A model system for the analysis of gene exclusion: cystic fibrosis and chromosome 19

BRANDON WAINWRIGHT, MARTIN FARRALL, EILA WATSON, AND ROBERT WILLIAMSON

From the Department of Biochemistry, St Mary's Hospital Medical School, University of London, London W2 1PG.

SUMMARY We have used multilocus analysis to exclude the cystic fibrosis locus from six polymorphic DNA markers covering most of chromosome 19. A substantial increase in the confidence for exclusion was obtained using the computer programme LINKAGE compared to analysis of pairwise lod scores. A structured approach to the analysis of linkage to autosomal recessive inherited diseases where the biochemical defect is not known is described.

Cystic fibrosis (CF) is the most common genetic disease in North European populations, with an incidence of 1 in 2000 live births in the United Kingdom. The biochemical basis of the mutation in CF is not known, although membrane transport phenomena associated with the disease have recently been described. Evidence from segregation analyses has confirmed that the inheritance of CF is consistent with its being a recessive defect caused by a mutation at a single autosomal locus.

The use of DNA probes for the analysis of genetic disease allows the determination of linkage between a restriction fragment length polymorphism (RFLP) and a disease locus even when the biochemical defect is unknown. This approach has already been used successfully by us to analyse sex linked genetic diseases such as Duchenne muscular dystrophy, as well as to study previously unassigned autosomal dominant diseases such as Huntington's disease.

There were, until recently, few clues as to the chromosomal localisation of the CF locus (CF). Mayo et al suggested that CF might be located on chromosome 4, based on data from cell hybrids expressing a ciliary dyskinesis factor. We have shown that the mutation causing CF is unlikely to be on chromosome 4 by exclusion mapping with several DNA and protein markers. Two different families have been reported showing chromosomal anomalies, one of which suggested that the CF mutation mapped to chromosome 5p, and the other suggested chromosome 13q34. We investigated the latter family initially by mapping the gene for factor X to 13q34 by gene dosage experiments and then excluded CF from linkage to factor X. We have also demonstrated that certain candidate genes, such as complement component 3, which had been proposed as the locus of the CF mutation, do not segregate with the disease.

Since this research was completed, Eiberg et al have reported linkage between CF and PON (paraoxonase) and several groups, including our own, have reported that the mutation causing cystic fibrosis is located on the long arm of chromosome 7 and shows linkage to 3-11 and TCRB to COLIA2 to MET and to DOCR1-917.

Although linkage between markers and CF has now been shown, our data for chromosome 19 illustrate the use of multipoint computer analyses which analyse multiple linked markers along a single chromosome to maximise data from the few informative families that can be obtained. It is therefore a paradigm for any single gene autosomal recessive disease where the biochemical defect is unknown. Multipoint linkage is of particular value as it usefully extends the exclusion regions between the markers to exclude most of the chromosome.

Materials and methods

GENE PROBES LOCATED ON CHROMOSOME 19

The cloned DNA sequences used in this study are listed in table 1 together with regional assignments of the markers and details of the RFLP defined by...
each clone. A cDNA encoding the LDL receptor gene was kindly provided by Dr D Russell. The gene for complement 3 was obtained from Dr G Fey, the cDNA clone for apolipoprotein C2 (APOC2) was isolated in this laboratory, and the cytochrome p450 cDNA recombinant was provided by Dr E Shepherd. Random probes D19S9 and D19S6 were provided by Drs D Shaw and R Bartlett respectively.

**Cystic Fibrosis Families**

Twelve European families with at least three affected children and up to four unaffected sibs and two families with two affected children and two normal sibs were used in this study. Blood was collected from all children and parents and DNA prepared by standard methods.

**Hybridisation Analysis**

DNA 5 μg from each family member was digested with the appropriate restriction enzyme according to the manufacturer’s instructions (BRL Ltd, Irvine, Scotland), fractionated by electrophoresis on 0.8% agarose, and transferred to Zetapore membranes (AMF Cuno) as described by Maniatis et al. Hybridisation and autoradiography were carried out as described previously.

**Linkage Analysis**

Pairwise lod scores were calculated using the LIPEP computer programme adapted by Ott for use with microcomputers. The allele frequencies used are given in table 1. The multipoint linkage computer package LINKAGE was used to test for linkage to cystic fibrosis at locations within a fixed map of known linked markers, as discussed by Farrall et al. The order of loci LDLR-C3-APOC2-CYP1-D19S6 was deduced from physical assignments, two point lod scores, and multipoint calculations.

There were insufficient data to place D19S9 accurately within the other markers. Estimates of recombination fractions between markers were taken from HGM8 and our own multipoint estimates. Each map distance is calculated from the upper limit (95% confidence interval) of the recombination fraction estimate by using Haldane’s mapping function. The use of this mapping function effectively disregards the influence of interference in subsequent analyses. By using the upper limits of the recombination fraction estimates and ignoring interference, the estimates of exclusion are cautious. Multipoint likelihoods may be conveniently presented as a location map, in the manner of Lathrop et al. For pairwise analyses, a lod score of -2.0 calculated for combined sexes was considered to exclude linkage significantly between CF and each marker at the relevant recombination fraction. For multipoint calculations, a location score of -9.2 (lod equivalent = -2.0) for CF within the fixed map of markers was similarly considered to exclude linkage.

**Results**

Table 2 shows the pairwise lod scores obtained for each probe and CF. Fig 1 shows the physical location of the markers used in this study. Fig 2 shows the multipoint likelihood map for CF within the linked markers LDLR, C3, APOC2, and CYP1. From the pairwise results we can exclude CF from moderate linkage to LDLR and C3, and tight linkage to APOC2, D19S9, CYP1, and D19S6. Multipoint analysis shows that the exclusion regions may be extended between LDLR, C3, APOC2, and CYP1.

The total excluded genetic length for multipoint analysis is 88 centiMorgans (cM) (LDLR-C3-APOC2-CYP1-D19S6). An estimate of the total

### Table 1 Polymorphisms used in this study.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Regional assignment</th>
<th>RFLP enzyme</th>
<th>No of alleles</th>
<th>Frequency of rare allele</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDLR</td>
<td>p13.2—p13.1</td>
<td>Prall</td>
<td>2</td>
<td>0.27</td>
<td>26</td>
</tr>
<tr>
<td>C3</td>
<td>p13.3—p13.2</td>
<td>Ssd</td>
<td>2</td>
<td>0.39</td>
<td>12</td>
</tr>
<tr>
<td>APOC2</td>
<td>cen—q13.3</td>
<td>TaqII</td>
<td>2</td>
<td>0.40</td>
<td>27</td>
</tr>
<tr>
<td>D19S9</td>
<td>p13.2—p13.2</td>
<td>LacR1</td>
<td>2</td>
<td>0.19</td>
<td>21</td>
</tr>
<tr>
<td>CYP1</td>
<td>q13.1—q13.3</td>
<td>Ssd</td>
<td>2</td>
<td>0.47</td>
<td>28</td>
</tr>
<tr>
<td>D19S6</td>
<td>q13.3—qter</td>
<td>Tpwl</td>
<td>2</td>
<td>0.48</td>
<td>29</td>
</tr>
</tbody>
</table>

### Table 2 Lod scores at selected recombination fractions between CF and six chromosome 19 markers.

<table>
<thead>
<tr>
<th>Marker</th>
<th>No of markers</th>
<th>Recombination fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>LDLR</td>
<td>-3.70</td>
<td>-2.31</td>
</tr>
<tr>
<td>C3</td>
<td>-4.40</td>
<td>-2.21</td>
</tr>
<tr>
<td>APOC2</td>
<td>-1.83</td>
<td>-1.02</td>
</tr>
<tr>
<td>D19S9</td>
<td>-2.20</td>
<td>-2.20</td>
</tr>
<tr>
<td>CYP1</td>
<td>-5.62</td>
<td>-2.91</td>
</tr>
<tr>
<td>D19S6</td>
<td>0.12</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>-1.40</td>
<td>-0.78</td>
</tr>
<tr>
<td></td>
<td>-3.22</td>
<td>-1.40</td>
</tr>
<tr>
<td></td>
<td>-0.66</td>
<td>-0.15</td>
</tr>
<tr>
<td></td>
<td>-1.52</td>
<td>-0.95</td>
</tr>
<tr>
<td></td>
<td>-2.31</td>
<td>-1.10</td>
</tr>
<tr>
<td></td>
<td>-0.31</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>-2.40</td>
<td>-1.62</td>
</tr>
<tr>
<td></td>
<td>-3.10</td>
<td>-1.31</td>
</tr>
<tr>
<td></td>
<td>-2.18</td>
<td>-1.10</td>
</tr>
<tr>
<td></td>
<td>-0.55</td>
<td>-0.28</td>
</tr>
<tr>
<td></td>
<td>-1.60</td>
<td>-0.39</td>
</tr>
</tbody>
</table>

m = male, f = female, c = combined.

**Calculations for:**

- Males only $l_{ma}$ (theta = 0.5)
- Females only $l_{fa}$ (theta = 0.5)
- Combined sexes $l_{mc}$ (theta = theta, theta = theta)
A model system for the analysis of gene exclusion: cystic fibrosis and chromosome 19

excluded genetic length derived from pairwise analyses is 102 cM. This is calculated by doubling the exclusion region around each locus and summing the individual regions.

**Discussion**

By analysing the segregation of cystic fibrosis with informative polymorphic DNA markers we have excluded CF from a major proportion of chromosome 19. This has been achieved by the use of a fixed genetic map and multipoint analysis. The total exclusion limits achieved by multilocus analysis may be compared to that calculated by adding the exclusion limits derived from pairwise scores at each locus. For these data pairwise methods appear to exclude a larger area. However, adding pairwise scores is only valid if the scores for each pair of loci are independent. The data presented here are not independent; the same meioses contribute to several lod tables. Hence, multipoint analysis, which allows correctly for the interdependence of data at several loci, is the only valid method for estimating the total exclusion region for this data set. We note that the confidence for exclusion between markers using multipoint analysis is substantial, an example of the increase in information derived from each family when multipoint methods are used.

In the case of many autosomal recessive disorders of unknown biochemical aetiology, the search for a linkage is made more difficult by the absence of a reliable heterozygote detection system, thereby reducing the statistical power of each disease family; in general, only the parents and affected sibs contribute significantly to the lod score. In such circumstances a 'random' probe approach is less powerful than a structured approach, which involves the construction of genetic maps followed by the systematic exclusion of a disease locus from all points between a set of markers. This becomes particularly important if there is reason to suspect genetic heterogeneity.

The chromosomal location of CF is now known and therefore this paper is not relevant to exclusion mapping for the CF locus. However, it illustrates the power of LINKAGE in increasing the exclusion length using ordered markers on a single chromosome, and provides a model for this approach to an autosomal recessive disease. This model will be useful for the analysis of rare recessive diseases where the maximal information for linkage may be derived from the few families available for testing.

We wish to thank the many clinicians and families who have made this work possible by their enthusiasm and cooperation, Drs Bartlett, Fey, Humphries,
Russell, Shaw, and Shepherd for the use of their cloned gene probes, and Dr Peter Scambler for helpful discussions. This work was supported generously by the UK Cystic Fibrosis Research Trust and the Medical Research Council. MF is a Foulkes Foundation fellow.

References


Correspondence and requests for reprints to Professor R Williamson, Department of Biochemistry, St Mary's Hospital Medical School, London W2 1PG.