Alleles of a Polymorphic ETV6 Binding Site in *DCDC2* Confer Risk of Reading and Language Impairment

Natalie R. Powers,[1] John D. Eicher,[1] Falk Butter,[3] Yong Kong,[4,5] Laura L. Miller,[6] Susan M. Ring,[6] Matthias Mann,[3] and Jeffrey R. Gruen[1,2,7,*]

Reading disability (RD) and language impairment (LI) are common learning disabilities that make acquisition and utilization of reading and verbal language skills, respectively, difficult for affected individuals. Both disorders have a substantial genetic component with complex inheritance. Despite decades of study, reading and language, like many other complex traits, consistently evade identification of causative and functional variants. We previously identified a putative functional risk variant, named BV677278 for its GenBank accession number, for RD in *DCDC2*. This variant consists of an intronic microdeletion and a highly polymorphic short tandem repeat (STR) within its breakpoints. We have also shown this STR to bind to an unknown nuclear protein with high specificity. Here, we replicate BV677278's association with RD, expand its association to LI, identify the BV677278-binding protein as the transcription factor ETV6, and provide compelling genetic evidence that BV677278 is a regulatory element that influences reading and language skills. We also provide evidence that BV677278 interacts nonadditively with *KIAA0319*, an RD-associated gene, to adversely affect several reading and cognitive phenotypes. On the basis of these data, we propose a new name for BV677278: "READ1" or "regulatory element associated with dyslexia 1."

Introduction

Specific learning disabilities (LDs) are disorders characterized by unexpected difficulty with a specific mode of learning, despite adequate intelligence and educational opportunity. LDs can involve reading, math, writing, and speech skills, among others, but the most common involve language. The National Institute of Child Health and Development (NICHD) estimates that as many as 15%–20% of Americans might be affected by an LD (NICHD website, see Web Resources), of which reading disability (RD) is the most common.[1] RD, also known as dyslexia (MIM 600202), is a specific impairment in processing written language.[2] Another LD, language impairment (LI [MIM 606711]), is characterized by difficulty processing and expressing spoken language.[3] These LDs are frequently comorbid, and children with LI have an increased risk of developing RD.[3] Because reading and language skills are fundamental to academic success, affected individuals are at risk for adverse psychological outcomes, as well as limited educational and occupational prospects.[2] Additionally, the prevalence of these LDs makes the cost of remediation burdensome to the educational system (National Center for Education Statistics website, see Web Resources). Intervention is more effective the earlier it is administered,[2] making early detection of high-risk individuals an attractive prospect.

Both RD and LI have a substantial genetic component, and inheritance for both disorders is typical of a complex trait.[3,4] Linkage and candidate-gene studies have identified risk loci and genes for RD and LI but have provided little insight into molecular mechanisms. The most replicated RD locus is the ~1.5 Mb DYX2 locus in chromosomal region 6p22; both linkage and association studies have repeatedly confirmed its involvement in RD.[4] Intriguingly, two of the most validated RD-associated genes, *DCDC2* (MIM 605755) and *KIAA0319* (MIM 609269), reside in DYX2 within 150 kb of each other.[4] Variants in both genes have been associated with RD in multiple independent genetic studies.[4] Likewise, both genes have been shown via in utero RNAi knockdown studies to be involved in neuronal migration during development,[5,6] although the extent of their importance in humans is unknown. Aberrant neuronal migration, however, is hypothesized to be a principal pathophysiology underlying RD.[7] *DCDC2*, a member of the doublecortin gene family, encodes a microtubule-binding protein. Rare mutations in one of its ancestral paralogs, *DCX* (MIM 300121), cause the Mendelian neuronal migration disorder X-linked lissencephaly (MIM 300067).[8] *KIAA0319* encodes a transmembrane protein whose function is currently unknown but that is thought to have a role in signaling.[9] In 2005, we reported in intron 2 of *DCDC2* a 2,445 bp microdeletion with a putative functional element within its breakpoints.[10] This element (GenBank accession number BV677278) is a highly polymorphic, purine-rich, compound short tandem repeat (STR). In the 2005 study, we showed that when the microdeletion was combined with several of the rarer BV677278

[1]Department of Genetics, Yale University, 464 Congress Avenue, Suite 243, New Haven, CT 06520, USA; [2]Department of Pediatrics, Yale University, 464 Congress Avenue, Suite 208, New Haven, CT 06520, USA; [3]Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany; [4]Department of Molecular Biophysics and Biochemistry, Yale University, 333 Cedar Street, New Haven, CT 06510, USA; [5]W.M. Keck Foundation Biotechnology Resource Laboratory, Yale University, 333 Cedar Street, New Haven, CT 06510, USA; [6]School of Social and Community Medicine, University of Bristol, Rooms OF10 and OF18, Oakfield House, Oakfield Grove, Clifton, Bristol BS8 2BN, UK; [7]Department of Investigative Medicine, Yale University, 464 Congress Avenue, Suite 208, New Haven, CT 06520, USA
*Correspondence: jeffrey.gruen@yale.edu
http://dx.doi.org/10.1016/j.ajhg.2013.05.008.

alleles, this "compound allele" showed strong association with an endophenotype for RD. That study unfortunately lacked the power to assess most of the BV677278 alleles individually as a result of their low allele frequencies, leaving open the question of whether the microdeletion or one or more of the BV677278 alleles was responsible for the signal. Subsequent association studies of the microdeletion have been inconclusive,[11–13] and until now, only one has been undertaken for BV677278, but it did not find evidence of association.[11] However, we recently showed that BV677278 binds a brain-expressed nuclear protein with very high specificity and that it is capable of modulating reporter-gene expression from the *DCDC2* promoter in an allele-specific manner.[14] We also recently showed that activation patterns in reading-related areas of the brain, as measured by functional MRI, are influenced by BV677278 alleles.[15]

It is currently unknown whether variants in *DCDC2*, *KIAA0319*, both, or neither are responsible for the DYX2 signal—principally because of a lack of power in previous studies. To address this question, we designed a tagSNP panel to densely cover the DYX2 locus and performed haplotype-based association analysis of reading and language in a large, extensively phenotyped birth cohort: the Avon Longitudinal Study of Parents and Children (ALSPAC).[16,17] This analysis revealed that in the same six-marker haplotype block, one haplotype associates with impaired reading ability and another associates with impaired language ability. Both of these haplotypes are in very strong linkage disequilibrium with an allele of BV677278. We also used SILAC (stable isotope labeling of amino acids in cell culture)-based mass spectrometry and ChIP (chromatin immunoprecipitation) to identify the BV677278-binding protein as the potent transcriptional regulator ETV6. Our results replicate and expand the previous association between BV677278 and reading and now language, provide strong circumstantial evidence that BV677278 is a regulatory element that exerts its effect through ETV6, and show that at least two of its alleles confer risk of a deleterious effect on reading and language. We also show that these two BV677278 "risk alleles" interact genetically with a known RD risk haplotype in *KIAA0319* in a nonadditive manner to influence reading, language, and IQ. Because BV677278 has these effects, we have renamed it "READ1," which stands for "regulatory element associated with dyslexia 1." It will be referred to hereafter in this paper as READ1.

Material and Methods

Subjects, Subject Recruitment, Data and DNA Collection, and Data Management

Subject recruitment and collection of phenotype data and DNA for the ALSPAC cohort was done by the ALSPAC team, as described elsewhere.[17] A detailed description of the phenotypes used in this study is available in Table S1, available online. The ALSPAC is a prospective birth cohort based in the Avon region of the United Kingdom. It consists mainly of children of northern European descent and born in 1991 and 1992. Children were recruited before birth; recruitment of their pregnant mothers 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 (see Web Resources).[17] The children of the ALSPAC cohort have been extensively phenotyped from before birth to early adulthood. An update on the status of the cohort was published recently.[17] The reading, language, and cognitive measures used for this study were collected when the children were 7, 8, and 9 years old. DNA samples from 10,259 of these children were available for genotyping, and of this subset, the number of children who completed the language and cognitive measures varies by measure but is generally 5,200–5,600 subjects.

Ethical Approval

Ethical approval for the study 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.[16]

DYX2 TagSNP Panel Design and Genotyping

TagSNPs designed to capture the common variation in the DYX2 locus were selected with the association study design server of Han et al. (see Web Resources).[18] SNPs were genotyped on the Sequenom platform in collaboration with the Yale Center for Genome Analysis as per standard protocols. The call rate and descriptive statistics for the SNPs reported in this article are listed in Table S3. rs4504469, rs2038137, and rs2143340 were genotyped by Scerri et al. as previously described.[19]

Haplotype-Based Association Analysis

Linkage disequilibrium was assessed and haplotypes were defined with Haploview version 4.2.[20] Markers that deviated substantially from Hardy-Weinberg equilibrium or that had a call rate $< 85\%$ were not used for haplotype analysis. We used the four-gamete-rule option in Haploview to demarcate haplotype blocks, which resulted in 44 haplotype blocks covering the DYX2 locus. We performed association analysis with individual haplotypes that had frequencies of 0.01 or greater (208 total) by using PLINK version 1.07.[21] Individuals who did not identify as non-Hispanic white, who had a total IQ below 75, or whose DNA sample returned an average call rate below 85% for SNPs that passed quality control were excluded from association analysis. To correct for multiple testing, we applied a Bonferroni correction with the alpha level set at 0.05 and treated each of the 208 haplotypes as an individual test; the threshold level was therefore $0.05/208 = 2.4038 \times 10^{-4}$.

READ1 Genotyping

Individuals who were positive for the *DCDC2* haplotypes of interest and could be phased unequivocally (with PLINK's –hap-phase function) were genotyped for the READ1 STR. READ1 was genotyped by PCR amplification, purification of PCR products with ExoSAP-IT enzyme mix, and Sanger sequencing. Sanger sequencing was performed at the Yale W.M. Keck DNA Sequencing Facility as per their standard sequencing protocol. Alleles were called by an in-house C language program developed for this purpose. Primer sequences and details of the amplification reaction are shown in Tables S8–S10. The allele-calling program is available upon request.

Microdeletion Genotyping

Individuals with the *DCDC2* haplotypes of interest were also genotyped for the 2,445 bp *DCDC2* microdeletion described previously.[10] This naturally occurring deletion encompasses the entire READ1 STR within its breakpoints, so it must be genotyped in addition to READ1 so that an accurate genotype can be achieved for apparent READ1 homozygotes. The microdeletion was genotyped by allele-specific PCR and agarose-gel electrophoresis with the use of a three-primer reaction that generates a ~600 bp amplicon from intact chromosomes and a ~200 bp amplicon from chromosomes with the deletion, allowing heterozygotes and both homozygotes to be readily distinguishable from one another. PCR products were electrophoresed on 1% agarose gels with the use of standard 1X TBE buffer with ethidium bromide (0.2 μg/ml) via standard methods at 100–150 V, depending on gel size. Gels were imaged on a UV transilluminator and documented with a Bio-Rad Gel Doc XR imaging system. Genotypes were called from the gels manually. Primer sequences and details of the amplification reaction are shown in Tables S8–S10.

Protein Identification by SILAC-Based Mass Spectrometry

Raji (human Burkitt lymphoma, ATCC CCL-86) and HeLa S3 (human cervical carcinoma, ATCC CCL-2.2) cells were SILAC labeled with Lys8 and Arg10 (Eurisotop) or their naturally occurring counterparts, Lys0 and Arg10 (Sigma), respectively, as described.[22] Heavy nuclear lysate prepared from these cells was incubated with a biotinylated oligonucleotide probe that was identical to a READ1 segment and that had been previously shown to bind a nuclear protein with high specificity.[14] Light nuclear lysate was incubated with a biotinylated scrambled probe previously shown not to bind the nuclear protein of interest.[14] The resulting oligonucleotide-protein complexes were pulled down with streptavidin-conjugated beads and subjected to quantitative mass spectrometry, as described previously.[23] The reverse experiment was also done (binding probe with light lysate and scrambled probe with heavy lysate), resulting in the two-dimensional interaction plots in Figures 2A and 2B. Details are described below.

Raji cells were labeled for at least eight generations in DMEM (-Arg, -Lys) medium containing 10% dialyzed fetal bovine serum (GIBCO) supplemented with 58 mg/l 13C615N4 L-arginine and 34 mg/l 13C615N2 L-lysine (Eurisotop) or the corresponding nonlabeled amino acids. For Raji, cell extracts were prepared as described in Wu et al.[24] HeLa S3 cells were SILAC labeled in RPMI 1640 (-Arg, -Lys) medium containing 10% dialyzed fetal bovine serum (GIBCO) supplemented with 84 mg/l 13C615N4 L-arginine and 40 mg/l 13C615N2 L-lysine (Eurisotop) or the corresponding nonlabeled amino acids. For HeLa S3, three consecutive batches of cells were independently harvested and cell extracts were prepared as described by Dignam et al.[25] SILAC, DNA pull-down of proteins, and quantitative mass spectrometry were performed as previously described[22] with the Raji and HeLa cell lines. The binding pull-down probe was a concatamer of two copies of the EMSA3 probe used in the EMSA experiments we reported in 2011, and the scrambled probe was a concatamer of two copies of the EMSA3-Scram1 probe from the same experiments.[14] The sequences of the oligonucleotides we used to make these probes are shown in Table S11. Twenty-five micrograms of annealed, concatenated, and desthiobiotinylated DNA probes was bound to 75 μl of Dynabeads MyOne C1 (Life Technologies). Excess oligonucleotides were removed, and beads were incubated with 400 μg of SILAC-labeled nuclear extracts in protein-binding buffer (150 mM NaCl, 50 mM Tris-HCl [pH 8], 0.5% NP-40, 10 mM $MgCl_2$, protease inhibitor cocktail; Roche). After 1 hr on a rotation wheel at 4°C, the beads were washed three times and combined, and DNA-protein complexes were eluted in protein binding buffer containing 16 mM biotin. The supernatant was precipitated with 4 v/v of ethanol overnight, and the proteins were pelleted by maximum centrifugation on a table-top microcentrifuge. The pellet was resolubilized in a solution of 8 M urea and 50 mM Tris (pH 8.0), reduced with 1 mM DTT, alkylated with 3 mM iodoacetamide, and subsequently digested with trypsin (Promega) in 50 mM ammonium bicarbonate buffer (pH8) at room temperature overnight. Samples were stored on stage tips and eluted prior to use. Peptides were separated with a 140 min gradient from 5% to 60% acetonitrile (EasyHPLC, Thermo Fisher) with a 75 μm 15 cm capillary packed with 3.0 μm C18 beads (Dr. Maisch) directly mounted to a LTQ-Orbitrap mass spectrometer (Thermo Fisher). The instrument was operated in a data-dependent top-ten acquisition modus. The raw data were searched with the MaxQuant software (version 1.2.0.18) suite against the complete International Protein Index human database (version 3.68; 87,061 entries). Enzyme search specificity was trypsin/p with two allowed miscleavages. Carbamidomethylation was set as fixed modification, whereas methionine oxidation and protein N-acetylation were considered variable modifications. The search was performed with an initial mass tolerance of 7 ppm mass accuracy for the precursor ion and 0.5 Da for the tandem-mass-spectrometry spectra.

ChIP-qPCR

To perform the ChIP assays reported in this article, we used the AbCam ChIP kit (catalog # ab500) according to the manufacturer's instructions but with several modifications. The modified protocol is available upon request. We used 5 μg of α-ETV6 antibody (sc-166835X, Santa Cruz Biotech) per reaction and 2 μg of control α-H3 antibody (ab1791, Abcam) per reaction. For quantitative PCR (qPCR), we used the QIAGEN QuantiTect SYBR Green qPCR kit for ChIP with qPCR (ChIP-qPCR) and followed the manufacturer's instructions. We used 25 pg of template per reaction and performed all reactions in triplicate. Primer sequences and details of the amplification reaction are shown in Tables S8–S10. Quality-control data for qPCR are shown in Figure S3. We calculated fold enrichment with respect to the no-antibody control (a complete ChIP reaction with only beads and no antibody) by raising 2 to the negative power of the difference between the cycle threshold (Ct) of an experimental condition (Ct Exp) and its respective no-antibody control (Ct NoAntibody):

$$\text{fold enrichment} = 2^{-[\text{Ct Exp} - \text{Ct NoAntibody}]}.$$

Results

Two Six-Marker Haplotypes in *DCDC2* Are Associated with Reduced Performance on Reading and Language Measures

During our analysis, we discovered a six-marker haplotype block within *DCDC2*; in this block, two haplotypes—CGCGAG and GACGAG—associated with very poor performance on a phoneme-deletion task and a composite language measure, respectively (Table 1). For this analysis,

Table 1. Association and Linkage-Disequilibrium Data for *DCDC2* Risk Haplotypes

	Haplotype	
	CGCGAG	**GACGAG**
Association Data		
Phenotype	phoneme-deletion task (RD)	WOLD-NWR (LI)
Cases (n)	89	270
Controls (n)	5,225	5,240
Haplotype frequency	0.0236	0.0364
Odds ratio	3.20	1.91
p value	$\mathbf{6.068 \times 10^{-5}}$	2.84×10^{-4}
Linkage-Disequilibrium Data		
Individuals (n)	226	392
% allele 5	92.0	12.0
% allele 6	7.5	77.6
% clade 1	94.3	91.3

Phenotypes are described in Table S1. Cases are defined by a score of ≤ 2 SDs below the mean. p values that survived Bonferroni correction for multiple testing ($\alpha = 0.05$) are bolded. "% allele 5" and "% allele 6" mean the percentage of individuals who were positive for the denoted haplotype and also had at least one copy of the denoted allele or group of alleles. Clade 1, the phylogenetic allele branch that includes alleles 5 and 6, is described in Figure S1.

we defined RD cases as individuals scoring ≥ 2 SDs below the mean on the phoneme-deletion task and LI cases as individuals scoring ≥ 2 SDs below the mean on either of two language measures, WOLD (Wechsler Objective Learning Dimensions) or NWR (nonword repetition). The phoneme-deletion task measures phonological awareness, which is widely considered to be the core deficit in RD.[2] The WOLD verbal comprehension and NWR tasks that comprise the WOLD-NWR composite language measure are used for assessing deficient language skills; children with LI show consistently poor performance on these measures[26,27] (see Table S1 for more information on these phenotypic measures). We defined cases in this way to examine association between DYX2 haplotypes and severe RD and LI. The two haplotypes showed strong association with their respective phenotypes; the association between CGCGAG and RD survived Bonferroni correction for multiple testing, and the GACGAG-LI p value was just below the threshold. However, the associations by themselves were not strong enough to rule out type 1 error, partly because of the low frequencies of the haplotypes and the low number of cases. Interestingly, however, the effect of these haplotypes was strong enough to significantly reduce mean performance on relevant phenotypic measures. Compared with CGCGAG-negative individuals, CGCGAG-positive individuals, on average, showed significantly poorer performance on eight reading-related measures. Likewise, relative to GACGAG-negative individuals, GACGAG-positive individuals showed significantly lower mean performance on the WOLD-NWR composite language measure (Table 2). This quantitative effect indicated to us that this finding is not a false positive and prompted us to pursue it further. Additionally, this haplotype block resides on the chromosome in close proximity to READ1, a putatively functional compound STR we reported previously as BV677278[10] (Figure 1B). The polymorphism of READ1 is derived from five discrete repeat units that vary in number (Figure 1A and Table S4). Like many repetitive elements, this STR appears to evolve rapidly, as indicated by its high degree of polymorphism among primate species and within *Homo sapiens* (Figure 1C, Figure S2, and Table S4). Although the risk haplotype block is close to READ1, and therefore to the 2,445 bp *DCDC2* microdeletion when it is present, the SNPs that compose it all lie outside the deleted region. The integrity of this haplotype block is therefore unaffected by the presence of the microdeletion, which for the purpose of this analysis we treated as an additional allele of READ1.

The *DCDC2* Risk Haplotypes Show Strong Linkage Disequilibrium with Two Alleles of READ1

Because the associated haplotype block is adjacent to READ1 (Figure 1B), we questioned whether the two risk haplotypes could be capturing association arising from functional READ1 alleles via linkage disequilibrium. To address this question, we subjected all individuals positive for these haplotypes to READ1 genotyping by Sanger sequencing. Of the CGCGAG-positive subjects, 92% were also positive for READ1 allele 5. Likewise, 78% of GACGAG-positive subjects were also positive for READ1 allele 6 (Table 1). Alleles 5 and 6 are similar in structure to each other and cluster phylogenetically to the same clade (Table S4 and Figure S1). Indeed, the genomes of nearly all individuals positive for one of these two haplotypes also harbored an allele from this clade (Table 1). These results further implicate READ1 as a RD risk variant and expand it as a possible LI risk variant[10] and, together with its apparent regulatory capacity,[14] suggest that these READ1 alleles are responsible for the risk haplotypes' effects.

READ1 Specifically Binds the Transcription Factor ETV6

To gain mechanistic insight into the function of READ1, we used quantitative mass spectrometry to identify the protein(s) that bind to this locus.[22] To this end, we incubated both a biotinylated oligonucleotide probe (with sequence derived from the READ1 STR) that we previously showed to bind a nuclear protein and a scrambled nonbinding control probe with nuclear extracts that had been SILAC labeled (Table S11).[14] SILAC labeling involves culturing two parallel populations of cells—one with media containing amino acids labeled with heavy isotopes of carbon and nitrogen and the other with naturally occurring isotopes. After the label is incorporated, proteins from the two populations ("heavy" and "light") can be differentiated from each other by quantitative mass spectrometry.

Table 2. Mean Performance on Reading and Cognitive Measures in Individuals Positive for and Negative for the *DCDC2* Risk Haplotype

	Mean Performance (SD)					
	CGCGAG (RD) Haplotype			GACGAG (LI) Haplotype		
Reading or Cognitive Measure	Positive	Negative	p Value	Positive	Negative	p Value
Reading at 7 years	27.34 (9.04)	29.01 (8.77)	0.005*	29.09 (8.62)	28.92 (8.80)	0.728
	n = 232	n = 929		n = 358	n = 4,803	
Spelling at 7 years	24.38 (13.46)	26.29 (12.33)	0.023*	25.56 (12.77)	26.26 (12.36)	0.305
	n = 229	n = 4,896		n= 355	n = 4,770	
Phoneme-deletion task	19.30 (10.00)	20.80 (9.17)	0.016*	20.61 (9.20)	20.74 (9.21)	0.796
	n = 230	n = 4,909		n = 357	n = 4,782	
Reading at 9 years	7.37 (2.71)	7.73 (2.27)	0.020*	7.75 (2.33)	7.72 (2.29)	0.754
	n = 228	n = 4,914		n = 359	n = 4,783	
NW reading at 9 years	5.05 (2.58)	5.38 (2.36)	0.043*	5.47 (2.36)	5.36 (2.44)	0.391
	n = 228	n = 4,911		n = 359	n = 4,780	
Spelling at 9 years	10.03 (2.58)	10.50 (3.23)	0.031*	10.48 (3.25)	10.48 (3.26)	0.987
	n = 228	n = 4,904		n = 357	n = 4,775	
Speed	105.44 (11.76)	106.34 (12.10)	0.299	106.71 (11.77)	106.27 (12.11)	0.524
	n = 207	n = 4,430		n = 326	n = 4,311	
Accuracy	102.77 (14.00)	105.22 (13.10)	0.009*	105.18 (13.24)	105.11 (13.15)	0.919
	n = 208	n = 4,438		n = 329	n = 4,317	
Reading comprehension	99.74 (11.67)	101.54 (11.37)	0.026*	101.73 (11.82)	101.44 (11.35)	0.663
	n = 208	n = 4,438		n = 329	n = 4,317	
Verbal IQ	107.35 (15.70)	108.97 (15.67)	0.113	108.38 (15.90)	108.94 (15.65)	0.497
	n = 245	n = 5,334		n = 388	n = 5,191	
Performance IQ	101.23 (14.96)	100.28 (16.16)	0.366	101.10 (15.72)	101.19 (16.14)	0.913
	n = 245	n = 5,334		n = 388	n = 5,191	
Total IQ	104.58 (14.22)	106.05 (15.26)	0.138	105.62 (14.95)	106.01 (15.23)	0.623
	n = 245	n = 5,334		n = 388	n = 5,191	
NWR	7.54 (1.94)	7.58 (1.91)	0.724	7.40 (1.91)	7.55 (1.91)	0.136
	n = 245	n = 5,276		n = 384	n = 5,137	
WOLD	7.11 (2.56)	7.33 (2.44)	0.178	7.12 (2.60)	7.33 (2.43)	0.104
	n = 245	n = 5,270		n = 383	n = 5,132	
NWR-WOLD	−0.031 (0.82)	0.00 (0.78)	0.532	−0.08 (0.77)	0.01 (0.78)	0.041*
	n = 245	n = 5,281		n = 384	n = 5,142	

The SD is shown in parentheses next to each mean. The number of subjects in each category is shown below each mean. The p values are from Student's t tests comparing the means of individuals positive for and negative for each haplotype; p values less than 0.05 are marked with an asterisk. Phenotypes are described in Table S1. Abbreviations are as follows: NW, nonword; NWR, nonword-repetition task; WOLD, Wechsler Objective Learning Dimensions verbal comprehension task; and NWR-WOLD, the average *Z* score of performance on NWR and WOLD verbal comprehension tasks.

We incubated the heavy nuclear extract with the READ1 probe and incubated the light nuclear extract with the control probe. We then pulled down the probes with streptavidin-conjugated beads, subjected the resulting protein mixture to quantitative mass spectrometry, and looked for proteins significantly enriched by pull-down with the READ1 probe compared to the control probe (high heavy-to-light ratio). The experiment was conducted with nuclear extracts derived from either HeLa cells or Raji cells and was repeated with a label switch resulting in a two-dimensional interaction plot. These experiments yielded a single candidate shared by both HeLa and Raji cells: the transcription factor ETV6 (Figures 2A and 2B). To confirm the READ1-ETV6 interaction, we performed ChIP-qPCR by using α-ETV6 antibody in both the Raji cell line and the human neuroepithelioma cell line

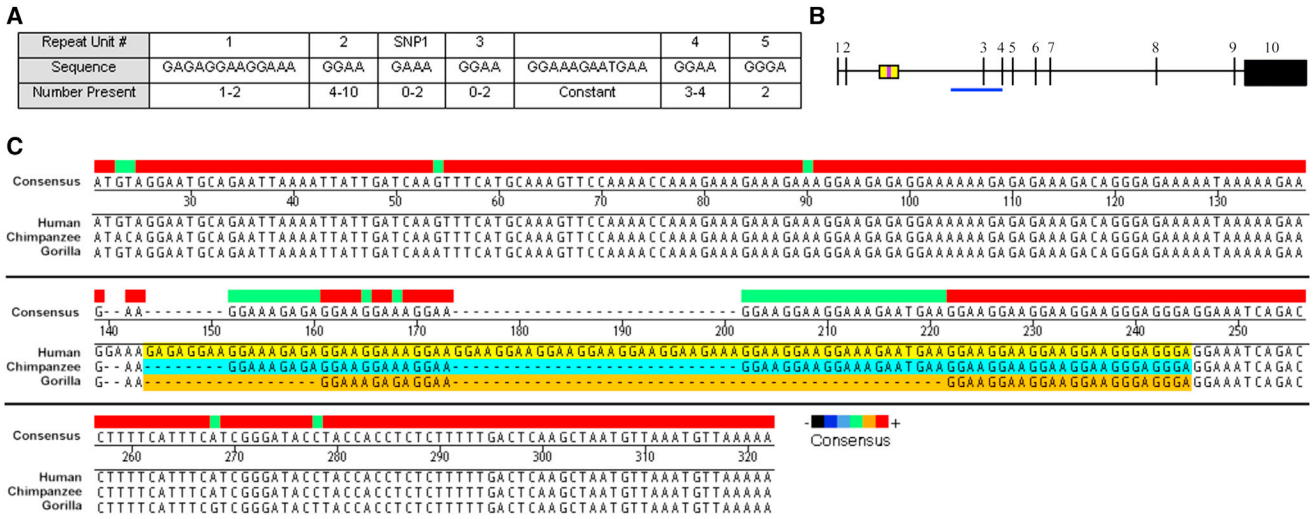

Figure 1. READ1 Is a Highly Polymorphic STR Located Near the *DCDC2* Risk Haplotype Block
(A) Structure of the READ1 STR.
(B) Location of the *DCDC2* risk-haplotype block (blue line) relative to the microdeletion (yellow box) and the READ1 STR (purple line) it encompasses. Exons are numbered.
(C) Alignment of READ1 and flanking sequence from human, chimpanzee, and gorilla *DCDC2*. READ1 is highlighted for all three species. Note the relative conservation of the flanking sequence compared to READ1.

SK-N-MC (ATCC HTB-10). Immunoprecipitation with the α-ETV6 antibody showed marked enrichment of the READ1 amplicon over the no-antibody control, but not for the control amplicon derived from the gene encoding β-actin (*ACTB* [MIM 102630]), in both cell lines (Figures 2C and 2D and Tables S6A and S6B). Immunoprecipitation with the positive-control antibody, which recognizes a histone H3 variant abundant in actively transcribing genes, enriched both amplicons, as would be expected from two actively transcribing genes (*DCDC2* and *ACTB*). These results demonstrate that ETV6 binds the READ1 region in vivo in both a human lymphoblastoma and a human neuroepithelioma cell line.

The *DCDC2* Risk Haplotypes Show a Synergistic Genetic Interaction with a Known RD Risk Haplotype in *KIAA0319*

Together with our previous findings, these data implicate READ1 as a regulatory element. Luciferase assays suggest that READ1 is capable of modulating expression from the *DCDC2* promoter, but it might regulate other genes.[14] A three-marker risk haplotype encompassing the 5′ half and upstream sequence of *KIAA0319* has been consistently associated with RD and lowered reading performance.[19,28–30] Expression of *KIAA0319* is lower in several cell lines that have this haplotype than in cell lines that do not.[31] Additionally, expression from the *KIAA0319* promoter in two human neural cell lines (including SK-N-MC) is reduced by the minor allele of a SNP that resides in the *KIAA0319* promoter and that is associated with this haplotype.[32] We therefore questioned whether READ1 might interact genetically with the *KIAA0319* risk haplotype and examined the effect of having both a *DCDC2* risk haplotype (CGCGAG or GACGAG) and the *KIAA0319* risk haplotype on several reading, language, and cognitive measures. Strikingly, subjects positive for risk haplotypes in both genes showed markedly worse mean performance (up to 0.40 SD) on nearly all measures examined (Figure 3A). This reduction in performance in subjects with a risk haplotype in both genes is, for most of the phenotypes examined, greater than the sum of those in subjects with a risk haplotype in only one gene or the other, indicating a synergistic interaction between these two variants. This result corroborates a previous report, which provided statistical evidence that *DCDC2* and *KIAA0319* interact to influence RD risk.[33]

Discussion

Given the remarkable similarity between the human exome and those of other higher primates, it has been hypothesized that rapidly evolving regulatory elements are responsible for the large phenotypic differences that we observe. The ENCODE Consortium's recently published results, which showed most of the noncoding genome to be active and much of the active proportion to be regulatory, lend circumstantial support to this hypothesis.[34] Here, we report evidence of just such a regulatory element affecting reading and language, two exclusively human phenotypes. READ1 appears to have expanded rapidly from gorilla to human, although the sequence flanking it is quite conserved (Figure 1C), and its presence, length, and sequence vary widely among primate species (Figure S2).

READ1 specifically binds ETV6, a transcription factor encoded by a proto-oncogene and also known as TEL

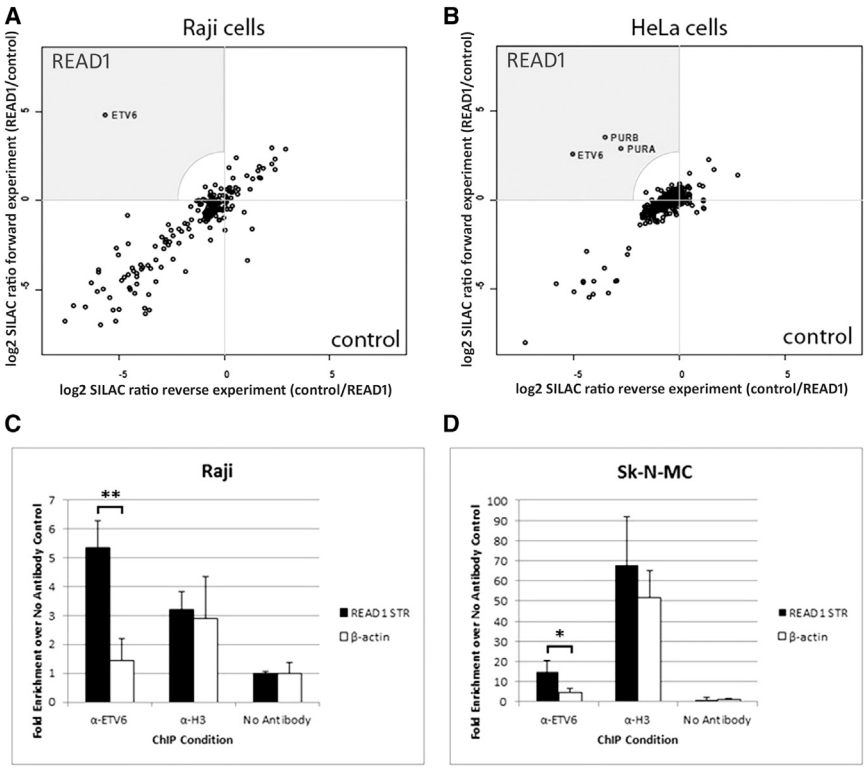

Figure 2. READ1 Binds the Transcription Factor ETV6

(A and B) SILAC results for Raji and HeLa cells and a two-dimensional interaction plot of enrichment for forward and reverse experiments.

(C and D) ChIP results for the Raji (C) and Sk-N-MC (D) cell lines. "α-H3" is the positive-control antibody to a histone H3 variant enriched in actively transcribing genes, and "β-actin" is the control amplicon from *ACTB*, which encodes β-actin. Error bars represent the SD among three replicates. The single asterisk represents a p value below 0.05, and the double asterisks represents a p value below 0.01 (one-tailed t test; see Tables S6A and S6B).

(translocation ETS leukemia). *ETV6* (MIM 600618) is known to be essential for hematopoiesis in bone marrow.[35] It is best known, however, for its tendency to form oncogenic fusions—often with *RUNX1* (MIM 151386)—which are frequently seen in leukemia.[36] Because much of the study of *ETV6* has been focused on these hematopoietic and oncogenic capacities, less is known about its role in other tissues, including the brain. Expression microarray showed *ETV6* to be expressed in both the fetal and the adult human brain (data are accessible at the NCBI Gene Expression Omnibus, accession numbers GDS3113 and GSE7905, see Web Resources),[37] and we confirmed ETV6 accumulation in the cell lines we used for the ChIP experiments in this study by immunoblot analysis (Figure S4). Interestingly, an in situ hybridization study done in mice as part of the Brain Gene Expression Map project[38] (see Web Resources) showed an intriguing pattern of *Etv6* expression in the murine brain. At embryonic day 15, *Etv6* expression appeared specifically upregulated in the ventricular zone, which houses proliferating neuronal precursors that will eventually migrate to the cerebral cortex.[39] At postnatal day 7, this pattern of specifically higher expression shifted to the cerebral cortex. *Etv6* expression then appeared to decrease and become regionally nonspecific in the adult mouse brain. This pattern, which mirrors the developmental expression pattern of the *KIAA0319* mouse ortholog (*D130043K22Rik*),[31] suggests that *Etv6* is expressed in migrating neurons during cortical development. If this is true, and if it is recapitulated in humans, it indicates that ETV6 had an established presence in migratory neurons before READ1 appeared in the genome and that READ1 was thus allowed to effectively commandeer ETV6—that is, to target ETV6 regulation to one or more genes that it did not regulate previously. If these target genes are themselves involved in neuronal migration, like *DCDC2* and *KIAA0319*, this would give READ1 alleles the potential to affect the neuronal migration process.

ETV6's effect on transcription is generally repressive via recruitment of a corepressor complex.[40] Monomeric ETV6 has essentially no affinity for its binding sequence; it must at least dimerize to bind DNA.[41] There is evidence that ETV6 polymerizes in vivo and that the length of the polymer is dependent on the number and spacing of binding sites.[42] This property suggests the possibility that different alleles of READ1 bind ETV6 polymers of different lengths depending on the number of suitably spaced ETV6 binding sites and that these differences change the regulatory power of the complex (Figure 3B). Supporting this idea is the structural similarity of alleles 5 and 6: relative to the most common allele, both have the same GGAA insertion in repeat unit 2 (Table S4). GGAA is the core binding sequence of ETV6,[41] and this insertion could recruit an additional ETV6 monomer to the complex.

However, whether ETV6 represses transcription in this context and which genes it targets are uncertain. Our previously reported luciferase assays appear to indicate that some READ1 alleles activate transcription from the *DCDC2* promoter and that alleles with very different structures (e.g., 3 and 5, Table S4) activate transcription to a similar extent.[14] However, READ1's genetic interaction with the *KIAA0319* risk haplotype and its dramatic effect on phenotype suggest *KIAA0319* as a target gene in vivo. The *KIAA0319* risk haplotype is known to be associated with reduced *KIAA0319* expression, at least in human neuronal cell lines (including SK-N-MC), suggesting that it could segregate with a promoter or promoter-proximal

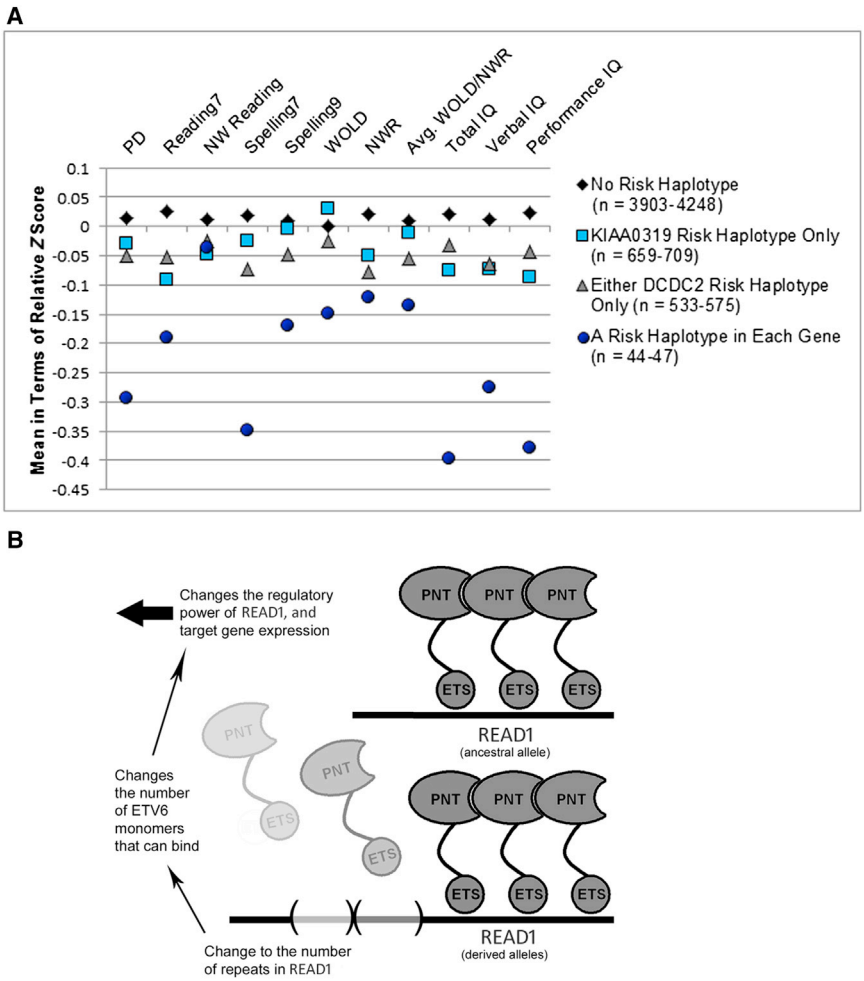

Figure 3. The *DCDC2* Risk Haplotypes Interact Synergistically with the *KIAA0319* Risk Haplotype

(A) Effect of genotype for the *DCDC2* and *KIAA0319* risk haplotypes on various reading, language, and cognitive phenotypes (described in detail in Table S1). Data points represent the mean of each group and were converted to a *Z* score relative to the mean of the ALSPAC sample population. Units of the y axis are fractions of a SD. Abbreviations are as follows: PD, phoneme-deletion task; Reading7, single-word reading at age 7 years; NW Reading, nonword reading at age 9 years; Spelling7 and Spelling9, spelling at ages 7 and 9 years, respectively; WOLD, Wechsler Objective Learning Dimensions verbal comprehension task; and NWR: nonword-repetition task.

(B) Hypothetical model of differential effects of READ1 alleles. ETV6 monomers must at least homodimerize through their pointed (PNT) domains to bind DNA through their ETS domains, and they are thought to homopolymerize in vivo. Indels of READ1 repeat units could change the size of the ETV6 polymer and thus affect target-gene expression.

variant that increases repression (or decreases activation) by READ1 and thereby result in reduced gene expression and possible phenotypic consequences. That we also observed reduced IQ with the *DCDC2-KIAA0319* interaction (Figure 3A) might reflect pleiotropic pathology at the cellular level (e.g., disrupted neuronal migration), or it might simply reflect the importance of language in measuring IQ. READ1 genotyping in all members of the ALSPAC cohort and subsequent combinatorial analysis, together with chromatin-conformation experiments, will further illuminate READ1's mechanism of action.

The effects of the *DCDC2* and *KIAA0319* risk haplotypes on reading and cognitive phenotypes appear to be synergistic. This lends credence to the "phantom-heritability" hypothesis, which explains the so-called missing heritability of continuous traits as resulting from nonadditive interactions between risk variants.[43] Also supporting this idea is the fact that although subjects with the *DCDC2* risk haplotypes showed reduced average performance on phenotypic measures, the SDs for these measures were generally similar to those of subjects negative for these risk haplotypes (Table 2). This implies that the magnitude of effect of the risk haplotypes on phenotype lies on a continuum and is dependent on other, interacting risk variants, as well as environmental and stochastic factors. This is what would be intuitively expected of a polymorphic regulatory element. Confirming or refuting this hypothesis will require much further work, but if it is found to be true, it could add to the inherent complexity regarding the predictive value of genetic variants for continuous traits. Finally, these results might suggest a partial explanation for the missing efficacy of genome-wide association studies (GWASs). If rapidly evolving regulatory elements are indeed substantially responsible for continuous phenotypic variation, they would be expected, like READ1, to show a higher degree of polymorphism than the average SNP. This would make identifying them difficult by standard single-marker analyses in GWASs, reinforcing the importance of multimarker, pathway, and gene-gene interaction analyses in the study of complex traits.

Supplemental Data

Supplemental Data include 4 figures and 11 tables and can be found with this article online at http://www.cell.com/AJHG.

Acknowledgments

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole Avon Longitudinal Study of Parents and Children (ALSPAC) team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers,

receptionists, and nurses. The UK Medical Research Council and the Wellcome Trust (grant 092731) and the University of Bristol provide core support for the ALSPAC. We are also grateful to the Yale Center for Genome Analysis for SNP genotyping services and to Silvia Paracchini for her generous contribution of genotype data for rs4504469, rs2038137, and rs2143340. This research was specifically funded by the National Institutes of Health (grants R01 NS043530 to J.R.G., P50 HD027802 to Y.K. and J.R.G., and F31 DC012270 to J.D.E.). The authors declare competing financial interests: Yale University has applied for a patent covering the complex tandem repeat and deletion in BV677278 (inventor: J.R.G.) and sublicensed it to JS Genetics Inc. J.R.G. is a founder and equity holder of JS Genetics Inc. and is a member of its scientific advisory board.

Received: January 9, 2013
Revised: April 30, 2013
Accepted: May 10, 2013
Published: June 6, 2013

Web Resources

The URLs for data presented herein are as follows:

Association Study Design Server, http://design.cs.ucla.edu
Avon Longitudinal Study of Parents and Children, http://www.bristol.ac.uk/alspac
Brain Gene Expression Map, http://www.stjudebgem.org
Brain Gene Expression Map ETV6 data, http://www.stjudebgem.org//web/view/probe/viewProbeDetails.php?id=2371
Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/
Learning Disabilities: Overview, http://www.nichd.nih.gov/health/topics/learning/Pages/default.aspx#f1
National Assessment of Educational Progress: The Nation's Report Card, Reading 2007, http://nces.ed.gov/nationsreportcard/pubs/main2007/2007496.asp
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org

References

1. Handler, S.M., and Fierson, W.M.; Section on Ophthalmology; Council on Children with Disabilities; American Academy of Ophthalmology; American Association for Pediatric Ophthalmology and Strabismus; American Association of Certified Orthoptists. (2011). Learning disabilities, dyslexia, and vision. Pediatrics *127*, e818–e856.
2. Peterson, R.L., and Pennington, B.F. (2012). Developmental dyslexia. Lancet *379*, 1997–2007.
3. Pennington, B.F., and Bishop, D.V. (2009). Relations among speech, language, and reading disorders. Annu. Rev. Psychol. *60*, 283–306.
4. Scerri, T.S., and Schulte-Körne, G. (2010). Genetics of developmental dyslexia. Eur. Child Adolesc. Psychiatry *19*, 179–197.
5. Meng, H., Hager, K., Held, M., Page, G.P., Olson, R.K., Pennington, B.F., DeFries, J.C., Smith, S.D., and Gruen, J.R. (2005). TDT-association analysis of EKN1 and dyslexia in a Colorado twin cohort. Hum. Genet. *118*, 87–90.
6. Peschansky, V.J., Burbridge, T.J., Volz, A.J., Fiondella, C., Wissner-Gross, Z., Galaburda, A.M., Lo Turco, J.J., and Rosen, G.D. (2010). 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 *20*, 884–897.
7. Poelmans, G., Buitelaar, J.K., Pauls, D.L., and Franke, B. (2011). A theoretical molecular network for dyslexia: integrating available genetic findings. Mol. Psychiatry *16*, 365–382.
8. Liu, J.S. (2011). Molecular genetics of neuronal migration disorders. Curr. Neurol. Neurosci. Rep. *11*, 171–178.
9. Velayos-Baeza, A., Levecque, C., Kobayashi, K., Holloway, Z.G., and Monaco, A.P. (2010). The dyslexia-associated KIAA0319 protein undergoes proteolytic processing with gamma-secretase-independent intramembrane cleavage. J. Biol. Chem. *285*, 40148–40162.
10. Meng, H., Smith, S.D., Hager, K., Held, M., Liu, J., Olson, R.K., Pennington, B.F., DeFries, J.C., Gelernter, J., O'Reilly-Pol, T., et al. (2005). DCDC2 is associated with reading disability and modulates neuronal development in the brain. Proc. Natl. Acad. Sci. USA *102*, 17053–17058.
11. Ludwig, K.U., Schumacher, J., Schulte-Körne, G., König, I.R., Warnke, A., Plume, E., Anthoni, H., Peyrard-Janvid, M., Meng, H., Ziegler, A., et al. (2008). Investigation of the DCDC2 intron 2 deletion/compound short tandem repeat polymorphism in a large German dyslexia sample. Psychiatr. Genet. *18*, 310–312.
12. Marino, C., Meng, H., Mascheretti, S., Rusconi, M., Cope, N., Giorda, R., Molteni, M., and Gruen, J.R. (2012). DCDC2 genetic variants and susceptibility to developmental dyslexia. Psychiatr. Genet. *22*, 25–30.
13. Wilcke, A., Weissfuss, J., Kirsten, H., Wolfram, G., Boltze, J., and Ahnert, P. (2009). The role of gene DCDC2 in German dyslexics. Ann. Dyslexia *59*, 1–11.
14. Meng, H., Powers, N.R., Tang, L., Cope, N.A., Zhang, P.X., Fuleihan, R., Gibson, C., Page, G.P., and Gruen, J.R. (2011). A dyslexia-associated variant in DCDC2 changes gene expression. Behav. Genet. *41*, 58–66.
15. Cope, N., Eicher, J.D., Meng, H., Gibson, C.J., Hager, K., Lacadie, C., Fulbright, R.K., Constable, R.T., Page, G.P., and Gruen, J.R. (2012). Variants in the DYX2 locus are associated with altered brain activation in reading-related brain regions in subjects with reading disability. Neuroimage *63*, 148–156.
16. Golding, J., Pembrey, M., and Jones, R.; ALSPAC Study Team. (2001). ALSPAC—the Avon Longitudinal Study of Parents and Children. I. Study methodology. Paediatr. Perinat. Epidemiol. *15*, 74–87.
17. Boyd, A., Golding, J., Macleod, J., Lawlor, D.A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S., and Davey Smith, G. (2013). Cohort Profile: the 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. Int. J. Epidemiol. *42*, 111–127.
18. Han, B., Kang, H.M., Seo, M.S., Zaitlen, N., and Eskin, E. (2008). Efficient association study design via power-optimized tag SNP selection. Ann. Hum. Genet. *72*, 834–847.
19. Scerri, T.S., Morris, A.P., Buckingham, L.L., Newbury, D.F., Miller, L.L., Monaco, A.P., Bishop, D.V., and Paracchini, S. (2011). DCDC2, KIAA0319 and CMIP are associated with reading-related traits. Biol. Psychiatry *70*, 237–245.
20. Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics *21*, 263–265.
21. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome

association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
22. Mittler, G., Butter, F., and Mann, M. (2009). A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. Genome Res. *19*, 284–293.
23. Butter, F., Kappei, D., Buchholz, F., Vermeulen, M., and Mann, M. (2010). A domesticated transposon mediates the effects of a single-nucleotide polymorphism responsible for enhanced muscle growth. EMBO Rep. *11*, 305–311.
24. Wu, K.K. (2006). Analysis of protein-DNA binding by streptavidin-agarose pulldown. Methods Mol. Biol. *338*, 281–290.
25. Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. *11*, 1475–1489.
26. Bishop, D.V., North, T., and Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: evidence from a twin study. J. Child Psychol. Psychiatry *37*, 391–403.
27. Newbury, D.F., Winchester, L., Addis, L., Paracchini, S., Buckingham, L.L., Clark, A., Cohen, W., Cowie, H., Dworzynski, K., Everitt, A., et al. (2009). CMIP and ATP2C2 modulate phonological short-term memory in language impairment. Am. J. Hum. Genet. *85*, 264–272.
28. Francks, C., Paracchini, S., Smith, S.D., Richardson, A.J., Scerri, T.S., Cardon, L.R., Marlow, A.J., MacPhie, I.L., Walter, J., Pennington, B.F., et al. (2004). 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. *75*, 1046–1058.
29. Luciano, M., Lind, P.A., Duffy, D.L., Castles, A., Wright, M.J., Montgomery, G.W., Martin, N.G., and Bates, T.C. (2007). A haplotype spanning KIAA0319 and TTRAP is associated with normal variation in reading and spelling ability. Biol. Psychiatry *62*, 811–817.
30. Paracchini, S., Steer, C.D., Buckingham, L.L., Morris, A.P., Ring, S., Scerri, T., Stein, J., Pembrey, M.E., Ragoussis, J., Golding, J., and Monaco, A.P. (2008). Association of the KIAA0319 dyslexia susceptibility gene with reading skills in the general population. Am. J. Psychiatry *165*, 1576–1584.
31. Paracchini, S., Thomas, A., Castro, S., Lai, C., Paramasivam, M., Wang, Y., Keating, B.J., Taylor, J.M., Hacking, D.F., Scerri, T., et al. (2006). The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. Hum. Mol. Genet. *15*, 1659–1666.
32. Dennis, M.Y., Paracchini, S., Scerri, T.S., Prokunina-Olsson, L., Knight, J.C., Wade-Martins, R., Coggill, P., Beck, S., Green, E.D., and Monaco, A.P. (2009). A common variant associated with dyslexia reduces expression of the KIAA0319 gene. PLoS Genet. *5*, e1000436.
33. Harold, D., Paracchini, S., Scerri, T., Dennis, M., Cope, N., Hill, G., Moskvina, V., Walter, J., Richardson, A.J., Owen, M.J., et al. (2006). Further evidence that the KIAA0319 gene confers susceptibility to developmental dyslexia. Mol. Psychiatry *11*, 1085–1091, 1061.
34. Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.
35. Wang, L.C., Swat, W., Fujiwara, Y., Davidson, L., Visvader, J., Kuo, F., Alt, F.W., Gilliland, D.G., Golub, T.R., and Orkin, S.H. (1998). The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow. Genes Dev. *12*, 2392–2402.
36. Fuka, G., Kauer, M., Kofler, R., Haas, O.A., and Panzer-Grümayer, R. (2011). The leukemia-specific fusion gene ETV6/RUNX1 perturbs distinct key biological functions primarily by gene repression. PLoS ONE *6*, e26348.
37. Dezso, Z., Nikolsky, Y., Sviridov, E., Shi, W., Serebriyskaya, T., Dosymbekov, D., Bugrim, A., Rakhmatulin, E., Brennan, R.J., Guryanov, A., et al. (2008). A comprehensive functional analysis of tissue specificity of human gene expression. BMC Biol. *6*, 49.
38. Magdaleno, S., Jensen, P., Brumwell, C.L., Seal, A., Lehman, K., Asbury, A., Cheung, T., Cornelius, T., Batten, D.M., Eden, C., et al. (2006). BGEM: an in situ hybridization database of gene expression in the embryonic and adult mouse nervous system. PLoS Biol. *4*, e86.
39. Noctor, S.C., Flint, A.C., Weissman, T.A., Wong, W.S., Clinton, B.K., and Kriegstein, A.R. (2002). Dividing precursor cells of the embryonic cortical ventricular zone have morphological and molecular characteristics of radial glia. J. Neurosci. *22*, 3161–3173.
40. Wang, L., and Hiebert, S.W. (2001). TEL contacts multiple corepressors and specifically associates with histone deacetylase-3. Oncogene *20*, 3716–3725.
41. Green, S.M., Coyne, H.J., 3rd, McIntosh, L.P., and Graves, B.J. (2010). DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. J. Biol. Chem. *285*, 18496–18504.
42. Kim, C.A., Phillips, M.L., Kim, W., Gingery, M., Tran, H.H., Robinson, M.A., Faham, S., and Bowie, J.U. (2001). Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. EMBO J. *20*, 4173–4182.
43. Zuk, O., Hechter, E., Sunyaev, S.R., and Lander, E.S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. Proc. Natl. Acad. Sci. USA *109*, 1193–1198.