Genetic Variation in the KIAA0319 5' Region as a Possible Contributor to Dyslexia

Adrienne Elbert, Maureen W. Lovett, Tasha Cate-Carter, Ashley Pitch, Elizabeth N. Kerr, and Cathy L. Barr

Version Post-Print/Accepted Manuscript


Publisher's Statement The final published version is available at Springer via https://dx.doi.org/10.1007/s10519-010-9434-1.

How to cite TSpace items

Always cite the published version, so the author(s) will receive recognition through services that track citation counts, e.g. Scopus. If you need to cite the page number of the TSpace version (original manuscript or accepted manuscript) because you cannot access the published version, then cite the TSpace version in addition to the published version using the permanent URI (handle) found on the record page.
Genetic Variation in the KIAA0319 5’ Region as a Possible Contributor to Dyslexia

Adrienne Elbert · Maureen W. Lovett · Tasha Cate-Carter · Ashley Pitch · Elizabeth N. Kerr · Cathy L. Barr

Received: 30 September 2010 / Accepted: 15 December 2010 © Springer Science+Business Media, LLC 2010

Abstract Reading disabilities (RD) have been linked and associated with markers on chromosome 6p with results from multiple independent samples pointing to KIAA0319 as a risk gene and specifically, the 5′ region of this gene. Here we focus genetic studies on a 2.3 kb region spanning the predicted promoter, the first untranslated exon, and part of the first intron, a region we identified as a region of open chromatin. Using DNA from probands with RD, we screened for genetic variants and tested select variants for association. We identified 17 DNA variants in this sample of probands, 16 of which were previously reported in public databases and one previously identified in a screen of this region. Based on the allele frequencies in the probands compared to public databases, and on possible functional consequences of the variation, we selected seven variants to test for association in a sample of families with RD, in addition to four variants which had been tested previously. We also tested two markers 5′ of this region that were previously reported as associated. The strongest evidence for association was observed with alleles of the microsatellite marker located in the first untranslated exon and haplotypes of that marker. These results support previous studies indicating the 5′ region of the KIAA0319 gene as the location of risk alleles contributing to RD.

Keywords Dyslexia · Genetics · Reading Disabilities · KIAA0319 · Transcription

Introduction

Reading Disabilities (RD), or Developmental Dyslexia, is a complex disorder that results from genetic and environmental factors. RD is characterized by difficulties with accurate and/or fluent word recognition and by poor spelling and decoding abilities and is estimated to impair 5–12% of school-aged children in North America (Shaywitz et al. 1999). The National Institute of Child Health and Human Development (NICHD), as reviewed in Grigorenko et al. (2006), describes RD as resulting from deficits in the cognitive constituents: phonological awareness, phonological coding, orthographic coding and rapid serial naming.

The heritability of reading difficulties ranges from 30 to 72% (DeFries and Gillis 1993; Wadsworth et al. 2000; Harlaar et al. 2005; Hawke et al. 2006; Wadsworth et al. 2007). Genes influencing reading are distributed throughout the genome, with evidence for linkage of RD, reading skills or related phenotypes reported to at least 10 chromosome regions (Paracchini et al. 2007; Scerri and Schulte-Korne 2010). Of these, the chromosome 6p region has been the most studied with linkage to RD or reading measures reported in multiple samples (Smith and Kimberling 1991; Cardon et al. 1994; Grigorenko et al. 1997; Fisher et al. 1999; Gayan et al. 1999; Grigorenko et al. 2000). Early association studies in the 6p RD linkage region have refined the candidate locus to a
region containing five brain-expressed genes: neurensin-1 (NRSN1), doublecortin domain containing 2 (DCDC2), KIAA0319, TRAF and TNF associated protein (TTRAP) and thioesterase superfamily 2 (THEM2) (Kaplan et al. 2002; Turic et al. 2003; Deffenbacher et al. 2004). Two of the candidate genes, KIAA0319 and DCDC2, have since shown the strongest evidence for association to RD and reading measures (Francks et al. 2004; Cope et al. 2005; Meng et al. 2005; Harold et al. 2006; Schumacher et al. 2006; Luciano et al. 2007; Paracchini et al. 2008; Couto et al. 2009a, b; Dennis et al. 2009; Wileke et al. 2009).

Both KIAA0319 and DCDC2 have roles in neuronal migration (DCDC2: (Meng et al. 2005; Burbridge et al. 2008); KIAA0319: (Paracchini et al. 2006; Peschansky et al. 2010)) and thus are attractive candidates based on previous evidence from post-mortem studies in individuals with RD that revealed the presence of ectopias and dysplasias indicative of neuronal migration abnormalities (Galaburda et al. 1985; Humphreys et al. 1990). Therefore, DNA variation in developmental genes governing the process of neuronal migration has been suggested as a mechanism contributing to RD (Galaburda et al. 1985).

Previous studies supporting KIAA0319 as a RD risk gene have found the most significant association to markers in the 5' end of KIAA0319, suggesting that the RD causal DNA variant(s) may have a role in regulating gene expression (Francks et al. 2004; Cope et al. 2005; Harold et al. 2006). This is consistent with Francks et al. (2004), who found no evidence for a protein-coding change in KIAA0319 that accounted for the association to RD and the finding of a decrease in KIAA0319 expression in cell lines carrying a risk haplotype (Paracchini et al. 2006). This haplotype spans a 77 kb region across KIAA0319 and adjacent gene TTRAP, represented by the alleles G-C-C of markers rs4504469, rs2038137 and rs2143340. Recently, the minor alleles (A) of rs9461045 and (G) of rs3212236 were shown to alter reporter gene expression from a luciferase construct in neuronal and non-neuronal cell lines (Dennis et al. 2009). Both the A allele of rs9461045 and the G allele of rs3212236 were associated with RD and multiple measures of reading ability and reside on the previously identified RD risk haplotype (Francks et al. 2004). The G allele of rs9461045 was shown to create a binding site for the transcriptional repressor OCT-1 (Dennis et al. 2009). Thus this allele was suggested as the risk allele contributing to RD.

However, there are inconsistencies in the KIAA0319 haplotype association findings across studies. Cope et al. (2005) could not replicate the association to the G-C-C haplotype, and instead identified a RD risk haplotype G-A, consisting of different KIAA0319 markers, rs4504469 with rs6935076. Luciano et al. (2007) also tested haplotypes consisting of the markers rs4504469, rs2038137, and rs2143340, and found evidence for association of the haplotypes G-C-T and A-C-T to poor reading, but not G-C-C. In Couto et al. (2009a, b), the RD risk haplotype described by Francks et al. (2004) showed no association with RD, and instead the haplotype A-A-T showed association to RD. This is inconsistent with both Francks et al. (2004) and Cope et al. (2005), in which the A-A-T haplotype was associated with better reading performance. These inconsistencies suggest that the risk variant resides on multiple haplotypes or there may be multiple RD causal variants residing on different haplotypes.

To identify RD risk variants in KIAA0319, we investigated changes occurring in the gene's regulatory elements. Regions of open chromatin mark regulatory elements and DNA variation within these regions may alter gene expression regulation. One way to predict the location of putative DNA regulatory elements is by mapping the locations of DNA-bound regulatory proteins or modified histones by chromatin immunoprecipitation (ChIP) (Ren et al. 2000). The N-terminal residues of histones 3 and 4 are exposed from the nucleosome and subject to modifications such as methylation, acetylation, ubiquitination and phosphorylation, resulting in a ‘histone code’ that facilitates or impedes the access of transcription factors to the DNA (Strahl and Allis 2000; Hon et al. 2009).

Previously, we identified a putative regulatory element in KIAA0319 which may harbour regulatory variants (Couto et al. 2009a, b). The 2.7 kb acetylated histone peak that spans the promoter and first untranslated exon, and part of intron 1 of KIAA0319 was identified by chromatin immunoprecipitation (ChIP) combined with detection by microarrays (ChIP-chip) using an antibody recognizing H3Ac at K9/K14 in a human retinoblastoma cell-line Y79. H3K9ac and H3K14ac are well-studied markers of accessible chromatin, which often contain regulatory elements (Liang et al. 2004; Bernstein et al. 2005; Lim et al. 2005; Kratz et al. 2010). This acetylated histone peak in KIAA0319 was also observed in six mouse tissues (Couto et al. 2009a, b), and as a 2.3 kb peak in two additional human cell lines, a neuroblastoma SH-SY5Y and a glioblastoma, A-172 by ChIP-sequencing (unpublished).

Interestingly, this region is at the centre of KIAA0319 association findings and overlaps four markers previously associated with RD: JAA4 (Kaplan et al. 2002), rs2235676 (Francks et al. 2004), rs9477247 (Francks et al. 2004) and rs2038137 (Francks et al. 2004; Harold et al. 2006). In addition, the marker rs6935076, which is associated to RD in two samples (Cope et al. 2005; Couto et al. 2009a, b) is less than 600 bp away from the acetylated histone region, and therefore may be in linkage disequilibrium with a regulatory variant therein.

The purpose of the current study was to screen this acetylated histone peak region identified in KIAA0319 for
putative functional variants. Our aim was to identify putative RD causal variant(s) that could better explain haplotype associations from previous RD studies. We also tested the previously reported functional variants rs9461045 and rs3212236 for association in our RD sample.

Methods

Participants and assessment

The recruitment and assessment of families participating in this study have been previously described (Luca et al. 2007; Couto et al. 2008, 2009a, b). Briefly, subjects with reading difficulties (aged 6–16) and their families (including siblings aged 6–16) were recruited from local schools in southern Ontario with the majority from the greater Toronto area. Children were excluded if there was evidence of neurological or chronic illness, psychosis, bipolar affective disorder, Tourette syndrome or chronic tics. Children were also excluded if scores on the Performance and Verbal Scales of Wechsler Intelligence Scale for Children (WISC) III were below 80. To qualify, probands had to have English as their first language, or to have been in an English-education system for at least 5 years.

The reading measures used in this study have been described previously (Wigg et al. 2004; Luca et al. 2007; Couto et al. 2008, 2009a, b) and the distribution of reading scores for the first 232 families of this study can be found in Luca et al. (2007). Information about behaviour and possible symptoms of neurological, medical, and psychiatric disorders were obtained using a structured interview with parents (Children’s Interview for Psychiatric Syndromes) (Weller et al. 2000) and a semi-structured interview with teachers (TTI) (Tannock et al. 2002), supplemented with standardized questionnaires (Conners Parent and Teacher Rating Scales Revised (Conners 1997) and Ontario Child Health Survey Scales-Revised (Boyle et al. 1993)). Single-word reading, phonological decoding skills, and spelling were assessed using the third edition of the Wide Range Achievement Test (WRAT-3), the Woodcock Reading Mastery Tests-Revised (WRMT-R), and the Woodcock Reading Mastery Tests-Revised (WRMT-R)—Word Attack subtest, and (iii) WRMT-R—Word Identification subtest. Subjects were considered to be affected with categorical RD if they scored (i) 1.5 standard deviations below average on two of the three tests, or (ii) 1.0 standard deviation below average on all three tests. Of the 352 families recruited to the study, 156 probands and 25 siblings met these criteria and were used in the analyses described in this study. These criteria identified a subset of individuals in our sample whose scores on the three core reading measures, on average, fall within the lower 5% tail of normally distributed reading ability in the general population.

The majority of the families recruited (68.1%) claim European ancestry. Another 26% of the parents describe their ancestry as Canadian Caucasian (no ethnicity specified), 1.8% as Latin American, 2.9% as non-European, and 1.2% as mixed non-European. The study was approved by the research ethics board of the Hospital for Sick Children and all participants provided written consent to participate.

Screening of putative regulatory element by PCR and DNA sequencing

DNA from subjects, siblings and parents was extracted from peripheral blood lymphocytes using a high salt extraction method as previously described (Miller et al. 2004). Forty-eight Caucasian probands that met the categorical criteria for RD were used to screen the 5’ region of open chromatin identified by Couto et al. (2009a, b). This region spans chr6: 24752922–24755657 (Mar. 2006 build NCBI36/hg18) and was screened using four sets of overlapping primers (Supplementary Table S1).

PCR reactions were performed in a total volume of 20 μl, with 60 ng of genomic DNA, 50 ng of each primer (forward and reverse), 2 μl of PCR buffer, 0.2 mM dNTP, 1.5 mM MgCl2 and 5 U of Taq polymerase. PCR thermal cycling stages were as follows: (i) 94°C for 6 min (initial denaturing of DNA), (ii) 35 cycles of 94°C for 20 s, X°C for 20 s, and 72°C for 45 s (where X°C is the annealing temperature of the specific primers), (iii) 72°C for 5 min (extension).

Prior to DNA sequencing, 5 μl of the PCR reaction was treated with exonuclease I (0.2 μl, 20 U/μl) (New England Biolabs, NEB, Ipswich, MA) and Antarctic phosphatase (0.55 μl, 5 U/μl) (NEB) in 0.55 μl H2O, for 15 min at 37°C, then for 15 min at 80°C.

Polymorphism genotyping

SNP genotyping was conducted using TaqMan assays from Applied Biosystems Inc. (ABI, Foster City, CA)
(Supplementary Tables S2 and S3). The PCR reaction mix for the TaqMan assays contained: 30 ng of DNA, 2.5 μl of TaqMan Genotyping PCR Master Mix, and 0.05 μl of allelic discrimination mix (which contained 36 μM of each primer and 8 μM of each probe) in a total reaction volume of 5 μl. The PCR thermal cycling stages were as follows: (i) 50°C for 2 min, (ii) 95°C for 10 min, then (iii) 60 cycles of: 95°C for 15 s, and 59°C for 1 min. The generated genotype data was interpreted by the ABI 7900-HT Sequence Detection System (DS3), using the allelic discrimination end-point analysis mode of the SDS software (version 2.0; Applied Biosystems).

The microsatellite marker, JA04, was amplified using a fluorescently labeled forward primer (Supplementary Table S4), then genotyped using the Applied Biosystems AB3730 DNA Analyzer. Genotyping results were analyzed with the GeneMapper software (Applied Biosystems, Foster City, CA).

284 Statistical tests

The TDT statistic in Haplovip (v.4.0) (Barrett et al. 2005) was used to test for association of biallelic markers, while the Extended-TDT (ETDT) was used for the analysis of the multi-allelic microsatellite (Sham and Curtis 1995). The Haplovip program (v.4.0) was also used to analyze and visualize linkage disequilibrium (LD) blocks using the Gabriel et al. algorithm (Gabriel et al. 2002; Barrett et al. 2005). Haplotype transmission was analyzed using the TRANSMIT program (Clayton and Jones 1999). Merlin was used to check the genotype data for unlikely crossover events indicating erroneous genotypes (Abecasis et al. 2002). This data set was free of any detectable Mendelian errors and there were no crossovers between markers. None of the markers used in the analyses deviated from Hardy–Weinberg equilibrium (HWE).

We estimated the power of the sample using the genetic power calculator available at http://pngu.mgh.harvard.edu/~purcell/cgi-bin/dtdt.cgi. Since the assumption of the screening studies in KIAA0319 was that we are testing putative causative polymorphisms, the test marker was assumed to be the causative marker (identical allele frequencies, in complete LD). Assuming an additive or dominant inheritance model, our family-based RD sample has acceptable power (0.8) to detect a genotype relative risk of 2.0 for a causative marker with an allele frequency ranging between 0.13 and 0.27.

The program Evolutionary-based Haplotype Analysis Package (eHAP, Seltman et al. 2003) was used to determine the phylogenetic relationship between haplotypes based on the genotype data. This tool was used to build a cladogram (Fig. 1).

![Fig. 1 Phylogenetic tree of polymorphisms in the KIAA0319 5'-region. Phylogenetic analysis determined the evolutionary relationship between the haplotypes, consisting of the markers rs6935076 (G/A), rs2038137 (C/A), rs71815143 (ins/del), rs3756821 (G/A) and rs9461045 (G/A). The ancestral haplotype (G-C-del-G-G) was selected based on chimpanzee genomic sequence. The risk haplotype (G-A-ins-G-G) is derived from a rare haplotype G-A-del-G-G (freq = 0.02). The range of JA04 allele lengths present on the haplotypes are also depicted. Only the risk haplotype carries the longer alleles (24–30GT). The risk allele of JA04 (25GT) and of rs71815143 (ins) are only present on the risk haplotype.](image-url)
that could explain the evidence for association of this gene to dyslexia. To identify polymorphisms, a region of KIAA0319 previously identified as a region of open chromatin spanning the promoter, first untranslated exon and part of intron 1 (chr6: 24753560–24755536, Mar. 2006) build NCBI36/hg18) was sequenced using overlapping PCR products amplified from 48 Caucasian RD probands. Seventeen polymorphisms were observed (Table 1), consisting of 14 annotated SNPs reported in public databases, one SNP previously reported in a screen of this region (Dennis et al. 2009), a microsatellite, and a 26 bp deletion (rs71815143). This 26 bp deletion consisted of a repetitive

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Summary of identified polymorphisms, predicted TFBS and TDT results (excluding JA04)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polymorphism (UCSC Mar 2006)</td>
<td>Alleles (strand)</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>rs1883592</td>
<td>G 0.41</td>
</tr>
<tr>
<td>chr6: 24755208</td>
<td>A 0.59</td>
</tr>
<tr>
<td>rs9467247</td>
<td>G 0.78</td>
</tr>
<tr>
<td>chr6: 24755198</td>
<td>T 0.22</td>
</tr>
<tr>
<td>rs1883593</td>
<td>T 0.96</td>
</tr>
<tr>
<td>chr6: 24755170</td>
<td>A 0.04</td>
</tr>
<tr>
<td>rs1883594</td>
<td>C 0.29</td>
</tr>
<tr>
<td>chr6: 24755163</td>
<td>G 0.71</td>
</tr>
<tr>
<td>rs35882364 and rs35853270</td>
<td>TT 0.36</td>
</tr>
<tr>
<td>chr6: 24755133</td>
<td>T 0.64</td>
</tr>
<tr>
<td>rs3756821</td>
<td>G 0.48</td>
</tr>
<tr>
<td>chr6: 24754800</td>
<td>A 0.52</td>
</tr>
<tr>
<td>rs6456625</td>
<td>G 0.33</td>
</tr>
<tr>
<td>chr6: 24754483</td>
<td>A 0.67</td>
</tr>
<tr>
<td>rs71815143</td>
<td>Insertion 0.33</td>
</tr>
<tr>
<td>chr6: 24754499–24754524</td>
<td>Deletion 0.67</td>
</tr>
<tr>
<td>rs3756821</td>
<td>Deletion 0.67</td>
</tr>
<tr>
<td>rs62400773</td>
<td>G 1.00</td>
</tr>
<tr>
<td>chr6: 24754176</td>
<td>A 0.0</td>
</tr>
<tr>
<td>rs2038135</td>
<td>A 0.38</td>
</tr>
<tr>
<td>rs6456625</td>
<td>C 0.62</td>
</tr>
<tr>
<td>rs2038136</td>
<td>C 0.40</td>
</tr>
<tr>
<td>chr6: 24753926</td>
<td>G 0.60</td>
</tr>
<tr>
<td>rs2038137</td>
<td>G 0.38</td>
</tr>
<tr>
<td>chr6: 24753922</td>
<td>C 0.62</td>
</tr>
<tr>
<td>rs2235676</td>
<td>C 0.94</td>
</tr>
<tr>
<td>chr6: 24753676</td>
<td>A 0.06</td>
</tr>
<tr>
<td>rs13220596</td>
<td>G 0.98</td>
</tr>
<tr>
<td>chr6: 24753654</td>
<td>A 0.02</td>
</tr>
<tr>
<td>rs2235677</td>
<td>T 0.38</td>
</tr>
<tr>
<td>chr6: 24753859</td>
<td>C 0.62</td>
</tr>
</tbody>
</table>

Brain-relevant transcription factors are marked by asterisks

‘Couto’ indicates polymorphisms genotyped in a previous study by Couto et al. (2009a, b)
sequence with a periodicity of 13 bp. The SNP rs6456625 (C/T), and the unannotated SNP (C/T; chr6: 24754493 (Mar 2006 NCBI36/hg18)) occurred within this deletion polymorphism.

Allele frequencies for the polymorphisms were calculated from the sequencing data in the 48 RD probands. Allele frequencies were available for seven of these polymorphisms in dbSNP (Table 1). Four of the SNPs had previously been tested for association in our RD sample rs9467247, rs2038137, rs2223677 and rs22235676 (Couto et al. 2009a, b), and showed no significant association with RD (Table 1). Of the remaining DNA variants, markers were selected for genotyping in the RD family sample, if there was (i) a minor allele frequency greater than 0.05 (rs1883593, rs62400773, and rs13220596 excluded), (ii) a difference in allele frequencies between probands and those reported in dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/) for CEU, (Caucasian individuals with Northern and Western European ancestry from the CEPH collection) (rs2038135 and rs2038136 not selected for genotyping). The marker rs2038135 was also in strong LD with rs2038137, which was not significantly associated with RD.

The SNPs rs35882364/rs35853270 were also not selected for genotyping. The SNPs rs35882364 and rs35853270 are adjacent single thymidine deletions and there were no allele frequencies available in the database for these SNPs. It is unclear whether rs35882364/rs35853270 represent two SNPs (as reported in the database), or one SNP. In the screened probands, the sequence at this locus was observed as “TT” or “T0”, where ‘0’ represents a deleted T. Since the “00” genotype was not observed, it is possible that one of the two SNPs is not polymorphic, or that both SNPs have low MAF. Therefore, no genotyping assay could be designed and these SNPs were not tested for association in the RD family sample. Moreover, there were no allele-specific TFBS predicted changes for these polymorphisms.

Since the functional risk allele may affect transcription factor binding affinity, the TFBS changes for all of the polymorphisms were predicted in silico using TESS. It is important to note that TESS utilizes (1) databases containing experimentally validated TFBS sequences to predict the TFBS, and (2) tools to evaluate the significance of the predicted sites, but that neither of these allow for total accuracy in TFBS predictions. Some TFBS will be missed, due to the incompleteness of experimentally validated TFBS sequences, while other TFBS will be predicted, but not be biologically relevant. To assess which TFBS are most likely to be biologically functional, the transcription factors were filtered for function/expression in the brain (Table 1). Brain-relevant TFBS changes were predicted to occur for most polymorphisms including: Sp1, HOX5, FOXL1, FOXC1, C/EBP, POU5F1, Stat6, s-Ets-2, RAR, AP-1 and GATA2 (Table 1). Of the polymorphisms selected for genotyping, several were of particular interest because of the predicted change in TFBS. These include the HOXA5 binding site predicted for the minor allele (C) of rs1883594, but not the major allele (G). HOXA5 is a transcriptional regulator during development and is expressed early in embryonic neurons (Joksimovic et al. 2005). The insertion allele of rs71815143 is predicted to contain three Sp1 binding sites and a CCAAT/enhancer binding protein (C/EBP) binding site, which are not predicted for the deletion allele. Sp1 is a ubiquitous transcriptional activator, except in the CNS, where it is expressed exclusively in early forebrain neurons (Mao et al. 2009). C/EBP recruits transcriptional activators to enhancer and promotor regions (Culella et al. 2007).

Association analysis

Seven polymorphisms within the screened region of KIAA0319 (rs1883592, rs1883594, rs3756821, rs71815143, rs6456625 and 6: 24754493 in Table 1, and JA04 in Table 2), along with two SNPs outside the region (rs9461045 and rs3212236 in Table 3) were tested for association to categorical RD as single markers in the entire sample of families meeting the criteria (n = 156 nuclear families including 25 affected siblings). The markers rs9461045 and rs3212236 were identified as functional RD risk variants in a previous study (Dennis et al. 2009).

The most significant evidence for association was with the marker JA04, in particular we observed biased transmission of allele 14 (25GT) ($\chi^2 = 7.367, p = 0.0067$). The A allele of rs1883592, the G allele of rs3756821 and the insertion allele of rs71815143 showed trends for biased transmission ($p = 0.081, p = 0.080$ and $p = 0.061$ respectively) to RD probands, whereas the marker rs1883594 and 6: 24754493 showed no evidence for association (Table 1). The markers rs9461045 and rs3212236, previously reported as associated by Dennis et al. (2009), showed no evidence for association to RD in our sample (Table 3). The p-values presented have not been corrected for multiple testing.

Haplotype analysis

We examined haplotypes of the polymorphisms tested in this study with additional markers to determine if these combinations could explain the previous reported haplotype findings, and further to determine if the risk is related to combinations of alleles as opposed to single alleles. The JA04 microsatellite was selected for haplotype analysis because it showed the strongest statistical evidence for association to RD and was predicted to function in gene
transcription regulation. Although rs71815143 did not show strong statistical evidence for association to RD, this variant was included in haplotype analysis because of the predicted effects of this deletion on TFBS. No haplotype combinations were analyzed other than the ones presented here. Transmit TDT analysis was used to test whether the alleles of JA04 and rs71815143 contributed to the previously identified RD risk haplotypes generated by the markers rs4504469, rs2038137 and rs2143340 (Francks haplotype) and rs4504469 and rs6935076 (Cope haplotype), which were genotyped in our RD sample in a previous study (Couto et al. 2009a, b). The results of the haplotype analyses can be found in Tables 4 and 5.

Results showed that the associated alleles of JA04 and rs71815143 contributed only to one Francks haplotype (Table 4). This haplotype consisted of the alleles A-A-T, which had a trend for association in a previous study (p = 0.091, Couto et al. unpublished). Including the two markers JA04 and rs71815143 in this haplotype (A-A-25GT-ins-T) increased the evidence for association of this haplotype to RD (p = 0.0093).

Couto et al. (2009a, b) had previously identified a risk haplotype (A-G; p = 0.018) consisting of the markers reported by Cope et al. (2005). Including the markers JA04 and rs71815143 (A-G-25GT-ins) increased the evidence for association of this haplotype (p = 0.0010) (Table 5). This was the only haplotype carrying the alleles biased in transmission in the single marker analyses of JA04 and rs71815143. The haplotype consisting of alternative alleles (G-A-11GT-del) showed a trend for under-transmission to RD (p = 0.08).

The LD across the markers in the 5' region was determined in our RD family-based sample using Haploview (version 4.0). The markers rs9461045 and rs3212236 which are 5' of KIAA0319, and the marker rs6935076, which is located 600 bp distal to the acetylated H3 peak in the first intron, were included in the analysis. The marker rs6935076 was associated in this RD sample in a previous study (Couto et al. 2009a, b). Table 6 shows the strong LD across the region, thus the region can be analyzed as one LD block.

To determine whether a combination of alleles across the putative regulatory element showed association with RD, haplotype analysis was performed for markers in this LD block. Haploview identified tag-markers: rs2038137, rs3756821, and rs9461045. In addition, we included the markers with evidence for association: rs6935076, JA04 and rs71815143. The haplotype that is biased in transmission (p = 0.0013) carries the risk alleles of the individually associated markers (Table 7).

Strong LD as measured by D' across the associated region makes it difficult to determine which markers distinguish the risk haplotype from unassociated haplotypes.
Therefore, eHap was used to determine the phylogenetic relationships between the polymorphisms in the affected RD probands. The analyzed haplotypes consisted of the markers: rs6935076, rs2038137, rs71815143, rs3756821 and rs9461045 (Fig. 1).

The results show that the alleles distinguishing the associated haplotype from an ancestral haplotype are the insertion allele of rs71815143 and alleles 13–19 of JA04. However, the ancestral haplotype was only observed at a frequency of 0.02, thus the frequency is too low for association testing. Therefore, it is unclear whether the ancestral haplotype is also associated with RD.

### Discussion

The purpose of this study was to identify putative functional polymorphism(s) that explain the previous association to RD. We hypothesized that causal risk variants in KIAA0319 function by altering transcription, and are located in the region of open chromatin identified by Couto et al. (2009a, b). In the current study, this region was screened, and putative functional variants were selected and tested for association to RD.

Two putative functional variants in the 5' region of KIAA0319 were identified: a microsatellite (JA04) and a
An unannotated polymorphism was identified at 6: 24754493 within the deletion allele of rs71815143. Although the insertion allele of rs71815143 showed a trend for association (p = 0.061), the polymorphism at 6: 24754493 in the deletion allele of rs71815143 showed no association with RD. This SNP was also reported in the Dennis et al. (2009) study, and no evidence for association was reported.

Although rs71815143 was not statistically associated with RD (p = 0.061), TFBS predictions suggest it may have allele-specific affects on gene expression. The risk allele of rs71815143 may function to augment risk for RD by influencing KIAA0319 expression. The risk allele was predicted to bind Sp1 and C/EBP. C/EBP activates gene expression by binding promoters and enhancer elements, and Sp1 is a transcriptional activator known to activate RNA synthesis from selected promoters by RNA polymerase II (Briggs et al. 1986; Landschulz et al. 1988). In addition, Sp1 can bind and activate gene expression from TATA-less promoters (Badran et al. 2005), such as the promoter of KIAA0319.

Allele 14 (25 GT repeats) of the microsatellite showed the most evidence for association to RD (p = 0.0067). This may not replicate the previous finding of Kaplan et al. (2002) who reported association with the most common allele of JA04, which would be 11GT repeats in our sample. However, we cannot be absolutely sure that the most common allele in that sample is the same as the most common in ours because the allele size was not specified in that paper.

Microsatellites have been implicated in the regulation of transcription (Hamada et al. 1984; Naylor and Clark 1990; Gebhardt et al. 1999; Rothenburg et al. 2001; Vinces et al. 2009). The occurrence of these repeats in the 5' region of genes more often than expected by chance has led some to speculate that they are involved in rapid gene evolution and responsiveness to changing environmental conditions (King et al. 1997). There are numerous ways that the microsatellite may increase or decrease gene expression: (i) relief of DNA supercoiling; (ii) altering nucleosome placement; or (iii) forming a joint to bridge the promoter with more distal regulatory elements. Any combinations of these scenarios could be occurring in the KIAA0319 gene. Alternating purine/pyrimidine repeats, such as GT repeats, are conducive to a transition from the normal, B form of DNA, to the Z form. Unlike the B form, Z-DNA double helix is left-handed, contains 12 bp per turn (instead of 10.4 bp) and exists only transiently. Different lengths of GT repeats form different amounts of Z-DNA, at different rates, suggesting that polymorphic dinucleotide repeats may be one of the genetic factors contributing to variation in gene expression. The transition to one helical turn of Z-DNA can relieve 1.8 helical turns of right-handed
Author Proof

...induced Z-DNA transition by inducing superhelical torsion of the nucleosomes, and gene activation. BRG1 binding the GT repeat converted to Z-DNA, disrupting the position of SWI2/SNF2-like chromatin-remodeling complex BAF within a nucleosome. Upon recruitment of BRG1, a subunit of SWI2/SNF2-like chromatin-remodeling complex BAF the GT repeat converted to Z-DNA, disrupting the position of the nucleosomes, and gene activation. BRG1 binding induced Z-DNA transition by inducing superhelical torsion (Liu et al. 2001, 2006).

In this study, the 25GT repeat allele of the microsatellite allele showed evidence for association to RD. However, it is unclear how the 25GT length, and not other lengths, could increase RD risk. Studies on the B-to-Z DNA transition suggest that longer alleles may form longer stretches of Z-DNA and elicit a stronger effect on gene expression. Another possibility is that the 25GT length represents a threshold at which the repeat is long enough to transition to Z-DNA in an energetically favourable manner. However, the TDT results for the microsatellite did not support that longer alleles of the microsatellite are associated with RD.

In spite of their predicted effects on transcription, evidence from the current study suggests that rs71815143 and JA04 are not functional variants that explain the association in other RD samples. This is because neither of these markers could resolve the inconsistency in haplotype associations in previous studies. Instead, it is likely these variants are in strong LD with the causative allele(s).

We expect that the causative allele will segregate onto multiple risk haplotypes. This would resolve the inconsistency in the haplotype findings in the literature, as one explanation is that the functional risk allele segregates on multiple haplotypes at different frequencies, in different RD samples. The associated alleles of the microsatellite (25GT) and rs71815143 (insertion allele) only segregated with the A-A-T (Francks) haplotype and the A-G (Cope) haplotype. These haplotypes have the alternative alleles of those previously identified by the other groups, suggesting that neither of these markers can explain the association finding in the Francks et al. (2004) or Cope et al. (2005) studies.

Luciano et al. (2007) had suggested that the haplotype of the alternative alleles, G-C-T, of the markers used in the Francks haplotype was the major risk haplotype, as some evidence for over-transmission was observed in Luciano et al. (2007), Cope et al. (2005) and Francks et al. (2004). This haplotype carries the non-risk alleles of rs71815143 (deletion) and the microsatellite (11GT and 12GT), and showed no bias in transmission in our RD sample. Therefore, these variants are not likely to be the common functional risk variants in the other RD samples studied.

It is possible that the risk alleles reside outside the region screened here. The Dennis et al. (2009) study identified an RD-associated SNP (rs9461045) that creates a binding site for the transcriptional repressor OCT-1. New elements such as this created by the risk alleles would only be identified in chromatin from tissues or cell lines carrying the risk alleles and with markers of repressed chromatin.

In summary, several putative functional variants were identified including a microsatellite (JA04) and rs71815143, of which JA04 showed evidence for association to RD. The microsatellite may create risk for RD by transitioning from B to Z-DNA, which can increase or decrease gene expression, depending on the gene context. The risk alleles of these two markers both contributed to a risk haplotype in our sample; however this risk haplotype is inconsistent with the haplotypes associated in other RD samples. Therefore, these two putative functional variants are not likely to be the main functional risk variants. The functional variants identified by Dennis et al. (2009) were not associated in our sample, and therefore we find no
support in our families for these alleles as the functional variants contributing to RD.

The $p$-values for the single marker association tests would not withstand a correction for multiple testing. Given the strong LD observed across the $5'$ region of KIAA0319, haplotype analysis is a more appropriate test to determine whether there is evidence for association to RD in this region. The haplotype analysis results suggest that there is evidence for significant haplotype associations to RD, even in light of multiple testing.

We found association for RD with haplotypes of polymorphisms in the $5'$ region. Although the main functional risk variant across different samples has not been identified, the results of this study, and those of others (Paracchini et al. 2006; Dennis et al. 2009) suggest that the functional RD risk variants reside in the $5'$ region and alter gene expression of KIAA0319. It is becoming more evident with complex traits/disorders that functional risk variants modify gene expression, as opposed to changing protein coding sequences (Jais 2005; Nica and Dermitzakis 2008). Studies of KIAA0319 in RD can serve as a model for the study of other complex traits and disorders. Future studies in KIAA0319 will determine the functional effects of these polymorphisms and haplotypes of these alleles on gene expression.

References


Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behav Genet

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behavioral Genetics

Behav Genet


The FBAT Program v1.4.1 http://www.biostat.harvard.edu/~fbat/default.html


