CrossMark

ORIGINAL INVESTIGATION

Enrichment of putatively damaging rare variants in the DYX2 locus and the reading-related genes *CCDC136* and *FLNC*

Andrew K. Adams[1] · Shelley D. Smith[2] · Dongnhu T. Truong[3] · Erik G. Willcutt[4,5] ·
Richard K. Olson[4,5] · John C. DeFries[4,5] · Bruce F. Pennington[6] · Jeffrey R. Gruen[1,7]

Abstract Eleven loci with prior evidence for association with reading and language phenotypes were sequenced in 96 unrelated subjects with significant impairment in reading performance drawn from the Colorado Learning Disability Research Center collection. Out of 148 total individual missense variants identified, the chromosome 7 genes *CCDC136* and *FLNC* contained 19. In addition, a region corresponding to the well-known DYX2 locus for RD contained 74 missense variants. Both allele sets were filtered for a minor allele frequency ≤0.01 and high Polyphen-2 scores. To determine if observations of these alleles are occurring more frequently in our cases than expected by chance in aggregate, counts from our sample were compared to the number of observations in the European subset of the 1000 Genomes Project using Fisher's exact test. Significant P values were achieved for both *CCDC136/FLNC* ($P = 0.0098$) and the DYX2 locus ($P = 0.012$). Taken together, this evidence further supports the influence of these regions on reading performance. These results also support the influence of rare variants in reading disability.

Introduction

Reading disability (RD; dyslexia) is characterized by difficulty in reading despite normal intelligence and adequate access to educational opportunities. With a prevalence ranging from 5 to 17%, RD is the most common learning disability (Shaywitz and Shaywitz 2005). It persists throughout life, and has significant detrimental effects on educational achievement and long-term socioeconomic status (Schatschneider and Torgesen 2004). Twin and family studies show a large genetic component with estimates of heritability ranging from 0.34 to 0.76 (DeFries et al. 1987; Hawke et al. 2006). Despite this high heritability, studies have identified a limited number of associated genetic variants of small effect size.

The most frequently replicated associations with RD have been with two genes, *DCDC2* and *KIAA0319*, encoded in the DYX2 locus on 6p22 (Fig. 1) (Cardon et al. 1994; Carrion-Castillo et al. 2013; Cope et al. 2005; Gayán et al. 1999; Grigorenko et al. 1997; Harold et al. 2006; Meng et al. 2005; Schumacher et al. 2006; Turic et al. 2003). In addition to RD, *DCDC2* has been associated with specific language impairment and different reading and language phenotypes in diverse populations and languages (Poelmans et al. 2011). *KIAA0319* has been associated with both RD and general language performance phenotypes (Scerri et al. 2011). *DYX1C1*, encoded in the DYX1 locus on 15q, was first associated with RD by chromosomal breakpoint mapping in one family, with support from an association study on 23 additional families, 33 mixed status couples, and 100

✉ Jeffrey R. Gruen
jeffrey.gruen@yale.edu

[1] Department of Genetics, Yale University, New Haven, CT, USA

[2] Munroe Meyer Institute, University of Nebraska Medical Center, Omaha, NE, USA

[3] Department of Pediatrics, Yale University, New Haven, CT, USA

[4] Institute for Behavioral Genetics, University of Colorado, Boulder, CO, USA

[5] Department of Psychology and Neuroscience, University of Colorado, Boulder, CO, USA

[6] Department of Psychology, University of Denver, Denver, CO, USA

[7] Department of Pediatrics and the Investigative Medicine Program, Yale University, New Haven, CT, USA

6p22.3

Fig. 1 Depiction of the *DCDC2/KIAA0319* region sequenced. Gene sizes are proportional per sizes listed in human genome build GRCh37/hg19. Intergenic distances are approximate, total region length approximately 1.04 Mbp

population controls (Dahdouh et al. 2009; Taipale et al. 2003). Evidence for replication of other genes is weaker. *PCDH9*, encoded on 13q21.32, falls within a linkage peak first associated with RD in a multipoint linkage analysis (Bartlett et al. 2002). *SEMA6D*, encoded on 15q21.1, was identified by chromosomal breakpoint mapping in one subject with RD (Ercan-Sencicek et al. 2012). *COMT*, encoded on 22q11.21, has been associated with several reading-related phenotypes (Grigorenko et al. 2007; Landi et al. 2013). *CCDC136*, *FLNC* (encoded on 7q32.1), and *RBFOX2* (encoded on 22q12.3) were associated with reading performance in a genome-wide association study (GWAS) of 1862 subjects selected from three family-based cohorts (Gialluisi et al. 2014).

RD frequently co-occurs with specific language impairment (SLI), defined as delayed onset of language acquisition, often including impairment in receptive vocabulary. 43% of children with SLI develop RD when they are later exposed to a reading curriculum in school (Snowling et al. 2000). The prevalence of SLI in kindergarten age children from the United States is approximately 7% (Tomblin et al. 1997). Like RD, the heritability for SLI is high, ranging from 0.36 to 0.97 in twin studies (Bishop and Hayiou-Thomas 2008). While there have been several linkage and association studies of SLI, evidence for two genes, *FOXP2* and *CNTNAP2*, are the most compelling. Through a series of chromosomal breakpoint mapping studies performed on a large, multigenerational family (called "KE"), *FOXP2* was initially found to be associated with verbal dyspraxia, a type of SLI (Fisher et al. 1998). *CNTNAP2* has been associated with nonsense word repetition in 184 families with a history of SLI. Nonsense word repetition is a quantitative assessment of phonological processing, which is frequently impaired in SLI (Bishop et al. 1996). Both *FOXP2* and *CNTNAP2* were associated with nonsense word repetition in a sample of 188 dyslexic families (Peter et al. 2011). Interestingly, FOXP2 is a transcription factor expressed in brain that binds and regulates expression of *CNTNAP2* (Vernes et al. 2008). The regulatory interaction between the FOXP2 protein and *CNTNAP2* supports the clinical findings that both genes are involved in similar reading and language phenotypes.

Attention deficit hyperactivity disorder (ADHD) also commonly co-occurs with RD. 80% of individuals with ADHD meet diagnostic criteria for at least one additional learning disorder; 40% of males with RD also have ADHD (Germano et al. 2010; Willcutt and Pennington 2000). While several genes have previously been associated with ADHD, *DBH* and *DRD2* have also been associated with reading and language (Smith et al. 2003). *DRD2* was associated with ADHD in a structural equation modeling study using 236 subjects, while later studies showed association with verbal language (Eicher et al. 2013; Rowe et al. 1999).

Previous association studies have focused on common variation, which accounts for only a small portion of the previously described heritability for RD. As costs for sequencing continue to fall, rare variants have recently become of interest, because they generally have larger effect sizes than non-coding SNPs, and they explain a larger portion of the observed heritability in complex traits. Projects such as the 1000 Genomes Project (The Genomes Project 2015), Exome Aggregation Consortium (ExAC) (Lek et al. 2016), and more recently the genome Aggregation Database (gnomAD; Lek et al. 2016) have catalogued a large portion of the genetic variation present in human populations at low frequency, protein altering, locations which is critical for determining whether they are enriched in any sample selected for a particular trait.

Resequencing studies have tended to show an enrichment of rare variants relative to controls in genes previously identified by GWAS. For example, in a 2010 study of hypertriglyceridemia (HTG), Johansen et al. demonstrated an enrichment of rare, putatively damaging mutations, with a 3-to-6-fold enrichment of missense and nonsense mutations in *GCKR*, *APOB*, *LPL*, and *APOA5* in 438 cases compared to 327 controls. These genes were initially identified by a GWAS of 463-affected subjects (and 1197 controls) predicated on a diagnosis of Fredrickson hyperlipoproteinemia as a categorical trait (Johansen et al. 2010).

To date, there have been two rare variant studies of RD. One study used imputed rare variants to study mismatch negativity in 200 children with RD, and found an association with four variants in DYX2 located midway between *DCDC2* and *KIAA0319* (Czamara et al. 2011). A second study of 376 subjects focused on copy number variants, but demonstrated no significant difference in the burden of CNVs between cases with RD and non-RD controls (Girirajan et al. 2011).

In this study, we seek to utilize high-throughput deep sequencing in 96 unrelated individuals with RD to investigate whether there is an enrichment in rare, putatively damaging variants in genes previously associated with reading or language performance by linkage and association analyses.

Methods

Subjects

Subjects were selected from the Colorado Learning Disability Research Center (CLDRC) nuclear family collections assessed for reading performance. The CLDRC maintains a collection of DNA samples from over 800 RD families phenotyped with a wide array of reading and language performance assessments (DeFries et al. 1997). For sequencing, 96 affected, unrelated subjects of European ancestry were selected based on a CLDRC-derived discriminant score indicating impairment in reading ability. Cases were defined as having a discriminant score of less than −0.60. Discriminant score coefficients were generated from performance on the spelling, reading comprehension, and the reading recognition subtests of the peabody individual achievement test. Discriminant function coefficients were then used to calculate a quantitative score used to maximally differentiate RD subjects from the general population (Wadsworth et al. 1989). Scores ranged from −0.60 to −4.65, with a mean of −2.04. Sixty-three subjects were male and 33 subjects were female, with ages ranging from 8.14 to 17.53 years at testing (mean 11.63). Fifteen included subjects were comorbid for ADHD.

All protocols were reviewed and approved by relevant Institutional Review Boards at the University of Colorado, University of Nebraska Medical Center, and Yale University. Subjects and families provided informed consent for use of both phenotype and genetic information at time of sample collection. In the absence of sequencing from non-RD controls, variant counts were compared to the European subset of the 1000 Genomes Project (1KGE) through the publicly available browser hosted by Ensembl (http://www.ensembl.org). 1KGE participants were apparently normal individuals, but were not specifically screened for reading or language phenotypes (The Genomes Project 2015).

Target selection and sequencing

Eleven genomic regions were targeted for a total of 6.4 Mb sequenced. Regions for sequencing were selected based on published genetic association and/or preliminary association with selected quantitative measures of reading performance on the full CLDRC sample (Table 1). Briefly, a targeted association analysis was performed in 1428 individuals from 337 families using 332 SNPs contained within 26 candidate genes plus 1 kbp on each side to include UTRs. Phenotypes for analysis included individual and composite measures of word reading, rapid automatized naming, orthographic coding, and comprehension. Seven genes contained one or more SNPs that survived correction for multiple testing with the False Discovery Rate method, with corrected $P <0.05$ for at least one phenotype (manuscript in preparation). Results from this analysis (including both uncorrected and corrected P values) were combined with literature support to obtain the final list of genes for sequencing. *CNTNAP2*, *COMT*,

Table 1 Loci sequenced for this study

Locus	Chr. Loc.	Sequenced coordinates	Size (Mb)	Rationale for selection
DCDC2/KIAA0319	6p22.3	24,124,528–25,166,900	1.04	Reading candidate[a] (Meng et al. 2005)
FOXP2	7q31.1	113,708,455–114,351,800	0.643	Reading candidate[a] (Peter et al. 2011)
CCDC136/FLNC	7q32.1	128,417,100–128,501,990	0.085	Reading candidate[a] (Gialluisi et al. 2014)
CNTNAP2	7q35-q36.1	145,792,593–148,183,542	2.4	Reading and language candidate[a,b] (Vernes et al. 2008)
DBH	9q34.2	136,498,622–136,524,722	0.026	Preliminary association with reading (data not shown)
DRD2	11q23.2	113,279,322–113,350,600	0.071	Language candidate[b] (Eicher et al. 2013)
PCDH9	13q21.32	66,843,132–67,813,032	0.97	Language candidate[b] (Bartlett et al. 2002)
SEMA6D	15q21.1	47,458,034–48,078,585	0.628	Language candidate[b] (Ercan-Sencicek et al. 2012)
DYX1C1	15q21.3	55,647,446–55,819,300	0.303	RD candidate[c] (Taipale et al. 2003)
COMT	22q11.21	36,129,602–36,432,802	0.309	Reading and language candidate[a,b] (Grigorenko et al. 2007)
RBFOX2	22q12.3	19,922,500–19,961,200	0.172	Reading candidate[a] (Gialluisi et al. 2014)

Each locus is named for the major associated gene(s) present. Columns are chromosomal location, sequenced coordinates in NCBIv37/hg19, approximate size of the locus in Mb, and rationale for selection. Genes selected based on literature support include corresponding citation

[a] Association with reading component phenotype or preliminary association evidence

[b] Association with a language phenotype or SLI

[c] Association with RD as a categorical trait

DBH, *DCDC2-KIAA0319*, *DRD2*, *DYX1C1*, and *FOXP2* had nominally significant results (uncorrected $P < 0.05$) for multiple SNPs in the CLDRC. Association with multiple SNPs from the *CCDC136/FLNC* locus and *RBFOX2* was described in a published GWAS that included 729 CLDRC subjects, including 51 of our affected subjects (Gialluisi et al. 2014). *PCDH9* and *SEMA6D* were included from extant publications (Bartlett et al. 2002; Ercan-Sencicek et al. 2012). Support was limited for other RD candidates including *C2ORF3* and *ROBO1* in the preliminary association analysis and was excluded for this reason. *CYP19A1* was excluded, because the previous analysis of multiple SNPs in this gene in the CLDRC cohort did not show association (Anthoni et al. 2012).

DNA sequencing was performed at the Yale Center for Genome Analysis (YCGA). Regions for sequencing were isolated and amplified with the Roche NimbleGen Seq-Cap EZ Target capture protocol. NimbleGen SeqCap is a system that allows for the isolation and amplification of a library corresponding to targeted sequence regions (Roche Sequencing, Pleasanton, CA). Libraries were sequenced on an Illumina Hi-Seq 2500 generating 75 bp, paired end reads (Illumina, San Diego, CA). Base calls and corresponding quality scores were generated by the YCGA.

Variant identification and analysis

Unaligned reads with corresponding quality scores were retrieved from the YCGA in FASTQ format. Downstream analysis of sequence data was performed on the Yale high-performance computing cluster in a Linux environment using an in-house sequence analysis pipeline modified for the analysis of targeted regions (ycga.yale.edu). Reads were first aligned to human reference genome build NCBI37 (hg19) using the Burrows–Wheeler Aligner memory efficient algorithm (bwa-mem) (Li 2013; Li et al. 2009). Aligned BAM files were then processed according to the Broad Institute's best practices for sequence data analysis. The Genome Analysis Toolkit (GATK) HaplotypeCaller identified variants from the reference genome (DePristo et al. 2011; McKenna et al. 2010). Because of the limited information provided by the size of the targeted regions, hard filtering instead of variant quality score recalibration was applied to variants (McKenna et al. 2010). A VCF file containing variants was downloaded from the Yale computing cluster and read with SNP and Variation Suite version 8 (Golden Helix, Inc., Bozeman, MT). Variants altering protein sequence (counts are reported in Table 2) were filtered for only those with minor allele frequencies (MAF) of less than 1%. Variant allele frequencies were determined using both the European subset of the 1KGE and the non-Finnish European subset of the gnomAD database (Tables 3, 4). Novel variants were also included in downstream analyses

Table 2 Observed missense variant counts in each sequenced locus

Locus	# of Genes	# of Variants
DCDC2/KIAA0319	9	74
FOXP2	1	3
CCDC136/FLNC	2	19
CNTNAP2	1	2
DBH	1	10
DRD2	1	7
PCDH9	1	7
SEMA6D	1	6
DYX1C1	2	10
COMT	2	8
RBFOX2	4	2
Total	25	148

Loci are named for the major associated gene(s) present. The number of genes in each individual locus is also reported

(Table 3). The remaining variants were assayed for putative damage to the corresponding proteins using SIFT (Kumar et al. 2009). Variants were further filtered based on Poly-phen-2 scores >0.7, indicating a high likelihood of significant damage to protein function (Adzhubei et al. 2010). The higher the Polyphen-2 score, the more likely the variant is damaging to the protein, with scores ranging from 0 to 1. Variant counts in RD case subjects were compared to counts of the same variants in the 1KGE using a two-tailed Fisher's exact test implemented in R (Team R Core 2016).

Linkage disequilibrium

Data for the *CCDC136/FLNC* region from the 503 European subjects from the 1KGE were downloaded from the 1000 Genomes Project data browser hosted at Ensembl. (http://www.ensembl.org). Data were loaded into Haploview and D' was calculated for all pairwise comparisons of markers (Barrett et al. 2005). Heatmaps were generated (Fig. 2), with LD blocks determined using the solid spine of LD method, filtering out variants with a minor allele frequency of less than 0.05.

Results

Variants

In total, 148 missense variants were identified in 96 RD cases. Within the 11 sequenced regions, two regions contained a disproportionate number of single nucleotide variants (SNVs) (Table 2). *CCDC136/FLNC* encoded on chromosome 7 contained 19 missense variants. The *DCDC2–KIAA0319* region (roughly contiguous with the

Table 3 Variants in *CCDC136* and *FLNC* with MAF <0.01 and PolyPhen-2 scores greater than 0.75

| Variant (11 total)	Chr	Position	Gene	Rs ID	AA change	SIFT	PolyPhen	CDS Obs	1KGE Obs	gnomAD
7:128441492-SNV	7	128441492	*CCDC136*	rs185493260	Arg/His	0	1	3	5	2.43×10^{-3}
7:128434498-SNV	7	128434498	*CCDC136*	rs143537881	Gly/Asp	0.01	0.923	2	6	1.03×10^{-2}
7:128449696-SNV	7	128449696	*CCDC136*	rs142055253	Lys/Glu	0	0.792	1	3	2.38×10^{-3}
7:128446877-Frameshift	7	128446877	*CCDC136*	Novel	N/A	N/A	N/A	1	0	N/A
7:128483880-SNV	7	128483880	*FLNC*	rs768103657	Gly/Arg	0	1	1	0	7.41×10^{-5}
7:128480184-SNV	7	128480184	*FLNC*	rs189525930	Gly/Arg	0	0.981	1	3	1.29×10^{-3}
7:128485240-SNV	7	128485240	*FLNC*	rs146953558	Arg/Cys	0.02	0.953	1	7	9.52×10^{-3}
7:128483367-SNV	7	128483367	*FLNC*	rs374983276	Arg/Cys	0	0.932	1	0	2.48×10^{-4}
7:128491324-SNV	7	128491324	*FLNC*	rs181067717	Arg/Cys	0.01	0.901	2	5	8.17×10^{-3}
7:128481578-SNV	7	128481578	*FLNC*	rs34972246	Asp/Ala	0	0.898	1	5	4.97×10^{-3}
7:128494547-SNV	7	128494547	*FLNC*	rs202223616	Glu/Gln	0.11	0.81	1	1	1.15×10^{-3}
							Variant	15	35	
							Reference	177	971	

Each variant is listed with chromosome, genomic location in hg19 coordinates, gene, rsID (if one has been assigned), SIFT score, PolyPhen-2 score, number of observations in our RD data set (CDS), number of observations in European subjects (*N* = 503) from the 1000 Genomes Project, and allele frequency in the gnomAD database. Amino acid changes are presented using the standard three-letter codes as reference/alternate. Variants without an rsID are novel to this study. Allele counts (variant and reference) are presented at the bottom of the table. Two-tailed Fisher's exact test *P* value = 0.0098

Table 4 Variants in *DCDC2/KIAA0319* locus with MAF <0.01 and PolyPhen-2 scores greater than 0.75

Variant (15 total)	Chr	Position	Gene	Rs ID	AA change	SIFT	PolyPhen	CDS Obs	1KGE Obs	gnomAD
6:24520644-SNV	6	24520644	*ALDH5A1*	rs149482918	Ala/Thr	0.03	0.986	1	0	7.89×10^{-6}
6:24520719-SNV	6	24520719	*ALDH5A1*	rs115784602	Val/Met	0.01	0.955	1	4	4.96×10^{-3}
6:24178856-SNV	6	24178856	*DCDC2*	rs766600237	Pro/Leu	0.03	0.874	1	0	3.17×10^{-5}
6:24174966-SNV	6	24174966	*DCDC2*	rs145154884	Val/Met	0	0.791	1	1	5.53×10^{-5}
6:24848281-SNV	6	24848281	*FAM65B*	rs375981215	Thr/Lys	0	1	1	0	4.73×10^{-5}
6:24495234-SNV	6	24495234	*GPLD1*	rs200793796	Cys/Gly	0.02	0.991	2	10	9.39×10^{-3}
6:24472814-SNV	6	24472814	*GPLD1*	rs62400471	Arg/Cys	0.02	0.912	1	1	6.39×10^{-4}
6:24476431-SNV	6	24476431	*GPLD1*	rs61754637	Asn/Ser	0.04	0.771	1	3	2.10×10^{-3}
6:24559241-SNV	6	24559241	*KIAA0319*	rs773691612	Gly/Arg	0.24	0.956	1	0	3.24×10^{-5}
6:24596526-SNV	6	24596526	*KIAA0319*	rs148566925	Pro/Thr	0.06	0.869	1	4	4.87×10^{-3}
6:24566953-SNV	6	24566953	*KIAA0319*	rs113411083	Arg/Trp	0.08	0.835	1	6	4.68×10^{-3}
6:24596675-SNV	6	24596675	*KIAA0319*	rs201302072	Arg/Leu	0.15	0.778	1	0	1.58×10^{-4}
6:24412567-SNV	6	24412567	*MRS2*	rs200245617	Val/Ile	0.12	0.973	1	0	4.39×10^{-4}
6:24667009-SNV	6	24667009	*TDP2*	rs146127613	Leu/Val	0.01	0.923	1	4	3.00×10^{-3}
6:24667074-SNV	6	24667074	*TDP2*	rs61760186	Cys/Tyr	0.35	0.781	1	5	4.47×10^{-3}
							Variant	16	38	
							Reference	176	968	

Each variant is listed with chromosome, genomic location in hg19 coordinates, gene, rsID (if one has been assigned), SIFT score, PolyPhen-2 score, number of observations in our RD data set (CDS), number of observations in European subjects (*N* = 503) from the 1000 Genomes Project, and allele frequency in the gnomAD database. Amino acid changes are presented using standard three-letter codes as reference/alternate. Variants and reference alleles are presented as counts out of number of observed alleles. Two-tailed Fisher's exact test *P* value = 0.0118

DYX2 locus) encoded on chromosome 6p22 contained 74 missense variants.

To restrict the analysis to rare SNVs, only those with MAF ≤0.01 were considered in downstream analyses. Ten of the eleven variants with MAF ≤0.01 in *CCDC136/ FLNC* (Table 3) were previously observed missense mutations causing putatively damaging changes to the relevant protein with Polyphen-2 scores ranging from 0.792 to 1. A Polyphen-2 score greater than 0.70 was generally considered damaging to proteins. Four variants were observed in *CCDC136*, including a one-base frameshift variant (7:128446877-Frameshift) not found in the 1KGE.

Fig. 2 Heatmap of LD across *CCDC136/FLNC*. Coloring is based on D' values with *darker gray* indicating high LD between marker pairs. Genes are aligned to markers on the heatmap, but are not accurately scaled. The major LD block joining *CCDC136* and *FLNC* is outlined in *black*
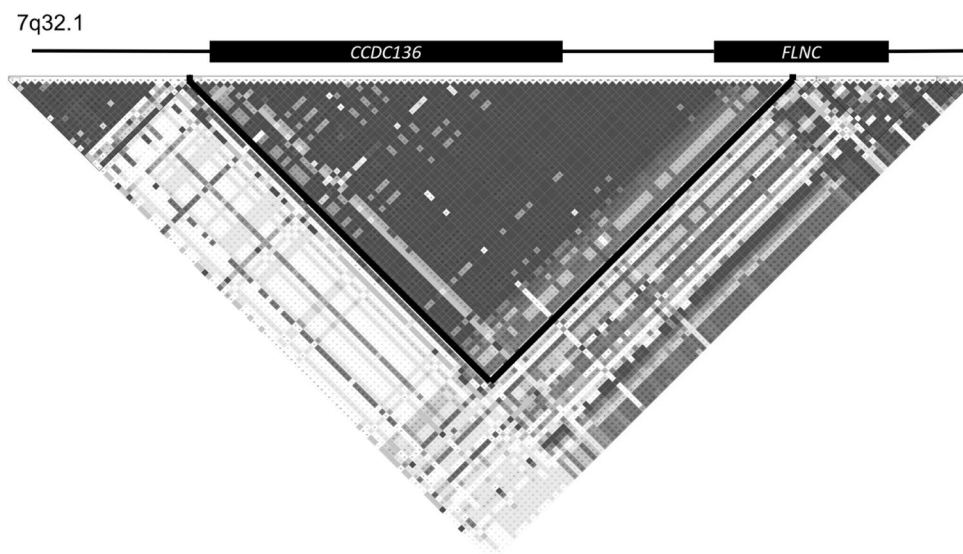
The remaining seven variants were found in *FLNC* with Polyphen-2 scores ranging from 0.81 to 1. Within DYX2, there were 16 observations of 15 variants with an MAF ≤0.01 and Polyphen-2 scores ranging from 0.771 to 1 (Table 4). Six of these 15 variants were not observed in the 1KGE, but were observed in the gnomAD database and are annotated in dbSNP (Table 4).

Enrichment analysis

Variant counts found in Tables 3 and 4 were used to test for enrichment of rare, damaging variants in cases relative to 1KGE. Within the *CCDC136/FLNC* region, 15 out of 192 chromosomes (7.8%) in RD cases contained one of the rare putatively damaging variants, compared to 35 out of 1006 chromosomes (3.5%) in the 1KGE, $P = 0.0098$ by two-tailed Fisher's exact test (Table 3). Within the DYX2 region, 16 out of 192 chromosomes (8.3%) in RD cases contained one of the rare putatively damaging variants, compared to 38 out of 1006 chromosomes (3.8%) in the 1KGE, $P = 0.012$ by two-tailed Fisher's exact test (Table 4).

Number of variants per subject ranged from 0 to 7, with no consistent correlation between discriminant score and number of variants. Eighteen subjects had no variants, 17 had one variant, 25 subjects had two variants, 14 subjects had three variants, 14 subjects had four variants, five subjects had five variants, two subjects had six variants, and one subject had seven variants. The worst performing subject had 0 variants that met all filtering criteria, and the subject with the most variants had a discriminant score of −2.41.

Discussion

In this study, 11 loci were sequenced in 96 unrelated RD subjects with significant impairment in reading performance. The loci were chosen, because they were previously associated with RD in the literature and/or in a preliminary association analysis (data not shown). In total, 148 missense variants were observed. Both the *CCDC136/FLNC* region and DYX2 locus showed a significant enrichment for rare (MAF ≤0.01), putatively damaging SNVs.

CCDC136/FLNC

A subset of 51 CLDRC RD subjects from this study was used in the published GWAS (total $N = 1862$) predicated on the first principal component from several quantitative reading and language exams (Gialluisi et al. 2014). This GWAS was the first to demonstrate association between *CCDC136/FLNC* and reading or language performance, with a weak, but replicated association between rs59197085 and both genes.

CCDC136 and *FLNC* are tandemly encoded in a genomic segment of just 48,000 bps that span four blocks of linkage disequilibrium (Fig. 2). The central block of linkage disequilibrium, however, spans the entirety of *CCDC136* and the 5′ half of *FLNC*, making it difficult to distinguish genetic association from one gene or the other. This justifies considering both genes as a single genetic locus in the previous GWAS as well as in the current study.

CCDC136 (also known as *NAG6*) is a coiled-coil domain containing protein described as a potential tumor suppressor in gastric cancer (Jiang et al. 2000). There is evidence for brain expression in both the GTEx and Brainspan databases.

In GTEx, human cortex has a median RPKM value (a normalized value indicating transcript abundance) of 14.250 based on an analysis of 114 human samples, while cerebellum has a median RPKM of 27.700 based on 125 human samples (Lonsdale et al. 2013). Brainspan demonstrates evidence for expression in humans, beginning at approximately 8 postconception weeks (pcw) and continuing throughout life (BrainSpan 2011). Data from the ExAC project suggest that *CCDC136* is fairly mutation tolerant with 314 observed missense variants compared to an expected 283.1, corresponding to a non-significant constraint metric of −0.90 (Lek et al. 2016).

FLNC (Filamin C) is a cytoskeletal component, responsible for crosslinking actin. Filamins are especially prominent in muscle, including cardiac muscle. Previous studies have implicated *FLNC* in a collection of myopathies and familial cardiomyopathies (Brodehl et al. 2016; Duff et al. 2011; Jiang et al. 2000; Shatunov et al. 2009; Thompson et al. 2000; Valdes-Mas et al. 2014; Vorgerd et al. 2005; Williams et al. 2005). Evidence for brain expression is limited, especially relative to *CCDC136*, with no expression observed in GTEx and only limited expression in Brainspan. Only a single sample in Brainspan demonstrates evidence for neuronal expression occurring at approximately 8–9 pcw, with expression tapering off to relatively low levels postnatally. ExAC missense variant tolerance information suggests that *FLNC* is not variation tolerant with an observed 865 missense variants compared to an expected 1191.1, for a corresponding constraint metric of 4.62.

Comparing the contrasting expression and functional evidence for both genes described above, it is unclear which gene may be playing the more important role in RD. Due to sample size and regional/timepoint specification, making conclusions solely based on Brainspan and/or GTEx is difficult. *FLNC* may be expressed at a non-assayed timepoint or brain location in development, and thus not observed in either database. The mutation tolerance data from ExAC suggest that variants in *FLNC* are actively being removed from the genome, though whether this is due to the known cardiac consequences of *FLNC* mutations or an unknown role in language processing is not clear. *CCDC136* shows evidence for brain expression but no significant evidence for mutation intolerance. Further study is required to disentangle the role of each gene on reading performance.

DCDC2/KIAA0319

DYX2 (contiguous with the sequenced *DCDC2/KIAA0319* region), the second genomic locus associated with RD, was first identified in a 1994 QTL study by Cardon et al. that utilized 27 of the same RD twinships/families in the current study (Cardon et al. 1994). Within DYX2, two genes, *DCDC2* and *KIAA0319*, have demonstrated consistent association with reading performance (as a quantitative trait) or RD (as a categorical trait). Knockout or knockdown studies of specific DYX2 genes in rodent models have demonstrated defects in auditory discrimination and neural spike timing, suggesting possible pathophysiologic mechanisms (Centanni et al. 2016; Che et al. 2014). RNAi studies in embryonic rats have shown that disruption of *Dcdc2* leads to neuronal migration defects (Burbridge et al. 2008). *DCDC2* localizes to cilia, and has been implicated in both autosomal (dominant or recessive) human kidney disease and autosomal recessive deafness (Massinen et al. 2011; Schueler et al. 2015). Typically, RD presents with a complex inheritance pattern indicative of a polygenic disorder. Also of note is that our group and others have shown that READ1 (Regulatory Element Associated with Dyslexia 1), a transcription regulatory element encoded in the second intron of *DCDC2*, is strongly associated with reading performance (Powers et al. 2013). Mass spectroscopy studies (SILAC) show that the transcription factor ETV6 binds READ1 sequence, while chromatin conformation capture (3C) experiments demonstrate close physical proximity between READ1 and the nearby *KIAA0319* promoter (KIAHap) in cultured human cells (Powers et al. 2016). *KIAA0319* encodes a membrane protein with a highly glycosylated extracellular domain suspected to play a role in signaling and neuronal adhesion (Velayos-Baeza et al. 2010). Evidence for human brain expression is strong in both GTEx and Brainspan. *KIAA0319* is expressed across all developmental timepoints in Brainspan, and has median RPKM values of 5.898 in the cortex ($N = 114$) and 5.996 in the cerebellum ($N = 125$). *DCDC2* has limited evidence for expression in Brainspan. Additional genes within the DYX2 locus (falling between *DCDC2* and *KIAA0319*) have shown varying degrees of association, though whether this is a true association or a consequence of linkage disequilibrium in the region is unknown.

It is probable that unrecognized RD is present in 1KGE subjects, which would inflate the observed variant counts among controls and diminish the significance in RD cases. Assuming that the 1KGE are representative of the general population (reading performance is a normally distributed trait), we would expect some of the 1KGE to be as impaired as our average case subject. Using controls screened for reading performance would likely strengthen the observed enrichments, but probably not by very much, because as noted above, we selected for less frequent extremes of the RD phenotype as cases. However, this may limit the generalizability of our results. For this study, utilizing subjects from a single ethnic group minimizes confounding due to population-level differences in the frequency of rare variants.

There is a limited, but growing, body of evidence in support of rare variants playing a role in RD. In a 2011 study, Czamara et al. demonstrated an association between late mismatch negativity (a neural trait shown to be affected in

RD cases) and four imputed rare variants (three with significant LD to one another) in the DYX2 region (Czamara et al. 2011). Further studies of rare forms of variation, including CNV studies, have been inconclusive, showing no significant difference between cases and controls in global CNV burden (Girirajan et al. 2011). Our results add to the evidence that rare variants are an important consideration in the genetic etiology of RD, and deserve further study in both family and population designs. Additional study may reveal that reading performance hinges on a complex mix of genetic variants ranging from rare pseudo-Mendelian variants having large effects (typically through disrupting proteins) to common non-coding variants affecting gene expression or cryptic non-coding RNAs.

Previous association studies have revealed few functional (protein altering) variants associated with complex traits. Most variants identified through GWAS occur outside of coding regions of genes and have no clear functional consequences. This has led to two main hypotheses over the meaning of GWAS results. The first is that GWAS-associated SNPs are found in undescribed or undefined regulatory regions, acting to subtly alter transcription or splicing of nearby associated genes. The second hypothesis, and the one investigated in this study, is that common variants typically used in GWAS are acting to tag a collection of rare variants in coding regions that are either not genotyped, or excluded from other analyses due to low minor allele frequencies in the general population. These rare variants change the protein-coding sequence and provide the "true" effect leading (at least in part) to the phenotype upon which the GWAS was predicated (Manolio et al. 2009). While no single variant can be associated with RD based on this study due to sample size limitations, there is evidence of a statistically significant increase in the frequency of specific observed mutations in RD cases.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and server for predicting damaging missense mutations. Nat Publ Group 7:248–249

Anthoni H, Sucheston LE, Lewis BA, Tapia-Paez I, Fan X, Zucchelli M, Taipale M, Stein CM, Hokkanen ME, Castren E, Pennington BF, Smith SD, Olson RK, Tomblin JB, Schulte-Korne G, Nothen M, Schumacher J, Muller-Myhsok B, Hoffmann P, Gilger JW, Hynd GW, Nopola-Hemmi J, Leppanen PH, Lyytinen H, Schoumans J, Nordenskjold M, Spencer J, Stanic D, Boon WC, Simpson E, Makela S, Gustafsson JA, Peyrard-Janvid M, Iyengar S, Kere J (2012) The aromatase gene CYP19A1: several genetic and functional lines of evidence supporting a role in reading, speech and language. Behav Genet 42:509–527. doi:10.1007/s10519-012-9532-3

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265. doi:10.1093/bioinformatics/bth457

Bartlett CW, Flax JF, Logue MW, Vieland VJ, Bassett AS, Tallal P, Brzustowicz LM (2002) A major susceptibility locus for specific language impairment is located on 13q21. Am J Hum Genet 71:45–55. doi:10.1086/341095

Bishop DVM, Hayiou-Thomas ME (2008) Heritability of specific language impairment depends on diagnostic criteria. Genes Brain Behav 7:365–372. doi:10.1111/j.1601-183X.2007.00360.x

Bishop DVM, North T, Donlan C (1996) Nonword repetition as a behavioural marker for inherited language impairment: evidence from a twin study. J Child Psychol Psychiatry 37:391–403. doi:10.1111/j.1469-7610.1996.tb01420.x

BrainSpan (2011) BrainSpan: Atlas of the Developing Human Brain

Brodehl A, Ferrier RA, Hamilton SJ, Greenway SC, Brundler MA, Yu W, Gibson WT, McKinnon ML, McGillivray B, Alvarez N, Giuffre M, Schwartzentruber J, Gerull B (2016) Mutations in FLNC are associated with familial restrictive cardiomyopathy. Hum Mutat 37:269–279. doi:10.1002/humu.22942

Burbridge TJ, Wang Y, Volz AJ, Peschansky VJ, Lisann L, Galaburda AM, Lo Turco JJ, Rosen GD (2008) Postnatal analysis of the effect of embryonic knockdown and overexpression of candidate dyslexia susceptibility gene homolog Dcdc2 in the rat. Neuroscience 152:723–733

Cardon LR, Smith SD, Fulker DW, Kimberling WJ (1994) Quantitative trait locus for reading disability on chromosome 6. Science 266(5183):276–279

Carrion-Castillo A, Franke B, Fisher SE (2013) Molecular genetics of dyslexia: an overview. Dyslexia 19:214–240. doi:10.1002/dys.1464

Centanni TM, Booker AB, Chen F, Sloan AM, Carraway RS, Rennaker RL, LoTurco JJ, Kilgard MP (2016) Knockdown of dyslexia-gene *Dcdc2* interferes with speech sound discrimination in continuous streams. J Neurosci 36:4895–4906. doi:10.1523/jneurosci.4202-15.2016

Che A, Girgenti MJ, LoTurco J (2014) The dyslexia-associated gene DCDC2 is required for spike-timing precision in mouse neocortex. Biol Psychiatry 76:387–396. doi:10.1016/j.biopsych.2013.08.018

Cope N, Harold D, Hill G, Moskvina V, Stevenson J, Holmans P, Owen MJ, O'Donovan MC, Williams J (2005) Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. Am J Hum Genet 76:581–591

Czamara D, Bruder J, Becker J, Bartling J, Hoffmann P, Ludwig KU, Muller-Myhsok B, Schulte-Korne G (2011) Association of a rare variant with mismatch negativity in a region between KIAA0319 and DCDC2 in dyslexia. Behav Genet 41:110–119. doi:10.1007/s10519-010-9413-6

Dahdouh F, Anthoni H, Tapia-Paez I, Peyrard-Janvid M, Schulte-Korne G, Warnke A, Remschmidt H, Ziegler A, Kere J, Muller-Myhsok B, Nothen MM, Schumacher J, Zucchelli M (2009) Further evidence for DYX1C1 as a susceptibility factor for dyslexia. Psychiatr Genet 19:59–63. doi:10.1097/YPG.0b013e32832080e1

DeFries JC, Fulker DW, LaBuda MC (1987) Evidence for a genetic aetiology in reading disability of twins. Nature 329:537–539. doi:10.1038/329537a0

DeFries J, Filipek P, Fulker D, Olson R, Pennington B, Smith S, Wise B (1997) Colorado learning disabilities research center. Learn Disabil 8:7–19

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498

Duff RM, Tay V, Hackman P, Ravenscroft G, McLean C, Kennedy P, Steinbach A, Schoffler W, van der Ven PF, Furst DO, Song J, Djinovic-Carugo K, Penttila S, Raheem O, Reardon K, Malandrini A, Gambelli S, Villanova M, Nowak KJ, Williams DR, Landers JE, Brown RH Jr, Udd B, Laing NG (2011) Mutations in the N-terminal actin-binding domain of filamin C cause a distal myopathy. Am J Hum Genet 88:729–740. doi:10.1016/j.ajhg.2011.04.021

Eicher JD, Powers NR, Cho K, Miller LL, Mueller KL, Ring SM, Tomblin JB, Gruen JR (2013) Associations of prenatal nicotine exposure and the dopamine related genes ANKK1 and DRD2 to verbal language. PLoS One 8:e63762. doi:10.1371/journal.pone.0063762

Ercan-Sencicek AG, Davis Wright NR, Sanders SJ, Oakman N, Valdes L, Bakkaloglu B, Doyle N, Yrigollen CM, Morgan TM, Grigorenko EL (2012) A balanced t(10;15) translocation in a male patient with developmental language disorder. Eur J Med Genet 55:128–131. doi:10.1016/j.ejmg.2011.12.005

Fisher SE, Vargha-Khadem F, Watkins KE, Monaco AP, Pembrey ME (1998) Localisation of a gene implicated in a severe speech and language disorder. Nat Genet 18:168–170

Gayán J, Smith SD, Cherny SS, Cardon LR, Fulker DW, Brower AM, Olson RK, Pennington BF, DeFries JC (1999) Quantitative-trait locus for specific language and reading deficits on chromosome 6p. Am J Hum Genet 64:157–164

Germano E, Gagliano A, Curatolo P (2010) Comorbidity of ADHD and dyslexia. Dev Neuropsychol 35:475–493. doi:10.1080/87565641.2010.494748

Gialluisi A, Newbury DF, Wilcutt EG, Olson RK, DeFries JC, Brandler WM, Pennington BF, Smith SD, Scerri TS, Simpson NH, Consortium TS, Luciano M, Evans DM, Bates TC, Stein JF, Talcott JB, Monaco AP, Paracchini S, Francks C, Fisher SE (2014) Genome-wide screening for DNA variants associated with reading and language traits. Genes Brain Behav 13:686–701

Girirajan S, Brkanac Z, Coe BP, Baker C, Vives L, Vu TH, Shafer N, Bernier R, Ferrero GB, Silengo M, Warren ST, Moreno CS, Fichera M, Romano C, Raskind WH, Eichler EE (2011) Relative burden of large CNVs on a range of neurodevelopmental phenotypes. PLoS Genet 7:e1002334. doi:10.1371/journal.pgen.1002334

Grigorenko EL, Wood FB, Meyer MS, Hart LA, Speed WC, Shuster A, Pauls DL (1997) Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. Am J Hum Genet 60:27–39

Grigorenko EL, Deyoung CG, Getchell M, Haeffel GJ, Klinteberg BA, Koposov RA, Oreland L, Pakstis AJ, Ruchkin VV, Yrigollen CM (2007) Exploring interactive effects of genes and environments in etiology of individual differences in reading comprehension. Dev Psychopathol 19:1089–1103. doi:10.1017/s0954579407000557

Harold D, Paracchini S, Scerri T, Dennis M, Cope N, Hill G, Moskvina V, Walter J, Richardson AJ, Owen MJ, Stein JF, Green ED, Azpos O, Donovan MC, Williams J, Monaco AP (2006) Further evidence that the KIAA0319 gene confers susceptibility to developmental dyslexia. Mol Psychiatry 11:1085–1091

Hawke JL, Wadsworth SJ, DeFries JC (2006) Genetic influences on reading difficulties in boys and girls: the Colorado twin study. Dyslexia 12:21–29

Jiang N, Zhan F, Tan G, Deng L, Zhou M, Cao L, Qiu Y, Xie Y, Li G (2000) A cDNA located on chromosome 7q32 shows loss of expression in epithelial cell line of nasopharyngeal carcinoma. Chin Med J (Engl) 113:650–653

Johansen CT, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, Martins RA, Kennedy BA, Hassell RG, Visser ME, Schwartz SM, Voight BF, Elosua R, Salomaa V, Apos O, Donnell CJ, Dallinga-Thie GM, Anand SS, Yusuf S, Huff MW, Kathiresan S, Hegele RA (2010) Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat Publ Group 42:684–687

Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073–1081

Landi N, Frost SJ, Mencl WE, Preston JL, Jacobsen LK, Lee M, Yrigollen C, Pugh KR, Grigorenko EL (2013) The COMT Val/Met polymorphism is associated with reading related skills and consistent patterns of functional neural activation. Dev Sci 16:13–23. doi:10.1111/j.1467-7687.2012.01180.x

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won H-H, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation Consortium (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285–291. doi:10.1038/nature19057 (http://www.nature.com/nature/journal/v536/n7616/abs/nature19057.html. Supplementary-information)

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. biorXiv 00: 3

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079

Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, Foster B, Moser M, Karasik E, Gillard B, Ramsey K, Sullivan S, Bridge J, Magazine H, Syron J, Fleming J, Siminoff L, Traino H, Mosavel M, Barker L, Jewell S, Rohrer D, Maxim D, Filkins D, Harbach P, Cortadillo E, Berghuis B, Turner L, Hudson E, Feenstra K, Sobin L, Robb J, Branton P, Korzeniewski G, Shive C, Tabor D, Qi L, Groch K, Nampally S, Buia S, Zimmerman A, Smith A, Burges R, Robinson K, Valentino K, Bradbury D, Cosentino M, Diaz-Mayoral N, Kennedy M, Engel T, Williams P, Erickson K, Ardlie K, Winckler W, Getz G, DeLuca D, MacArthur D, Kellis M, Thomson A, Young T, Gelfand E, Donovan M, Meng Y, Grant G, Mash D, Marcus Y, Basile M, Liu J, Zhu J, Tu Z, Cox NJ, Nicolae DL, Gamazon ER, Im HK, Konkashbaev A, Pritchard J, Stevens M, Flutre T, Wen

X, Dermitzakis ET, Lappalainen T, Guigo R, Monlong J, Sammeth M, Koller D, Battle A, Mostafavi S, McCarthy M, Rivas M, Maller J, Rusyn I, Nobel A, Wright F, Shabalin A, Feolo M, Sharopova N et al (2013) The genotype-tissue expression (GTEx) project. Nat Genet 45:580–585. doi:10.1038/ng.2653

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TFC, McCarroll SA, Visscher PM (2009) Finding the missing heritability of complex diseases. Nature 461:747–753

Massinen S, Hokkanen M-E, Matsson H, Tammimies K, Tapia-Páez I, Dahlström-Heuser V, Kuja-Panula J, Burghoorn J, Jeppsson KE, Swoboda P, Peyrard-Janvid M, Toftgård R, Castrén E, Kere J (2011) Increased expression of the dyslexia candidate gene *DCDC2* affects length and signaling of primary cilia in neurons. PLoS One 6:e20580

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303

Meng H, Smith SD, Hager K, Held M, Liu J, Olson RK, Pennington BF, DeFries JC, Gelernter J, O'Reilly-Pol T, Somlo S, Skudlarski P, Shaywitz SE, Shaywitz BA, Marchione K, Wang Y, Paramasivam M, LoTurco JJ, Page GP, Gruen JR (2005) DCDC2 is associated with reading disability and modulates neuronal development in the brain. Proc Natl Acad Sci USA 102:17053–17058. doi:10.1073/pnas.0508591102

Peter B, Raskind WH, Matsushita M, Lisowski M, Vu T, Berninger VW, Wijsman EM, 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. J Neurodev Disord 3:39–49. doi:10.1007/s11689-010-9065-0

Poelmans G, Buitelaar JK, Pauls DL, Franke B (2011) A theoretical molecular network for dyslexia: integrating available genetic findings. Mol Psychiatry 16:365–382. doi:10.1038/mp.2010.105

Powers NR, Eicher JD, Butter F, Kong Y, Miller LL, Ring SM, Mann M, Gruen JR (2013) Alleles of a polymorphic ETV6 binding site in DCDC2 confer risk of reading and language impairment. Am J Hum Genet 93:19–28. doi:10.1016/j.ajhg.2013.05.008

Powers NR, Eicher JD, Miller LL, Kong Y, Smith SD, Pennington BF, Willcutt EG, Olson RK, Ring SM, Gruen JR (2016) The regulatory element READ1 epistatically influences reading and language, with both deleterious and protective alleles. J Med Genet 53:163–171

Rowe DC, Van den Oord EJ, Stever C, Giedinghagen LN, Gard JM, Cleveland HH, Gilson M, Terris ST, Mohr JH, Sherman S, Abramowitz A, Waldman ID (1999) The DRD2 TaqI polymorphism and symptoms of attention deficit hyperactivity disorder. Mol Psychiatry 4:580–586

Scerri TS, Morris AP, Buckingham LL, Newbury DF, Miller LL, Monaco AP, Bishop DVM, Paracchini S (2011) DCDC2, KIAA0319 and CMIP are associated with reading-related traits. BPS 70:237–245

Schatschneider C, Torgesen JK (2004) Using our current understanding of dyslexia to support early identification and intervention. J Child Neurol 19:759–765

Schueler M, Braun DA, Chandrasekar G, Gee HY, Klasson TD, Halbritter J, Bieder A, Porath JD, Airik R, Zhou W, LoTurco JJ, Che A, Otto EA, Böckenhauer D, Sebire NJ, Honzik T, Harris PC, Koon SJ, Gunay-Aygun M, Saunier S, Zerres K, Bruechle NO, Drenth JPH, Pelletier L, Tapia-Páez I, Lifton RP, Giles RH, Kere

J, Hildebrandt F (2015) DCDC2 mutations cause a renal-hepatic ciliopathy by disrupting wnt signaling. Am J Hum Genet 96:81–92

Schumacher J, Anthoni H, Dahdouh F, König IR, Hillmer AM, Kluck N, Manthey M, Plume E, Warnke A, Remschmidt H, Hülsmann J, Cichon S, Lindgren CM, Propping P, Zucchelli M, Ziegler A, Peyrard-Janvid M, Schulte-Körne G, Nöthen MM, Kere J (2006) Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. Am J Hum Genet 78:52–62

Shatunov A, Olive M, Odgerel Z, Stadelmann-Nessler C, Irlbacher K, van Landeghem F, Bayarsaikhan M, Lee HS, Goudeau B, Chinnery PF, Straub V, Hilton-Jones D, Damian MS, Kaminska A, Vicart P, Bushby K, Dalakas MC, Sambuughin N, Ferrer I, Goebel HH, Goldfarb LG (2009) In-frame deletion in the seventh immunoglobulin-like repeat of filamin C in a family with myofibrillar myopathy. Eur J Hum Genet 17:656–663. doi:10.1038/ejhg.2008.226

Shaywitz SE, Shaywitz BA (2005) Dyslexia (specific reading disability). Biol Psychiat 57:1301–1309

Smith KM, Daly M, Fischer M, Yiannoutsos CT, Bauer L, Barkley R, Navia BA (2003) Association of the dopamine beta hydroxylase gene with attention deficit hyperactivity disorder: genetic analysis of the Milwaukee longitudinal study. Am J Med Genet B Neuropsychiatr Genet 119B:77–85. doi:10.1002/ajmg.b.20005

Snowling M, Bishop DVM, Stothard SE (2000) Is preschool language impairment a risk factor for dyslexia in adolescence? J Child Psychol Psychiatry 41:587–600. doi:10.1111/1469-7610.00651

Taipale M, Kaminen N, Nopola-Hemmi J, Haltia T, Myllyluoma B, Lyytinen H, Muller K, Kaaranen M, Lindsberg PJ, Hannula-Jouppi K, Kere J (2003) A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. Proc Natl Acad Sci USA 100:11553–11558. doi:10.1073/pnas.1833911100

Team R Core (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

The Genomes Project C (2015) A global reference for human genetic variation. Nature 526:68–74. doi:10.1038/nature15393 (http://www.nature.com/nature/journal/v526/n7571/abs/nature15393.html. Supplementary-information)

Thompson TG, Chan YM, Hack AA, Brosius M, Rajala M, Lidov HG, McNally EM, Watkins S, Kunkel LM (2000) Filamin 2 (FLN2): a muscle-specific sarcoglycan interacting protein. J Cell Biol 148:115–126

Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M (1997) Prevalence of specific language impairment in kindergarten children. J Speech Lang Hear Res 40:1245–1260

Turic D, Robinson L, Duke M, Morris DW, 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 MJ, Apos O, Donovan MC, Williams J (2003) Linkage disequilibrium mapping provides further evidence of a gene for reading disability on chromosome 6p21.3–22. Mol Psychiatry 8:176–185

Valdes-Mas R, Gutierrez-Fernandez A, Gomez J, Coto E, Astudillo A, Puente DA, Reguero JR, Alvarez V, Moris C, Leon D, Martin M, Puente XS, Lopez-Otin C (2014) Mutations in filamin C cause a new form of familial hypertrophic cardiomyopathy. Nat Commun 5:5326. doi:10.1038/ncomms6326

Velayos-Baeza A, Levecque C, Kobayashi K, Holloway ZG, Monaco AP (2010) The dyslexia-associated KIAA0319 protein undergoes proteolytic processing with {gamma}-secretase-independent intramembrane cleavage. J Biol Chem 285:40148–40162. doi:10.1074/jbc.M110.145961

Vernes SC, Newbury DF, Abrahams BS, Winchester L, Nicod J, Groszer M, Alarcón M, Oliver PL, Davies KE, Geschwind DH, Monaco AP, Fisher SE (2008) A functional genetic link between

distinct developmental language disorders. N Engl J Med 359:2337–2345. doi:10.1056/NEJMoa0802828

Vorgerd M, van der Ven PF, Bruchertseifer V, Lowe T, Kley RA, Schroder R, Lochmuller H, Himmel M, Koehler K, Furst DO, Huebner A (2005) A mutation in the dimerization domain of fil-amin c causes a novel type of autosomal dominant myofibrillar myopathy. Am J Hum Genet 77:297–304. doi:10.1086/431959

Wadsworth SJ, Gillis JJ, DeFries JC, Fulker DW (1989) Differential genetic aetiology of reading disability as a function of age. Ir J Psychol 10:509–520. doi:10.1080/03033910.1989.10557766

Willcutt EG, Pennington BF (2000) Psychiatric comorbidity in children and adolescents with reading disability. J Child Psychol Psychiatry 41:1039–1048

Williams DR, Reardon K, Roberts L, Dennet X, Duff R, Laing NG, Byrne E (2005) A new dominant distal myopathy affecting pos-terior leg and anterior upper limb muscles. Neurology 64:1245–1254. doi:10.1212/01.wnl.0000156524.95261.b9