Genome-wide association study of shared components of reading disability and language impairment

J. D. Eicher
Yale University

Et al.

Let us know how access to this document benefits you.
Follow this and additional works at: https://escholarship.umassmed.edu/iddrc_pubs

Part of the Behavioral Neurobiology Commons, Communication Sciences and Disorders Commons, Genomics Commons, Mental Disorders Commons, Psychiatry Commons, and the Psycholinguistics and Neurolinguistics Commons

Repository Citation

Creative Commons License
This work is licensed under a Creative Commons Attribution 3.0 License.
This material is brought to you by eScholarship@UMassChan. It has been accepted for inclusion in Intellectual and Developmental Disabilities Research Center Publications by an authorized administrator of eScholarship@UMassChan. For more information, please contact Lisa.Palmer@umassmed.edu.
Genome-wide association study of shared components of reading disability and language impairment


September 2013, accepted for publication 9 September 2013

The development of reading and verbal language skills through early childhood and into adolescence is vital to a child’s academic performance, self-perception of cognitive abilities, and development of sociability. Reading disability (RD) and language impairment (LI) are common and prevent affected individuals from developing adequate communication skills, leaving them at risk for adverse academic, socioeconomic, and psychiatric outcomes. Both RD and LI are complex traits that frequently co-occur, leading us to hypothesize that these disorders share genetic etiologies. To test this, we performed a genome-wide association study on individuals affected with both RD and LI in the Avon Longitudinal Study of Parents and Children. The strongest associations were seen with markers in ZNF385D (OR = 1.81, P = 5.45 × 10⁻⁷) and COL4A2 (OR = 1.71, P = 7.59 × 10⁻⁷). Markers with high NDS4 showed the strongest associations with LI individually (OR = 1.827, P = 1.40 × 10⁻⁷). We replicated association of ZNF385D using a 16 fiber tract to examine the implications of replicated markers. ZNF385D was a predictor of overall fiber tract volume in both hemispheres, as well as global brain volume. Here, we present evidence for ZNF385D as a candidate gene for RD and LI. The implication of transcription factor ZNF385D in RD and LI underscores the importance of transcriptional regulation in the development of higher order neurocognitive traits. Further study is necessary to discern target genes of ZNF385D and how it functions in neural development ontolinguistic language.

Key words: ALSPAC, dyslexia GWAS, language impairment, PING, reading disability, ZNF385D

Received 21 June 2013, revised 16 August 2013 and 6 September 2013, accepted for publication 9 September 2013

The development of reading and verbal language skills through early childhood and into adolescence is vital to a child's academic performance, self-perception of cognitive abilities, and development of sociability. Reading disability (RD) and language impairment (LI) are common and prevent affected individuals from developing adequate communication skills, leaving them at risk for adverse academic, socioeconomic, and psychiatric outcomes. Both RD and LI are complex traits that frequently co-occur, leading us to hypothesize that these disorders share genetic etiologies. To test this, we performed a genome-wide association study on individuals affected with both RD and LI in the Avon Longitudinal Study of Parents and Children. The strongest associations were seen with markers in ZNF385D (OR = 1.81, P = 5.45 × 10⁻⁷) and COL4A2 (OR = 1.71, P = 7.59 × 10⁻⁷). Markers with high NDS4 showed the strongest associations with LI individually (OR = 1.827, P = 1.40 × 10⁻⁷). We replicated association of ZNF385D using a 16 fiber tract to examine the implications of replicated markers. ZNF385D was a predictor of overall fiber tract volume in both hemispheres, as well as global brain volume. Here, we present evidence for ZNF385D as a candidate gene for RD and LI. The implication of transcription factor ZNF385D in RD and LI underscores the importance of transcriptional regulation in the development of higher order neurocognitive traits. Further study is necessary to discern target genes of ZNF385D and how it functions in neural development ontolinguistic language.

Key words: ALSPAC, dyslexia GWAS, language impairment, PING, reading disability, ZNF385D

Received 21 June 2013, revised 16 August 2013 and 6 September 2013, accepted for publication 9 September 2013

The development of reading and verbal language skills through early childhood and into adolescence is vital to a child's academic performance, self-perception of cognitive abilities, and development of sociability. Reading disability (RD) and language impairment (LI) are common and prevent affected individuals from developing adequate communication skills, leaving them at risk for adverse academic, socioeconomic, and psychiatric outcomes. Both RD and LI are complex traits that frequently co-occur, leading us to hypothesize that these disorders share genetic etiologies. To test this, we performed a genome-wide association study on individuals affected with both RD and LI in the Avon Longitudinal Study of Parents and Children. The strongest associations were seen with markers in ZNF385D (OR = 1.81, P = 5.45 × 10⁻⁷) and COL4A2 (OR = 1.71, P = 7.59 × 10⁻⁷). Markers with high NDS4 showed the strongest associations with LI individually (OR = 1.827, P = 1.40 × 10⁻⁷). We replicated association of ZNF385D using a 16 fiber tract to examine the implications of replicated markers. ZNF385D was a predictor of overall fiber tract volume in both hemispheres, as well as global brain volume. Here, we present evidence for ZNF385D as a candidate gene for RD and LI. The implication of transcription factor ZNF385D in RD and LI underscores the importance of transcriptional regulation in the development of higher order neurocognitive traits. Further study is necessary to discern target genes of ZNF385D and how it functions in neural development ontolinguistic language.
with LI are more likely to develop RD later in childhood (Pennington 2006). Additionally, children with RD and/or LI exhibit defects in any of the same neurocognitive domains, including phonological processing, com prehension, fluency and phonological short-term memory (Catts et al. 2005; Gathercole & Baddeley 1990; Pennington 2006; Pennington & Bishop 2009; Wise et al. 2007).

The relatedness between RD and LI goes deeper than similar clinical presentations. RD and LI share numerous risk factors and associated genes, as both are complex disorders with substantial genetic contributions (Pennington & Bishop 2009; Scofield & Schreiber-Kone 2010). Linkage, candidate gene association and genome-wide association studies have identified genes that contribute to RD and/or LI (Graham & Fisher 2013; Newbury et al. 2009, 2011; Phee et al. 2012; Rice et al. 2009; Sceleti et al. 2011). Some of these risk genes, including DCDC2, KIAA0319, FOXP2, CNTNAP2 and CM2, contribute to both RD and LI (Newbury et al. 2009, 2011; Phee et al. 2012; Rice et al. 2009; Sceleti et al. 2011; Wise et al. 2011). These studies suggest that RD and LI share certain risk genes that influence core language processes. However, genome-wide association studies (GWAS) on reading and language abilities have been limited. Recently, Lucano et al. (2013) reported a GWAS on quantitative performance on reading- and language-related measures. The strongest associations were seen between ABC2 and nonword repetition. These analyses identified novel loci and provided further performance on written and verbal language tasks, but do not address disorder status (i.e., RD or LI) nor the common etiology of RD and LI.

Neuropsychological aging studies of written and verbal language have identified various brain regions and measures in proportion for fluent language and alzheimer’s impaired individuals (Shaw et al. 2008; Vandemeulen et al. 2012). Some argue that these aging differences may represent a mediatory step between genetic risk variants and the ultimate clinical phenotype (Ether & Gruen 2013). Thus, recent studies have used these neuropsychological measures as endophenotypes in their analyses. These neuropsychological studies have associated RD and LI risk genes— including FOXP2, CNTNAP2, KIAA0319, DCDC2 and C2orf7—with various brain aging phenotypes—including brain activation patterns, white and grey matter volumes and fiber tract volumes (Cope et al. 2012; Dakki et al. 2012; Ether & Gruen 2013; Ligeon et al. 2003; Phee et al. 2012; Scott-Valdez et al. 2010; Sceleti et al. 2012; Tan et al. 2010; Wise et al. 2011).

The goal of this investigation is to identify novel genes that contribute to the overlap of RD and LI by performing a GWAS on subjects with both RD and LI in an extensively phenotyped birth cohort—the Avon Longitudinal Study of Parents and Children (ALSPAC). The large number of neurocognitive assessments in the ALSPAC allows for the simultaneous analysis of RD and LI. By doing so, we aim to identify new genes that contribute to both RD and LI. We then replicate our results in the Pediatric Language Neurocognition Genetics (PING) study using oral reading and age-appropriate vocabulary measures. For replication, we use a test for associations with fiber tract volumes previously implicated in language.

**Materials and Methods**

**Avon Longitudinal Study of Parents and Children**

Subject recruitment and collection of phenotype and genetic data for the ALSPAC cohort were completed by the ALSPAC team. The ALSPAC is a prospective population-based birth cohort based on the Avon region of the UK. It consists of a birth cohort of children of northern European descent, born in 1991 and 1992. Children were recruited before birth; recruitment of their parents resulted in a total of 15,458 fetuses, of whom 14,701 were alive at 1 year of age. Details regarding the participants, recruitment and study methodologies are described in detail elsewhere (http://www.bristol.ac.uk/alspac) (Boyd et al. 2012; Golding et al. 2001). The children of the ALSPAC have been extensively phenotyped from before birth to early adulthood. Ethical approval was obtained from the ALSPAC Ethics and Law Committee, Local UK Research Ethics Committees, and the Yale Human Investigation Committee.

**Reading and Language Measures**

The reading, language and cognitive measures used for this study were collected at ages 7, 8, and 9 years. Subjects defined as RD cases were based on the Wechsler Intelligence Scale for Children (WISC-III Total IQ, computed at age 8 years), were excluded from the presented analyses (Wise et al. 2012). Reading and language measures in the ALSPAC include a phoneme deletion task at age 7, single-word reading at ages 7 and 9 years, single nonword reading at age 9 years, and reading passage comprehension at age 9 years. The phoneme deletion task measures phoneme awareness, which is considered to be a core deficit in both RD and LI (Pennington 2006; Pennington & Bishop 2009). For the phoneme deletion task, also known as the Auditory Analysis Test, the child listens to a word spoken aloud, and then is asked to name a specific phoneme from that word to make a new word (Rosner & Sim 1971). Single-word reading was assessed at age 7 using the reading subtest of the Wechsler Objective Reading (W O R D). At age 9, single-word and nonword reading were assessed by asking the child to read 10 real words and 10 nonwords aloud from a subset of a larger list of words and nonwords taken from research conducted by Terezinha Nunes and colleagues (Rust et al. 1993). Reading comprehension scores were assessed at age 8, using the Neale Analysis of Reading Ability (NARA-II) (Neale 1997). Two additional language measures, nonword repetition and verbal comprehension tasks, were completed during clinical interviews at age 9 years. An adaptation of the Nonword Repetition Task (W R T), in which subjects repeated recordings of nonwords, was used to assess short-term phonological processing (Gathercole & Baddeley 1996). Children also completed the W echsler Objecti ve Language D isorder ( W O L D) test—a verbal comprehension task, where they answer word comprehension questions about a paragraph read aloud by an examiner (describing a presented picture). W echsler 1985. Z-scores were calculated for each subject on each individual measure.

**Case definitions**

We aimed to capture permanently poor performers in various reading and verbal language domains as RD and LI cases in our case definitions (Table 1). Therefore, we defined RD cases as having a z-score less than or equal to −1 on at least 3 of the 5 following tasks: single-word reading at age 7 years, phoneme deletion at age 7 years, single-word reading at age 9 years, and reading comprehension at age 9 years. These were 527 subjects defined as RD cases. We defined LI cases as having a z-score less than or equal to −1 on at least 2 of the 3 following tasks: phoneme deletion at age 7 years, verbal comprehension at age 9 years, and nonword repetition at age 9 years. These were 337 subjects defined as LI cases. As phoneme awareness is important in both RD and LI, we chose to include it as a part of the case definition for both RD and LI to reflect clinical presentation. These were 174 individuals affected with both RD and LI, with a male to female ratio of 1:2.1. In the further characterization of observed

**References**

Gatho et al. 2008; Vandermosen et al. 2012. comorbidity of RD and LI. not address disorder status (i.e., RD or LI) nor the common etiology of RD and LI.
associations, we created subsets of cases with no comorbidity. There were 163 LI cases excluding those with comorbid RD, and 353 RD cases excluding those with comorbid LI (Fig. 1). For all analyses, controls were defined as ALSPAC subjects of European ancestry who consented to all the necessary neuropsychological assessments but did not meet the criteria for case status.

### Genotyping and analysis

Subjects were genotyped on Illuma Human660 Quad bead arrays (San Diego, CA, USA). Subjects were excluded if the percentage of missing genotypes was greater than 2% (n = 6). To prevent possible population stratification, only subjects of European ancestry were included. In our primary analysis of RD and LI individually, there were 174 cases and 4117 controls. These analyses were completed to determine whether a single disorder RD or LI was driving association signals in the comorbid RD and LI analysis (Fig. 1). We also examined the associations of markers within several previously defined RD and/or LI risk genes, including those recently reported in Luciano et al. (2013), in order to present their results with these phenotypic definitions. These genes included: ABC13, ATP2C2, C0907916, CP M, CNTNAP2, DAZAP1, DCCD2, DYS1C1, FOXP2, KIAA0319, KRA0319L, PRKCN, ROBO1 and TDP2.

Gene-based analyses were performed on each phenotype (comorbid RD and LI, as well as RD and LI individually) using the VEGAS program, as described by the Luciano et al. study (Li et al. 2010; Luciano et al., 2013). To correct for multiple testing, we set a Bonferroni corrected threshold of $\alpha = 2.84 \times 10^{-6} = 0.05/17610$ genes tested.

### PING replication analyses

Replication analyses were completed in the PING study. Details on the inclusion, ascertainment, neuropsychological, genetic and neuromotor aging methods and data acquisition in the PING study are described in detail elsewhere, but are summarized briefly below (Alshooofi et al. 2010; Brown et al. 2012; Pfeil et al. 2012; Pfeil et al. 2013). The PING study is a longitudinal cohort of typically developing children between the ages of 3 and 20 years. Subjects were screened for history of ADHD, developmental, psychiatric, and neuromotor development problems, and all participants were included in the study. The human research protections program and institutional review boards at the 10 institutions (Well Cornell Medical College, University of California at Davis, University of Kentucky, Trinity Institute, Massachusetts General Hospital, University of California at Los Angeles, University of California at San Diego, University of Massachusetts Medical School, University of Southern California and Yale University) participating in the PING study approved all procedures and consent forms. For individuals under 18 years of age, parental informed consent and child assent (for those 7–17 years of age) were obtained. All participants, age 18 years and older, gave their written informed consent.

Subjects were genotyped on the Illumina Human660 Quad bead arrays (San Diego, CA, USA), with markers used for replication analyses passing quality control filters (sample call rate $\geq 95\%$, SNP call rate $\geq 95\%$, minor allele frequency $\geq 5\%$, W chromosome constructed reference panel as described elsewhere, Brown et al. 2012; Pfeil et al., 2013; W Alho et al. 2012). To assess ancestry and admixture proportions in the PING participants, we used a supervised clustering approach and were entered into the ADMIXTURE software (Alexander et al. 2009) and clustered participant data into six clusters corresponding to six major continental populations: African, Central Asian, East Asian, European, Native American and Oceanian. In the present analysis, we used ancestry proportions in the PING participants to determine ancestry and admixture proportions in the PING participants. These data were described elsewhere (Brown et al. 2012; Pfeil et al., 2012; W Alho et al. 2012). To prevent possible population stratification, only subjects with a European genetic ancestry factor (GAF) $\geq 0.6$ were included in genetic analysis of behavior. These 440 individuals of European ancestry (mean age of 11.5 years, 53.0% male) were analyzed using quantitative performance on the Oral Reading Recognition and Picture Vocabulary scores with $\alpha = 2.84 \times 10^{-6} = 0.05/17610$ genes tested.

Now reading comprehension age 9 years

RD cases had a z-score of less than -1 or equal to -1 on at least 3 of the 5 reading measures. LI cases had a z-score of less than -1 or equal to -1 on at least 2 of the 3 language measures.

**Table 1: Reading and language measures used to define RD and LI cases**

<table>
<thead>
<tr>
<th>Measure</th>
<th>RD (n = 527)</th>
<th>LI (n = 337)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phoneme deletion age 7 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-word reading age 7 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-word reading age 9 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Now reading comprehension age 9 years</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RD: Reading, LI: Language

**Figure 1:** Numbers of RD and LI cases in the ALSPAC cohort following the case definitions in Table 1. There were 174 subjects with comorbid RD and LI. There were 163 subjects with LI without comorbid RD, and 353 subjects with RD without comorbid LI.
PNG in aging analyses

PNG imaging techniques, data acquisition and analyses are discussed in detail elsewhere and briefly below (Brown et al. 2012; Pfeifer et al. 2012; W. Abhold et al. 2012). Across the 10 sites and 12 scanners, a standardized multiple modality high-resolution structural MRI protocol was in place, involving 3D T1- and T2-weighted volumes and a set of diffusion-weighted scans. At the University of California at San Diego, data were obtained on a GE 3 T Signa scanner and a 3 T Discovery 750x scanner (IR Healthcare, W. Sussex, U. S. A.) using eight-channel phased array head coils. The protocol included a conventional three-plane localizer, a sagittal 3D inversion recovery spoiled gradient echo T1-weighted volume, and form factor using grayscale/matter contrast echos (echo time = 3.5 ms; repetition time = 8.1 ms; inversion time = 640 ms; flip angle = 8°; receiver bandwidth = 31.25 kHz; FOV = 24 cm; frequency = 256; phase = 192; slice thickness = 1.2 mm), and two axial 2D diffusion tensor imaging DTI popular scans (b-value = 1000, TE = 83 ms; b-value = 13,000 ms; frequency = 96; phase = 96; slice thickness = 2.5 mm). Acquisition protocols with pulse sequence parameters identical or near identical to those protocols used at the University of California at San Diego were installed on scanners at the other nine sites. Data were acquired on all scanners to estimate motion rates and measures and correct for scanner-specific gradient coil nonlinear warping. In age files DICO format were processed with an automated processing stream written in MATLAB (Math, U. S. A.) and C++ by the UCSF Multi-Modal Imaging Laboratory. T1-weighted structural images were corrected for distortions caused by gradient nonlinearities, co-registered, averaged and rigidly resampled into an atlas with an atlas brain. In age preprocessing and analysis were performed using a fully automated set of tools available in the FreeSurfer software suite (http://surfer.nimr.mgh.harvard.edu) as well as an atlas-based method for the delineation and labeling of WM fiber tracts (Fischl 2012).

Diffusion tensor imaging

Diffusion weighted ages were corrected for eddy current distortion using a least-squares inverse and iterative conjugate gradient descent method to solve for the 12 scaling and translation parameters describing eddy current distortions across the entire diffusion MRI scan, explicitly taking into account the orientations and amplitudes of the diffusion gradient (Zhang et al. 2006). Head motion was corrected by registering each diffusion weighted image to a corresponding image synthesized from a tensor fit to the data (Rajger et al. 2009). Diffusion MRI data were corrected for eddy current and intensity distortions caused by B0 magnetic field inhomogeneities using the reversion gradient method (Olland et al. 2010). Distortions caused by gradient nonlinearities were corrected by applying a predefined, scanner-specific, nonlinear transformation atlas (Goorick et al. 2006). Diffusion weighted images were automatically aligned to a standard orientation relative to the T1-weighted images with isotropic 2 mm voxels. Cubic interpolation was used for all image processing steps. Conventional DTI methods were used to calculate diffusion measures (Basser et al. 1994; Pajevic et al. 1996). Scanning duration for the DTI sequence was 4.24 mn. While matter fiber tracts were biased using a probabilistic-atlas-based segmented method (Hagler et al. 2009). Volumes containing primarily gray matter or cerebrospinal fluid, identified using FreeSurfer’s automated atlas on brain segmentation, were excluded from analysis (Fischl et al. 2010). Fiber tract volumes were calculated as the sum of hemispheres with probability greater than 0.88, the value that provided optimal correspondence in volume between atlas-derived regions of interest and manually traced fiber tracts.

Statistical analyses

In aging-genetics analyses we performed an individual of European genetic ancestry. Scanner, age, handedness, socioeconomic status and sex were included as covariates in all analyses (Bavish et al. 2012; Pfeifer et al. 2012; W. Abhold et al. 2012). The 10 strongest GWAS associations with comorbid RD and LI in ALSpac were presented in Table 2. The strongest associations were observed with ZNF385D (OR = 1.81, P = 5.45 x 10^-7) and COL4A2 (OR = 1.71, P = 7.59 x 10^-7) (Table 2). Next, we examined RD and LI individually—with no comorbid cases included—determining whether one disorder was driving these associations. The 10 strongest associations for RD cases and LI cases individually are presented in Tables 3 and 4, respectively. The strongest associations with LI were with markers in NDT4 (OR = 1.83, P = 1.40 x 10^-7) (Table 3). A marker on chromosome 10 (OR = 1.43, P = 5.16 x 10^-6), chromosome 8 (OR = 1.70, P = 5.85 x 10^-6) and the OP3 gene (OR = 1.53, P = 6.92 x 10^-6) had the strongest associations with RD (Table 4). A marker with P < 0.01 that has previously been implicated in RD and/or LI is presented in Table S1. Supporting this, a variant related to the structural phenotype was with a marker seen for KRAI319 with comorbid RD and LI (rs16889556, P = 0.005177), PFOX2 with comorbid RD and LI (rs1530680, P = 0.001702), CNT-NA2 with RD (rs6915437, P = 0.0000642) and DCCD2 with RD (rs793834, P = 0.0002679) (Table S1a–S1c). Gene-based analyses were completed on each phenotype from comorbid RD and LI, and LI individually and LI individually, and the 10 strongest gene-based associations are presented in Table S2. None of the gene-based associations survived correction for multiple testing; however, the strongest association was seen with: GPRSH2, OR5H6 and RRAWA with comorbid RD and LI (OR = 1.97, P = 4.1E-05) and NAK2, DLRC1 and NARS with LI (OR = 1.97, P = 4.1E-05) with the OR5H2, OR5H6 and RRAWA genes previously implicated in RD. A marker with the strongest P-values in discovery analyses in ZNF385D, COL4A2 and NDT4 were carried forward for replication analysis in PNG. We observed replication of two markers with ZNF385D and perform analysis on the Picture Vocabulary Test (P = 0.00245 and 0.004173) (Table 5). However, no markers did not replicate with the Oral Reading Recognition Test (P > 0.05).

In aging genetics of ZNF385D

To follow-up on the replicated associations of ZNF385D, we examined the effects of these variants on fiber tract volumes previously implicated in written and verbal language. Before doing so, we determined that fiber tract volumes was a predictor of performance on the Oral Reading Recognition and Picture Vocabulary Tests (P = 2.24). In the subjects of only European genetic ancestry, ZNF385D genotypes were predictors of overall fiber tract volume as well as fiber tract volumes in the right and left hemispheres (Table 6). ZNF385D SNPs were also predictors bilaterally with the inferior longitudinal fasciculus (ILF), inferior fronto-occipital fasciculus (FO) and temporal superior longitudinal fasciculus (SLF) in this subset (Table 6). To discern...
whether these associations between \textit{ZNF385D} and fiber tract volumes reflect global brain volume differences among genotype, we next examined the relationship of \textit{ZNF385D} with both total brain segmentation and total cortical volumes. We found associations for both measures with \textit{rs1679255} ($P = 0.000072$ and $0.00027$, respectively) and \textit{rs12636438} ($P = 0.000259$ and $0.000069$, respectively). The effects appeared to be additive in nature, with heterozygous individuals having intermediate phenotypes relative to those homozygous for the major allele and to those homozygous for the minor allele. In fact, when these total brain volume measures were inserted into the model as covariates, \textit{ZNF385D} associations with DTI fiber tract volumes were no longer present.

\section*{Discussion}

In this investigation, we sought to identify genes that contribute to the co-occurrence of RD and LI in our discovery analyses, we found associations of \textit{ZNF385D} and \textit{COL4A2} in both \textit{RD} and \textit{LI} cases, and of \textit{NDST4} in \textit{LI} cases. Next, we observed associations of \textit{ZNF385D} with performance on a vocabulary measure, but not on an oral reading measure, in PING. Association with performance on a vocabulary measure, although not statistically significant, does provide further evidence for the contribution of \textit{ZNF385D} to language. To gain functional understanding, we interrogated the effects of replicated \textit{ZNF385D} markers on the volume of language-related fiber tracts. \textit{ZNF385D} markers associated bilaterally with overall fiber tract volumes and overall brain volume.

Studies have shown that RD and LI share genetic contributors (Trzaskowski et al. 2013). However, specific genes that contribute to both RD and LI have only recently begun to be examined. These studies have used a candidate gene approach to examine the shared genetic etiology. Such an approach has been successful in showing the shared contribution of \textit{DCDC2}, \textit{KRA0319}, \textit{POXP2}, \textit{CNTNAP2}, among others, to both RD and LI (Eicher & Gruen 2013; Graham & Fisher 2013; Newbury et al. 2009, 2011; Pinel et al. 2012; Rice et al. 2009; Scerri et al. 2011). In fact, markers within \textit{KRA0319}, \textit{POXP2}, and \textit{CNTNAP2} (along with \textit{BC0307918}) showed nominal association with comorbid RD and LI in our analyses ($P < 0.01$). RD/LI risk genes also showed a tendency to associate with LI individually (\textit{DCDC2}, \textit{KRA0319}, and \textit{CNTNAP2}) and with RD individually (\textit{CNTNAP2} and \textit{CMIP}) ($P < 0.01$). The lack of replication for other RD/LI risk genes and differences between this study and those of Scerri et al. (2011) and Luciano et al. (2013) are likely a result of different case definitions and sample sizes, as we designed our case classifications to capture a wide range of reading- and language-impaired subjects, as opposed to using highly specific neurocognitive measures.

A glaring omission in the genetic investigations of RD and LI is the lack of hypothesis-free methods. These methods
**Table 4: Associations with RD cases in ALSPAC, excluding comorbid LI cases (n= 353)**

<table>
<thead>
<tr>
<th>Marker</th>
<th>Chr</th>
<th>Base pair</th>
<th>M</th>
<th>MAF Aff</th>
<th>MAF Unaff</th>
<th>Gene</th>
<th>Odds ratio</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs180950</td>
<td>10</td>
<td>115697957</td>
<td>G</td>
<td>0.456</td>
<td>0.369</td>
<td>-</td>
<td>1.431</td>
<td>5.16 × 10−6</td>
</tr>
<tr>
<td>rs2590673</td>
<td>8</td>
<td>126037337</td>
<td>G</td>
<td>0.133</td>
<td>0.083</td>
<td>-</td>
<td>1.697</td>
<td>5.85 × 10−6</td>
</tr>
<tr>
<td>rs892100</td>
<td>19</td>
<td>50772522</td>
<td>C</td>
<td>0.228</td>
<td>0.162</td>
<td>PAFA</td>
<td>1.526</td>
<td>6.92 × 10−6</td>
</tr>
<tr>
<td>rs1792745</td>
<td>18</td>
<td>51955991</td>
<td>T</td>
<td>0.187</td>
<td>0.129</td>
<td>KRA5</td>
<td>1.558</td>
<td>1.22 × 10−5</td>
</tr>
<tr>
<td>rs12546767</td>
<td>8</td>
<td>126151747</td>
<td>C</td>
<td>0.152</td>
<td>0.099</td>
<td>-</td>
<td>1.618</td>
<td>1.32 × 10−5</td>
</tr>
<tr>
<td>rs12634033</td>
<td>3</td>
<td>146524529</td>
<td>C</td>
<td>0.135</td>
<td>0.087</td>
<td>-</td>
<td>1.646</td>
<td>1.80 × 10−5</td>
</tr>
<tr>
<td>rs892270</td>
<td>12</td>
<td>105002956</td>
<td>G</td>
<td>0.534</td>
<td>0.451</td>
<td>NUAK1</td>
<td>1.395</td>
<td>2.16 × 10−5</td>
</tr>
<tr>
<td>rs10887149</td>
<td>10</td>
<td>124156994</td>
<td>A</td>
<td>0.278</td>
<td>0.357</td>
<td>PLEKHA1</td>
<td>0.690</td>
<td>2.25 × 10−5</td>
</tr>
<tr>
<td>rs10041417</td>
<td>5</td>
<td>33218502</td>
<td>T</td>
<td>0.226</td>
<td>0.164</td>
<td>-</td>
<td>1.489</td>
<td>2.58 × 10−5</td>
</tr>
<tr>
<td>rs6792971</td>
<td>3</td>
<td>68468217</td>
<td>C</td>
<td>0.111</td>
<td>0.068</td>
<td>FAM19A1</td>
<td>1.703</td>
<td>2.59 × 10−5</td>
</tr>
</tbody>
</table>

Chr, chromosome; MAF Aff, minor allele frequency in affected subjects; MAF Unaff, minor allele frequency in unaffected subjects.

**Table 5: Replication of associations in PING (n= 440)**

<table>
<thead>
<tr>
<th>Marker</th>
<th>Chr</th>
<th>Base pair</th>
<th>M</th>
<th>Gene</th>
<th>Beta</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12636438</td>
<td>3</td>
<td>68468217</td>
<td>C</td>
<td>ZNF385D</td>
<td>-0.1867</td>
<td>0.9452</td>
</tr>
<tr>
<td>rs1679255</td>
<td>12</td>
<td>105002956</td>
<td>G</td>
<td>ZNF385D</td>
<td>-1.84</td>
<td>0.5016</td>
</tr>
<tr>
<td>rs9521789</td>
<td>5</td>
<td>33218502</td>
<td>T</td>
<td>COL4A2</td>
<td>0.3411</td>
<td>0.7332</td>
</tr>
<tr>
<td>rs476739</td>
<td>8</td>
<td>126037337</td>
<td>G</td>
<td>NDST4</td>
<td>0.5406</td>
<td>0.5891</td>
</tr>
<tr>
<td>rs505277</td>
<td>12</td>
<td>105002956</td>
<td>G</td>
<td>NDST4</td>
<td>0.5406</td>
<td>0.5891</td>
</tr>
<tr>
<td>rs7695228</td>
<td>3</td>
<td>68468217</td>
<td>C</td>
<td>NDST4</td>
<td>0.6258</td>
<td>0.5318</td>
</tr>
<tr>
<td>rs867036</td>
<td>3</td>
<td>68468217</td>
<td>G</td>
<td>NDST4</td>
<td>0.2605</td>
<td>0.7946</td>
</tr>
<tr>
<td>rs867035</td>
<td>3</td>
<td>68468217</td>
<td>G</td>
<td>NDST4</td>
<td>0.2961</td>
<td>0.7673</td>
</tr>
<tr>
<td>rs1940309</td>
<td>3</td>
<td>68468217</td>
<td>G</td>
<td>NDST4</td>
<td>0.6049</td>
<td>0.5456</td>
</tr>
</tbody>
</table>

MAF, minor allele frequency in full PING sample.

*P value less than FDR-adjusted statistical threshold (FDR-adjusted threshold = 0.05 × 2/19 = 0.00526).

**P value less than FDR-adjusted statistical threshold (FDR-adjusted threshold = 0.05 × 1/20 = 0.00250).
Figure 2: Association of total fiber tract volumes and neurocognitive tasks. Relationship of total DTI fiber tract volume with performance on (a) Picture Vocabulary Test and (b) Oral Reading Test. Total DTI fiber tract volumes were predictors of performance on both vocabulary \( P = 0.000602 \) and reading \( P = 0.03596 \) following correction for age, handedness, gender, scanner device used and socioeconomic status.

In an attempt to replicate these associations, particularly due to the known involvement of COL4A2 in pancephaly and white matter lesions (Verbeek et al. 2012; Yoneda et al. 2012), gene-based analyses did not reveal any associations that survived correction for multiple testing. Nonetheless, there were intriguing gene associations that should be investigated in future studies. For instance, with LI, there were suggestive associations with genes on chromosome 19—IL4I1, ATF5, NUP62 and STBL111—which may correspond to the SLI linkage peak (Monaco 2007; SLI Consortium 2002), Luciano et al. (2013) found a similar accumulation of suggestively associated genes approximately 5 Mba away from our genes. Additionally, MAP4, a microtubule assembly gene, was the strongest associated gene with RD. There is evidence that microtubule function plays a key role in reading development, as aberrant neuronal migration is thought to contribute to the etiology of RD and other RD candidate genes are thought to interact with microtubules (e.g., DCDC2 and ACOT13) (Cheng et al. 2006). Although intriguing, these suggestive findings must be validated in an independent cohort.

The strongest observed associations in this study were with ZNF385D. ZNF385D has previously been implicated in schizophrenia and attention deficit hyperactivity disorder (ADHD) (Poelmans et al. 2011; Xu et al. 2013). Both schizophrenia and ADHD are neuropsychiatric disorders thought to have come in pance ments in com on with RD and LI, including com prehension and sem antic processing (Gäger et al. 1992; Liet al. 2009; W. D. Smith et al. 2005). Additionally, our observed association of ZNF385D with global brain volume emay indicate that ZNF385D influences various neurocognitive traits through its effect on the entire brain.
Table 6: ZNF385D associations with DT fiber tract volumes in subjects with 100% European genetic ancestry (n = 332)

<table>
<thead>
<tr>
<th>Fiber tract</th>
<th>Slope</th>
<th>P value</th>
<th>Slope</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>-322.9</td>
<td>0.044*</td>
<td>-371.9</td>
<td>0.023*</td>
</tr>
<tr>
<td>Right All</td>
<td>-1731.4</td>
<td>0.039*</td>
<td>-1965</td>
<td>0.017*</td>
</tr>
<tr>
<td>Left All</td>
<td>-1616.3</td>
<td>0.055</td>
<td>-1775.6</td>
<td>0.033*</td>
</tr>
<tr>
<td>Right LF</td>
<td>-251.3</td>
<td>0.011*</td>
<td>-234.4</td>
<td>0.016*</td>
</tr>
<tr>
<td>Left LF</td>
<td>-256.9</td>
<td>0.0088**</td>
<td>-254.6</td>
<td>0.009**</td>
</tr>
<tr>
<td>Right FO</td>
<td>-200.8</td>
<td>0.032*</td>
<td>-190</td>
<td>0.041*</td>
</tr>
<tr>
<td>Left FO</td>
<td>-221</td>
<td>0.012*</td>
<td>-226.3</td>
<td>0.009**</td>
</tr>
<tr>
<td>Right SLF</td>
<td>-168.1</td>
<td>0.06</td>
<td>-206</td>
<td>0.02*</td>
</tr>
<tr>
<td>Left SLF</td>
<td>-195.5</td>
<td>0.022*</td>
<td>-212.9</td>
<td>0.013*</td>
</tr>
<tr>
<td>Right ISL</td>
<td>-170.8</td>
<td>0.011*</td>
<td>-180.7</td>
<td>0.0068**</td>
</tr>
<tr>
<td>Left ISL</td>
<td>-163.1</td>
<td>0.023*</td>
<td>-169.9</td>
<td>0.016*</td>
</tr>
<tr>
<td>Right pSLF</td>
<td>-151.7</td>
<td>0.037</td>
<td>-182.4</td>
<td>0.034*</td>
</tr>
<tr>
<td>Left pSLF</td>
<td>-112.2</td>
<td>0.18</td>
<td>-125.3</td>
<td>0.131</td>
</tr>
<tr>
<td>Right pSC</td>
<td>-148.5</td>
<td>0.052</td>
<td>-146.6</td>
<td>0.029*</td>
</tr>
<tr>
<td>Left pSC</td>
<td>-34.54</td>
<td>0.66</td>
<td>-54.3</td>
<td>0.48</td>
</tr>
<tr>
<td>CC</td>
<td>-977.1</td>
<td>0.15</td>
<td>-1187.1</td>
<td>0.081</td>
</tr>
</tbody>
</table>

All, all fiber tracts; CC, corpus callosum; pSLF, parietal superior longitudinal fasciculus; SLF, superior longitudinal fasciculus; pSC, posterior superior cerebellar cortex.

*P 0.05, **P 0.01.

There is little known regarding the function of ZNF385D, although its zinc finger domain suggests it is a transcriptional regulator. The in postance of transcriptional regulation in written and verbal language is not a new concept. The most widely studied language gene, FOXP2, is a potent transcription factor that has been shown to regulate another language gene, CNTNAP2 (Vernes et al. 2007; Vernes et al. 2011). Additionally, ZNF385D may have been harnessed as reading/language in ALSPAC and PING due to its transcriptional activity (Couto et al. 2010; Dennis et al. 2009; Meng et al. 2011) and possibly interact (Ludwig et al. 2008; Poon et al. 2013), although more evidence is needed to demystify these functions. ZNF385D variants now join this list of putative transcriptional variants that influence written and verbal language skills. The characterization of target genes of ZNF385D and its transcriptional effects on these targets will be an important next step. Additionally, the identification of target genes may generate therapeutic candidates for treatment and remediation of RD and LI. To gain further insight into ZNF385D, we performed both gene-gene and gene-environment analyses of ZNF385D and fiber tract volumes as language-related traits. ZNF385D appears to modulate fiber tract and total brain volume, which may subsequently affect the connectivity and functionality of brain regions involved in the efficient, fluent integration of written and verbal language. Thus, identification of target genes and how the modulation of their expression during neural development explains differences in fiber tract and total brain volume, which may be involved in the neurodevelopmental and behavioral consequences of ZNF385D, but also for the development of core language skills in children.

References


**Acknowledgment**

We thank all the families and participants who took part in these studies. We also wish to acknowledge the many who were responsible for their help in recruiting them, and the whole ALSFAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, and receptionists and nurses. The UK medical research council and the Wellcome Trust supported the ALSFAC study (National Institutes of Health Grant ref: R01 NS043530). The ALSFAC study was funded by the National Institute on Drug Abuse and the Eunice Kennedy Shriver National Institute of Child Health & Human Development. The authors gratefully thank the children, adolescents, adults, and parents who participated in the study. Data collection and sharing for this project were funded by the National Institutes of Health (National Institutes of Health Grant RC2DA029475). ALSFAC was supported by the National Instiutes of Health Grant: R01 NS043530 awarded to J.E. and J.E.R., and F31 DC012270 awarded to J.D.E.