Putative functional alleles of DYX1C1 are not associated with dyslexia susceptibility in a large sample of sibling pairs from the UK

T S Scerri, S E Fisher, C Francks, I L MacPhie, S Paracchini, A J Richardson, J F Stein, A P Monaco

Developmental dyslexia is diagnosed as a specific impairment in reading ability, despite adequate intelligence and educational opportunity, that affects approximately 5% of schoolchildren. Much evidence has been accumulated from twin and family based studies to indicate that dyslexia can have a hereditary basis, but that the genetic aetiology is complex, involving multiple risk factors. Linkage analysis has identified numerous genomic regions that may harbour susceptibility genes influencing dyslexia, including on chromosomes 1, 2, 3, 6, 15, and 18, with varying degrees of reproducibility. The first of these linkages was reported two decades ago, to the centromere of chromosome 15. Although subsequent studies failed to replicate linkage to this specific region, there is evidence for linkage elsewhere on chromosome 15, particularly at 15q21 (DYXI, OMIM 127700). Recently, DYX1C1 (also known as EKN1) was proposed as the gene underlying the putative effect on 15q21. This was initially based on studies of a balanced translocation, t(2;15)(q11;q21), co-segregating with reading problems within a single nuclear family from Finland. The 15q21 breakpoint in this family directly disrupts DYX1C1, in an interval that includes exons 8 and 9 (fig 1). Investigation of DYX1C1 in individuals from 20 additional Finnish families with multiple cases of dyslexia led to the identification of eight single nucleotide polymorphisms (SNPs). Two of these SNPs were found to associate with dyslexia in these families, and in additional Finnish affected cases and controls. It was proposed that these two associated SNPs altered the expression or function of DYX1C1, one by altering a transcription factor binding site, the other as a result of a premature truncation of the protein product by four amino acids, thereby leading to increased risk of developing dyslexia.

In the present study we aimed to investigate whether these allelic variants in DYX1C1 represent important risk factors in common cases of developmental dyslexia. In our ongoing quantitative trait locus (QTL) study of dyslexia, we have collected 264 nuclear sib-pair families from the UK. Each family contains at least one proband with a strict diagnosis of dyslexia, and every proband and co-sibling has been administered a battery of quantitative psychometric tests. These families include the original 89 from our QTL based genome-wide linkage scan for dyslexia, plus a further 84 which have been investigated for replication of linkage to 18p11.2 and an additional 91 which have not been previously included in molecular genetic studies. The large size of our total sample, combined with the use of quantitative measures of reading disability, provides us with a powerful means to identify genetic associations with reading related deficits. Our analyses of DYX1C1 indicate that this gene is unlikely to make a significant contribution to deficits in reading related abilities in this large sample of families. Moreover our pattern of results casts doubt on the hypothesis that the putative functional SNPs which were previously identified lead to increased risk of reading impairment.

METHODS
Family ascertainment
The sample of 264 nuclear families analysed in this study consists of 173 families described in previous reports and an additional set of 91 families similarly ascertained through the dyslexia clinic at the Royal Berkshire Hospital, Reading. The

Key points
- Two variants within the gene DYX1C1, located on chromosome 15q21, have recently been reported to associate with the reading disorder dyslexia in a Finnish population. In the present study we investigated these sequence variants of DYX1C1 for association to dyslexia within a large sibling pair sample collected in the UK.
- Eight sequence variants within DYX1C1 were genotyped in 1154 individuals from 264 nuclear families, each containing at least one proband with dyslexia and one or more siblings. Each proband and their siblings had been administered six quantitative reading related psychometric tests. Quantitative trait association analysis was used to test for a putative effect of DYX1C1 variants.
- Of the eight sequence variants, only one (1249G→T) showed nominally significant association with any of the quantitative measures. The more common allele, 1249G, was associated with poorer performance for an orthographic coding test (p value of 0.0212 without correction for multiple testing), whereas in the Finnish population, the rarer 1249T allele was found to be associated with dyslexia and was thought to be a potentially functional polymorphism.
- We conclude that the DYX1C1 alleles previously associated with dyslexia are not associated with the trait in our sample, and are in fact associated with somewhat better performance on our tests of reading related abilities. This implies that neither of the proposed functional variants of DYX1C1 increases the risk of dyslexia in our UK families.
A test involving reading of irregular words (OC-irreg). We included standard tests of single word reading (READ) and spelling (SPELL), accompanied by tests aimed at measuring a variety of reading and language related cognitive abilities. Phenome awareness (PA), the ability to reflect on and manipulate the separate speech units that make up a word, was assessed via performance on a “spoonerism” task. Phonological decoding (PD), the ability to convert a sequence of written symbols into their corresponding phonemes, was measured with a non-word reading test. Orthographic coding (OC), the ability to recognise orthographic representations of whole words and retrieve appropriate phonological representations from a mental lexicon, was assessed with two complementary tests, a forced choice task (OC-choice) and a test involving reading of irregular words (OC-irreg). We have shown previously\cite{11,12} that these quantitative traits are significantly familial in our UK sample. More details of phenotype measurements, including standardisation methods and descriptive statistics, have been extensively described elsewhere.\cite{10,11}

**Phenotype measurements**

A battery of psychometric tests was administered to each proband, as described in earlier publications.\cite{11,12} These included standard tests of single word reading (READ) and spelling (SPELL), accompanied by tests aimed at measuring a variety of reading and language related cognitive abilities. Phenome awareness (PA), defined as the ability to reflect on and manipulate the separate speech units that make up a word, was assessed via performance on a “spoonerism” task. Phonological decoding (PD), the ability to convert a sequence of written symbols into their corresponding phonemes, was measured with a non-word reading test. Orthographic coding (OC), the ability to recognise orthographic representations of whole words and retrieve appropriate phonological representations from a mental lexicon, was assessed with two complementary tests, a forced choice task (OC-choice) and a test involving reading of irregular words (OC-irreg). We have shown previously\cite{11,12} that these quantitative traits are significantly familial in our UK sample. More details of phenotype measurements, including standardisation methods and descriptive statistics, have been extensively described elsewhere.\cite{10,11}

**Genotyping**

SNPs were genotyped for all individuals (parents and children) using the MassEXTEND assay from Sequenom, according to the manufacturer’s instructions. This technique is based on the analysis of allele specific primer extension products using matrix assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. All individuals were also genotyped for highly polymorphic microsatellite markers by semi-automated fluorescent genotyping techniques, and quality checked as previously described.\cite{10,11} All primer sequences are available on request.

**Data analysis**

Data were checked for possible genotyping errors using MERLIN (0.9.12b).\cite{14} Allele frequencies were tested with PEDSTATS, to ensure that markers were in Hardy-Weinberg equilibrium. MERLIN was then used to construct the most likely haplotypes. Pair wise estimates of D’ were calculated from the founder haplotypes to signify the extent of linkage disequilibrium (LD) between each pair of markers in our sample. Marker-trait associations were tested with the computer program QTDT (2.4.2a)\cite{15} for all quantitative traits to each marker individually, or as part of a haplotype, by utilising information from all siblings in the sample (“total association”). In this analysis, maximum likelihood modelling is first used to fit a null variance components model to quantitative trait data in sibships, in which variance is partitioned into unshared environmental, shared environmental/polygenic, and QT specific components, with a single trait mean. The shared environmental/polygenic component is modelled using the coefficient of relatedness between relatives (0.5 for siblings), and the QT component is modelled using multipoint identity by descent sharing between sibling pairs at the locus. Subsequently, a full association model is fitted that includes the same variance components as the null model, but includes the mean effects of the SNP alleles. The likelihood ratio test statistic is distributed asymptotically as a chi-square with 1 df. This “total association” test can be biased by population stratification, so we also verified our results by performing a direct test of stratification within QTDT.

To limit multiple testing when analysing a multi-allelic marker, QTDT provides an option to calculate a global p value, as well as individual p values for each allele of that marker. Multi-point linkage analysis was performed with the computer package GENEHUNTER (2.1_r2 beta),\cite{16} either under the traditional Haseman-Elston model or a VC framework.

**RESULTS**

We genotyped all 1153 individuals from our entire set of 264 families for the eight SNPs reported by Taipale et al in their Finnish sample (table 1),\cite{8} and of these only six were found to be polymorphic in our sample. The two sequence variants that we did not find to be polymorphic (that is, –2G→A and 4C→T) corresponded to the least frequent within the Finnish sample.

For the six polymorphic SNPs, we were able to acquire experimentally 95% of all possible genotypic data in our sample. Assuming our experimental genotype calls to be correct, a further 2% of genotypes could be inferred with certainty by utilising haplotypes generated by MERLIN. To aid error detection and the generation of haplotypes, and to facilitate linkage and association modelling, multiple
microsatellite markers on chromosome 15 were also genotyped for each individual. These primarily consisted of a core set of four markers flanking each side of DYX1C1 (including D15S1012, D15S132, D15S143, and D15S978 centromeric to DYX1C1, and D15S117, D15S153, D15S131, and D15S130 telomeric to DYX1C1; fig 1). In addition, nine other markers previously genotyped as part of the earlier genome-wide scan were incorporated into the analyses.

We also conducted a similar set of analyses in a subgroup of the 264 UK families. This new subgroup was composed exclusively of families that contained at least one proband whose average psychometric score was less than 1 standard deviation (SD) below the mean of a normative population. This process eliminated 130 of the 264 families. The remaining 134 families, comprising a total of 603 individuals, therefore contained at least one child that on average performed poorly across all quantitative traits. Testing this subgroup for total association across all phenotypes and all six SNPs revealed only a single significant p value of 0.0076 for the OC-choice measure with marker 1249G→T, and again it was the common 1249G allele that was associated with poorer performance, that is, the opposite result to that found by Taipale et al. Association for the −3G→A:1249G→T haplotypic marker was tested in the subgroup and a global p value of 0.0140 was obtained for association with OC-choice. Further analysis yielded p values of 0.0104 and 0.0182 for haplotypes −3G:1249G and −3G:1249T, respectively, with allelic effects in the same directions as for the total sample.

Multi-point linkage analysis performed with GENEHUNTER yielded a peak LOD score of 1.0 (with the PD measure) between markers D15S132 and D15S143 (see fig I for their positions relative to DYX1C1) for the complete sample under the VC framework. No linkage was observed in the subgroup with either the Haseman-Elston or VC methods (all LOD scores <0.5).

**DISCUSSION**

The present study was performed in response to a recent proposal that putative functional variants of DYX1C1 are risk factors for common forms of developmental dyslexia. We wished to determine whether we could find supporting evidence for this by analysing our sample of 264 families,
each containing at least one dyslexic proband and one or more phenotypic siblings. Our large sample, together with our use of a battery of heritable quantitative psychometric measures, has previously yielded sufficient power to detect replicable linkage effects. It is therefore likely that our sample provides sufficient power to detect allelic effects with association analysis, which is generally more powerful than linkage analysis. We find that the proposed functional risk alleles of *DYX1C1* are not associated with dyslexia in our sample. If anything we see the opposite; inheritance of these alleles is correlated with increases, not decreases, in performance on reading and language related tasks (although this relationship is only significant for one measure, and not significant after adjusting for multiple testing). These findings are difficult to reconcile with the hypothesis that these specific variants are truly of functional relevance for dyslexia.

The proposal of *DYX1C1* as a candidate gene for involvement in dyslexia was based on several lines of evidence. Initially, the breakpoint of a translocation t(2;15)(q11;q21) co-segregating with reading problems within a single Finnish family was discovered, and later refined to an interval including exons 8 and 9 of *DYX1C1*. This region of the gene encodes a tetratricopeptide repeat (TPR) domain that might be involved in protein-protein interactions. Taipale et al went on to sequence the 10 exons of *DYX1C1* in 20 unrelated dyslexic individuals and identified eight sequence variants (table 1). We tested all eight of these in our complete UK sample, but found only six to be polymorphic, each being within the range of allele frequencies reported by the Finnish group (table 1).

After comparing allele frequencies in 109 dyslexics and 195 controls from Finland, Taipale et al reported that the −3A and 1249T alleles showed significant associations with dyslexia, both individually (p values of 0.016 and 0.048, after Bonferroni correction, respectively) and as the haplotype −3A:1249T (p value of 0.015). In the Finnish study, none of the other identified SNPs yielded results approaching significance, even without correction for multiple testing. The −3A variant is in a putative Elk-1 binding site of *DYX1C1*, and the 1249T allele causes a truncation of four residues from the C terminus of the *DYX1C1* protein, close to the TPR. Taipale et al therefore hypothesised that these two variants are functionally important. It is important to note that this hypothesis was based solely on association analyses and bioinformatic predictions; no studies have yet been performed to directly assess the functional significance of either variant. It is not known whether Elk-1 binds to the −3G→A region of *DYX1C1* in vivo, nor whether the −3A variant alters binding efficiency. Moreover, it is possible that a truncation of only four amino acids from the C terminus of this protein may have no functional effect.

The present study did not find consistent significant association across six reading and language related measures with either of these variants. Indeed, dyslexia is a heterogeneous neurological syndrome, so we need not necessarily find an association with every psychometric measure. However, the only significant associations we did find, for both single markers and two marker haplotypes, are in the reverse directions to those reported by Taipale et al. There are a number of different possibilities that might account for the discrepancy between our findings and those of the Finnish study. Although disruption of *DYX1C1* may be responsible for the problems of the family carrying the translocation, the effects of the gene may not extrapolate to common forms of developmental dyslexia. A comparable situation was reported for the role of FOXP2 in speech and language disorders (OMIM 605317); while the gene is clearly involved in a severe, rare form of impairment, its effects do not generalise to common language related disorders. This would imply that the initial associations of the −3A and 1249T alleles with developmental dyslexia were false positives, in part due to stochastic effects related to small sample size. It is worth noting that the Finnish study carried out a case control comparison for detecting association, but a significant proportion of their sample were family based; the total 109 individuals with dyslexia were derived from only 23 unrelated families and 33 unrelated dyslexic-nondyslexic couples. Non-independence of alleles in different related individuals from the same families may distort evidence for association. Taipale et al did report a positive TDT result for the proposed risk haplotype (p = 0.025), but this was based on an extremely small number of informative trios (n = 9) selected from the same families that had yielded the initial association.

Alternatively, the opposite direction of allelic associations between our studies may imply that a susceptibility locus for dyslexia (perhaps *DYX1*) is in LD with these *DYX1C1* variants, and that our conflicting results are due in part to our two distinct samples, Finnish and UK based. Each sample may contain the same 15q susceptibility gene for dyslexia, but on different haplotype backgrounds. It therefore remains possible that the true functional variants are as yet unidentified SNPs affecting *DYX1C1* function. However, the involvement of unidentified variants of *DYX1C1* in increasing risk in our sample appears unlikely given our association data. Previous studies of the *DYX1* susceptibility locus on 15q21 have found linkage and association 8 Mb proximal to *DYX1C1*, around markers D15S132 and D15S143 (fig 1).

![Table 3](https://www.jmedgenet.com/)

<table>
<thead>
<tr>
<th>SNP</th>
<th>−164C→T</th>
<th>−3G→A</th>
<th>271G→A</th>
<th>572G→A</th>
<th>1249G→T</th>
<th>1259C→G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total association p values (bottom left)</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td><em>Total association p values (underlined) were calculated with a VC model, and only p values &lt;0.1 are reported (correction for multiple testing was not applied).</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D’ values (not underlined)</td>
<td>0.0918</td>
<td>0.0351</td>
<td>0.0659</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*Table 3 Analysis of two marker haplotypes within *DYX1C1***

---

856 Letter to JMG

www.jmedgenet.com
distance of about 7 cM that includes \textit{DYX1C1} (fig 1). If this 15q breakpoint does not disrupt \textit{DYX1C1}, then it will likely be proximal to it, and hence closer to D15S143 where peak linkage is reported.\textsuperscript{7 7}

A recent genome-wide scan looking for evidence of transmission distortion identified 15q21.3 as having the highest Z score, being more than 3.0.\textsuperscript{10} Under the null hypothesis of loci obeying Mendelian inheritance, the Z scores in this analysis would have a mean of 0 and SD of 1. Given that 15q21.3 was found to deviate from the mean by more than 3 SD, it implies that this genomic region is preferentially shared amongst siblings, and so could have a distorting effect on any association or linkage analysis carried out in this region. As \textit{DYX1C1} is in the middle of 15q21.3, extra caution is therefore required when interpreting association findings for this gene to any trait.

We note that a report is now in press entitled “Support for \textit{EKN1} as the susceptibility locus for dyslexia on 15q21.”\textsuperscript{11} Wigg \textit{et al} have interpreted their association data, based on a sample of Canadian families, as supportive of an involvement of \textit{DYX1C1} in dyslexia. However, they report results similar to ours, that is, a biased transmission of the −3G/1249G haplotype to children with poorer reading related skills (the opposite finding to the original report by Taipale \textit{et al}).\textsuperscript{4} Thus the data of Wigg \textit{et al}, like ours, cast doubt on the functional significance of the −3A and 1249T alleles proposed to underlie dyslexia susceptibility.

In conclusion, our study indicates that the previously suggested functional variants of \textit{DYX1C1} are unlikely to increase risk of dyslexia. While disruption of this gene may indeed be implicated in rare cases, such as in the translocation family identified by Taipale \textit{et al},\textsuperscript{4} allonic variants of \textit{DYX1C1} do not appear to play a major role in our large sample of families from the UK. It remains possible that undiscovered \textit{DYX1C1} variants, in LD with those studied here, account for the association reported in the Finnish study. Alternatively, another gene in 15q21 may be contributing to risk in common cases of developmental dyslexia. Thus, further studies involving analyses of other genes in this region remain essential for dissecting the genetic aetiology of dyslexia.

Acknowledgements

We wish to thank all of the families who have participated in this study, and especially Janet Walter for collecting many of their samples and carrying out many of the psychometric tests, Dr Toril Fagerheim for providing useful suggestions in the writing of this report, and also Dr Joel B Talcott and Dr Kathleen E Taylor for their assistance with administering the phenotypic data. Dr Fisher is a Royal Society Research Fellow and Professor Monaco is a Wellcome Trust Principal Research Fellow.

Correspondence to: Professor Anthony P Monaco, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, UK, anthony.monaco@well.ox.ac.uk

Received 10 January 2004

References