<table>
<thead>
<tr>
<th>Title</th>
<th>Genetic analysis of cystic fibrosis using linked DNA markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Tsui, LC; Buetow, K; Buchwald, M</td>
</tr>
<tr>
<td>Citation</td>
<td>American Journal Of Human Genetics, 1986, v. 39 n. 6, p. 720-728</td>
</tr>
<tr>
<td>Issued Date</td>
<td>1986</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10722/42284">http://hdl.handle.net/10722/42284</a></td>
</tr>
<tr>
<td>Rights</td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.; American journal of human genetics. Copyright © University of Chicago Press.</td>
</tr>
</tbody>
</table>
Genetic Analysis of Cystic Fibrosis Using Linked DNA Markers

LAP-CHEE TSUI,1 KEN BUETOW,2 AND MANUEL BUCHWALD1

SUMMARY

Genetic linkage has been analyzed between cystic fibrosis (CF) and a number of markers on the long arm of chromosome 7, including D7S15, COL1A2, PON, MET, D7S8, and TCRB, using a cohort of 47 Canadian and 13 Danish CF families. The analysis confirms the previous observations that both MET and D7S8 are closely linked to CF. Based on the result from one family, MET appears to be more proximal to the centromere than CF. Our analysis also suggests that genetic heterogeneity may account for the high recombination fraction between CF and D7S8 observed in another family. In addition, a strong linkage disequilibrium has been observed between CF and the two closely flanking markers.

INTRODUCTION

Genetic linkage analysis based on the simple autosomal recessive mode of inheritance of cystic fibrosis (CF) has recently resulted in the identification and chromosomal localization of the disease locus [1–11]. Eiberg et al. [1–3] first detected a linkage between CF and PON (a genetic determinant for serum paraoxonase activity) [1–3]. We demonstrated a linkage between CF and a randomly isolated DNA marker, D7S15 (formerly DOCR1-917) [4], which subsequently led to the suggestion that CF is on chromosome 7 [5]. This assign-

Received July 24, 1986. 
This work was supported by grants from the Canadian Cystic Fibrosis Foundation, the National Institutes of Health, the North Dakota Cystic Fibrosis Foundation, and the Sellers Fund from the Hospital for Sick Children, Toronto. L.-C. T. is a research scholar of the CCFF. K. B. is a recipient of a postdoctoral traineeship from the National Institutes of Health.
1 Department of Genetics, Research Institute, The Hospital for Sick Children, Toronto, Ontario M5G 1X8 Canada.
2 Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, PA 19111.
© 1986 by the American Society of Human Genetics. All rights reserved. 0002-9297/86/3906-0006S02.00

ment was confirmed in other chronic DNA marker receptor β gene known chromosomal assignment of.

In order to extend the possible geneotype analysis, a genetic mapping linkage analysis of other DNA markers was performed. The majority of 47 Canadian and 47 Danish CF samples were provided by M. The DNA sample agarose gel blots for D7S15 (La cDNA) [15] and D7S8 (La system, and two described [1–3]. B. La Du and colleagues developed an order [5] of mapping system, and the described [1–3]. B. La Du.

The maximal markers were determined v by calculations. The order of J.-M. Laloue and colleagues for the previously

Linkage Analysis

The results of this study suggest that the genetic markers for CF previously
ment was confirmed by the demonstration of linkage between CF and a number of other chromosome 7 DNA markers, including the met oncogene (MET) [6], DNA marker D7S8 [7], the proc02 (I) collagen gene (COL1A2) [8, 9], the T-cell receptor β gene (TCRB) [8], and DNA marker 7C22 [10]. Furthermore, the known chromosomal locations of these DNA markers have allowed a tentative assignment of CF to band q31 [11].

In order to obtain a more accurate description of the CF locus and to investigate the possibility of genetic heterogeneity in CF, we and others [12] have initiated a group effort to study the linkage relationship between CF and the genetic markers previously described [1–9]. Here, we give an update of the linkage data based primarily on the study of 47 Canadian families each with two or more affected children. We also discuss the implications of this data set in regards to genetic heterogeneity and linkage disequilibrium.

MATERIALS AND METHODS

The majority of the linkage data presented in this report were derived from a cohort of 47 Canadian CF families. The structure of these families has been described [11]. DNA samples were prepared either directly from peripheral blood or from established lymphoblast lines as described [4]. The Danish family data were derived from DNA samples provided by M. Schwartz (Copenhagen) (13 families). The linkage data derived from the 47 Canadian and 13 Danish families are referred to as the Toronto families in the joint analysis [12]. Toronto family 10 is also known as GM 1078 and family 17 as GM 1076. The additional family data for D7S15 were derived from blots containing DNA samples from A. Bowcock (Stanford) (nine families), P. Scambler (London) (four families), and M. Leppert (Salt Lake City) (five families).

The DNA restriction fragment length polymorphisms (RFLPs) were determined by agarose gel blot hybridization analysis essentially as described [4, 9]. The DNA probes for D7S15 (Lam4-917) [4], COL1A2 (NJ-3, NJ-1, and HF-32) [13, 14], TCRB (a Jurkat cDNA) [15], MET (pmerH) [6, 16] and pmetD ([16], and R. White, personal communication) and D7S8 (p3.11) [7, 17] have been described. PON was scored as a codominant system, and the paraoxonase assays were performed in the laboratory of H. Eiberg as described [1–3]. Confirmation of the PON data on selected families was provided by B. La Du.

The maximal likelihood estimate for the recombination fraction between genetic markers was obtained using the lod score analysis method [18]. The lod (z) score calculations were performed using the LIPEQ [19] computer program provided by J. Ott. The order of genetic loci was examined using the LINKAGE program [20] provided by J.-M. Lalouel. Test of homogeneity based on linkage data was carried out using the HOMOG program [21] provided by J. Ott. The general procedures and methods of linkage analysis have been described [4, 11, 22].

RESULTS AND DISCUSSION

Linkage Analysis

The results of pairwise two-point linkage analyses between CF and various genetic markers are shown in table 1. Although the maximal likelihood estimates for θ derived from the updated family data differ slightly from those previously reported [4, 9, 11], the new values are well within the confidence
<table>
<thead>
<tr>
<th>Marker Loci</th>
<th>No. Informative Families</th>
<th>LOD (2) Scores at Recombination Fraction (θ) of</th>
<th>Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>.01</td>
<td>.05</td>
</tr>
<tr>
<td>CF-D7S15</td>
<td>67</td>
<td>-21.52</td>
<td>-0.77</td>
</tr>
<tr>
<td>CF-PON</td>
<td>31</td>
<td>-1.83</td>
<td>2.02</td>
</tr>
<tr>
<td>CF-COL1A2</td>
<td>31</td>
<td>-15.78</td>
<td>-1.50</td>
</tr>
<tr>
<td>CF-TCRB</td>
<td>46</td>
<td>13.80</td>
<td>-4.73</td>
</tr>
<tr>
<td>CF-MET</td>
<td>47</td>
<td>21.05</td>
<td>19.73</td>
</tr>
<tr>
<td>CF-D7S8</td>
<td>20</td>
<td>3.36</td>
<td>6.48</td>
</tr>
<tr>
<td>D7S15-PON</td>
<td>30</td>
<td>11.02</td>
<td>13.25</td>
</tr>
<tr>
<td>D7S15-COL1A2</td>
<td>18</td>
<td>-8.15</td>
<td>-1.93</td>
</tr>
<tr>
<td>D7S15-TCRB</td>
<td>28</td>
<td>-16.17</td>
<td>-2.69</td>
</tr>
<tr>
<td>D7S15-MET</td>
<td>21</td>
<td>-17.01</td>
<td>-4.04</td>
</tr>
<tr>
<td>D7S8-COL1A2</td>
<td>21</td>
<td>-1.90</td>
<td>1.47</td>
</tr>
<tr>
<td>COL1A2-PON</td>
<td>22</td>
<td>-14.80</td>
<td>-6.68</td>
</tr>
<tr>
<td>COL1A2-TCRB</td>
<td>29</td>
<td>-9.38</td>
<td>-0.33</td>
</tr>
<tr>
<td>COL1A2-MET</td>
<td>32</td>
<td>-14.16</td>
<td>-3.05</td>
</tr>
<tr>
<td>TCRB-PON</td>
<td>8</td>
<td>-7.08</td>
<td>-3.15</td>
</tr>
<tr>
<td>TCRB-MET</td>
<td>12</td>
<td>-3.11</td>
<td>0.49</td>
</tr>
<tr>
<td>TCRB-D7S8</td>
<td>17</td>
<td>-3.04</td>
<td>1.10</td>
</tr>
<tr>
<td>MET-PON</td>
<td>16</td>
<td>-8.13</td>
<td>-1.85</td>
</tr>
<tr>
<td>MET-D7S8</td>
<td>27</td>
<td>23.65</td>
<td>21.42</td>
</tr>
<tr>
<td>D7S8-PON</td>
<td>15</td>
<td>-9.24</td>
<td>-2.91</td>
</tr>
</tbody>
</table>

* D7S15, PON, COL1A2, MET, and D7S8 data derived from a cohort of 47 Canadian and 13 Danish CF families; including three from HCGMR, four from London, five from Salt Lake City, and nine from Stanford-Berkeley.
limits of the previous estimates. In addition, the result of multipoint point linkage analyses (data not shown) is consistent with the previous observation that the most probable order for CF and the other genetic markers is COL1A2:D7S15:PON:CF:TCRB [4, 9, 11]. The analyses also show that both MET and D7S8 are closely linked to CF in the Canadian and Danish families.

Only two apparent crossovers in 125 meioses were detected between CF and MET by direct counting of the informative chromosomes in the CF children and two between CF and D7S8 out of 120. The families in which recombinants were detected are shown in figure 1. Family 17 showed a single recombination between CF and MET, and family 19 revealed one between CF and MET and two between CF and D7S8. The identity of each individual in these two families was confirmed by resampling of blood and reexamination of the RFLPs for each probe using the new samples. In fact, no unexpected alleles were observed in any of these children from the more than 60 different genetic markers tested, including several highly polymorphic DNA markers (data not shown), thus arguing strongly against the possibility of a false parentage.

The first child (17-3) in family 17 apparently carries a paternal chromosome that had recombined between CF and MET. Based on the order of COL1A2, D7S8, PON, and CF derived from previous studies [4, 9] and the haplotype information derived from the grandparents (data not shown), evidence is highly suggestive that MET maps between CF and PON. Unfortunately, the relative position for D7S8 could not be derived from this family as the father was not informative for the analysis. A more extensive analysis on the order of these closely linked loci is presented in the joint study [12].

Genetic Heterogeneity

Genetic heterogeneity in CF has been considered as one of the possible explanations for the high frequency of the disease [23]. Although it is clear that the majority of CF mutations must be at the same locus on the long arm of
<table>
<thead>
<tr>
<th>Hypotheses*</th>
<th>Max. lod</th>
<th>α</th>
<th>θ</th>
<th>Source</th>
<th>d.f.</th>
<th>$\chi^2$</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF vs. D7S8</td>
<td></td>
<td></td>
<td></td>
<td>H2 vs. H1</td>
<td>1</td>
<td>4.267</td>
<td>.0194</td>
</tr>
<tr>
<td>H2</td>
<td>52.171</td>
<td>.97</td>
<td>.01</td>
<td>H1 vs. H0</td>
<td>1</td>
<td>100.075</td>
<td>.0000</td>
</tr>
<tr>
<td>H1</td>
<td>50.037</td>
<td>(1)</td>
<td>.01</td>
<td>Total</td>
<td>2</td>
<td>104.342</td>
<td>.0000</td>
</tr>
<tr>
<td>H0</td>
<td>(0)</td>
<td>(.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CF vs. MET</td>
<td></td>
<td></td>
<td></td>
<td>H2 vs. H1</td>
<td>1</td>
<td>0.105</td>
<td>.3727</td>
</tr>
<tr>
<td>H2</td>
<td>52.784</td>
<td>.98</td>
<td>.01</td>
<td>H1 vs. H0</td>
<td>1</td>
<td>105.463</td>
<td>.0000</td>
</tr>
<tr>
<td>H1</td>
<td>52.732</td>
<td>(1)</td>
<td>(.5)</td>
<td>Total</td>
<td>2</td>
<td>105.568</td>
<td>.0000</td>
</tr>
<tr>
<td>H0</td>
<td>(0)</td>
<td>(0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Hypothesis H2 assumes α being the fraction of families that carry a CF mutation linked to the test marker; H1 assumes that all families carry a CF mutation at the same locus; H0 is the null hypothesis that the test marker is not linked to CF. The α values and the corresponding maximal likelihood estimate for θ are derived from the linkage data based on the Canadian and Danish families using the TMAP program [22].
chromosome 7, it remains possible that a different CF mutant locus is being segregated in a small number of families. In this regard, family 19 seems to be particularly interesting.

Given the close genetic distance between CF and D7S8, it is difficult to interpret the inheritance of these two markers in family 19 where there are two apparent crossovers in a total of four meioses (see fig. 1). In addition, at least one double crossover would have had to occur to account for the marker inheritance patterns. One possible explanation is that family 19 segregates mutant CF genes at a second locus. A test of homogeneity based on the linkage data between CF and D7S8 from 47 informative Canadian and Danish families was therefore performed using the HOMOG computer program [21]. As shown in table 2, the hypothesis of heterogeneity (H2) appears to be statistically significant ($\chi^2 = 4.27, \text{d.f.} = 1; P = .02$). The analysis also suggests that there is an “unlinked group” comprising 3% of the families and the posterior probability of family 19 to be in the “linked” group is .05. Therefore, family 19 was included in this study only and deleted from the joint analysis. While these results are suggestive of heterogeneity in CF, confirmation of this hypothesis must await the identification and molecular characterization of the CF gene itself.

We also examined whether genetic heterogeneity was implicated by the clinical heterogeneity known to exist among CF patients [24]. Since both patients in family 19 do not require pancreatic enzyme supplements in their diet, they apparently do not have the malabsorption problem typical of the majority of patients. Upon analyzing all the families with patients not requiring enzymes (table 3), family 19 seems to be the only exception in the group; all other

<table>
<thead>
<tr>
<th>Marker Loci</th>
<th>Family</th>
<th>.00</th>
<th>.10</th>
<th>.20</th>
<th>.30</th>
<th>.40</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF vs. MET</td>
<td>3</td>
<td>2.31</td>
<td>1.81</td>
<td>1.29</td>
<td>0.74</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>2.43</td>
<td>1.9</td>
<td>1.34</td>
<td>0.76</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.30</td>
<td>0.21</td>
<td>0.13</td>
<td>0.06</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>$\infty$</td>
<td>-0.44</td>
<td>-0.19</td>
<td>-0.08</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>0.30</td>
<td>0.21</td>
<td>0.13</td>
<td>0.06</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>0.68</td>
<td>0.52</td>
<td>0.35</td>
<td>0.19</td>
<td>0.05</td>
</tr>
<tr>
<td>Total</td>
<td>$\infty$</td>
<td>4.21</td>
<td>3.05</td>
<td>1.73</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>CF vs. D7S8</td>
<td>3</td>
<td>2.31</td>
<td>1.81</td>
<td>1.29</td>
<td>0.74</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>2.43</td>
<td>1.9</td>
<td>1.34</td>
<td>0.76</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.60</td>
<td>0.43</td>
<td>0.27</td>
<td>0.13</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>$\infty$</td>
<td>-0.89</td>
<td>-0.39</td>
<td>-0.15</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>0.30</td>
<td>0.21</td>
<td>0.13</td>
<td>0.06</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>0.68</td>
<td>0.52</td>
<td>0.35</td>
<td>0.19</td>
<td>0.05</td>
</tr>
<tr>
<td>Total</td>
<td>$\infty$</td>
<td>3.98</td>
<td>2.99</td>
<td>1.72</td>
<td>0.53</td>
<td></td>
</tr>
</tbody>
</table>
families, including two large families each with four affected and multiple unaf-
fected children, showed strong linkage to both MET and D7S8. Division of
families based on geographic region or ethnic background also did not reveal
any differences (data not shown).

**Linkage Disequilibrium**

Given the tight linkage between CF and MET and D7S8, wenext examined
for the presence of nonrandom association in this region of chromosome 7.
Initial pairwise analyses revealed a strong linkage disequilibrium between CF
and MET and between CF and D7S8 [11]. To investigate this more closely, a
single restriction site was chosen from each of the three probes to construct
haplotypes consisting of three polymorphic sites (TaqI for both *pme1H* and
*pme1D*, which are greater than 10 kbases (kb) apart; *MspI* for D7S8). The
haplotype information was then derived where possible for each of the two
chromosomes in parents of CF as described [11]. As shown in table 4, there is a
striking difference in the haplotype distribution among the CF and normal
chromosomes in the Toronto (TOR) data set. In the combined TOR data, the
haplotype “1 1 1” is found to be represented over four times more fre-
frequently in the CF chromosome pool than in the normal (42% in CF vs. 11% in
normals). Conversely, the high frequency normal chromosome “1 1 1”
(36% in normals) contains the CF mutation in relatively low frequency (16%).
In an attempt to determine whether this observation was an artifact due to the
pooling of different populations, the French Canadian and Danish families were
analyzed separately from the rest of the TOR families. The result, as shown in
table 4, shows that the distribution of the haplotypes across subdivisions is
similar to that seen in the pooled data.

Linkage disequilibrium between the CF locus and marker haplotypes can be
assessed using a method suggested by Hedrick and Thomson [25]. In this
procedure, the homogeneity $\chi^2$ divided by the sample size and degrees of
freedom to obtain a standardized measure of linkage disequilibrium, $r^2$. The

---

**Table 4**

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>TaqI</th>
<th>TaqI</th>
<th>MspI</th>
<th>Total</th>
<th>French-Canadian</th>
<th>Other-Canadian</th>
<th>Danish</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Normal CF</td>
<td>Normal CF</td>
<td>Normal CF</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>36</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
<td>85</td>
<td>14</td>
</tr>
</tbody>
</table>

---

We thank A linkage disequi-
Zsiga, and Stel-
for sending the
U.S.A. for spec-
made the colla

---

1. **EIBERG H.,**
   Eighth Int
   40:623-19

2. **EIBERG H.,**
   ships of pi
   1985

3. **SCHMIEDEL**
   fibrosis an
   1985

4. **TSUI L-C.,**
   genetically
   1985

5. **KNOWLTON**
   marker M
   1985

6. **WHITE R.,**
   cystic fibri
   1985

7. **WAINWRIG**
   locus to 1
   1985

8. **SCAMBLEI**
   gene to c
   1985

9. **BUCHWAI**
   prosa2 (1)
   239, 1985

10. **SCAMBLEI**
    informati
    1985

11. **TSUI L-C.**
    locus on
square root of this value, r, is the correlation between the locus of interest and the marker haplotypes. The $r$ value observed for the TOR sample was .18 ($\chi^2 = 23.90, N = 155, P = .0001$).

Since linkage disequilibrium is not uncommon for closely linked genetic loci, the high degree of association between CF and the "$-$ $-$ $-$" haplotype described above may suggest that CF is in fact very close in physical distance from MET and D7S8. Furthermore, differences observed in the CF and normal haplotype distribution have implications in genetic counseling. Based on the TOR data set, individuals who are homozygous for the high frequency haplotype ("$-$ $-$ $-$") are expected to have an approximately 20% chance of being a CF carrier, as opposed to an average population risk of 1/20. A more detailed discussion on the subject of linkage disequilibrium is presented in the accompanying joint paper [12].

ACKNOWLEDGMENTS

We thank Aravinda Chakravati and Jeffrey C. Murray for helpful discussions on linkage disequilibrium; Natasa Plavsic, Danuta Markiewicz, Dana Kennedy, Martha Zsga, and Stefanie Zengerling for DNA analysis; Richard Rozmahel for assistance in computing; Ray White for providing the pmerH and pmerD probes; and Jorg Schmidtke for sending the pJ3.11 probe. We especially thank the Cystic Fibrosis Foundation of the U.S.A. for sponsoring a two-day meeting in Toronto (December 16–17, 1985), which made the collaborative study among the CF research groups possible.

REFERENCES