ORIGINAL ARTICLE

The regulatory element READ1 epistatically influences reading and language, with both deleterious and protective alleles

Natalie R Powers,[1,2] John D Eicher,[1] Laura L Miller,[3] Yong Kong,[4,5] Shelley D Smith,[6] Bruce F Pennington,[7] Erik G Willcutt,[8,9] Richard K Olson,[8,9] Susan M Ring,[3,10] Jeffrey R Gruen[1,2,11]

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2015-103418).

For numbered affiliations see end of article.

Correspondence to
Dr Jeffrey R Gruen, Yale Child Health Research Center, Yale University School of Medicine, 464 Congress Avenue, New Haven, CT, 06520-8081, USA; jeffrey.gruen@yale.edu

Received 27 July 2015
Revised 22 September 2015
Accepted 14 October 2015
Published Online First 12 December 2015

To cite: Powers NR, Eicher JD, Miller LL, *et al*. *J Med Genet* 2016;**53**:163–171.

ABSTRACT

Background Reading disability (RD) and language impairment (LI) are heritable learning disabilities that obstruct acquisition and use of written and spoken language, respectively. We previously reported that two risk haplotypes, each in strong linkage disequilibrium (LD) with an allele of READ1, a polymorphic compound short tandem repeat within intron 2 of risk gene *DCDC2*, are associated with RD and LI. Additionally, we showed a non-additive genetic interaction between READ1 and KIAHap, a previously reported risk haplotype in risk gene *KIAA0319*, and that READ1 binds the transcriptional regulator ETV6.

Objective To examine the hypothesis that READ1 is a transcriptional regulator of *KIAA0319*.

Methods We characterised associations between READ1 alleles and RD and LI in a large European cohort, and also assessed interactions between READ1 and KIAHap and their effect on performance on measures of reading, language and IQ. We also used family-based data to characterise the genetic interaction, and chromatin conformation capture (3C) to investigate the possibility of a physical interaction between READ1 and KIAHap.

Results and conclusions READ1 and KIAHap show interdependence—READ1 risk alleles synergise with KIAHap, whereas READ1 protective alleles act epistatically to negate the effects of KIAHap. The family data suggest that these variants interact *in trans* genetically, while the 3C results show that a region of *DCDC2* containing READ1 interacts physically with the region upstream of *KIAA0319*. These data support a model in which READ1 regulates *KIAA0319* expression through KIAHap and in which the additive effects of READ1 and KIAHap alleles are responsible for the *trans* genetic interaction.

INTRODUCTION

Reading disability (RD) and language impairment (LI) are common, heritable learning disabilities, each involving a specific learning modality. RD, which is commonly known as dyslexia, is defined as an unexpected difficulty in processing written language in the presence of general cognitive ability that should be sufficient for proficient literacy.[1] LI is defined as an unexpected difficulty of the same type, but with verbal language instead of written.[2] The two disorders are closely related, involving many of the same underlying neurological processes and are frequently comorbid.[3 4] RD and LI are also highly heritable, but inheritance is complex.[3 5] Although the genetic component of both disorders has been extensively studied, few causal or functional variants have been identified. Because of the fundamental importance of language and literacy to education, affected children are often academically impeded relative to their unaffected peers, which can lead to a variety of adverse psychological, social and socioeconomic outcomes.[1 3] As RD and LI are both highly prevalent,[3 4] these adverse outcomes have an impact on society as a whole—through their cost to the health and educational systems as well as the lost potential of many affected people, whose difficulties with reading, language, or both, mask their talents and erode their confidence. RD and LI can be treated and although response to treatment varies widely, it is generally more effective at younger ages and when tailored to the individual.[2 6] A thorough understanding of the genetic components will permit better and earlier identification of individuals at risk for RD and LI and perhaps, eventually, for a priori matching of each individual to the intervention most likely to be effective.

Among the RD risk loci that have been discovered so far, the best-supported and most intriguing locus is DYX2 on chromosome 6p21.3. Several genes in this locus have been associated with RD, but two genes, *DCDC2* and *KIAA0319*, are by far the most replicated.[5] Because these genes reside within 200 kb of each other, it was previously unknown which gene was responsible for the linkage and association of DYX2 with RD. However, emerging evidence from human, animal and cellular studies suggests that both *DCDC2* and *KIAA0319* contribute to RD.[7–18] We recently showed that risk variants in both genes interact with each other in a non-additive manner to influence phenotype.[19] That study, which is summarised below and which we build upon in this study, further implicated both *DCDC2* and *KIAA0319* in reading, language and IQ and identified the source of at least some of the contribution to RD and LI risk from the DYX2 locus.

In our previous study, we used a haplotype-based strategy to scan SNPs densely covering the DYX2 locus for associations with RD and LI in the Avon

Longitudinal Study of Parents and Children (ALSPAC), a longitudinal birth cohort based in the former county of Avon, UK.[8 20] Using the extensive phenotypic and genetic data from approximately 5500 children of European descent in ALSPAC, we identified two six-marker risk haplotypes in the same haplotype block in *DCDC2*.[19] One of these haplotypes was associated strongly with severe RD, the other, with severe LI. Each of these risk haplotypes was in strong linkage disequilibrium (LD) with an allele of READ1 (regulatory element associated with dyslexia 1; GenBank accession No BV677278), a compound short tandem repeat in intron 2 of *DCDC2*. READ1 is a highly polymorphic, human-specific variant, with six common alleles and 34 rare alleles described so far. A naturally occurring, 2445 bp microdeletion encompassing READ1 also exists in human populations. READ1 alleles vary primarily by the number of each of five discrete repeat units and, consequently, also vary in length. Online supplementary table S1 lists all known READ1 alleles and gives details of their structures and allele frequencies in the ALSPAC.

Our previous study added to the existing literature examining the association of READ1 and the microdeletion with RD and other related endophenotypes.[9 15 21–27] Functionally, we hypothesised that READ1 serves a transcriptional regulatory role, as it specifically binds the transcriptional repressor ETV6[19] and can modulate the activity of the *DCDC2* promoter, as shown by a luciferase reporter experiment.[15] Because ETV6 must homodimerise to bind DNA,[28] and because of evidence showing that it is capable of homopolymerisation,[29] we speculated that allele structure and length—and therefore number of ETV6 binding sites—determines the regulatory power of a READ1 allele and its effect on phenotype. The two READ1 alleles in LD with the *DCDC2* risk haplotypes, alleles 5 and 6, both contain a GGAA insertion relative to the most common allele (see online supplementary table S1) and it is possible that this insertion creates a binding site for an additional ETV6 monomer.

In light of these observations, we questioned whether there might be a genetic interaction between the two *DCDC2* risk haplotypes and a known risk haplotype in *KIAA0319*, the other major RD risk gene in the DYX2 locus. The *KIAA0319* risk haplotype, which will be referred to hereafter as KIAHap for brevity, resides in a 3-marker, 77 kb haplotype block that spans approximately the 5′ half of *KIAA0319*, including its promoter, some of its upstream sequence and some of its neighbouring gene *TDP2*.[30] KIAHap and other haplotypes and individual markers in the same 77 kb interval, have been repeatedly associated with RD, subclinical reading performance and verbal IQ.[8 17 31–33] Interestingly, there is evidence that KIAHap influences *KIAA0319* expression.[16 34] We showed that individuals with at least one copy of a *DCDC2* risk haplotype *and* at least one copy of KIAHap, on average, performed worse than individuals with only one or the other (or neither), on reading, language and IQ measures.[19] These interaction effects were greater than would be expected if the risk variants acted additively and suggested to us a regulatory interaction between READ1 and *KIAA0319*.

In order to further characterise READ1 in relation to reading and language and to examine the effects of all READ1 alleles, we genotyped and analysed READ1 by Sanger sequencing in the entire ALSPAC cohort (we had previously only genotyped READ1 in individuals with the risk haplotypes). To investigate how READ1 and KIAHap are transmitted relative to each other, we also genotyped a family-based, European-ancestry cohort from the Colorado Learning Disabilities Research Center (CLDRC). In ALSPAC, the associations of alleles 5 and 6 with severe RD and LI mirrored the associations of their respective *DCDC2* risk haplotypes in our previous study,[19] alone and when grouped with rarer alleles of similar structure. Interestingly, another class of alleles emerged that appears both to protect against severe RD and to epistatically mask the deleterious effect of KIAHap on reading and IQ measures when present. By examining transmission of READ1 and KIAHap in the CLDRC family-based cohort, we provide circumstantial evidence that KIAHap and a given READ1 allele do not have to be *in cis* (on the same chromosome) to interact genetically. Finally, we provide evidence by chromatin conformation capture (3C) that READ1 and a region upstream of *KIAA0319* interact physically. The data reported here provide further support for the role of READ1 as a regulatory element and raise many fascinating questions about its mechanism of action.

METHODS

Subjects, recruitment and DNA collection

Subject recruitment and collection of phenotype data and DNA for the ALSPAC cohort was completed by the ALSPAC team, as described elsewhere.[20] A detailed description of the phenotypes and case–control criteria used in this study for ALSPAC is available in online supplementary tables S2A and S2B. The ALSPAC is a birth cohort based in the Avon region of the UK, consisting mainly of children of northern European descent, born in 1991 and 1992. Recruitment of pregnant mothers resulted in a total of 15 458 fetuses, of whom 14 701 were alive at 1 year of age. Details of the participants, recruitment and study methodologies are given in detail elsewhere.[20 35] Please note that the study website contains details of all the data that are available through a fully searchable data dictionary (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary).

The CLDRC cohort consists of 1201 European–American individuals in 293 nuclear families. Families were recruited to the study if at least one child had a history of reading problems.[9 36]

Phenotypes and exclusion criteria for this study are given in the online supplementary methods and in supplementary table S2.

Statistical analysis

Association analysis for this study was done using SNP and Variation Suite (SVS) V.8.1.0 (Golden Helix), using a standard regression-based association test under an allelic model. A Bonferroni correction was applied to correct for multiple testing—11 tests for each phenotype. Means, SDs were obtained and an analysis of variance was performed using SPSS Statistics (IBM).

Genotyping and 3C

Detailed methods for READ1 and SNP genotyping and the 3C experiment can be found in the online supplementary methods.

RESULTS

READ1 includes both deleterious and protective alleles for RD/LI

Upon completion of READ1 genotyping in the ALSPAC, we repeated the association analysis with severe RD and severe LI previously performed with the *DCDC2* risk haplotypes. A description of the case–control definitions is given in online supplementary table S2B; they are identical to those we used in our previous study.[19] For alleles 3, 4, 5, 6, 10 and the 2445 bp microdeletion encompassing READ1, all of which are relatively

common minor alleles in Europeans (minor allele frequency (MAF) >0.035; see online supplementary table S1), we examined association with individual alleles. We also combined these with some of the rare alleles into 'composite alleles', in which we grouped alleles based on structural or phylogenetic similarity. For example, related alleles clustered in the same clade in a phylogenetic tree we derived previously[19] from a ClustalW multiple alignment, under standard parameters. Online supplementary table S3 gives details of the constituents and rationales for the composite alleles. Since our previous study, the number of READ1 alleles observed has expanded from 22 to 40 (plus the microdeletion), most of which are rare.

Table 1 shows associations of READ1 with severe RD and LI. As expected, allele 5 is associated with severe RD, and allele 6 with severe LI. However, when alleles 5 and 6 are combined, the resulting composite allele is associated with both phenotypes. The same is true of 'clade 1,' which includes alleles 5 and 6 and rare alleles that cluster with them phylogenetically; and 'long alleles,' which include alleles >105 bp in length regardless of structure. By contrast, association results for another group of alleles suggest a protective effect for severe RD—they show only nominal association, but with ORs well below 1 (table 1). These alleles are denoted 'RU1-1' because they contain only one iteration of the 13 bp repeat unit 1 (RU1-1, the first of READ1's five repeat units), whereas most READ1 alleles contain two (see online supplementary table S3). This deletion makes RU1-1 alleles shorter than most other READ1 alleles and presumably removes some binding sites for ETV6. Allele 3 is the only common RU1-1 allele seen in Europeans (MAF=0.0456; see online supplementary table S1). The 'short alleles' group, which contains alleles <90 bp regardless of structure, also shows this effect, but as most of these short alleles are also RU1-1 alleles, the two categories are almost identical (see online supplementary table S3).

Deleterious READ1 alleles synergise with KIAHap, whereas protective READ1 alleles epistatically negate its effect

Our previous and present association results in the ALSPAC cohort prompted us to examine the effects of READ1 protective and deleterious allele classes on reading, language and IQ phenotypes in the presence and absence of KIAHap. We therefore compared mean performance on reading, language and IQ phenotypes, among individuals with different combinations of READ1 and KIAHap alleles. We performed this analysis with allele 3, allele 5, allele 6, the clade 1 alleles and the RU1-1 alleles, as these were the main classes of risk (alleles 5, 6, clade 1) and protective (allele 3, RU1-1) alleles (figure 1, table 1).

Consistent with the association results and our previous study,[19] allele 5 interacts synergistically with KIAHap for reading phenotypes, as well as total and verbal IQ (figure 1A). Likewise, allele 6 interacts synergistically with KIAHap for non-word repetition (NWR), a common endophenotype for LI (figure 1B). WOLD (Wechsler Objective Language Dimensions), another measure used to assess LI, shows a synergistic interaction between KIAHap and both alleles 5 and 6. When alleles 5 and 6 are combined with the other rare alleles that cluster together phylogenetically into clade 1, the magnitude of the interaction is somewhat attenuated—possibly owing to the tendency of alleles 5 and 6 to associate with different phenotypes in this cohort (figure 1C). However, a one-way analysis of variance shows that mean differences between groups for the clade 1 composite allele reach statistical significance more often than do those for allele 5 or allele 6 alone (see online supplementary table S4), probably owing to the higher number of carriers and the consequent increase in statistical power.

Conversely, the effect of KIAHap for every phenotype except NWR appears to be epistatically negated in the presence of allele 3. Individuals with at least one copy of both KIAHap and allele 3, on average, perform above the population mean on all measures except NWR (figure 1D). When allele 3 is combined with the other, rare RU1-1 alleles, this trend is recapitulated for most measures (figure 1E). These interactions suggest an interdependent relationship between at least some READ1 alleles and KIAHap, where the effect of each depends on the genotype of the other.

Transmission patterns suggest that the READ1/KIAHap genetic interaction does not occur *in cis*

Because READ1 and KIAHap reside close together on the chromosome, we questioned whether the genetic interaction

Table 1 Associations of single and composite READ1 alleles with severe RD and severe LI in the ALSPAC cohort

	Severe RD		Severe LI	
READ1 Allele	**OR (95% CI)**	**p Value**	**OR (95% CI)**	**p Value**
Allele 3	**0.47** (0.17 to 1.27)	0.0913	0.77 (0.48 to 1.23)	0.2554
Allele 4	1.24 (0.78 to 1.99)	0.3766	0.78 (0.56 to 1.09)	0.1407
Allele 5	**2.54** (1.48 to 4.36)	**0.0025926**	0.84 (0.50 to 1.40)	0.4880
Allele 6	1.54 (0.87 to 2.73)	0.1639	**1.65** (1.18 to 2.30)	0.005955*
Allele 10	0.79 (0.36 to 1.67)	0.5063	0.90 (0.59 to 1.36)	0.6034
Microdeletion	0.86 (0.48 to 1.51)	0.5810	0.85 (0.60 to 1.21)	0.3618
Alleles 5 and 6	**2.04** (1.36 to 3.08)	**0.0015725**	**1.66** (1.28 to 2.17)	**0.0003556**
Clade 1 (contains 5/6)	**1.99** (1.33 to 2.97)	**0.0020036**	**1.73** (1.34 to 2.23)	**0.00007402**
RU1-1 alleles	**0.41** (0.15 to 1.12)	0.0442*	0.80 (0.52 to 1.23)	0.2923
Short alleles	**0.41** (0.15 to 1.12)	0.0448*	0.80 (0.52 to 1.23)	0.2923
Long alleles	**2.39** (1.42 to 4.04)	**0.0033829**	**1.68** (1.17 to 2.43)	0.008962*

The association results for single and composite alleles of READ1 and the microdeletion. Values are regression-based under an allelic model. p Values that survived Bonferroni correction for multiple testing (threshold=0.05/11=0.0045) are shown in bold, with nominal associations marked with an asterisk. The highest and lowest ORs are also shown in bold. The criterion for severe RD is a score ≥2 SDs below the mean on the phoneme deletion task; the criterion for severe LI is a score of ≥2 SDs below the mean on at least one of two oral language measures (see online supplementary table S2B). For a description of the composite alleles, see online supplementary table S3, and for a detailed description of the phenotypes, see online supplementary table S2A, B.
ALSPAC, Avon Longitudinal Study of Parents and Children; LI, language impairment; RD, reading disability.

was necessarily *cis*—that is, does a deleterious READ1 allele have to be on the same chromosome as KIAHap to interact genetically with it? To examine this question, we genotyped READ1 and KIAHap in the family-based CLDRC cohort and analysed transmission patterns to determine (in Europeans) how often each of the common alleles occurred *in cis* with KIAHap and how often it occurred without KIAHap. Table 2 shows the results in132 informative families (families in which at least one parent has a copy of KIAHap). Even in families selected for the presence of KIAHap, all of the common alleles occur alone more often than they occur *in cis* with KIAHap. Allele 5 is the most extreme case; out of 31 instances of allele 5 and KIAHap

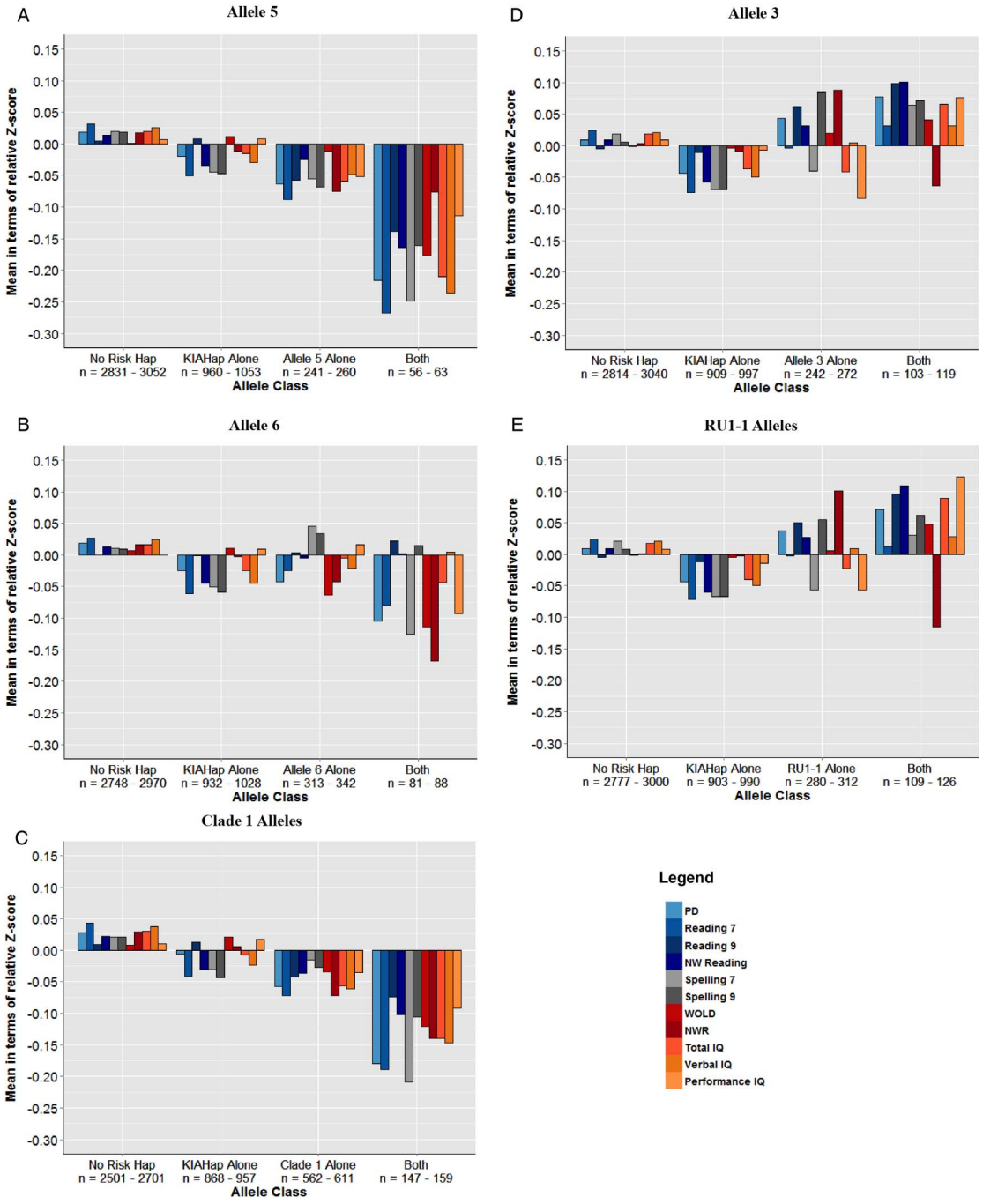


Figure 1 READ1–KIAHap interactions for single and composite alleles in the Avon Longitudinal Study of Parents and Children (ALSPAC). (A–E) These charts show the effect of the denoted READ1 single or composite allele on phenotype in the presence and absence of KIAHap. Each bar shows the z-score of the denoted allele class on the denoted measure, relative to the entire ALSPAC mean; units of the y-axis are fractions of a SD. Allele classes: KIAHap Alone, individuals positive for KIAHap but negative for the indicated READ1 allele; READ1 Allele Alone, individuals positive for the indicated READ1 allele but negative for KIAHap; No Risk Hap, individuals negative for both; Both, individuals positive for both. Phenotypes: PD, phoneme deletion task; Reading 7, single-word reading at age 7; Reading 9, single-word reading at age 9; NW Reading, non-word reading at age 9; Spelling 7, spelling at age 7; Spelling 9, spelling at age 9; WOLD, Wechsler Objective Language Dimensions verbal comprehension task; NWR, non-word repetition; Total, Verbal, and Performance IQ, Wechsler Intelligence Scale for Children (WISC-III). A description of these phenotypes is presented in online supplementary table S2A.

Table 2 Linkage disequilibrium between READ1 alleles and KIAHap in the CLDRC cohort

READ1 Allele	Cis with KIAHap	Trans with KIAHap	Freq. Cis	Freq. Trans
1	199	403	0.3306	0.6694
3	33	51	0.3929	0.6071
4	33	91	0.2661	0.7339
5	1	30	0.0323	0.9677
6	10	29	0.2564	0.7436
10	7	43	0.1400	0.8600
Del	44	64	0.4074	0.5926

Transmission data for 132 informative CLDRC families (KIAHap present in at least one parent) is shown. The table shows the number of instances of each common READ1 allele (and the microdeletion) that occurred on the same chromosome as KIAHap (*cis*) versus on the other chromosome (*trans*), in all individuals carrying both that allele and KIAHap. Frequencies of *cis* and *trans* are also shown. *Cis/trans* status was elucidated by pattern of transmission in each family.
CLDRC, Colorado Learning Disabilities Research Center.

occurring together, we only observed one instance of the two occurring *in cis*. However, among the single deleterious READ1 alleles, allele 5 shows the strongest synergistic effect with KIAHap for reading and IQ phenotypes (figure 1). This indicates, albeit circumstantially, that READ1 and KIAHap do not need to be *in cis* to interact genetically.

The presence of READ1 increases intrachromosomal interactions between *DCDC2* intron 2 and the *KIAA0319* upstream region

The observations that READ1 binds a transcription factor, that KIAHap spans the promoter region of *KIAA0319* and that they appear to exhibit interdependence on each other to affect phenotype, led us to inquire whether READ1 might have a direct regulatory interaction with *KIAA0319*. To examine this question, we used 3C to determine whether READ1 and *KIAA0319* interact physically. 3C covalently crosslinks DNA and any bound proteins in their native conformation. The fixed chromatin is then fragmented, diluted and treated with DNA ligase to join fragments that are proximal to each other. If two loci interact through a transcription factor or protein complex, they would be expected to generate fusion fragments more often than would be seen by chance. Relative amounts of fusion fragments are detected by qPCR with primers designed to amplify across ligation junctions.

Figure 2A depicts our approach graphically. To assess physical interactions in the presence and absence of READ1, we chose to study two lymphoblastoid cell lines—GM17831, which is homozygous for the 2445 bp microdeletion encompassing READ1; and Raji, which is homozygous intact for this 2445 bp region. Raji cells also contain a READ1 risk allele; the READ1 genotype of Raji cells is 4,5. We chose HindIII as the restriction enzyme because it generates a restriction fragment containing the entire 2445 bp microdeletion interval; the flanking HindIII sites are therefore still present in a cell line homozygous for the microdeletion. HindIII also generates three restriction fragments in and around the *KIAA0319* promoter (figure 2A). We used two anchor primers for this experiment: one flanking the HindIII site on the READ1 restriction fragment, the other flanking the HindIII site on a restriction fragment near the *NRSN1* promoter, outside any loop that would occur between READ1 and *KIAA0319*, as a control. Prey primers flank the three HindIII fragments near the *KIAA0319* transcription start site (KIAJ1, KIAJ2, KIAJ3), the region upstream of *DCDC2* (DCDC2), the region upstream of both *GPLD1* and *ALDH5A1* and the *KIAA0319* 3′ untranslated region (KIA3′) (figure 2A).

Figure 2B shows the combined results of two six-replicate qPCR experiments (12 experiments in all). We first calculated fusion fragment enrichment in Raji over GM17831, corrected for digestion efficiency and normalised to a control amplicon (ACTβ) that does not contain a HindIII site. For each prey primer, we then compared these values between the READ1 and NRSN1 anchor primers. If READ1 does not interact specifically with a given region of DYX2, there should be no difference in Raji/GM17831 enrichment between the anchor primers for the corresponding prey primer. That is, the presence (Raji) or absence (GM17831) of READ1 should not make a difference if it does not physically interact with that region of the locus. As shown in the figure, there is no difference between the READ1 and NRSN1 anchor primers upstream of the *DCDC2* promoter, or at KIAJ1 or KIAJ2, but a small difference is apparent in the region between the *DCDC2* and KIAJ1 fragments, which disappears at KIAJ1 and KIAJ2, then reappears much more strongly at KIAJ3. Interestingly, the KIAJ3 fragment contains rs9461045, a SNP previously proposed to be a functional variant in KIAHap and shown to cause an allelic reduction in *KIAA0319* expression in several cell lines.[16] These results suggest that READ1 increases the probability of this interaction when it is present in comparison with when it is absent and probably indicates a direct regulatory interaction between READ1 and the *KIAA0319* gene. They may also suggest that READ1 interacts with (and may regulate) other genes in the locus, including *GPLD1* and *ALDH5A1*, albeit much less strongly.

DISCUSSION

In previous work, we provided strong evidence that READ1 is a transcriptional regulatory element that interacts non-additively with KIAHap, a risk haplotype spanning the 5′ half of *KIAA0319*.[19] This evidence, though compelling, was indirect; the variants associating with RD and LI and interacting with KIAHap were not alleles of READ1 itself, but two six-SNP haplotypes in strong LD with two alleles of READ1. In this further study, we were able to examine the effects of all READ1 alleles in the large, ethnically homogeneous ALSPAC cohort. The results suggest at least two classes of READ1 alleles in European populations: deleterious and protective. The association results show this for severe RD, while it is suggested by genetic interaction analysis for most of the reading, language and IQ phenotypes considered in this study. This indicates that READ1 is a functional variant in the region and provides insight into its mechanism of action. First and foremost, whether an allele is deleterious or protective seems to depend on its length and/or structure; longer alleles with insertions in repeat unit 2 tend to be deleterious, while shorter alleles with a deletion of one copy of repeat unit 1 tend to be protective. As repeat unit 1 was the major *in vitro* ETV6 binding probe in our previous electrophoretic mobility shift assay (EMSA) and stable isotope labelling by amino acids in cell culture (SILAC) experiments,[19] this is consistent with our model that indels in READ1 change the size of the ETV6 homopolymer that can bind and thus alter the regulatory power of the allele.

Interestingly, the genetic interaction between READ1 and KIAHap is different for different classes of READ1 alleles. Deleterious READ1 alleles synergise with KIAHap to reduce performance on reading, language and IQ measures more than would be expected if these READ1 alleles and KIAHap acted additively. By contrast, protective READ1 alleles epistatically

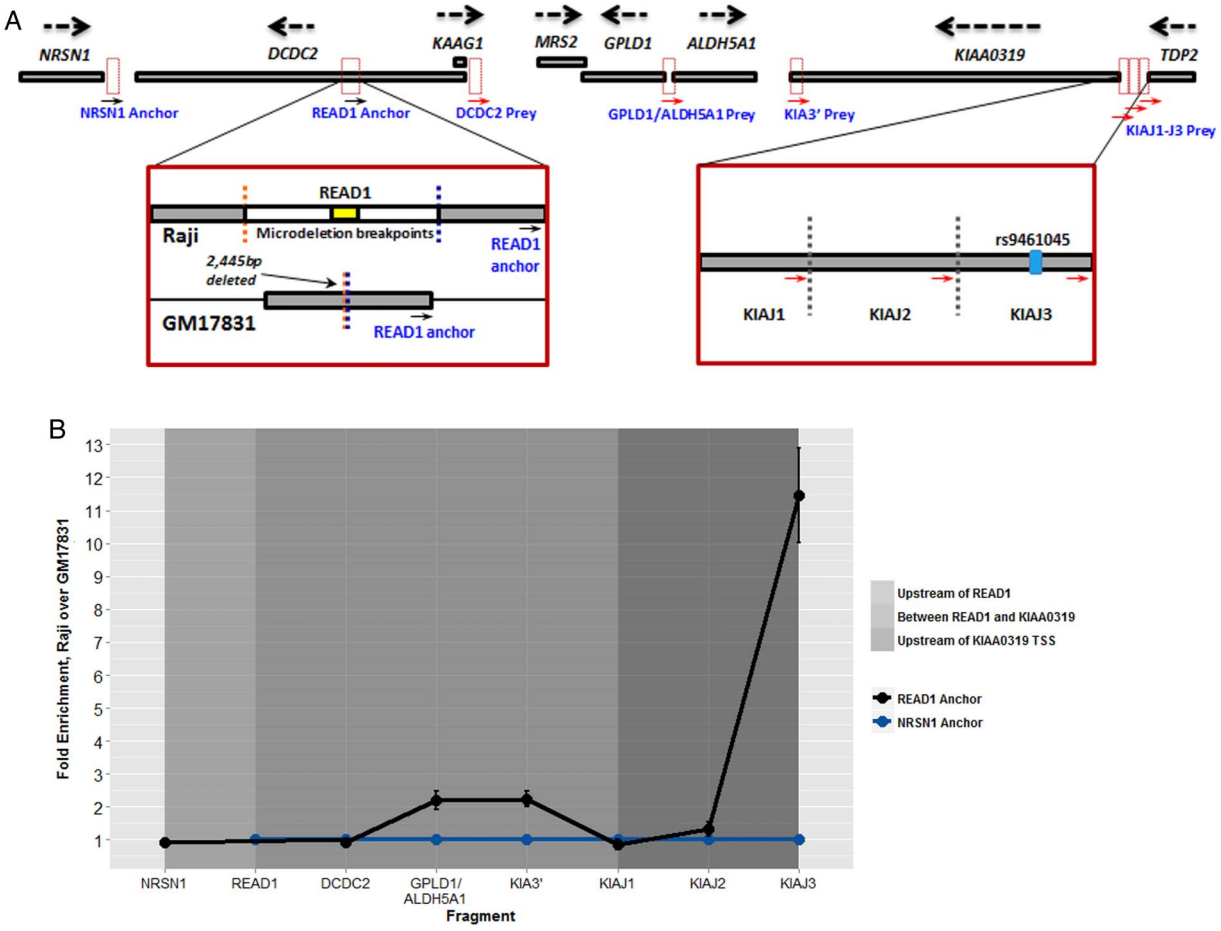


Figure 2 Chromatin conformation capture. (A) Schematic representation of our 3C strategy. The relevant region of the DYX2 locus is shown, with gene names in black font. Strand orientation of each gene is also shown. Dotted-line boxes show the positions of HindIII restriction fragments used for this experiment, and the positions of anchor and prey primers are indicated by arrows and labelled in blue font. Magnified views of the READ1 anchor primer and KIAJ1–J3 regions are shown. The positions of READ1, the breakpoints of the 2445 bp microdeletion (blue and orange dotted lines), and the READ1 anchor primer within the restriction fragment are shown in homozygous READ1-intact Raji cells, and GM17831 cells homozygous for the microdeletion. The primers KIAJ1–J3 flank three adjacent restriction fragments, which together encompass the intergenic region upstream of *KIAA0319* and downstream of *TDP2*. The presence of rs9461045 on the KIAJ3 fragment is noted. (B) 3C results. This graph shows enrichment of the indicated fusion fragment in Raji over GM17831, for the READ1 anchor primer relative to the control NRSN1 anchor primer. The y-axis indicates fold-enrichment of READ1-anchor fusion fragments (black line) normalised to NRSN1-anchor fusion fragments (blue line), which were set at 1. Error bars represent SE among two six-replicate qPCR experiments. Shaded areas mark the position of the included fragments relative to READ1. The prey primers shown on the x-axis are listed in the order in which they reside on the chromosome.

suppress the deleterious effect of KIAHap: performance on reading-related measures is typically at or above the population mean in the RU1-1-positive group, regardless of the presence or absence of KIAHap. Although this increase in performance is slight, it shows that the small deleterious population effect of KIAHap on reading performance does not occur when RU1-1 is present. In other words, for reading-related measures, KIAHap does not confer risk for poorer performance in the presence of an RU1-1 allele. Similarly, the deleterious READ1 alleles alone, like KIAHap alone, reduce mean performance only slightly, whereas their effects are greater in the presence of each other. This apparent genetic interdependence lends a contextual dimension to these 'risk variants': if used in the clinic to assess individual risk, they cannot be considered apart from each other.

Although the READ1–KIAHap genetic interaction shows strong general trends, there is some variability among phenotypes. For instance, single-word reading shows a somewhat attenuated effect at age 9 (reading 9) versus at age 7 (reading 7) (figure 1). This may be due to the measures themselves: ALSPAC's reading task at age 9 is abbreviated compared with that at age 7 and therefore may not capture reading ability with the same resolution. However, the effect of instruction is also likely to be important. At age 7, formal reading instruction is in many cases just beginning, while at age 9, the quality of instruction is expected to exert significantly greater influence on reading performance.[37] A stronger genetic effect at age 7 would be expected. There is also some disparity between the two language measures: NWR and verbal comprehension (WOLD). This is not unexpected, as these tasks measure different aspects of verbal language. NWR, in which the child listens to a non-word and repeats it to the examiner, measures receptive phonological working memory, as well as other language skills such as phonological processing and articulation.[38] WOLD, in which the child answers questions about a story read to him/her by the examiner, measures ability to derive meaning from spoken language.[39] This variability in the effect of the READ1–KIAHap genetic interaction points to the complex nature of reading, language and cognitive traits presented here.

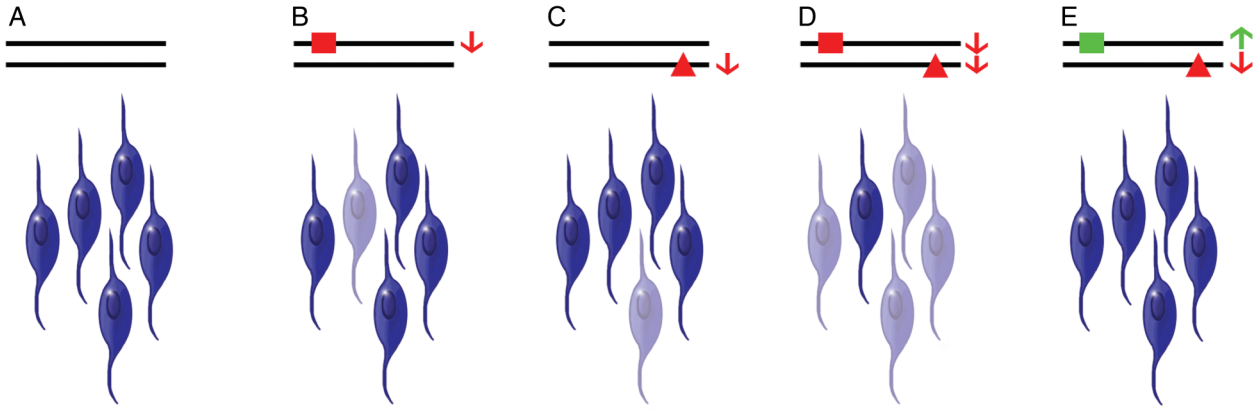


Figure 3 Theoretical model of the READ1–KIAHap genetic interaction. (A) In the presence of neutral READ1 alleles and the absence of KIAHap, most neural progenitor cells/neurons express *KIAA0319* above a minimal threshold (non-faded cells). (B and C) In the presence of a deleterious READ1 allele (red square) or KIAHap alone (red triangle), *KIAA0319* from the affected allele decreases, dropping overall *KIAA0319* expression in some cells below the minimal threshold (faded cells), slightly increasing the probability of problems with reading and language. (D) In the presence of both a deleterious READ1 allele and KIAHap *in trans*, *KIAA0319* expression from both alleles decreases, dropping overall *KIAA0319* expression below the minimal threshold in many cells, substantially increasing the probability of problems with reading and language. (E) A compensatory increase in *KIAA0319* expression due to a protective READ1 allele (green square) negates the decrease in expression due to KIAHap, maintaining overall *KIAA0319* expression above the minimal threshold in this case.

Several independent lines of evidence point to a direct regulatory interaction between READ1 and *KIAA0319*, including their genetic interaction, the different effects of structurally distinct alleles on this interaction, the binding of the potent transcriptional repressor ETV6 to READ1 and now, the physical interaction between READ1 and a promoter-proximal region of *KIAA0319* shown by 3C. More specifically, these data show higher fusion fragment enrichment in Raji (a homozygous READ1-intact lymphoblastoid cell line that contains a risk allele of READ1) relative to GM17831 (a homozygous READ1-deleted lymphoblastoid cell line) for the READ1 anchor primer versus the control NRSN1 anchor primer. This enrichment is also present for *GPLD1* and KIA3′—two fragments in the region between *DCDC2* and KIAJ1—but decays at the KIAJ1 and KIAJ2 fragments and then reappears very strongly at KIAJ3 (figure 2B). Intriguingly, we did not observe an interaction between READ1 and fragments containing the *KIAA0319* promoter (KIAJ1) or the *DCDC2* promoter (DCDC2). This appears to indicate that the functional variant responsible for the READ1–KIAHap genetic interaction is located upstream of *KIAA0319*. This region has been previously shown to affect *KIAA0319* expression by implication of an RD-associated allele of SNP rs9461045 with reduced *KIAA0319* expression in several cell lines.[16]

Although the genetic interaction between READ1 and KIAHap is clear, it is interesting that it is not necessarily a *cis* interaction. The transmission data in the CLDRC cohort show that allele 5, which synergises strongly with KIAHap for reading and IQ measures, rarely occurs on the same chromosome as KIAHap in Europeans (table 2). When considered with the 3C results, this creates an apparent paradox: the two variants interact *in cis* physically, but *in trans* genetically. Transvection would resolve this paradox, but while we cannot definitively rule it out, we consider this possibility unlikely because homologue pairing in somatic cells is strictly limited in mammals.[40] Our model, illustrated in figure 3, resolves the paradox in terms of total *KIAA0319* expression from both chromosomes. Under this model, READ1 regulates *KIAA0319* gene expression *in cis*—that is, each READ1 allele directly regulates only the copy of *KIAA0319* on the same chromosome as itself. KIAHap likewise alters *KIAA0319* expression *in cis*. If deleterious READ1 alleles and KIAHap do indeed decrease expression of *KIAA0319*, the *additive* effect of these deleterious variants could drop average *KIAA0319* expression below a tolerable threshold. If enough cells (eg, neurons or neural progenitors) express *KIAA0319* below this threshold, inadequate KIAA0319 will be elaborated, increasing the risk for reading and verbal language problems. Under our model (figure 3), the presence of both a READ1 risk variant and KIAHap would drop *KIAA0319* expression below this threshold in many more cells than the presence of only one or the other. The mechanism by which *KIAA0319* expression influences reading and language is unknown, but KIAA0319 is thought to be a signalling protein and is known to be involved in neuronal migration and dendrite outgrowth.[10 41]

The model explains the epistatic effect of the RU1-1 alleles over KIAHap the same way. These alleles, which have lost some ETV6 binding sites, may have lost enough of their repressive power to allow *KIAA0319* expression to be relatively higher, thereby compensating for reductions in expression caused by deleterious READ1 alleles and/or KIAHap. This model also allows for considerable phenotypic variation among individuals with the same genotype, as gene expression in individual cells can be influenced by many genetic, epigenetic, environmental and stochastic factors. In spite of this complexity, READ1 and KIAHap have a clear effect on population risk of RD in Europeans and may be useful in assessing individual risk if included in a model with environmental risk factors and other genetic risk variants.

Another perplexing facet of this interaction is that the long allele–deleterious/short allele–protective trend, while compelling, is certainly not the whole story. The RU1/RU2 region is also only part of the puzzle, as shown by the different effects of allele 5 and allele 6 on phenotype. These alleles differ by only 4 bp in RU4 (see online supplementary table S1), yet allele 5 has a stronger effect than allele 6, and also preferentially affects reading-related and IQ measures, whereas allele 6 mainly affects verbal language (figure 1A, B). When the two alleles are combined together and with the other, rare clade 1 alleles

(figure 1C), the magnitude of their synergistic effects appears somewhat attenuated, suggesting that alleles 5 and 6 are the main drivers for their respective phenotypes.

Taken together, the results presented here broadly suggest a model in which READ1 alleles differentially suppress *KIAA0319* expression through a direct, *cis*-regulatory interaction, the magnitude of which depends on the structure of the READ1 allele, and also on the presence or absence of a variant in LD with KIAHap, possibly rs9461045. Under our model, the additive effects on *KIAA0319* gene expression of READ1 and KIAHap genotypes on the two homologous chromosomes are responsible for the apparent *trans* genetic interaction. Physical interaction between READ1 and the *KIAA0319* upstream region appears to be restricted to the KIAJ3 restriction fragment, but there are interactions with other regions of the locus too, including the upstream regions of *GPLD1*/*ALDH5A1*, but surprisingly not *DCDC2*. This may imply that READ1 can regulate other genes in the locus and that its preference for its binding site upstream of *KIAA0319* can be altered by variants in these regions.

This study confirms and elaborates our previous work, but also raises many tantalising questions about the READ1–*KIAA0319* interaction. For instance, what other genetic and environmental factors can influence this interaction? Exactly what effect do these variants and their interaction have on neural development and how do they exert it? Much further work will be required to answer these and other questions, but the answers will provide a case of interacting regulatory variants that influence highly heritable complex traits in humans—a model that may well be broadly applicable to complex inheritance.

Author affiliations
[1]Investigate Medicine, Yale University, New Haven, Connecticut, USA
[2]Department of Pediatrics, Yale University, New Haven, Connecticut, USA
[3]School of Social and Community Medicine, University of Bristol, Bristol, UK
[4]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, USA
[5]W.M. Keck Foundation Biotechnology Resource Laboratory, Yale University, New Haven, Connecticut, USA
[6]Departments of Pediatrics and Developmental Neuroscience, University of Nebraska Medical Center, Omaha, Nebraska, USA
[7]Department of Psychology, University of Denver, Denver, Colorado, USA
[8]Institute for Behavioral Genetics, University of Colorado, Boulder, Colorado, USA
[9]Departments of Psychology and Neuroscience, University of Colorado, Boulder, Colorado, USA
[10]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
[11]Department of Investigative Medicine, Yale University, New Haven, Connecticut, USA

Acknowledgements We are extremely grateful to all members of the Avon Longitudinal Study of Parents and Children (ALSPAC) Study: the participants and their families who took part in this study, the midwives for their help in recruiting them and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. We thank the Yale Keck DNA sequencing facility for Sanger sequencing services, and Dr Silvia Paracchini for her generous contribution of genotype data in the ALSPAC Study for rs4504469, 2038137 and 2143340. We also thank Dr Laura DeMare for helpful technical discussions about the 3C protocol, and Dr Tae Hoon Kim for invaluable guidance and advice for the 3C experimental design.

Contributors NRP performed READ1 and microdeletion genotyping, and designed and performed the 3C experiments. NRP and JDE performed statistical analysis and drafted the manuscript. YK wrote the C++ program used to call READ1 genotypes from electropherograms. SDS and LLM were involved in recruitment, DNA collection and phenotyping of the ALSPAC cohort, and directly managed our collaboration with ALSPAC. LLM managed transfer of deidentified data between our group and ALSPAC. SDS, BFP, EGW and RKO were involved in recruitment, DNA collection and phenotyping of the Colorado Learning Disabilities Research Center cohort, and genotyped the SNPs for that cohort. JRG conceived and supervised the study, and helped with experiments and statistical analysis. All authors read and approved the manuscript.

Funding This work was supported by the National Institutes of Health (grant numbers R01 NS043530 to JRG; P50 HD027802 to YK and JRG; F31 DC012270 to JDE). The UK Medical Research Council and the Wellcome Trust (grant number 092731) and the University of Bristol provide core support for ALSPAC.

Competing interests The authors declare competing financial interests: Yale University has applied for a patent covering the complex tandem repeat and deletion in READ1, also known as BV677278 (inventor: JRG).

Patient consent Obtained.

Ethics approval Ethical approval for the ALSPAC cohort was obtained from the ALSPAC Ethics and Law Committee, the Local UK Research Ethics Committees and the Yale Human Investigation Committee. Informed consent for the study was obtained by the ALSPAC team. Ethical approval for recruitment and study methodologies for the CLDRC cohort was obtained from the Yale Human Investigation Committee and institutional review boards at the University of Denver, University of Colorado-Boulder and University of Nebraska Medical Center. This study adhered to the tenets of the Declaration of Helsinki.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The C++ program for calling READ1 alleles from electropherograms is available on request.

REFERENCES

1 Peterson RL, Pennington BF. Developmental dyslexia. *Lancet* 2012;379:1997–2007.
2 Rice ML. Language growth and genetics of specific language impairment. *Int J Speech Lang Pathol* 2013;15:223–33.
3 Pennington BF, Bishop DV. Relations among speech, language, and reading disorders. *Annu Rev Psychol* 2009;60:283–306.
4 Newbury DF, Fisher SE, Monaco AP. Recent advances in the genetics of language impairment. *Genome Med* 2010;2:6.
5 Raskind WH, Peter B, Richards T, Eckert MM, Berninger VW. The genetics of reading disabilities: from phenotypes to candidate genes. *Front Psychol* 2012;3:601.
6 van der Leij A. Dyslexia and early intervention: what did we learn from the Dutch Dyslexia Programme? *Dyslexia* 2013;19:241–55.
7 Scerri TS, Morris AP, Buckingham LL, Newbury DF, Miller LL, Monaco AP, Bishop DV, Paracchini S. DCDC2, KIAA0319 and CMIP are associated with reading-related traits. *Biol Psychiatry* 2011;70:237–45.
8 Eicher JD, Powers NR, Miller LL, Mueller KL, Mascheretti S, Marino C, Willcutt EG, DeFries JC, Olson RK, Smith SD, Pennington BF, Tomblin JB, Ring SM, Gruen JR. Characterization of the DYX2 locus on chromosome 6p22 with reading disability, language impairment, and IQ. *Hum Genet* 2014;133:869–81.
9 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. DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proc Natl Acad Sci USA* 2005;102:17053–8.
10 Peschansky VJ, Burbridge TJ, Volz AJ, Fiondella C, Wissner-Gross Z, Galaburda AM, Lo Turco JJ, Rosen GD. The effect of variation in expression of the candidate dyslexia susceptibility gene homolog Kiaa0319 on neuronal migration and dendritic morphology in the rat. *Cereb Cortex* 2010;20:884–97.
11 Adler WT, Platt MP, Mehlhorn AJ, Haight JL, Currier TA, Etchegaray MA, Galaburda AM, Rosen GD. Position of neocortical neurons transfected at different gestational ages with shRNA targeted against candidate dyslexia susceptibility genes. *PLoS ONE* 2013;8:e65179.
12 Centanni TM, Chen F, Booker AM, Engineer CT, Sloan AM, Rennaker RL, LoTurco JJ, Kilgard MP. Speech sound processing deficits and training-induced neural plasticity in rats with dyslexia gene knockdown. *PLoS ONE* 2014;9:e98439.
13 Che A, Girgenti MJ, LoTurco J. The dyslexia-associated gene dcdc2 is required for spike-timing precision in mouse neocortex. *Biol Psychiatry* 2014;76:387–96.
14 Truong DT, Che A, Rendall AR, Szalkowski CE, LoTurco JJ, Galaburda AM, Holly Fitch R. Mutation of Dcdc2 in mice leads to impairments in auditory processing and memory ability. *Genes Brain Behav* 2014;13:802–11.
15 Meng H, Powers NR, Tang L, Cope NA, Zhang PX, Fuleihan R, Gibson C, Page GP, Gruen JR. A dyslexia-associated variant in DCDC2 changes gene expression. *Behav Genet* 2011;41:58–66.
16 Dennis MY, Paracchini S, Scerri TS, Prokunina-Olsson L, Knight JC, Wade-Martins R, Coggill P, Beck S, Green ED, Monaco AP. A common variant associated with dyslexia reduces expression of the KIAA0319 gene. *PLoS Genet* 2009;5:e1000436.
17 Harold D, Paracchini S, Scerri T, Dennis M, Cope N, Hill G, Moskvina V, Walter J, Richardson AJ, Owen MJ, Stein JF, Green ED, O'Donovan MC, Williams J, Monaco

AP. Further evidence that the KIAA0319 gene confers susceptibility to developmental dyslexia. *Mol Psychiatry* 2006;11:1085–91, 61.
18 Ludwig KU, Roeske D, Schumacher J, Schulte-Korne G, Konig IR, Warnke A, Plume E, Ziegler A, Remschmidt H, Muller-Myhsok B, Nothen MM, Hoffmann P. Investigation of interaction between DCDC2 and KIAA0319 in a large German dyslexia sample. *J Neural Transm (Vienna)* 2008;115:1587–9.
19 Powers NR, Eicher JD, Butter F, Kong Y, Miller LL, Ring SM, Mann M, Gruen JR. Alleles of a polymorphic ETV6 binding site in DCDC2 confer risk of reading and language impairment. *Am J Hum Genet* 2013;93:19–28.
20 Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, Molloy L, Ness A, Ring S, Davey Smith G. Cohort profile: the 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* 2013;42:111–27.
21 Cope N, Eicher JD, Meng H, Gibson CJ, Hager K, Lacadie C, Fulbright RK, Constable RT, Page GP, Gruen JR. Variants in the DYX2 locus are associated with altered brain activation in reading-related brain regions in subjects with reading disability. *Neuroimage* 2012;63:148–56.
22 Ludwig KU, Schumacher J, Schulte-Körne G, König IR, Warnke A, Plume E, Anthoni H, Peyrard-Janvid M, Meng H, Ziegler A, Remschmidt H, Kere J, Gruen JR, Müller-Myhsok B, Nöthen MM, Hoffmann P. Investigation of the DCDC2 intron 2 deletion/compound short tandem repeat polymorphism in a large German dyslexia sample. *Psychiatr Genet* 2008;18:310–12.
23 Wilcke A, Weissfuss J, Kirsten H, Wolfram G, Boltze J, Ahnert P. The role of gene DCDC2 in German dyslexics. *Ann Dyslexia* 2009;59:1–11.
24 Meda SA, Gelernter J, Gruen JR, Calhoun VD, Meng H, Cope NA, Pearlson GD. Polymorphism of DCDC2 reveals differences in cortical morphology of healthy individuals-a preliminary voxel based morphometry study. *Brain Imaging Behav* 2008;2:21–6.
25 Marino C, Meng H, Mascheretti S, Rusconi M, Cope N, Giorda R, Molteni M, Gruen JR. DCDC2 genetic variants and susceptibility to developmental dyslexia. *Psychiatr Genet* 2012;22:25–30.
26 Marino C, Scifo P, Della Rosa PA, Mascheretti S, Facoetti A, Lorusso ML, Giorda R, Consonni M, Falini A, Molteni M, Gruen JR, Perani D. The DCDC2/intron 2 deletion and white matter disorganization: focus on developmental dyslexia. *Cortex* 2014;57:227–43.
27 Riva V, Marino C, Giorda R, Molteni M, Nobile M. The role of DCDC2 genetic variants and low socioeconomic status in vulnerability to attention problems. *Eur Child Adolesc Psychiatry* 2015;24:309–18.
28 Green SM, Coyne HJ III, McIntosh LP, Graves BJ. DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. *J Biol Chem* 2010;285:18496–504.
29 Kim CA, Phillips ML, Kim W, Gingery M, Tran HH, Robinson MA, Faham S, Bowie JU. Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. *EMBO J* 2001;20:4173–82.
30 Francks C, Paracchini S, Smith SD, Richardson AJ, Scerri TS, Cardon LR, Marlow AJ, MacPhie IL, Walter J, Pennington BF, Fisher SE, Olson RK, DeFries JC, Stein JF, Monaco AP. A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. *Am J Hum Genet* 2004;75:1046–58.
31 Luciano M, Lind PA, Duffy DL, Castles A, Wright MJ, Montgomery GW, Martin NG, Bates TC. A haplotype spanning KIAA0319 and TTRAP is associated with normal variation in reading and spelling ability. *Biol Psychiatry* 2007;62:811–17.
32 Paracchini S, Steer CD, Buckingham LL, Morris AP, Ring S, Scerri T, Stein J, Pembrey ME, Ragoussis J, Golding J, Monaco AP. Association of the KIAA0319 dyslexia susceptibility gene with reading skills in the general population. *Am J Psychiatry* 2008;165:1576–84.
33 Elbert A, Lovett MW, Cate-Carter T, Pitch A, Kerr EN, Barr CL. Genetic variation in the KIAA0319 5′ region as a possible contributor to dyslexia. *Behav Genet* 2011;41:77–89.
34 Paracchini S, Thomas A, Castro S, Lai C, Paramasivam M, Wang Y, Keating BJ, Taylor JM, Hacking DF, Scerri T, Francks C, Richardson AJ, Wade-Martins R, Stein JF, Knight JC, Copp AJ, Loturco J, Monaco AP. The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. *Hum Mol Genet* 2006;15:1659–66.
35 Golding J, Pembrey M, Jones R, ALSPAC Study Team. ALSPAC—the Avon Longitudinal Study of Parents and Children. I. Study methodology. *Paediatr Perinat Epidemiol* 2001;15:74–87.
36 Gayán J, Smith SD, Cherny SS, Cardon LR, Fulker DW, Brower AM, Olson RK, Pennington BF, DeFries JC. Quantitative-trait locus for specific language and reading deficits on chromosome 6p. *Am J Hum Genet* 1999;64:157–64.
37 Christopher ME, Hulslander J, Byrne B, Samuelsson S, Keenan JM, Pennington B, Defries JC, Wadsworth SJ, Willcutt E, Olson RK. Modeling the etiology of individual differences in early reading development: evidence for strong genetic influences. *Sci Stud Read* 2013;17:350–68.
38 Coady JA, Evans JL. Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *Int J Lang Commun Disord* 2008;43:1–40.
39 Rust J. *WOLD: Wechsler objective language dimensions manual*. London, UK: The Psychological Corporation, 1996.
40 Apte MS, Meller VH. Homologue pairing in flies and mammals: gene regulation when two are involved. *Genet Res Int* 2012;2012:430587.
41 Velayos-Baeza A, Levecque C, Kobayashi K, Holloway ZG, Monaco AP. The dyslexia-associated KIAA0319 protein undergoes proteolytic processing with {gamma}-secretase-independent intramembrane cleavage. *J Biol Chem* 2010;285:40148–62.

Supplementary Material

The Regulatory Element READ1 Epistatically Influences Reading and Language, with both Deleterious and Protective Alleles

Authors: Natalie R. Powers[1,2,], John D. Eicher[1,], Laura L. Miller[3], Yong Kong[4,5], Shelley D. Smith[6], Bruce F. Pennington[7], Erik G. Willcutt[8,9], Richard K. Olson[8,9], Susan M. Ring[3,10], Jeffrey R. Gruen[1,2,11,*]

*Corresponding author
E-mail: jeffrey.gruen@yale.edu
Telephone: (203) 737-2202
Fax: (203) 737-5972

Supplementary Methods

Phenotypes and Exclusion Criteria

The children of the ALSPAC have been extensively phenotyped from before birth to early adulthood. An update on the status of the cohort was published recently. [1] The reading, language, and cognitive measures used for this study were collected at ages 7, 8, and 9 years. DNA samples from 10,259 of these children were available for genotyping. Because this is a follow-up of our previous study, for the ALSPAC we used the same dataset and the same case/control definitions for severe RD and LI as reported in our previous study. Briefly, subjects were excluded if they scored below 75 on the WISC-III full-scale IQ measure, if they reported an ancestry other than European, or if their DNA samples did not meet a minimum standard of quality. Subjects were also excluded if they did not have an unequivocal READ1 genotype, or if phasing of KIAHap returned a posterior probability value below 0.95. After all exclusion criteria were applied, a final ALSPAC dataset of 4,428 subjects remained for analysis. Details of the phenotypes and case/control definitions for ALSPAC are presented in Supplementary Table 2A-B. The Colorado Learning Disabilities Research Center (CLDRC) cohort consists of families with twin pairs and their siblings; families were recruited to the study if at least one member of each twin pair had a history of reading problems. For this study, 1,188 individuals in 292 families were used for analysis, after exclusion of several families that showed Mendelian errors in a SNP dataset we used previously. In the case of monozygotic twins, only one member of each twin pair was used for transmission analysis in this study. A full description of the cohort and its phenotypic measures is available in Gayán et al. (1999). [2] For transmission assessment in the CLDRC, families were only included if at least one copy of KIAHap was present in at least one of the parents.

READ1 Genotyping

READ1 and the 2,445bp microdeletion encompassing it were genotyped by Sanger sequencing and allele-specific PCR, respectively. Primers and amplification protocols are described in detail elsewhere. [3] For microdeletion genotyping, PCR products were electrophoresed on large (150-250 well) 1% agarose gels at 150V for approximately 1.5h. Gels were imaged and documented with a Kodak 1D documentation system, and genotypes were called manually from the gels. For READ1 genotyping, purification of PCR products and Sanger sequencing were done by the Keck DNA Sequencing Lab at Yale University, according to standard protocols. Genotypes were called from the chromatograms using an in-house C++ program developed by YK (available upon request). In cases where the program detected errors, chromatograms were read and de-convoluted manually; such errors often yielded a new allele. Call rates in ALSPAC for READ1 and the microdeletion were 0.963 and 0.993, respectively.

SNP Genotyping

rs4504469, rs2038137, and rs2143340 were genotyped in ALSPAC by Scerri et al., as described. [4] Call rates for these three SNPs in our dataset were 0.891 0.900 and 0.896, respectively. For the CLDRC sample, the SNPs rs4504469, rs2038137, and rs2143340 were genotyped at the University of Nebraska Medical Center as part of an Illumina BeadXpress array. Parents and siblings were genotyped, and the calls were reviewed for genotyping quality and for Mendelian inconsistencies. Samples or SNPs with poor performance were excluded from further analysis. After cleaning, the call rates were 0.987, 0.960, and 0.993, respectively. Phasing of these SNPs to construct individual haplotypes was done using the –hap-phase function in PLINK. Only samples with high-confidence phasing (posterior probability greater than or equal to 0.95) were used for analysis.

Cell Culture

Raji cells (ATCC CCL-86) were grown at 37ºC, 5% CO_2, in RPMI 1640 medium supplemented with 10% FBS. GM17831 cells (Coriell) were grown under the same conditions in RPMI 1640 medium supplemented with 15% FBS.

Chromatin Conformation Capture (3C)

3C was done as previously described, [5] with several minor modifications. For each 3C prep, 100 million freshly harvested cells were fixed in 1% formaldehyde in 21mL PBS, at room temperature for 10min. Formaldehyde was then quenched for 5 min at room temperature after addition of 1.26mL of 2.5M glycine. After fixation, cells were pelleted, flash-frozen in liquid nitrogen, and stored at -80ºC until used for 3C. For 3C, cells were resuspended in 6 volumes of Lysis buffer 1 (10mM Tris-HCl ph=8.0, 10mM NaCl, 0.2% Igepal), and homogenized with 7 strokes in a Dounce homogenizer. They were then incubated on ice for 20min, and homogenized

again with 25 strokes in a Dounce homogenizer. The nuclei were collected by centrifugation (5min, 2500g, 4ºC), washed with 500ul 1X restriction Buffer 2.1 (NEB), collected by centrifugation (same settings), resuspended in 1X Buffer 2.1, and split into 10 aliquots, ~60µl each. To each aliquot, 312µl Buffer 2.1 and 38µl 1% SDS was added, and the aliquots were incubated with rotation at 65ºC for 15min. 44µl 10% Triton X-100 were added to each tube, and 400 units of HindIII restriction enzyme (NEB) were added to 9 of the tubes. The remaining tube was split into two aliquots, and 200 units of HindIII were added to one of them (digested control); the other tube served as an undigested control. All tubes were incubated overnight at 37ºC with rotation. The following morning, an additional 200 units of HindIII were added to the 9 3C digests, and an additional 100 units to the digested control, and the tubes were incubated at 37ºC for an additional 2 hours with rotation. The enzyme was then inactivated by addition of 43µl of 20% SDS to each of the 9 3C digests, and 20µl each to the digested and undigested controls, and incubating at 65ºC with rotation for 30 minutes. Ligation reactions were then set up. Each of the 9 reactions consisted of 1 3C digest, 745µl 10X ligation buffer (500mM Tris-HCl, pH=7.5, 100mM $MgCl_2$, and 100mM dithiothreitol), 745µl 10% Triton X-100, 8µl BSA (100mg/mL), 1µl T4 DNA ligase (30 Weiss units/µl), and 5.5mL H_2O. The reactions were mixed and allowed to proceed at 16ºC for 3 hours. Meanwhile, the digested and undigested controls were treated with 10µg RNAse A and incubated for 1 hour at 37ºC. When ligation was complete, the ligation reactions were treated with 50µl proteinase K (25mg/mL), and incubated overnight at 65ºC to reverse crosslinks and digest protein. The digested and undigested controls were subjected to the same treatment, with 10µl proteinase K. The following morning, an additional 25µl and an additional 5µl of proteinase K was added to each ligation reaction and to each of the controls, respectively; they were then incubated at 65ºC for an additional 2 hours. The digested and undigested controls were stored at -20ºC until further use. The ligation reactions were distributed among 11 MaxTract tubes (Qiagen) for DNA purification. An equal volume of phenol-chloroform-isoamyl alcohol (25:24:1) was added to each tube. The tubes were inverted several times, and spun 5 minutes at 1,500xg. The aqueous phases were decanted and combined into 4 30mL glass centrifuge tubes. To the solution in each tube, 0.7vol isopropanol and 0.1vol 3M sodium acetate (pH=5.2) was added, the tubes were mixed by inversion, and placed at -80ºC for 1 hour. They were the thawed at room temperature for 30min, and spun for 45 minutes at 14,000xg (4ºC). The supernatant was decanted, the pellets were dried at room temperature, and each was dissolved in 250µl 1X TE (10mM Tris-HCl, pH=8.0, 1mM EDTA). The DNA was combined into 1mL total, transferred to a 1.5mL tube, and treated with 100µg RNAse A. The RNAse reaction was allowed to proceed at 37ºC for 1 hour. The DNA was then split into 2 500µl aliquots, and each was added to a 2mL phase-lock tube. The digested and undigested controls were each also added to a phase-lock tube. An equal volume of phenol-chloroform-isoamyl alcohol (25:24:1) was added to each tube, the tubes were inverted to mix, and spun 5min at 16,000xg (room temperature). The same volume of chloroform was added to the aqueous phase of each tube, and the tubes were again inverted to mix and spun at 16,000xg for 5min. The aqueous phase was then collected and transferred to a fresh 2mL tube, and 0.1vol of 3M sodium

acetate (pH=5.2) and 2.5vol of ice cold 100% ethanol was added to each tube. The tubes were inverted to mix, and stored at -80ºC overnight to precipitate the DNA. The following morning, the tubes were spun for 45min, 16,000xg, at 4ºC. The 3C pellets were washed 5X with cold 70% ethanol; the digested and undigested controls, 2X. The pellets were then dried, and each 3C pellet was dissolved in 500µl 1X TE, while the digested and undigested control pellets were dissolved in 150µl 1X TE. The two dissolved 3C pellets were combined into one, and all three samples were quantified with PicoGreen (Life Technologies), as per manufacturer's instructions. Samples were then used as qPCR template. The digested and undigested controls were used to correct for digestion efficiency between the two cell lines, as previously described. [5]

qPCR

qPCR was done with the QuantiTect SYBRGreen qPCR kit from Qiagen, in 50µl reactions, as per manufacturer's instructions. Primers are listed in Supplementary Table 5. The qPCR reaction is as follows: 15 min at 95ºC, then 45 cycles of 30 sec at 95ºC followed by 30 sec at 60ºC followed by 1 min at 72ºC, then 6 min at 72ºC, and an indefinite hold at 4ºC. For qPCR reactions, 3C template DNA was diluted to a final concentration of 20ng/ul, and each primer was diluted to a final concentration of 0.25µM. qPCR results were normalized across templates to a control amplicon from the gene encoding β-actin (*ACTβ*). The *ACTβ* primers amplify across a region without a HindIII, BamH1, or BglII site.

β-globin Control Experiment

To assess the effectiveness of our 3C protocol, and to eliminate any systematic differences between Raji and GM17831 cells, we performed 3C, according to the above protocol, with a set of previously described intrachromosomal interactions in the β-globin locus. Vu et al. (2010) detail two interactions and one non-interaction with an LCR region in the locus (flanked by anchor primer C). [6] One is a strong local interaction with a nearby region (flanked by prey primer B), one is a weaker long-range interaction (flanked by prey primer A), and one is a non-interaction with a distant region (flanked by prey primer D). Globin primer sequences are listed in Supplementary Table 6. Because the globin primers flank BglII or BamH1 sites rather than HindIII sites, fixed cells were subjected to double-digests with these enzymes in NEB restriction buffer 3.1. For each 3C digest, 200 units of each enzyme were used (100 units of each for the digested control). All other aspects of the protocol are unchanged, except an annealing temperature of 65ºC rather than 60ºC was used for qPCR with the globin primers. The results of the control experiment are shown in Supplementary Figure 1. They agree with those of Vu et al. (2010), [6] and show no difference in enrichment between Raji and GM17831 cells at the β-globin locus—indicating an effective 3C protocol and suggesting that there is no systematic bias between the two cell lines.

Supplementary Tables and Figures

Allele	Repeat unit 1	Repeat unit 2	SNP1	Repeat unit 3	Const. Region	Repeat unit 4	Repeat Unit 5	Length (bp)	Allele Frequency
1	**(GAGAGGAAGGAAA)2**	**(GGAA)7**	**(GAAA)1**	**(GGAA)2**	**GGAAAGAATGAA**	**(GGAA)4**	**(GGGA)2**	**102**	**0.6286**
2	(GAGAGGAAGGAAA)1	(GGAA)9	(GAAA)0	(GGAA)0	GGAAAGAATGAA	(GGAA)4	(GGGA)2	85	0
3	**(GAGAGGAAGGAAA)1**	**(GGAA)6**	**(GAAA)1**	**(GGAA)2**	**GGAAAGAATGAA**	**(GGAA)4**	**(GGGA)2**	**85**	**0.0456**
4	**(GAGAGGAAGGAAA)2**	**(GGAA)6**	**(GAAA)1**	**(GGAA)2**	**GGAAAGAATGAA**	**(GGAA)4**	**(GGGA)2**	**98**	**0.0955**
5	**(GAGAGGAAGGAAA)2**	**(GGAA)8**	**(GAAA)1**	**(GGAA)2**	**GGAAAGAATGAA**	**(GGAA)4**	**(GGGA)2**	**106**	**0.0361**
6	**(GAGAGGAAGGAAA)2**	**(GGAA)8**	**(GAAA)1**	**(GGAA)2**	**GGAAAGAATGAA**	**(GGAA)3**	**(GGGA)2**	**102**	**0.0471**
7	(GAGAGGAAGGAAA)2	(GGAA)8	(GAAA)1	(GGAA)1	GGAAAGAATGAA	(GGAA)4	(GGGA)2	102	0
8	(GAGAGGAAGGAAA)2	(GGAA)7	(GAAA)0	(GGAA)0	GGAAAGAATGAA	(GGAA)4	(GGGA)2	90	$5.136x10^{-5}$
9	(GAGAGGAAGGAAA)1	(GGAA)7	(GAAA)1	(GGAA)2	GGAAAGAATGAA	(GGAA)4	(GGGA)2	89	0.0063
10	**(GAGAGGAAGGAAA)2**	**(GGAA)4**	**(GAAA)1**	**(GGAA)2**	**GGAAAGAATGAA**	**(GGAA)4**	**(GGGA)2**	**90**	**0.0467**
11	(GAGAGGAAGGAAA)2	(GGAA)7	(GAAA)1	(GGAA)2	GGAAAGAATGAA	(GGAA)3	(GGGA)2	98	0.0014
12	(GAGAGGAAGGAAA)1	(GGAA)8	(GAAA)1	(GGAA)2	GGAAAGAATGAA	(GGAA)3	(GGGA)2	89	$5.138x10^{-5}$
13	(GAGAGGAAGGAAA)2	(GGAA)9	(GAAA)1	(GGAA)2	GGAAAGAATGAA	(GGAA)3	(GGGA)2	106	0.0012
14	(GAGAGGAAGGAAA)2	(GGAA)9	(GAAA)1	(GGAA)2	GGAAAGAATGAA	(GGAA)4	(GGGA)2	110	0.0016
15	(GAGAGGAAGGAAA)2	(GGAA)5	(GAAA)2	(GGAA)2	GGAAAGAATGAA	(GGAA)4	(GGGA)2	98	0.0005
16	(GAGAGGAAGGAAA)2	(GGAA)5	(GAAA)1	(GGAA)2	GGAAAGAATGAA	(GGAA)4	(GGGA)2	94	0.0015
17	(GAGAGGAAGGAAA)2	(GGAA)4	(GAAA)2	(GGAA)2	GGAAAGAATGAA	(GGAA)4	(GGGA)2	94	0
18	(GAGAGGAAGGAAA)2	(GGAA)7	(GAAA)2	(GGAA)2	GGAAAGAATGAA	(GGAA)4	(GGGA)2	106	0.0003
19	(GAGAGGAAGGAAA)2	(GGAA)9	(GAAA)0	(GGAA)0	GGAAAGAATGAA	(GGAA)4	(GGGA)2	98	$5.138x10^{-5}$
20	(GAGAGGAAGGAAA)2	(GGAA)10	(GAAA)1	(GGAA)2	GGAAAGAATGAA	(GGAA)4	(GGGA)2	114	0.0004
21	(GAGAGGAAGGAAA)2	(GGAA)6	(GAAA)1	(GGAA)2	GGAAAGAATGAA	(GGAA)3	(GGGA)2	94	$5.138x10^{-5}$
22	(GAGAGGAAGGAAA)2	(GGAA)10	(GAAA)0	(GGAA)0	GGAAAGAATGAA	(GGAA)4	(GGGA)2	102	0.0027
23	(GAGAGGAAGGAAA)2	(GGAA)11	(GAAA)0	(GGAA)0	GGAAAGAATGAA	(GGAA)4	(GGGA)2	106	0.0001
24	(GAGAGGAAGGAAA)2	(GGAA)6	(GAAA)2	(GGAA)2	GGAAAGAATGAA	(GGAA)4	(GGGA)2	102	0.0002
25	(GAGAGGAAGGAAA)1	(GGAA)8	(GAAA)1	(GGAA)2	GGAAAGAATGAA	(GGAA)4	(GGGA)2	93	$5.138x10^{-5}$
26	(GAGAGGAAGGAAA)2	(GGAA)5	(GAAA)1	(GGAA)2	GGAAAGAATGAA	(GGAA)3	(GGGA)2	90	0.0001
27	(GAGAGGAAGGAAA)1	(GGAA)5	(GAAA)1	(GGAA)2	GGAAAGAATGAA	(GGAA)4	(GGGA)2	81	0.0001
28	(GAGAGGAAGGAAA)2	(GGAA)7	(GAAA)1	(GGAA)2	GGAAAGAATGAA	(GGAA)5	(GGGA)2	106	0.0002
29	(GAGAGGAAGGAAA)2	(GGAA)5+(GGGA)1+(GGAA)1	(GAAA)1	(GGAA)2	GGAAAGAATGAA	(GGAA)4	(GGGA)2	102	$5.138x10^{-5}$

30	(GAGAGGAAGGAAA)2	(GGAA)5	(GAAA)1	(GGAA)4	GGAAAGAATGAA	(GGAA)4	(GGGA)2	102	$5.138x10^{-5}$
31	(GAGAGGAAGGAAA)2	(GGAA)7	(GAAA)1	(GGAA)1+ (GGGA)1	GGAAAGAATGAA	(GGAA)4	(GGGA)2	102	$5.138x10^{-5}$
32	(GAGAGGAAGGAAA)2	(GGAA)8	(GAAA)0	(GGAA)0	GGAAAGAATGAA	(GGAA)4	(GGGA)2	94	0.0001
33	(GAGAGGAAGGAAA)2	(GGAA)6	(GAAA)0	(GGAA)0	GGAAAGAATGAA	(GGAA)3	(GGGA)2	82	$5.138x10^{-5}$
34	(GAGAGGAAGGAAA)2	(GGAA)7	(GAAA)2	(GGAA)2	GGAAAGAATGAA	(GGAA)3	(GGGA)2	102	$5.138x10^{-5}$
35	(GAGAGGAAGGAAA)1+(GAGAGGAAGAAAA)1	(GGAA)7	(GAAA)1	(GGAA)2	GGAAAGAATGAA	(GGAA)4	(GGGA)2	102	0
36	(GAGAGGAAGGAAA)1+(GAGAGGAAGGAA)1	(GGAA)9	(GAAA)1	(GGAA)2	GGAAAGAATGAA	(GGAA)4	(GGGA)2	109	0
37	(GAGAGGAAGGAAA)2	(GGAA)6	(GAAA)1	(GGAA)2	GGAAAGAATGAA	(GGAA)1+ (GAAA)1+ (GGAA)2	(GGGA)2	98	0
38	(GAGAGGAAGGAAA)1	(GGAA)10	(GAAA)0	(GGAA)0	GGAAAGAATGAA	(GGAA)4	(GGGA)2	89	0
39	(GAGAGGAAGGAAA)1	(GGAA)5	(GAAA)2	(GGAA)2	GGAAAGAATGAA	(GGAA)4	(GGGA)2	85	0
40	(GAGAGGAAGGAAA)2	(GGAA)2+(GGAAA)1+(GGAA)7	(GAAA)1	(GGAA)2	GGAAAGAATGAA	(GGAA)4	(GGGA)2	115	$5.138x10^{-5}$
Del	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**0.0831**

Supplementary Table 1: Structure, length, and allele frequency in ALSPAC of all READ1 alleles described to date. The six common alleles and the microdeletion are shown in bold.

A.

Phenotype	Description
Reading at 7	Wechsler Objective Reading Dimensions (WORD), single-word reading task, age 7
Reading at 9	Single-word reading task, age 9
Phoneme Del	Auditory Analysis task, age 7
Total IQ	Wechsler Intelligence Scale for Children (WISC), Third Edition, Full-Scale IQ, age 8
Verbal IQ	WISC Verbal IQ component, age 8
Performance IQ	WISC Performance IQ component, age 8
WOLD	Wechsler Objective Language Dimensions (WOLD), verbal comprehension task, age 8
NWR	Non-word repetition task, age 8
NW Read at 9	Non-word reading task, age 9
Spelling at 7	Single-word spelling task, age 7
Spelling at 9	Single-word spelling task, age 9

B.

Phenotype	Description
Severe RD	Cases defined as having a score less than or equal to 2 standard deviations below the mean on the phoneme deletion task
Severe LI	Cases defined as having a score less than or equal to 2 standard deviations below the mean on either the WOLD verbal comprehension task *or* the non-word repetition task

Supplementary Table 2: (A) List of phenotypes used in ALSPAC analyses. Reading measures in the ALSPAC include a phoneme deletion task at age 7, single-word reading at ages 7 and 9, spelling at ages 7 and 9, single non-word reading at age 9, and passage comprehension, speed and accuracy at age 9. The phoneme deletion task measures phoneme awareness, [7] which is widely considered to be a core deficit in RD. [8] For the phoneme deletion task the child listens to a word spoken aloud, and is then asked to remove a specific phoneme from that word to make a new word (e.g. what word is created when the /b/ sound is removed the word 'block'? 'Lock'). This task is also known as the Auditory Analysis Test, and was developed by Rosner and Simon. [9] Single-word reading was assessed at age 7 using the reading subtest of the Wechsler Objective Reading Dimensions (WORD). [10] At age 7 and 9, spelling was assessed; the child was asked to spell a set of 15 age-adjusted words. At age 9, single-word reading was again assessed by asking the child to read ten real words and ten non-words aloud. The words and non-words used are a subset of a larger list of words and non-words taken from research conducted by Terezinha Nunes and others at Oxford. [11] The non-word repetition (NWR) task was ascertained at 8 years of age. This is a verbal language measure wherein the child was asked to repeat recorded non-words. This task measures short-term phonological memory and processing; [12] children with LI consistently perform poorly. [13] Verbal, performance, and total IQ were assessed at age 8, using the Wechsler Intelligence Scale for Children (WISC-III). [14] **(B)** Case/control definitions used in association analysis (Table 1).

Allele	Description	READ1 Alleles	Allele Freq. (ALSPAC)
Clade1	Clusters in Clade 1	5, 6, 11, 12, 13, 14, 20, 21	0.0905
RU1-1	Only 1 copy of RU1	2, 3, 9, 12, 25, 27, 38	0.0521
Short	<90bp	2, 3, 9, 12, 27, 33, 38	0.0521
Long	>102bp	5, 13, 14, 18, 20, 23, 28, 35, 40	0.0400

Supplementary Table 3: Description and allele frequencies for composite READ1 alleles. 'Clade 1' was derived from a ClustalW multiple alignment (standard parameters), and is of interest to us because it contains alleles phylogenetically related to alleles 5 and 6. [3]

	Allele 5	Allele 6	Clade 1	Allele 3	Ru1-1
PD	0.161	0.395	**0.033**	0.379	0.398
Reading7	**0.015**	0.143	**0.004**	0.081	0.135
Reading9	0.569	0.997	0.528	0.542	0.585
NW Reading	0.350	0.506	0.229	0.204	0.169
Spelling7	0.059	0.207	**0.028**	0.097	0.100
Spelling9	0.139	0.292	0.199	0.094	0.146
WOLD	0.549	0.424	0.305	0.955	0.955
NWR	0.423	0.281	**0.038**	0.458	0.191
Total IQ	0.168	0.688	0.059	0.342	0.302
Verbal IQ	0.075	0.273	**0.019**	0.282	0.289
Performance IQ	0.638	0.820	0.448	0.407	0.357

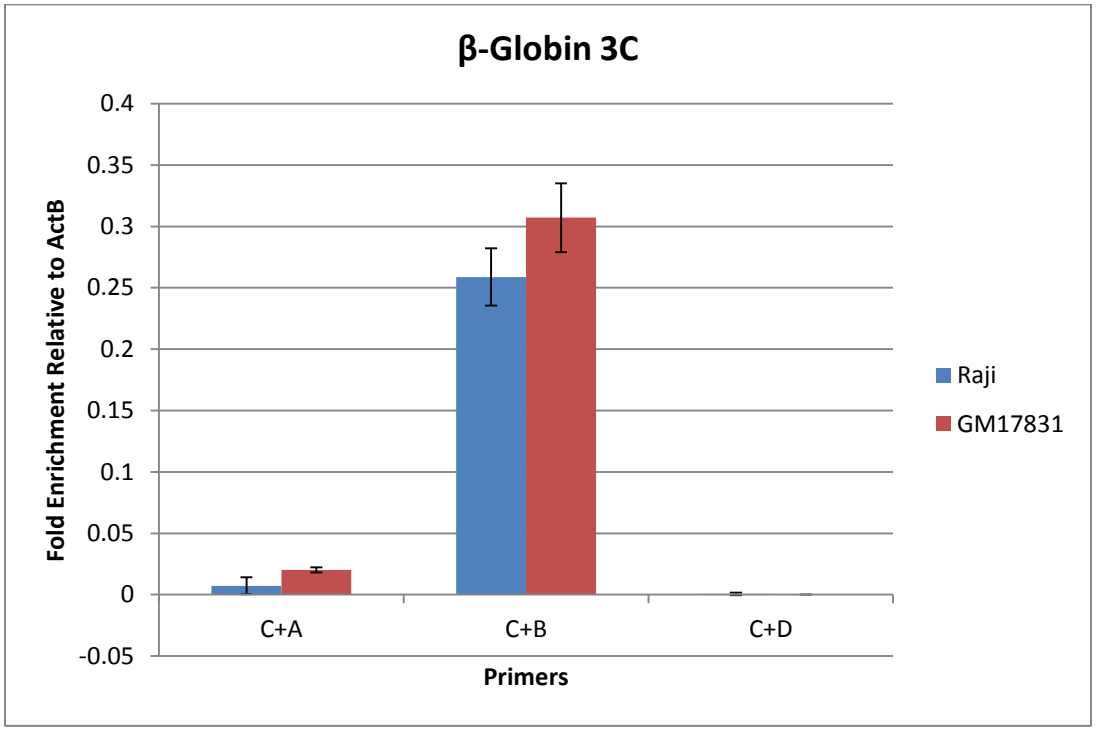
Supplementary Table 4: One-way ANOVA between groups for genetic interactions between READ1 risk and protective alleles and KIAHap (Figure1A-E). Values listed are p-values showing statistical significance of differences between means for the four genotype classes listed in Figure 1, for the indicated phenotype and READ1 single or composite allele. P-values below 0.05 are shown in bold.

Anchor Primers	READ1	5’-AGCCCTCCCTACTGACGGAAACACAT-3’ 5’-TTGCAGGGTGAAAATGAGGAGTTGAAAT-3’
	NRSN1	5’-TGCCCGGTACTCCCTCCAATCAGC-3’ 5’-CCAAGCCAAGGCCGCAGTGTTC-3’
Prey Primers	DCDC2	5’-AGTAAATGGACGCCTGCTGTGT-3’ 5’-GACTCTTTACTGGGGCTGTTACTATTCTCA-3’
	GPLD1/ALDH5A1	5’-AATATTTTTCTTTCTGCCCCACACC-3’ 5’-CCCAGCCTCTTCTCCCCCATTTT-3’
	KIA3’	5’-AGCTCCTCCTCCCCTTTCTATTG-3’ 5’-CATCTGTGGAGGTACGGAGTCTTG-3’
	KIAJ1	5’-TTTATCCTCCCGATTAATTTGTGACATTCC-3’ 5’-CAGAGCGCCTGGCCGAGAAATA-3’
	KIAJ2	5’-GGGCATTCCTCGCACATCTCATTA-3’ 5’-CCTCGGCCTGCCAAAGTGCTA-3’
	KIAJ3	5’-TGTCCCATGGTGCTATCAAACC-3’ 5’-TGCCAGCTGGATTCCAAACA-3’
Control Primers	ACTβ	5’-GCCCTAGGCACCAGGGTGTGA-3’ 5’-ACAGGGTGCTCCTCAGGGGC-3’

Supplementary Table 5: Primer sequences for 3C primers. Primers in black were used to assess fusion fragments for 3C template (anchor + prey). Primers in red are reverse primers with respect to their cognate 3C primers. 3C + reverse primers amplify across the relevant restriction site, and these short amplicons were used with the digested and undigested control template to assess digestion efficiency. Control primers do not amplify across a restriction site; they generate a short amplicon from the *ACTβ* gene, which was used to normalize across different qPCR templates.

Anchor Primer	Globin_C	5’-CGGTCATCCTCACGGTGACTAACGCA-3’
Prey Primers	Globin_A	5’-GACTCTTGAGGGCCTGACCTCGCTTAC-3’
	Globin_B	5’-GGTGGCAAAAGGCCTGTGCTGTTAGA-3’
	Globin_D	5’-AATGGCAATCACCACGATGGCCACA-3’

Supplementary Table 6: Primer sequences for control 3C experiment at the β-globin locus. [6]



Supplementary Figure 1: Results of the β-globin control 3C experiment. The y-axis shows fold-enrichment of the indicated fusion fragment relative to the control *ACTβ* primers, which were used to normalize across 3C templates. Error bars represent standard error among three replicates. These results agree with previously reported findings for this locus, [6] and indicate an effective 3C protocol.

Supplementary References

1. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, Molloy L, Ness A, Ring S, Davey Smith G. Cohort Profile: The 'Children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. International journal of epidemiology 2012;**42**(1):111-27 doi: 10.1093/ije/dys064[published Online First: Epub Date]|.
2. Gayan J, Smith SD, Cherny SS, Cardon LR, Fulker DW, Brower AM, Olson RK, Pennington BF, DeFries JC. Quantitative-trait locus for specific language and reading deficits on chromosome 6p. American journal of human genetics 1999;**64**(1):157-64 doi: 10.1086/302191[published Online First: Epub Date]|.
3. Powers NR, Eicher JD, Butter F, Kong Y, Miller LL, Ring SM, Mann M, Gruen JR. Alleles of a Polymorphic ETV6 Binding Site in DCDC2 Confer Risk of Reading and Language Impairment. American journal of human genetics 2013;**93**(1):19-28 doi: 10.1016/j.ajhg.2013.05.008[published Online First: Epub Date]|.
4. Scerri TS, Morris AP, Buckingham LL, Newbury DF, Miller LL, Monaco AP, Bishop DV, Paracchini S. DCDC2, KIAA0319 and CMIP are associated with reading-related traits. Biological psychiatry 2011;**70**(3):237-45 doi: 10.1016/j.biopsych.2011.02.005[published Online First: Epub Date]|.
5. Miele A, Dekker J. Mapping cis- and trans- chromatin interaction networks using chromosome conformation capture (3C). Methods Mol Biol 2009;**464**:105-21 doi: 10.1007/978-1-60327-461-6_7[published Online First: Epub Date]|.
6. Vu TH, Nguyen AH, Hoffman AR. Loss of IGF2 imprinting is associated with abrogation of long-range intrachromosomal interactions in human cancer cells. Human molecular genetics 2010;**19**(5):901-19 doi: 10.1093/hmg/ddp558[published Online First: Epub Date]|.
7. Hulme C, Goetz K, Gooch D, Adams J, Snowling MJ. Paired-associate learning, phoneme awareness, and learning to read. Journal of experimental child psychology 2007;**96**(2):150-66 doi: 10.1016/j.jecp.2006.09.002[published Online First: Epub Date]|.
8. Peterson RL, Pennington BF. Developmental dyslexia. Lancet 2012;**379**(9830):1997-2007 doi: 10.1016/S0140-6736(12)60198-6[published Online First: Epub Date]|.
9. Rosner J, Simon DP. The Auditory Analysis Test: An Initial Report. Journal of Learning Disabilities 1971;**4**(384):40-48 doi: 10.1177/002221947100400706[published Online First: Epub Date]|.
10. Rust J, Golombok S, Trickey G. *WORD: Wechsler Objective Reading Dimensions Manual*. Sidcup, UK: Psychological Corporation, 1993.
11. Nunes T, Bryant P, Olssen J. Learning Morphological and Phonological Spelling Rules: An Intervention Study. Scientific Studies of Reading 2003;**7**(3):289-307 doi: 10.1207/S1532799XSSR0703_6[published Online First: Epub Date]|.
12. Gathercole SE, Baddeley AD. *The Children's Test of Nonword Repetition*. London, UK: Psychological Corporation, 1996.
13. Bishop DV, North T, Donlan C. Nonword repetition as a behavioural marker for inherited language impairment: evidence from a twin study. Journal of child psychology and psychiatry, and allied disciplines 1996;**37**(4):391-403
14. Wechsler D. *Wechsler Intelligence Scale for Children, Fourth Edition*. London, UK: The Psychological Corporation, 2004.