Original articles

A new gene (DYX3) for dyslexia is located on chromosome 2

Toril Fagerheim, Peter Raeymaekers, Finn Egil Tønnessen, Marit Pedersen, Lisbeth Tranebjærg, Herbert A Lubs

Abstract

Developmental dyslexia is a specific reading disability affecting children and adults who otherwise possess normal intelligence, cognitive skills, and adequate schooling. Difficulties in spelling and reading may persist through adult life. Possible localisations of genes for dyslexia have been reported on chromosomes 15 (DYX1), 6p21.3-23 (DYX2), and 1p over the last 15 years. Only the localisation to 6p21.3-23 has been clearly confirmed and a genome search has not previously been carried out. We have investigated a large Norwegian family in which dyslexia is inherited as an autosomal dominant trait. A genome wide search for linkage analysis was performed under three different diagnostic models. Linkage analysis in the family identified a region in 2p15-p16 which cosegregated with dyslexia. Maximum lod scores of 3.54, 2.92, and 4.32 for the three different diagnostic models were obtained. These results were confirmed by a non-parametric multipoint GENEHUNTER analysis in which the most likely placement of the gene was in a 4 cM interval between markers D2S2352 and D2S1337. Localisation of a gene for dyslexia to 2p15-16, together with the confirmed linkage to 6p21.3-23, constitute strong evidence for genetic heterogeneity in dyslexia. Since no gene for dyslexia has been isolated, little is known about the molecular processes involved. The isolation and molecular characterisation of this newly reported gene on chromosome 2 (DYX3) and DYX1 will thus provide new and exciting insights into the processes involved in reading and spelling.

Keywords: developmental dyslexia; reading disability; linkage analysis; chromosome 2

Developmental dyslexia is a specific reading disability affecting children and adults who otherwise possess normal intelligence, cognitive skills, and adequate schooling. In 1950, Hallgren reported that more than 80% of children with dyslexia in Stockholm schools had other family members with dyslexia. Subsequent studies have shown that there is both a physical and genetic basis for many children and adults with severe reading disability and that the most frequent deficit is in phonological coding. This is manifested by decreased ability to pronounce letter strings (or pseudowords) correctly that have not been encountered previously. Recent studies using new methods have provided a clearer view of the difference in the processing of written words in dyslexics and non-impaired readers. In none of the studies, however, were the details of the family history known, and gene localisation studies were not carried out. Using whole brain magnetoencephalography, it was shown that in response to reading lists of real and non-words normal readers first activated the posterior occipital cortex in 150 ms, then activated the left temporoparietal region by 184 ms. Six dyslexics, however, failed to activate the temporoparietal region in the first 200 ms and instead activated the right or left inferior frontal gyrus. This region contains Broca’s area and its activation suggests the use of guessing as an alternative processing strategy. Using functional magnetic resonance imaging (fMR) and a carefully constructed, increasingly complex series of reading tests, 29 dyslexics were found to show relatively less posterior activation (including Wernicke’s area, the angular gyrus, and striate cortex), and increased activation anteriorly (inferior frontal gyrus), compared to non-impaired readers. These studies show the use of a different processing pathway in response to phonological tasks and the presence of an altered temporal response in dyslexics. In addition, impaired discrimination of both rapidly presented visual and auditory non-verbal information have been reported, and a more generalised defect in the rapid processing of information may also play a role in some or all instances of dyslexia.

Many children under 2 years of age with damage to the language areas of the left hemisphere can go on to develop satisfactory language ability. This observation implies that specialisation of the brain mechanism responsible for language function is not completed at
A new gene for dyslexia is located on chromosome 2

that age and that the plasticity of these complex neural networks permits their adaptive reorganisation. In contrast, the majority of teenagers who experience comparable brain damage never recover normal language function. A molecular test, or tests, for dyslexia, which would result from cloning genes for dyslexia, would allow earlier diagnosis of children at high risk for dyslexia. This in turn would permit institution of therapy while the language areas were at an earlier, more plastic stage of development.

The localisation of a susceptibility locus for dyslexia to 6p21.3 has been confirmed in three independent studies. The localisation to 6p, however, was not found in a large sib pair study, or in the large family studied by Sawyer et al. The linkage with chromosome 15p markers reported in 1983 was not initially confirmed, but recent reports suggest a localisation to D15S143, with lod scores of 3.15 and 1.38. This marker, however, is 49 cM distal to the centromere and thus does not conform the earlier localisation with 15p markers. A focus on 1p has also been considered because of a suggestive lod score at 1p34-36 and the identification of a family in which dyslexia segregated together with a 46,XYY(1;2)(p22;q22) translocation. By analogy to many other common inherited disorders, genetic heterogeneity is likely and could provide some explanation for the high frequency of dyslexia, but has not yet been clearly shown. Furthermore, the complexity of the reading process also suggests that many genes may be involved. In order to identify possible additional genes predisposing to dyslexia, we have carried out a genome search and performed linkage analysis in a large Norwegian family in which dyslexia is clearly inherited. The pedigree is shown in fig 1.

Methods

Tests for Dyslexia

The test instrument used to test for dyslexia in the Norwegian population was developed at the National Center for Reading Research. The assessments of both orthographic and phonological abilities are very similar to those used in the genetic studies of Olson et al and the Woodcock Reading Mastery Tests. However, a measure of reaction time was also included in three of the tests.

The present dyslexia test battery was based on selected subtests from a series of tests given to Norwegian schoolchildren with possible reading problems to test their reading level and ability (KOAS and KOAP). Tests were presented in a standardised form from a computer. The most informative five subtests from KOAS and KOAP were selected and supplemented with a standardised spelling test. Three tests were scored both by time of response and percentage correct (table 1).

Any history of reading and spelling problems in school was also recorded for the family members. In generation I and II, since the school system at that time was often not aware of dyslexia, family members were considered as affected when two tests were abnormal in the
absence of a history of reading problems. There was no history of recognised dyslexia in generation I. Only family members above the age of 8 years were tested. The tests derived from the KOAS and KOAP test battery were as follows.

1. KOAS 1 is a test of how well the subject can read words in isolation when given a generous amount of time; 72 common words are presented one at a time on the computer screen. The subject is asked to read the word aloud into a microphone as quickly as possible. The computer registers the amount of time the subject takes. The results are recorded both as percentage pronounced correctly and response time. The remaining tests are similarly administered and scored.

2. KOAS 2.1 tests how well the subject can read when given a very short amount of time; 72 common words, different from KOAS 1, are flashed on the screen one at a time, each presentation lasting only 100 msec. Otherwise this test is administered in the same way as KOAS 1.

3. KOAS 3.1 tests the subject’s ability to “read” non-words (experimentally designed letter strings such as “wug” or “mape” that are orthographically acceptable, but semantically empty); 36 non-words are presented.

4. KOAP 9A measures the subject’s ability to form words on the basis of individually presented phonemes. The test giver pronounces a series of sounds, for example, /b/, /æ/, and /g/, with 0.5 second pauses between the sounds. The subject is then asked to blend the sounds into a familiar word (“bag”). The test consists of 16 sequences of phonemes which are from three to nine phonemes long. (Reaction time is not measured.)

5. KOAP 9B is the same as KOAP 9A, but presents 16 sequences of phonemes that make up non-words.

6. Spelling from dictation. This is a standard list of 50 Norwegian words, both regular and irregular, scored as percentage correctly spelled.

ESTABLISHMENT OF CUT OFF POINTS

The distributions of the initial scores in family members who had a clearly positive history were compared with spouses and family members who had a clearly positive history. The distributions of the initial scores in family members above the age of 8 years were tested. The tests derived from the KOAS and KOAP test battery were as follows.

Table 1 Test results

<table>
<thead>
<tr>
<th>Test</th>
<th>Cut off scores</th>
<th>% correct</th>
<th>Milliseconds (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KS1- ordinary words</td>
<td>&lt;90</td>
<td>&gt;900</td>
<td></td>
</tr>
<tr>
<td>KS2.1- ordinary words</td>
<td>&lt;96</td>
<td>&gt;625</td>
<td></td>
</tr>
<tr>
<td>KS3.1- non-words</td>
<td>&lt;91</td>
<td>&gt;1925</td>
<td></td>
</tr>
<tr>
<td>KP9A- phonemes to words</td>
<td>&lt;70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KP9B- phonemes to non-words</td>
<td>&lt;56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dictation (spelling)</td>
<td>&lt;56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

show bimodality. Cut off points defining affected status were selected to include these extreme values and to minimise overlapping values. Using these cut off points only four of 198 scores in the overall study were scored as positive in the 27 unaffected family members or spouses. The known distributions of test scores in older school children (14-15 years) were also helpful in determining these cut off points. Because of this conservative definition of the cut off points, only two of nine test scores were required to be abnormal to establish an affected status. One, however, had to be a non-word test, since non-word tests are generally accepted as the most discriminate. Thus, the cut off points in adults were established empirically and the determination of affected status was primarily based upon non-overlapping, outlying scores.

Too few (affected or unaffected) children were available to establish comparable cut off points in children. Inspection of the distributions of scores in Norwegian school children did not show obvious cut off points, and there was no basis for comparing a group of those children with familial dyslexia with other children, since no family history or genetic data were available. Therefore, it was necessary to use the adult cut off points even though it was known that the test means in children changed significantly from grades 2 to 10. Consequently, it was anticipated that both false positive and negative results might occur in children. The cut off points for the tests are shown in Table 1.

GENOTYPING

The family was ascertained through the Norwegian Dyslexia Association in response to our request for large families for linkage analysis. Candidate regions were first screened using 12 markers on 1p, 11 markers on 6p, and 11 markers on 15q. A total of 307 highly informative di-, tri-, and tetranucleotide repeat markers were selected for an (average) 20 cM genome screen, used in all aspects of the genome wide screening study. PCR primers were labelled with either 6-FAM, TET, or HEX phosphoramidites, and pooled PCR products were run on a 377 sequencer (Applied Biosystems). The runs were followed by analysis with GENESCAN™ (version 2.0), and the results were entered in CYRILLIC (version 2.1) for generation of linkage files and haplotype analysis.

LINKAGE ANALYSIS

Parametric linkage analysis

Since there was no single reliable diagnostic test available, it was decided to carry out a categorical linkage analysis assuming an autosomal dominant inheritance pattern with reduced penetrance rather than a quantitative trait locus model. This seemed reasonable since an autosomal dominant model was compatible with the segregation of the trait in this family. It should be noted, however, that the inheritance of dyslexia is complex and that the effects of sex differences in penetrance, heterogeneity, absence of a definitive diagnostic test, and age compensation all complicate
A new gene for dyslexia is located on chromosome 2

Table 2 Maximum lod scores for two point linkage analysis of dyslexia with 2p15-p16 markers

<table>
<thead>
<tr>
<th>Marker</th>
<th>Model 1* Zmax</th>
<th>Model 2 Zmax</th>
<th>Model 3 Zmax</th>
<th>Distance in cM†</th>
</tr>
</thead>
<tbody>
<tr>
<td>D2S177</td>
<td>0.058 0.22</td>
<td>0.009 0.23</td>
<td>1.365 0.03</td>
<td>65</td>
</tr>
<tr>
<td>D2S1356</td>
<td>0.818 0.20</td>
<td>0.652 0.21</td>
<td>1.856 0.05</td>
<td>65</td>
</tr>
<tr>
<td>D2S2328</td>
<td>2.953 0.00</td>
<td>2.234 0.04</td>
<td>2.005 0.06</td>
<td>65</td>
</tr>
<tr>
<td>D2S2994</td>
<td>0.415 0.19</td>
<td>0.258 0.20</td>
<td>1.439 0.06</td>
<td>68.2</td>
</tr>
<tr>
<td>D2S119</td>
<td>0.915 0.17</td>
<td>0.381 0.22</td>
<td>1.949 0.00</td>
<td>69</td>
</tr>
<tr>
<td>D2S2998</td>
<td>1.195 0.18</td>
<td>0.635 0.19</td>
<td>2.814 0.06</td>
<td>70</td>
</tr>
<tr>
<td>D2S2200</td>
<td>0.689 0.19</td>
<td>0.234 0.24</td>
<td>1.458 0.09</td>
<td>73.3</td>
</tr>
<tr>
<td>D2S3278</td>
<td>0.697 0.20</td>
<td>0.484 0.20</td>
<td>1.771 0.06</td>
<td>74</td>
</tr>
<tr>
<td>D2S352</td>
<td>1.440 0.08</td>
<td>1.162 0.07</td>
<td>1.533 0.07</td>
<td>76</td>
</tr>
<tr>
<td>D2S2352</td>
<td>2.860 0.03</td>
<td>2.920 0.00</td>
<td>2.064 0.03</td>
<td>80</td>
</tr>
<tr>
<td>D2S2578</td>
<td>2.668 0.09</td>
<td>2.055 0.11</td>
<td>1.323 0.00</td>
<td>81</td>
</tr>
<tr>
<td>D2S2799</td>
<td>0.641 0.18</td>
<td>0.429 0.19</td>
<td>2.160 0.00</td>
<td>81</td>
</tr>
<tr>
<td>D2S2183</td>
<td>3.528 0.00</td>
<td>2.789 0.00</td>
<td>2.860 0.00</td>
<td>81</td>
</tr>
<tr>
<td>D2S1337</td>
<td>1.152 0.17</td>
<td>0.762 0.19</td>
<td>2.393 0.02</td>
<td>82</td>
</tr>
<tr>
<td>D2S2393</td>
<td>3.467 0.00</td>
<td>2.929 0.00</td>
<td>3.051 0.00</td>
<td>84</td>
</tr>
<tr>
<td>D2S357</td>
<td>1.550 0.15</td>
<td>0.965 0.18</td>
<td>3.010 0.02</td>
<td>84</td>
</tr>
<tr>
<td>D2S337</td>
<td>1.766 0.15</td>
<td>1.170 0.17</td>
<td>3.010 0.02</td>
<td>84</td>
</tr>
<tr>
<td>D2S3134</td>
<td>0.257 0.26</td>
<td>0.026 0.41</td>
<td>0.011 0.32</td>
<td>88</td>
</tr>
</tbody>
</table>

*Models 1–3 are defined in the text.
†Distances from p telomere according to the Généthon linkage map.

Results

A genome wide search beginning with the three reported possible loci for dyslexia (see Methods) was initiated in 36 members of the family, using a total of 307 microsatellite markers (Fagerheim et al, in preparation). This screening did not show significant linkage, but marker D2S1356 at 2p15-p16 showed a slightly positive lod score of 0.8. Analysis of 17 additional microsatellite markers in the region gave strong evidence in favour of linkage (table 2). The parametric linkage analysis was calculated under three different models. Model 1 included all adults and children depicted in fig 1. This model assumes sex dependent penetrance and phenocopy values and differential diagnostic weight factors as described in the Methods section. Model 2 included all normal readers and only the affected family members with both a history of reading problems and positive testing. Because of concern about the accuracy of the test results in classifying school children, as discussed above, a third analysis model excluded all persons under the age of 20 (IV.7 to IV.33). As indicated in table 1, maximum lod scores were obtained for the three models with markers D2S2183 (Z=3.53, °=0.00), D2S393 (Z=2.93, °=0.0), and D2S378 (Z=4.32, °=0.0), which were within a 3 cM region. Most importantly, the “Lander-Kruglyak” criteria of 3.3 for genome wide significance were exceeded in two of the three analysis models. In addition, correcting for multiple testing in our analysis calculated the threshold for reporting significant linkage as 3.47, and this was obtained for both models 1 and 3. Furthermore, lod scores in excess of 3 were obtained in models 1 and 3 with other markers from the 2p15-p16 region as well (D2S393 for model 1 and D2S357, D2S393, D2S378 for model 3). Parametric high density multipoint linkage analysis greatly increases the number of false negative mapping results in cases where the exact mode of inheritance of the disorder is unknown. Therefore, we
decided to perform a non-parametric multi-point linkage analysis (NPL) using the program GENEHUNTER. Although not entirely meeting the levels of genome wide significance, this independent method essentially confirmed the two point lod score results (fig 2). The lower lod score may have resulted from loss of information owing to exclusion of normal family members. NPL associated p values of 0.023 (model 1), 0.016 (model 2), and 0.0009 (model 3) were obtained in the region between D2S2352 and D2S1337. The distance between these markers is approximately 4 cM according to the latest Genethon linkage map. This localisation was confirmed by inspection of marker haplotypes, which identified two critical recombinants (II.1 and III.7) in adult patients. These placed the DX13 locus centromeric to D2S2352 and telomeric to D2S1337 (fig 1).

Discussion

The present family is one of the few large, well studied kindreds in which dyslexia is inherited as an autosomal dominant trait (fig 1). The dyslexia was generally mild to moderate in this family, and most of the 11 affected adults responded well to minimal therapy or compensated over time. A computer administered test battery was used for the diagnosis of dyslexia. In order to classify a family member as dyslexic, two of the nine test scores had to be positive (see Methods). A broader version of this test system is used throughout Norway to evaluate children with possible reading problems. Cut off points for this study were developed empirically by comparing results in clearly affected and clearly normal adults after 20 subjects had been tested. No unaffected and only three affected young children were initially available in the family, so that comparable empirical norms in children could not be developed. As anticipated, these norms worked well in adults but were less accurate in school children because the continuous development of reading ability during the school years results in both year specific means and wide standard deviations. A retrospective review of the haplotype/phenotype correlation in children showed several discrepancies (IV.7, IV.11, IV.32, and IV.33). IV.33, who had a strongly positive history of reading problems and possible attention deficit hyperactivity disorder (ADHD), tested positively on all tests for dyslexia, but did not have the affected haplotype. On the other hand, her brother, IV.32, had no history of reading disability and fell in the dyslexia range on only one test, a non-word test, but had the disease haplotype. He was classified as normal in the linkage analysis, but may represent an instance of non-penetration. These and other discrepancies clearly indicate the need for better tests in children. Since most of the study subjects were adults, however, the diagnostic classification system still allowed the detection of the present linkage. These discrepancies lowered the lod scores for models 1 and 2.

Few genes have been identified in 2p15-p16 (http://www.ncbi.nlm.nih.gov/genemap) and none is an obvious candidate gene for dyslexia. Calcineurin B, a calmodulin dependent protein phosphatase, which constitutes 1% of the total brain protein, was roughly mapped to the region by in situ hybridisation. It was therefore an obvious candidate gene, but sequencing of the whole coding region did not detect any changes in the sequence in two affected family members. Recent mapping information, however, has placed calcineurin B centromeric to the candidate region.

The finding of linkage to 2p15-p16 in this family and the confirmed linkage to 6p21.3 indicate that dyslexia is heterogeneous. Cloning of the DX13 gene will provide important insight into the nature and frequency of at least one gene that is involved in reading and spelling. Furthermore, it represents an opportunity to study the phenotypic variation associated with one gene, and to develop better routine testing by correlating test results with specific genotypes. It will also be of great interest to determine whether magnetoencephalography or fMR will provide evidence of clinical subtypes when combined with a precise genetic diagnosis. Specific remedial plans and a more predictable long term outcome may also be important benefits.

We are deeply indebted to all members of the family who made this research project possible. We thank Dr Wedelius and the Nordic Genome Initiative for providing us with the genetic markers and a special thanks to Dr Gorstholm Schaffner and coworkers at the Service Department at IMP, Vienna, for excellent help and service to TF at a critical stage of the project. This research project was funded in part by the Norwegian National Research Council to FET, HL, and TF (grant no 107367/330). The study was approved by the Research Ethical Committee of Health Region 3, Norway.