

Received 11 November; accepted 12 November 1985.

1. Talamo, R. C., Rosenstein, B. J. & Berninger, R. W. in *The Metabolic Basis of Inherited Disease* 5th edn (eds Stanbury, J. B. et al.) 1889-1917 (McGraw-Hill, New York, 1983).
2. Danks, D. M., Phelan, P. D. & Chapman, C. *Am. J. hum. Genet.* **36**, 1389-1402 (1984).
3. Romeo, G. et al. *Am. J. hum. Genet.* **37**, 338-349 (1985).
4. Tsui, L.-C. et al. *Science* (in the press).
5. Schmiegelow, K. et al. *Lancet* (in the press).
6. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. *Cell* **15**, 1157-1174 (1978).
7. Schumm, J. W. et al. *8th Human Gene Mapping Workshop*, Helsinki (1985); *Cytogenet. Cell Genet.* (in the press).
8. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. *Am. J. hum. Genet.* **32**, 314-331 (1980).
9. Nguyen Van Cong, Weil, D. & Finaz, C. *Annls Génét.* (in the press).
10. Grzeschik, K. H. *Somatic Cell Genet.* **2**, 401-410 (1976).
11. Jones, N. et al. *Science* **227**, 311-314 (1985).
12. Barker, P. E., Ruddle, F. H., Royer, H.-D., Acuto, O. & Reinherz, E. L. *Science* **226**, 348-349 (1984).
13. Collins, M. K. L. et al. *EMBO J.* **3**, 2347-2349 (1984).
14. Isobe, M., Erikson, J., Emanuel, B. S., Nowell, P. C. & Croce, C. *Science* **228**, 580-582 (1985).
15. Morton, C. C., Duby, A. D., Eddy, R. L., Shows, T. B. & Seidman, J. G. *Science* **228**, 582-584 (1985).
16. Shows, T. B. *Biochem. Genet.* **7**, 193-204 (1972).
17. Gedde-Dahl, T. et al. *8th Human Gene Mapping Workshop*, Helsinki (1985); *Cytogenet. Cell Genet.* (in the press).
18. Meera Khan, P. & Robson, E. B. *Human Gene Mapping* **5**, 39-46 (1979).
19. Wahl, G. M., Stern, M. & Stark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3683-3687 (1979).
20. Gross, M.-S. et al. *Annls Génét.* **25**, 87-91 (1982).
21. Weil, D. et al. *Annls Génét.* **24**, 89-92 (1981).
22. Retief, E., Parker, M. I. & Retief, A. E. *Hum. Genet.* **69**, 304-308 (1985).

A closely linked genetic marker for cystic fibrosis

Ray White*, Scott Woodward*, Mark Leppert*, Peter O'Connell*, Mark Hoff*, John Herbst†, Jean-Marc Lalouel*, Michael Dean‡ & George Vande Woude‡

* Howard Hughes Medical Institute and the Cystic Fibrosis Research Center, Department of Human Genetics and † Department of Pediatrics, University of Utah College of Medicine, Salt Lake City, Utah 84132, USA

‡ Litton Bionetics, Inc., Basic Research Program, NCI-Frederick Cancer Research Facility, Frederick, Maryland 21701, USA

Cystic fibrosis is a recessive genetic disorder, characterized clinically by chronic obstructive lung disease, pancreatic insufficiency and elevated sweat electrolytes; affected individuals rarely live past their early twenties. Cystic fibrosis is also one of the most common genetic diseases in the northern European population. The frequency of carriers of mutant alleles in some populations is estimated to be as high as 1 in 20, carrying a concomitant burden of about one affected child in 1,500 births. Because little is known of the essential biochemical defect caused by the mutant gene, a genetic linkage approach based on arbitrary genetic markers and family studies is indicated to determine the chromosomal location of the cystic fibrosis (CF) gene. We have now obtained evidence for tight linkage between the CF locus and a DNA sequence polymorphism at the *met* oncogene locus. This evidence, combined with the physical localization data for the *met* locus presented in the accompanying paper¹, places the CF locus in the middle third of the long arm of chromosome 7, probably between bands q21 and q31.

Appropriate families are required for genetic linkage analyses with recessive disease loci. For this study, we sampled and tested 13 families with multiple affected offspring. Siblings were characterized at cystic fibrosis clinical centres as affected or unaffected based on quantitative sweat chloride determination as well as other clinical features. Several of the families were from Utah, but most came from different regions in the United States.

We estimate that more than 100 loci have been tested for linkage in CF, including 21 probes from our own laboratory, with negative results. However, Eiberg et al. have reported evidence of linkage between the CF locus and the gene for the enzyme paroxonase², and recently Tsui and others have obtained evidence for genetic linkage with a DNA marker (L.-C. Tsui, personal communication). Both of these linkages are rather

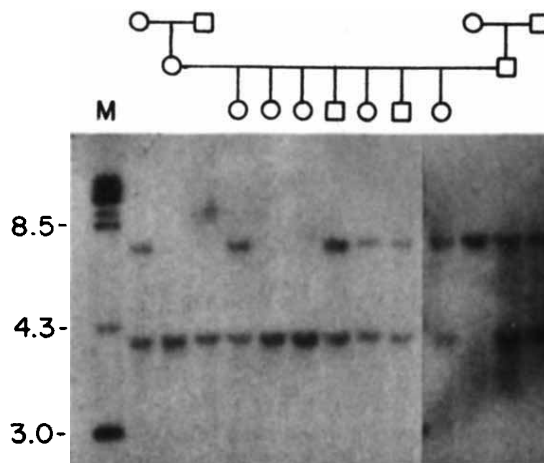


Fig. 1 Segregation of the *TaqI* polymorphism in a complete three-generation family. DNAs from the indicated individuals were digested with the restriction enzyme *TaqI*, electrophoresed in agarose and transferred by the method of Southern to a nylon filter (Micron Separations, Inc.). The filters were hybridized⁹ with the clone *metH* (refs 1, 3). The polymorphisms were found by examination of a panel of DNAs from six unrelated individuals digested with the restriction enzymes *MspI*, *TaqI*, *EcoRI*, *HindIII*, *PstI* or *BclI*. Only *MspI* and *TaqI* revealed polymorphism. Individuals digested with *TaqI* yielded fragment lengths of either 7.5 kb (allele 1) alone or 4.0 kb (allele 2) alone, or both 7.5 and 4.0 kb. Digestion with *MspI* revealed individuals with fragment lengths of 2.3 and 1.8 kb, or only one of the two. In six informative families with large sibships, we found no exceptions to mendelian inheritance. A survey of 60 unrelated individuals gave a frequency of 0.56 for the 7.5-kb *TaqI* allele and the 2.3-kb *MspI* fragment, and 0.44 for each of the two smaller fragments. In all cases where phase could be distinguished, the large fragments were found together as a single haplotype, indicating a high degree of linkage disequilibrium in this marker system.

loose (10% and 15% recombination, respectively) but they do begin to pave the way for a more precise localization. Klinger has also reported suggestive, but not definitive, evidence for linkage of cystic fibrosis to a genetic marker located on the short arm of chromosome 21 in two large Amish pedigrees. However, the linkage was not evident when a collection of outbred nuclear families was examined (Klinger, personal communication). To develop a genetic marker at the *met* locus, we cloned fragment H (refs 1, 3), a 1.6-kilobase (kb) *SalI/EcoRI* fragment, into pBR322. Figure 1 shows the inheritance of a polymorphic *TaqI* restriction fragment revealed by this probe in a complete three-generation family. An *MspI* fragment also showed polymorphism with this probe. Of 60 unaffected unrelated individuals examined, 36 were heterozygous. However, we have seen only two haplotypes thus far.

At least one parent was heterozygous at the *met* locus in 12 of the 13 CF families investigated. Linkage tests between the CF locus and the *met* locus were performed with the computer program LINKAGE (ref. 4). Our results are reported in Fig. 2 and Table 1. As shown in Fig. 2, the LOD score reaches a maximum of 8.65 at a recombination value of 0, indicating no evidence for recombination. A LOD score of 8.65 corresponds to odds of $4 \times 10^8:1$, favouring the hypothesis of complete linkage over that of independent segregation. We also verified that in each family the LOD score is maximal under the hypothesis of complete linkage. As in all estimation situations, our sample estimate of the recombination rate between these two loci may depart from the true, unknown value because of sampling error. As indicated in Fig. 2, however, the true recombination frequency is unlikely to be $>5\%$. Our data therefore indicate tight linkage between cystic fibrosis and the *met* gene in these families.

Table 1 shows the composition, the genotypes at the *met* locus and the LOD score for each family. Under the hypothesis of

Table 1 Summary of family data

Family	Parents	Offspring	LOD score
1409	12 × 11	3A (11)	0.60
1414	12 × 22	3A (12)	0.60
1415	12 × 12	3A (11) 1N (12)	1.33
1422	12 × 11	3A (12)	0.60
1425	12 × 11	5A (12) 2N (12, 11)	1.15
1426	22 × 12	3A (12)	0.60
1427	12 × 12	3A (11) 4N (12)	1.70
1438	11 × 12	2A (11)	0.30
1442	12 × 11	3A (11)	0.60
1446	11 × 12	3A (11)	0.60
774	11 × 12	2A (11)	0.30
1378	22 × 12	3A (22) 2N (22)	0.25
1436	11 × 22	3A (12)	0.00

Genotypes at the marker locus reported for each family member. The LOD score for each informative family at a recombination value of 0 is also reported, allowing a closer inspection of the contribution of each family to the linkage evidence. Assuming that all parents are heterozygous at the CF locus, all but two (families 1415, 1427) are single backcrosses with respect to the test locus. In such families, each affected offspring contributes terms equal to θ and $1 - \theta$ under each parental phase respectively. Under complete linkage, phase information can be derived from a single affected individual. It follows that each additional individual will contribute 2 units to the odds ratio, or equivalently 0.301 to the LOD score, while the penalty of having to infer phase amounts to 0.301 LOD unit in each family. In two instances of such crosses, unaffected siblings were available for testing. In family 1425, five affected siblings have genotype 12, whereas two unaffected siblings have genotypes 12 and 11, respectively. Without the unaffected siblings, the LOD score would have been 1.20 instead of 1.15. The two unaffected siblings, while compatible with the hypothesis of no recombination, have discordant genotypes at the test locus and contribute terms $(2 - \theta)(1 + \theta)$ under each parental phase, accounting for -0.05 in LOD units. In family 1378, the contribution of three affected offspring is offset by two unaffected siblings with the same genotype at the *met* locus, each unaffected sibling accounting for terms $(2 - \theta)$ and $(1 + \theta)$ under each phase. Family 1427 is a double intercross and contributes more support to the hypothesis of complete linkage than any other family. This results from the increased contribution of affected individuals of genotype 11 in this intercross as opposed to the backcrosses, as well as from the information provided by the four unaffected siblings of genotype 12. Once an affected individual has yielded phase, each additional affected sibling of genotype 11 contributes $(1 - \theta)^2$, or 0.602 LOD units, while each unaffected sibling of genotype 12 contributes $2(1 - \theta + \theta^2)$, or 0.125, to the overall LOD score. Family 1415 is similar, but includes only one unaffected sibling of genotype 12.

complete linkage, we can derive the parental phases from the genotypes of the affected offspring. This reveals that 19 of the 26 parental chromosomes carrying the CF allele also carry allele 1 at the *met* locus, while only 11 of 26 parental chromosomes carrying the normal allele at the CF locus carry allele 1 at the *met* locus, which is marginally significant ($\chi^2_1 = 3.86$, $p = 0.05$). The allelic distribution at the *met* locus among the latter chromosomes does not differ significantly from that observed among 60 chromosomes tested in a random Utah control panel ($\chi^2_1 = 0.98$). When these chromosomes carrying a normal allele at the CF locus are pooled and contrasted to the chromosomes carrying the CF deleterious allele, heterogeneity with respect to the *met* alleles 1 and 2 is no longer significant ($\chi^2_1 = 2.71$). Judgment regarding possible allelic association must await the characterization of additional CF families.

The apparent close genetic linkage found in family studies suggests an upper limit of 5% on the genetic distance between the CF locus and the *met* locus. The absence of recombinants between the CF locus and the *met* locus in our study does not support the hypothesis that cystic fibrosis can be caused by mutations at more than one locus. However, ours is still a limited sample and it is important to test many more families to place a more restrictive upper limit on the possible frequency of cystic fibrosis caused by mutations at other loci.

Although the *met* locus is potentially useful as a prenatal diagnostic marker for cystic fibrosis in families known to be at risk, several additional data must be developed before such application. First, more CF families with large sibships must be tested, to refine the genetic distance as well as to detect locus heterogeneity. Observation of a subset of families showing no linkage to the *met* locus could reveal locus heterogeneity. These data will be essential for providing quantitative estimates of

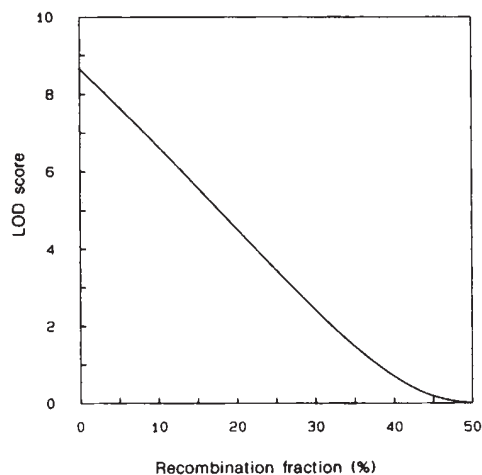


Fig. 2 LOD score for linkage between cystic fibrosis and the *met* locus as a function of recombination fraction. A recessive model with complete penetrance was assumed for cystic fibrosis, with a frequency of 0.025 for the deleterious allele. The results were unaltered over a wide range of allelic frequencies. The test marker was treated as a co-dominant, two-allele system, with gene frequencies of 0.56 and 0.44 for alleles 1 and 2, respectively. The *TaqI* polymorphism was scored as the primary genetic marker and was corroborated by the *MspI* polymorphism. The LOD score is computed as the decimal logarithm of the probability of the observations for a given value of the recombination fraction divided by the corresponding probability under the assumption of no linkage. This probability ratio measures our degree of belief in the hypothesis of linkage at the assumed value relative to the hypothesis of independent segregation. Rather than a standard error based on large sample theory, the use of an empirical confidence region based on the observed LOD score distribution has been recommended¹⁰. A one-unit support limit is obtained by finding that value of the recombination fraction for which the LOD score is decreased by one unit, defining a range of the recombination value for which the odds remain within 10 to 1 of the maximum. This yields an empirical bound for the recombination fraction of 0.051 in our sample. Values for recombinant fraction θ of 0, 0.01, 0.05, 0.10, 0.20, 0.30 and 0.40 produced respectively values for $z(\theta)$ of 8.65, 8.46, 7.66, 6.64, 4.52, 2.41 and 0.70.

individual risk. Second, the extent of polymorphism must be improved so that the marker locus has more than two alleles. At present, many individuals at risk would be unable to obtain useful information. Third, if, as we expect, some recombinants between the *met* and CF loci are found, it will be important to develop a genetic marker locus on the other side of the CF locus to detect recombination events in the region. Until these conditions are met, diagnostic applications are premature.

The fact that no recombinants between the *met* gene locus and the CF locus have yet been identified suggests that the two are indeed very close and even raises speculation that the *met* gene might actually be the cystic fibrosis gene. Furthermore, because the product of the *met* gene is a membrane-spanning receptor protein showing tyrosine kinase homology¹, characteristics which might also be expected of a gene causing a chloride ion-exchange defect, in retrospect *met* might have been considered a candidate gene for the disease. However, although we cannot exclude the *met* gene as a candidate for the CF gene, aetiological involvement of the *met* gene in cystic fibrosis is unlikely: the close linkage still permits a span of more than 1,000 kb within which to place the CF gene and as many as a hundred genes may be located within so wide a region. It is nonetheless interesting that the *met* gene bears homology with the insulin receptor¹, located on chromosome 19 very near the low-density lipoprotein (LDL) receptor²; this homology may suggest that these are members of a clustered family of genes. By this reasoning, although not itself the CF gene, *met* could

open the way to identifying a family of closely linked genes of which one may be the gene for cystic fibrosis.

The most significant aspect of the localization of the CF gene is the impetus it provides to efforts to identify and clone the gene. New techniques make it feasible to cover the large distances that are likely to be involved. Pulse-field gel electrophoresis⁶ and new, rapid 'walking' techniques⁷ suggest approaches to moving closer to the gene. Recent advances in cell culture technology that permit *in vitro* growth of cells which maintain the CF phenotype⁸ suggest the possible invention of protocols that could provide functional assays for regions that might encompass the gene. We hope that identification of the CF gene will permit identification and analysis of the biochemical pathway which is perturbed by mutations in the gene, and will suggest rational means of intervention and management of this disease.

We thank Barbara Ogden for assistance in family collection, Tara Bragg, John Cowan, Russell Eldridge, Mary Hadley, Leslie

Jerominski, Lesa Nelson and Leslie Sargent for technical assistance, Carol Wood of the Intermountain Cystic Fibrosis Center for assistance in family collection, the NIGMS Human Genetic Mutant Cell Repository for lymphoblastoid cell lines of family 774, and Ruth Foltz for editorial assistance in preparing this manuscript. M.D. is an American Cancer Society postdoctoral fellow.

Received 1 November; accepted 6 November 1985.

1. Dean, M. *et al.* *Nature* 318, 385-388 (1985).
2. Eiberg, H. *et al.* 8th int. Workshop Human Gene Mapping, Helsinki, Abstr. 100 (Karger, New York, in the press).
3. Cooper, C. S. *et al.* *Nature* 311, 29-33 (1984).
4. Lathrop, G. M. & Lalouel, J.-M. *Am. Soc. hum. Genet.* 36, 460-465 (1984).
5. 8th int. Workshop Human Gene Mapping, Helsinki, draft rep. 223 (Karger, New York, in the press).
6. Schwartz, D. & Cantor, C. *Cell* 37, 67-75 (1984).
7. Collins, F. & Weissman, S. *Proc. natn. Acad. Sci. U.S.A.* 81, 6812-6816 (1984).
8. Stutts, M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* 82, 6677-6681 (1985).
9. Barker, D., Holm, T. & White, R. *Am. J. hum. Genet.* 36, 1159-1171 (1984).
10. Conneally, P. *et al.* 8th Int. Workshop Human Gene Mapping, Helsinki (Karger, New York, in the press).

Localization of cystic fibrosis locus to human chromosome 7cen-q22

Brandon J. Wainwright*, Peter J. Scambler*, Jorg Schmidtke†, Eila A. Watson*, Hai-Yang Law*, Martin Farrall*, Howard J. Cooke‡, Hans Eiberg§ & Robert Williamson*

* Department of Biochemistry, St Mary's Hospital Medical School, University of London, London W2 1PG, UK

† Institut für Humangenetik der Universität, Gosslerstrasse 12d, D-3400 Gottingen, FRG

‡ MRC Mammalian Genome Unit, West Mains Road, Edinburgh EH9 3JJ, UK

§ Institute of Medical Genetics, Panum Institute, Copenhagen, Denmark

Cystic fibrosis (CF) is the most common genetic disease in Caucasian populations, with an incidence of 1 in 2,000 live births in the United Kingdom, and a carrier frequency of approximately 1 in 20. The biochemical basis of the disease is not known¹, although membrane transport phenomena associated with CF have been described recently². Consanguinity studies have shown that the inheritance of CF is consistent with it being a recessive defect caused by a mutation at a single autosomal locus³. Eiberg *et al.*⁴ have reported a genetic linkage between the CF locus and a polymorphic locus controlling activity of the serum aryl esterase paraoxonase (PON). The chromosomal location of PON, however, is not known⁵. Linkage to a DNA probe, DOCRI-917, was also recently found at a genetic distance of ~15 centimorgans (L.-C. Tsui and H. Donniss-Keller, personal communication), but no chromosomal localization was given. Here we report tight linkage between the CF locus and an anonymous DNA probe, pJ3.11, which has been assigned to chromosome 7cen-q22.

The use of DNA probes allows the determination of linkage between a restriction fragment length polymorphism (RFLP) and a mutated locus causing pathology, even when the biochemical defect is unknown. We have used this approach successfully to analyse sex-linked diseases such as Duchenne mus-

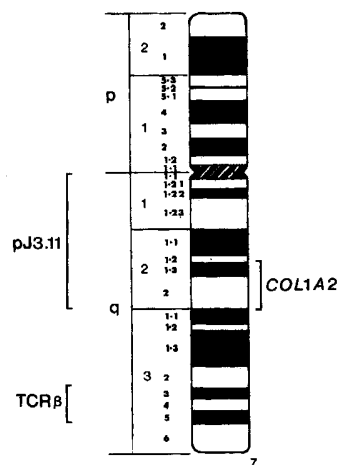


Fig. 1 Map of human chromosome 7, showing the chromosomal locations of probes *TCRβ* (ref. 9), *COL1A2* (ref. 25) and pJ3.11 (ref. 18).

cular dystrophy⁵; it has also been used to study previously unassigned autosomal dominant diseases such as Huntington's chorea⁶ and adult onset polycystic kidney disease⁷. However, although linkage has been attempted with DNA markers to autosomal recessive diseases of unknown biochemical aetiology, only exclusions have been achieved to date.

There have been few clues to the chromosomal localization of the CF locus. Mayo *et al.*⁸ suggested that the CF locus might be localized on chromosome 4, based on data from cell hybrids expressing a ciliary dyskinesia factor, but this was subsequently shown to be unlikely by exclusion mapping with several DNA and protein markers⁹. Two different families have been reported to show both chromosomal anomalies and CF; the study of one of these families suggested that the CF locus mapped to 5p, whereas in the other family it was mapped to 13q34 (ref. 10). We investigated the latter family by mapping the gene for clotting factor X to 13q34 (ref. 11) and then excluded CF from linkage to factor X¹². We have also excluded the candidate gene coding for complement component 3 by showing that alleles for an

Table 1 LOD scores at various recombination fractions (θ) for the relationships between the CF locus (*CF*) and the locus defined by pJ3.11 and between *CF* and *TCRβ*

Marker	No. of meioses	$\theta = 0$	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
<i>CF</i> versus J3.11	27	5.24	4.52	3.80	3.09	2.41	1.77	1.20	0.77	0.33	0.11
<i>TCRβ</i> versus <i>CF</i>	22	−∞	1.87	2.06	1.87	1.54	1.17	0.80	0.47	0.22	0.06

Combined LOD scores were calculated using the computer program package LINKAGE as described by Lathrop²⁴.