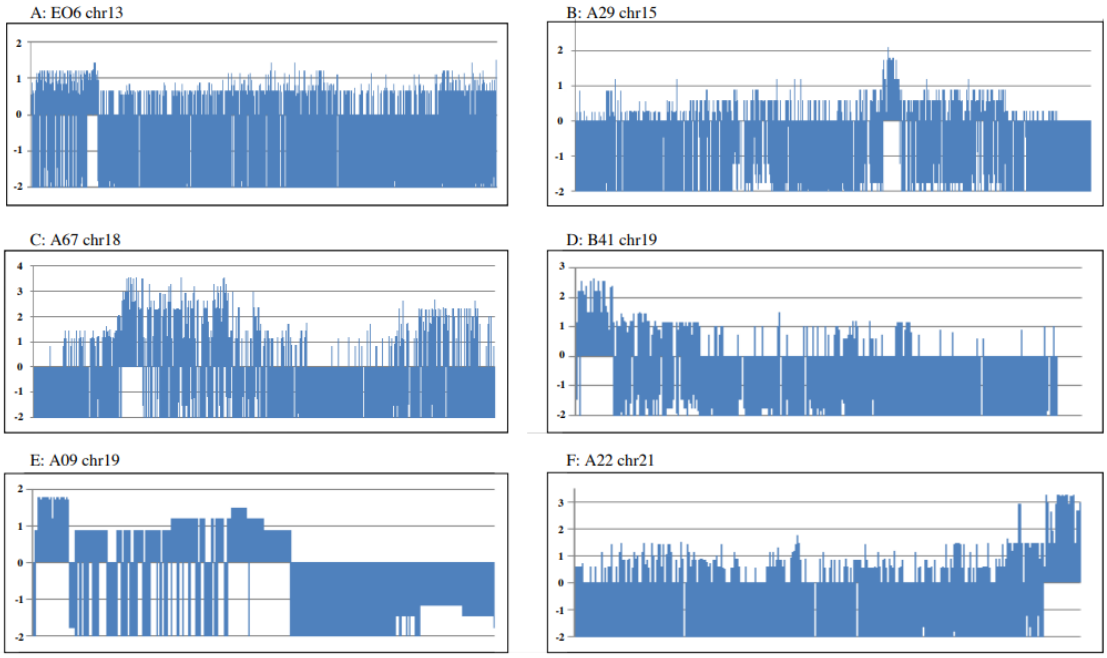


Figure 2: Graphic presentation of the LOD scores from the initial WGL analysis of the six families. Plotting LOD scores for all calculated genotype positions results in continues negative LOD scores for all chromosomal regions except where linkage is found.

All families had LOD scores between 1.52 and 3.57 (Table 1). By sampling additional family member for four families and subsequently genotyping of informative SNPs fine mapping of the linkage region and a final LOD score for each family was calculated (Table 1). Genotyping of additional family member in A09 excluded one of two linkage regions found by SNP array typing, and risk haplotype could be established in each family with well-defined boundaries for the DD loci.

3.2. Family E06

The DD trait mapped to a 1,6Mbp region at 13q12.3. The family comprised four generations with 20 individuals whereof 9 were reported with DD (Figure 1A). Eleven individuals were genotyped by SNP-arrays with a resulting LOD score of 1.52 (Figure 2A), and genotyping of additional 8 individuals for informative SNPs fine mapped the linkage region and a LOD score of 4.42 was calculated for a haplotype frequency p=0.001. Recombination in IV:3 and IV:7 delimited the region at rs1005969 and rs1472248718 (Figure 1A).

3.3. Family A29

The family represent three generations with 14 individuals and 7 reported with DD (Figure 1B). SNP-array analysis of 11 individuals revealed a positive LOD scores 15q23-q24.1 with Z=2.1, and by including three extra individuals and Sanger sequencing of informative markers (Figure 2B) resulted in a LOD score Z=3.01. The linkage region was 3.3Mbp delimited by the markers rs12593849 and rs59322618 due to recombination in II:1 and II:3.

3.4. Family A67

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

3.5. Family B41

A tree generations family of 31 individuals with 9 reported with DD (Figure 1D) were sampled. LOD score calculation using SNP-array data for 12 persons revealed a continuous region with a LOD score of Z=2.64 (Figure 2D). Including additional 7 family members tested for informative SNPs and an STS marker resulted in a final LOD score Z=4.72 for a haplotype frequency p=0.001. The region covered 6,4Mbp delimited by the markers rs689396 and rs2305806 due to recombination in II:1 and II:2.

3.6. Family A09

The family comprises 16 individuals in three generations with five members reported for dyslexia (Figure 1E). Genotyping of 8 individuals resulted in two continuous regions with a positive LOD score of Z=1.7 (Figure 2E). Genotyping of additional four individuals for informative SNP markers excluded one linkage regions leaving the region at 19p13.3 with a positive LOD score of maximum Z=2.64 for a haplotype frequency p=0.001. The region covered 1.5Mbp delimited by the markers rs17673260 and rs2159561 due to recombination in II:3 and III:3 and is embedded in the linkage region mapped for family B41.

3.7. Family A22

The family represent three generations with 21 individuals whereof 7 was reported with dyslexia (Figure 1F). 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 2F). 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 recombination in III:7.

3.8. The WGS Analyses

Two individuals in families E06 and A22 and one individual in A29, A67, B41 and A09 were chosen for whole genome sequencing. The variants in the linkage regions between the boundary SNPs were filtered for heterozygous SNVs and indels with MAF values <0.02. All variants located in repeated regions were excluded and the resulting SNVs and indels were analyzed by VEP for variants in regulatory regions. The analyses revealed no variants affecting coding gene regions wherefore intergenic and intron variants were analyzed for location regulatory regions. This resulted in a catalogue of 22 SNVs and one indels (Table 2).

Family	SNP id	Location (hg19/ GRCh37)	Alleles	MAF [1] (effect allele)	Consequence [2]	Gene [2]	Biotype [2]
E06	rs117556116	13:29371767	T/A	A:0.01969	Intergenic, regulatory region variant		CTCF-binding site
	rs573197999	13:30003980	C/T	T:0.00231	Intron, regulatory variant	*MTUS2*	Promoter
A29	rs8192373	15:72519018	T/C	C:0.01630	Intron, regulatory region variant	*PKM*	Promotor, CTCF-binding site
A67	rs143669678	18:11026113	G/A	A:0.00867	Intron, regulatory region variant	*PIEZO2*	CTCF-binding site, enhancer
	rs185745732	18:11259814	C/T	T:0.00141	Intergenic, regulatory region variant		Enhancer
B41	rs144512862	19:1377234	G/T	T:0.01182	3'UTR, Regulatory region variant	*PWWP3A*	Promoter flanking region
	rs116900972	19:1378011	C/G	G:0.01219	Downstream, regulatory region variant	*PWWP3A*	Promoter flanking region, CTCF-binding site
	rs147204443	19:1380013	G/T	G:0.01221	Intergenic, regulatory region variant		Promoter flanking region
	rs557485888	19:1445779	C/T	T:0.00538	Upstream, regulatory region variant	*APC2*	Promoter, CTCF-binding site, TF-binding site variant
	rs201353187	19:1475258	C/G	G:0.00616	Splice acceptor variant regulatory region variant	*C19orf25*	Protein coding, CTCF-binding site
	rs115178429	19:1882327	G/A	A:0.000647	Intron/upstream, regulatory region variant	*ABHD17A*	Promoter
	rs149364482	19:1987562	C/A	A:0.00615	Missense variant (S373I) Regulatory region variant	*BTBD2*	Protein coding (SIFT predicted benign) CTCF binding site

	rs117195808	19:3060851	G/A	A:0.00015	Intron, regulatory region variant	*TLE5*	Promoter
	rs1364917700	19:3666035	C/T	T:0.00000	Intron, regulator region variant	*PIP5K1C*	CTCF-binding site, TF-binding site variant
	rs1049614944	19:3643060	T/C	C:0.00000	Intron, regulator region variant	*PIPSKIC*	TF-bindings site variant
A09	rs3837993	19:730206	-/A	A:0.01467	Intron, regulatory region variant	*PALM*	Promoter flanking region
	rs12974027	19:1457509	T/C	C:0.01467	Intron, regulatory region variant	*APC2*	CTCF-binding site
	rs193271498	19:1874187	C/G	G:0.00065	Downstream gene variant Regulatory region variant	*ABHD17A*	Promoter
	rs146449301	19:2013973	G/C	C:0.00520	Intron, regulatory region variant	*BTBD2*	Promoter
	rs148452202	19:2527577	G/A	A:0.01467	Intron, regulatory region variant	*GNG7*	Promoter flanking region
A22	rs539002811	21:45175002	G/A	A:0.00518	Intron, regulatory region variant	*PDXK*	TF-binding site
	rs572129208	21:46584395	C/T	C:0.00542	Intron, regulatory region variant	*ADARB1*	Promoter flanking region
	rs560135812	21:46994131	C/T	T:0.00556	Intergenic, regulatory region variant	-	Promoter flanking region

(1) MAF values are from gnomAD v3.1.2 for the European non finish population.
(2) VEP analyses done for assembly GRCH37/hg19.

Table 2: Candidate SNPs for the Dyslexia Trait Found in the Linkage Regions

13q12.3 locus. Two DNA variants, rs573197999 and rs117556116, in family E06 II:2 and IV:3 was in regulatory regions in the vicinity of the gene *MTUS2*. rs117556116 is in a CTCF binding site upstream for the gene and rs573197999 is in a promotor region in intron 1 (Table 2). *MTSU2* is expressed in heart and is regional enhanced in the cerebral cortex and single-cell RNA data suggests expression in neurons (data HPA).

15q23-q24.1 locus. One SNV, rs8192373, was found in II:1 and II:5 in family A29 in intron 1 of the gene *PKM*. The SNV affects a predicted promoter region and the *PKM* gene is ubiquitously expressed.

18p11.21 locus. Two SNVs, rs143669678 and rs185745732, was found in a regulatory enhancer and CTCF-binding region. rs143669678 is in intron 2 of *PIEZO2* and rs185745732 is intergenic upstream for *PIEZO2*. The gene is expressed in the brain (GTEx and HPA) (Table 2).

19p13.3 locus. A total of 10 SNVs was found family B41 and five in family A09 (Table 2). None of the SNVs were recurrent in both families, but variants in three genes, *APC2*, *ABHD17A* and *BTBD2*, were found in B41 and A09. Two SNVs, rs557485888 in B41 and rs12974027 in family A09 were upstream or in an intron of the gene *APC2*, all in regulatory regions. One SNV, 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. Two SNVs, rs115178429 and rs193271498, were in regulatory regions of *ABHD17A*, one in intron 1 and one downstream, and *ABHD17A* is ubiquitously expressed including the brain. Two SNVs in or close to the gene *BTBD2*, included a missense variant (rs149364482; p.Ser373Ile/NP_060267.2, benign PolyPhen2) and rs146449301 in a regulatory region in intron 1 of *BTBD2. BTBD2* is ubiquitously expressed with low tissue specificity. The remaining SNVs were in regulatory regulator regions close to in the genes *PWWP3A, TLE5, PIP5K1C, TCF3* and *GNG7*, where *GNG7* is highly expressed in the basal ganglia (HPA) (Table 2).

19p13.3 locus. Family A22 revealed three SNVs of interest 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.

4. Discussion

The genetic etiology of dyslexia is complex demonstrated by many studies (for reviews see [14, 17, 31, and 32]). A small number of genes has been characterized as DD susceptibility genes without single deleterious mutations found segregating in families or as recurrent mutations in cohorts of individuals with DD. Functional consequences for variants in susceptibility genes have not been demonstrated, likely due to lack of specific cell or tissue activities. DD is supposed to be a neurodevelopmental disability and disturbed activities in the brain is the most obvious cause, and characterized as polygenetic with a strong genetic component and heritability estimated to 40-60% [17]. A potential molecular mechanism is supposed to be linked to neuronal migration [33]. Comorbidities between dyslexia and other neurodevelopmental disorders has been reported with overlapping loci for DD, autism, and ADHD [34, 35]. Other studies show families with autosomal dominant inheritance of DD and cases of incomplete penetrance, autosomal recessive inheritance has so far not been reported [6-9, 36, 37].

In the present study of six families with autosomal dominant DD, linkage analyses mapped five DD loci where one is known as the DYX6 locus. The rest of the loci are supported by other studies predominantly by linkage, association, or deletion mapping (Table 3). The locus at 13q12 in family E06 has been reported by Igo et al., 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 [36]. Luciano et al. and Truong et al. reported three SNPs, rs9508555, rs2892463 and rs7997649, in the linkage region associated with non-word repetition and DD [14, 37]. None of the three studies suggested candidate gene for the 13q12 locus and our WGS analyses failed to find variants in protein coding regions. Two candidate SNVs, one intergenic and one in an intron in *MTUS2* are the most promising candidates supported by expression of *MTUS2* in the brain with low regional specificity but mainly in neurons and synapses and in the heart (HPA).

Marker	Position (hg19)	Method	P value or LOD score Z	Genes proposed	Reference
Chromosome 13q12.3, family E06 (chr13:29,324,683-30,941,039)					
D13S1304/ ATA5A09	13:27-31,07 Mb	Linkage analyses	Z=2.94	-	[36]
rs2892463	13:30,347,835	Association	p=7.50E-06	-	[14]
rs9508555	13:30,372,036	Association	p=1.18E-05	-	[14]
rs7997649	13:30,494,296	Association	p=1.48E-06	-	[18]
Chromosome 15q23, family A29 (chr15: 70,362,585-73,666,730)					
CNV	15:72,154,000-72,325,595	CNV del (171,595bp)	p=0.0002	*MYO9A*	[38]
Chromosome 18p11.2 (DYX6), family A67 (chr18:10,905,079-11,901,888)					
D18S53	18:11,492,730-11,492,931	Linkage	p=1E-04	-	[39; 40]
rs7507114	18:13,539,693	Linkage	Z>2 (UK)		
Z=3,5 (USA)	PTPN2	[41]			
rs1846090	18:14,583,728	Linkage	Z=5.1	-	[42]
Chromosome 19p13.3, family B41 (chr19:1,364,306-2,827,300) and A09 (chr19: 366,412-6,755,007)					
-	19:1,118,914-1,837,061	Microdeletion		-	[43]
rs3786978 rs3786983 rs3760995	19:1,413,574 19:1,427,260 19:1,436,874	Association	p=7,49E-06 p=5,71E-06 p=2,69E-05	*DAZAP1*	[14]
CNV	19:556,985-602,852	deletion 45,868 bp	p=0.0115	*BSG, HCN2*	[38]
CNV	19:556,985-602,852	deletion 45,868 bp	p=0.0115	*BSG, HCN2*	[38]
Chromosome 21q22.3, family A22 (chr21:44,828,031-48,129,895)					
Deletion	21:44,376,016–48,080,911	Translocation, deletion	One person	-	[44]
rs73234886	21:46,130,548	Association	p=9E-06	*TSPEAR*	[16]
rs2255526	21:47,971,539	Association	p=2.3E-02	*DIPA2*	[45]

rs2839259-rs9982863	21:47,854,392-48,030,465	Deletion from rs2839259-qter	Small family	*PCNT, DIP2, S100B*	[46]
rs9722	21: 48,019,239	Association	p=0.016	*S100B*	[47]

Table 3: Reported Studies for Dyslexia in the Linkage Regions

The 15q23 locus in family A29 is supported by a CNV deletion from a study of more than 1300 DD cases [38]. The CNV deletion suggested *MYO9A* as a candidate, but the WGS analyses failed to identify variants in the gene. An intron variant in the *PKM* gene in a predicted promoter region and with CTCF-binding capacity is the most obvious regulatory candidate.

The locus 18p11.2 in family A67 mapped to in the DYX6 locus (OMIM 606616). Studies found linkages to D18S464 (p=0.00004) and D18S53 (p=0.0002), both close to the A67 linkage region, but a candidate gene is not suggested [39, 40]. The SNPs (rs7507114 and rs1846090) have been associated to DD with *PTPN2* as a candidate gene, but the WGS analysis of A67 failed to identify variants in or near the gene [41, 42]. Two variants in regulatory regions, one intergenic and one in an intron of *PIEZO2* were found. Support for *PIEZO2* is expression in the brain with low specificity and a function as part of an activated cation channel (Gene ID 63895). The locus telomeric 19p11.2 in the families B41 and A09 is reported in three different studies. A CNV deletion of 46kbp 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 [38]. identified several SNPs in the shared region for B41 and A09 and suggests *DAZAP1* as a DD candidate gene, but the WGS analyses in both families failed to find variants in or near the gene. Finally, a microdeletion has been reported in the shared region with dyslexia as part of the phenotype [14, 43]. Ten SNVs in family B41 and five in A09 all in regulatory regions fulfilled the filtration criteria. Three genes, *APC2, ABHD17A* and *BTBD2*, with candidate variants are shared by the two families (Table 2). All three genes are enriched expressed in the brain (HPA). *APC2* has been associated to intellectual developmental disorder (MRT74, OMIM 617169) and involved in cortical dysplasia (CDCBM10, OMIM 618677), none of the two genes have been associated with known disorders. Finally, the locus telomeric 21q22.3 in family A22 is reported in five other studies (Table 3). In two families with DD, translocation breakpoints have been reported with 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 [44, 46]. Other studies have shown association to DIP2A or S100B, and finally, a GWA study by Gialluisi et al., found association to rs73234886 in the TSPEAR gene and close to *KRTAP10-12* [45, 47]. The WGS analyses did failed to find any variants in regulatory region in any of these genes or deletions [16]. Intron variants are found in regulatory regions in the genes *PDXK* and *ADARB1*, both with low brain regional expression (Table 2).

Though the five mapped loci for DD found in this study are supported by other studies, none of the suggested candidate genes or deleted regions are supported by variants from the WGS analyses. The lack of protein coding mutations directed the analyses to suggest variants in regulator regions. Therefore, a catalog of variants with a MAF below 0.02 has been suggested and judged by nearby genes expressed in the brain. The inclusion criteria of a MAF value <0.02, was based on the assumption that a causative variant in single DD families must be relatively rare, might have led to exclusion of other variants involved in DD.

Few studies report the combination of linkage analyses of large families and the use of NGS of the mapped regions, but a strong candidate gene or regulatory region as the genetic cause has not been demonstrated [21, 22]. A large number of GWA studies done in the last decade suggest several genes involved in DD, but none of these were found in our study of the six families or in other reported family studies [16, 17, 20, 35]. Additional functional analyses are needed for variants affecting gene regulation, or if candidate genes in the future are suggested, to gain more and deep knowledge of the DD etiology. A better insight of a molecular mechanistic genetic model for DD is needed to analyze regulatory variants possible consequences for genes involved in DD. An approach suggested by Price et al., 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 [35]. Future genetic analyses, either family studies or GWAS studies, and a better understanding of the neuroanatomy of the language-related brain regions are needed to clarify the genetic components in DD and the overlapping comorbidities to other neurogenetic disorders.

Web Sources

SNP6-LINK package (https:www.rhododendron.dk/SNP6-LINK/)
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/)
Gene (https://www.ncbi.nlm.nih.gov/gene/)

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

References

1. Bishop, D. V., Laws, G., Adams, C., & Norbury, C. F. (2006). High heritability of speech and language impairments in 6-year-old twins demonstrated using parent and teacher report. Behavior Genetics, 36, 173-184.
2. Hawke, J. L., Wadsworth, S. J., & DeFries, J. C. (2006). Genetic influences on reading difficulties in boys and girls: the Colorado twin study. Dyslexia, 12(1), 21-29.
3. Aylward, E. H., Richards, T. L., Berninger, V. W., Nagy, W. E., Field, K. M., Grimme, A. C., ... & Cramer, S. C. (2003). Instructional treatment associated with changes in brain activation in children with dyslexia. Neurology, 61(2), 212-219.
4. Odegard, T. N., Ring, J., Smith, S., Biggan, J., & Black, J. (2008). Differentiating the neural response to intervention in children with developmental dyslexia. Annals of dyslexia, 58, 1-14.
5. Erbeli, F., Rice, M., & Paracchini, S. (2021). Insights into dyslexia genetics research from the last two decades. Brain Sciences, 12(1), 27.
6. Taipale, M., Kaminen, N., Nopola-Hemmi, J., Haltia, T., Myllyluoma, B., Lyytinen, H., ... & 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, 100(20), 11553-11558.
7. 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, 102(47), 17053-17058.
8. Cope, N., Harold, D., Hill, G., Moskvina, V., Stevenson, J., Holmans, P., ... & Williams, J. (2005). Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. The American Journal of Human Genetics, 76(4), 581-591.
9. 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 genetics, 1(4), e50.
10. Mascheretti, S., De Luca, A., Trezzi, V., Peruzzo, D., Nordio, A., Marino, C., & Arrigoni, F. (2017). Neurogenetics of developmental dyslexia: from genes to behavior through brain neuroimaging and cognitive and sensorial mechanisms. Translational psychiatry, 7(1), e987-e987.
11. Fisher, S. E., Francks, C., Marlow, A. J., MacPhie, I. L., Newbury, D. F., Cardon, L. R., ... & Monaco, A. P. (2002). Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. Nature genetics, 30(1), 86-91.
12. Svensson, I., Nilsson, S., Wahlström, J., Jernås, M., Carlsson, L. M., & Hjelmquist, E. (2011). Familial dyslexia in a large Swedish family: a whole genome linkage scan. Behavior genetics, 41, 43-49.
13. Meaburn, E. L., Harlaar, N., Craig, I. W., Schalkwyk, L. C., & Plomin, R. (2008). Quantitative trait locus association scan of early reading disability and ability using pooled DNA and 100K SNP microarrays in a sample of 5760 children. Molecular psychiatry, 13(7), 729-740.
14. Luciano, M., Evans, D. M., Hansell, N. K., Medland, S. E., Montgomery, G. W., Martin, N. G., ... & Bates, T. C. (2013). A genome-wide association study for reading and language abilities in two population cohorts. Genes, Brain and Behavior, 12(6), 645-652.
15. Gialluisi, A., Newbury, D. F., Wilcutt, E. G., Olson, R. K., DeFries, J. C., Brandler, W. M., ... & Fisher, S. E. (2014). Genome-wide screening for DNA variants associated with reading and language traits. Genes, Brain and Behavior, 13(7), 686-701.
16. Gialluisi, A., Andlauer, T. F., Mirza-Schreiber, N., Moll, K., Becker, J., Hoffmann, P., ... & Schulte-Körne, G. (2019). Genome-wide association scan identifies new variants associated with a cognitive predictor of dyslexia. Translational psychiatry, 9(1), 77.
17. Gialluisi, A., Andlauer, T. F., Mirza-Schreiber, N., Moll, K., Becker, J., Hoffmann, P., ... & Schulte-Koerne, G. (2021). Genome-wide association study reveals new insights into the

heritability and genetic correlates of developmental dyslexia. Molecular psychiatry, 26(7), 3004-3017.
18. Truong, D. T. (2019). Pediatric, Imaging, Neurocognition, and Genetics Consortium, Multivariate genome-wide association study of rapid automatised naming and rapid alternating stimulus in Hispanic American and African-American youth. J. Med. Genet, 56, 557-566.
19. Eising, E., Mirza-Schreiber, N., De Zeeuw, E. L., Wang, C. A., Truong, D. T., Allegrini, A. G., ... & 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.
20. Doust, C., Fontanillas, P., Eising, E., Gordon, S. D., Wang, Z., Alagöz, G., ... & Luciano, M. (2022). Discovery of 42 genome-wide significant loci associated with dyslexia. Nature genetics, 54(11), 1621-1629.
21. Carrion-Castillo, A., Estruch, S. B., Maassen, B., Franke, B., Francks, C., & Fisher, S. E. (2021). Whole-genome sequencing identifies functional noncoding variation in SEMA3C that cosegregates with dyslexia in a multigenerational family. Human genetics, 140(8), 1183-1200.
22. Grimm, T., Garshasbi, M., Puettmann, L., Chen, W., Ullmann, R., Müller-Myhsok, B., ... & Kuss, A. W. (2020). A novel locus and candidate gene for familial developmental dyslexia on chromosome 4q. Zeitschrift für Kinder-und Jugendpsychiatrie und Psychotherapie.
23. Eiberg, H., Nielsen, L. S., Klausen, J., Dahlén, M., Kristensen, M., Bisgaard, M. L., ... & Mohr, J. (1989). Linkage between serum cholinesterase 2 (*CHE2*) and γ-crystallin gene cluster (CRYG): assignment to chromosome 2. Clinical genetics, 35(5), 313-321.
24. Ott, J. (1976). A computer program for linkage analysis of general human pedigrees. American Journal of Human Genetics, 28(5), 528.
25. Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics, 25(14), 1754-1760.
26. DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., ... & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics, 43(5), 491-498.
27. Nassar, L. R., Barber, G. P., Benet-Pagès, A., Casper, J., Clawson, H., Diekhans, M., ... & Kent, W. J. (2023). The UCSC genome browser database: 2023 update. Nucleic acids research, 51(D1), D1188-D1195.
28. McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., ... & Cunningham, F. (2016). The ensembl variant effect predictor. Genome biology, 17(1), 1-14.
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, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., ... & Pontén, F. (2015). Tissue-based map of the human proteome. Science, 347(6220), 1260419.
31. 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.
32. Carrion-Castillo, A., Franke, B., & Fisher, S. E. (2013). Molecular genetics of dyslexia: an overview. Dyslexia, 19(4), 214-240.
33. Guidi, L. G., Velayos-Baeza, A., Martinez-Garay, I., Monaco, A. P., Paracchini, S., Bishop, D. V., & Molnár, Z. (2018). The neuronal migration hypothesis of dyslexia: A critical evaluation 30 years on. European Journal of Neuroscience, 48(10), 3212-3233.
34. Price, K. M., Wigg, K. G., Feng, Y., Blokland, K., Wilkinson, M., He, G., ... & Barr, C. L. (2020). Genome-wide association study of word reading: Overlap with risk genes for neurodevelopmental disorders. Genes, Brain and Behavior, 19(6), e12648.
35. Price, K. M., Wigg, K. G., Eising, E., Feng, Y., Blokland, K., Wilkinson, M., ... & Barr, C. L. (2022). Hypothesis-driven genome-wide association studies provide novel insights into genetics of reading disabilities. Translational psychiatry, 12(1), 495.
36. Igo Jr, R. P., Chapman, N. H., Berninger, V. W., Matsushita, M., Brkanac, Z., Rothstein, J. 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, 141(1), 15-27.
37. Truong, D. T., Shriberg, L. D., Smith, S. D., Chapman, K. L., Scheer-Cohen, A. R., DeMille, M. M. C., ... & Gruen, J. R. (2016). Multipoint genome-wide linkage scan for nonword repetition in a multigenerational family further supports chromosome 13q as a locus for verbal trait disorders. Human Genetics, 135, 1329-1341.
38. Chapman, J. (2011). Molecular genetics of developmental dyslexia. Cardiff University (United Kingdom).
39. Bates, T. C., Luciano, M., Castles, A., Coltheart, M., Wright, M. J., & Martin, N. G. (2007). Replication of reported linkages for dyslexia and spelling and suggestive evidence for novel regions on chromosomes 4 and 17. European Journal of Human Genetics, 15(2), 194-203.
40. Fisher, S. E., Francks, C., Marlow, A. J., MacPhie, I. L., Newbury, D. F., Cardon, L. R., ... & Monaco, A. P. (2002). Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. Nature genetics, 30(1), 86-91.
41. Scerri, T. S., Paracchini, S., Morris, A., MacPhie, I. L., Talcott, J., Stein, J., ... & Monaco, A. P. (2010). Identification of candidate genes for dyslexia susceptibility on chromosome 18. PloS one, 5(10), e13712.
42. Seshadri, S., DeStefano, A. L., Au, R., Massaro, J. M., Beiser, A. S., Kelly-Hayes, M., ... & Wolf, P. A. (2007). Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham Study. BMC medical genetics, 8(1), 1-14.
43. Palumbo, P., Palumbo, O., Leone, M. P., Stallone, R., Palladino,

T., Zelante, L., & Carella, M. (2016). Clinical and molecular characterization of a de novo 19p13. 3 microdeletion. Molecular Cytogenetics, 9, 1-7.
44. Coton, J., Labalme, A., Till, M., Bussy, G., Papoz, S. K., Lesca, G., ... & Rossi, M. (2018). Characterization of two familial cases presenting with a syndromic specific learning disorder and carrying (17q; 21q) unbalanced translocations. Clinical Case Reports, 6(5), 827.
45. Kong, R., Shao, S., Wang, J., Zhang, X., Guo, S., Zou, L., ... & Song, R. (2016). Genetic variant in *DIP2A* gene is associated with developmental dyslexia in Chinese population. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 171(2), 203-208.
46. Poelmans, G., Engelen, J. J. M., Van Lent-Albrechts, J., Smeets, H. J., Schoenmakers, E., Franke, B., ... & Schrander-Stumpel, C. T. R. M. (2009). Identification of novel dyslexia candidate genes through the analysis of a chromosomal deletion. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 150(1), 140-147.
47. 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(7), 399-401.