Am J Hum Genet 39:720–728, 1986

Genetic Analysis of Cystic Fibrosis Using Linked DNA Markers

LAP-CHEE TSUI,[1] KEN BUETOW,[2] AND MANUEL BUCHWALD[1]

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.

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 proα2 (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 (p*met*H) [6, 16] and p*met*D ([16], and R. White, personal communication) and D7S8 (pJ3.11) [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 LIPED [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

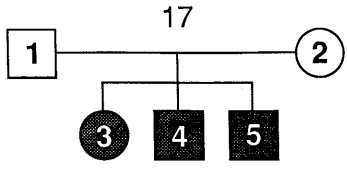
The results of pairwise two-point linkage analyses between CF and various genetic markers are shown in table 1. Although the maximal likelihood estimates for θ derived from the updated family data differ slightly from those previously reported [4, 9, 11], the new values are well within the confidence

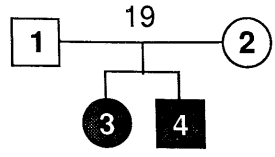
TABLE 1

LINKAGE RELATIONSHIPS BETWEEN CF AND FLANKING MARKERS

MARKER LOCI	No. INFORMATIVE FAMILIES*	LOD (Z) SCORES AT RECOMBINATION FRACTION (θ) OF									θ_{max}	z_{max}	CONFIDENCE INTERVALS
		.01	.05	.10	.15	.20	.25	.30	.35	.40			
CF-D7S15	67†	−21.52	−0.77	5.09	6.61	6.43	5.43	4.05	2.58	1.26	0.17	6.68	.11–.24
CF-PON	31	−1.83	2.02	2.85	2.80	2.42	1.90	1.34	0.82	0.39	0.12	2.88	...
CF-COL1A2	51	−15.78	−1.50	2.62	3.77	3.77	3.21	2.4	1.52	0.74	0.17	3.87	.11–.27
CF-TCRB	31	−14.80	−4.73	−1.32	0.08	0.66	0.81	0.73	0.52	0.27	0.25	0.81	...
CF-MET	46	23.77	22.31	19.33	16.12	12.9	9.74	6.76	4.1	1.93	0.01	23.77	.001–.05
CF-D7S8	47	21.05	19.73	16.97	14.04	11.13	8.35	5.78	3.50	1.66	0.01	21.05	.001–.04
D7S15-PON	20	5.36	6.48	6.15	5.45	4.59	3.66	2.71	1.78	0.94	0.06	6.47	.01–.15
D7S15-COL1A2	30	11.02	13.25	12.58	11.14	9.38	7.48	5.51	3.58	1.83	0.05	13.25	.02–.11
D7S15-TCRB	18	−8.15	−1.93	0.1	0.87	1.12	1.1	0.91	0.63	0.33	0.22	1.14	...
D7S15-MET	28	−16.17	−2.69	1.53	2.99	3.32	3.05	2.43	1.65	0.85	0.20	3.32	.12–.30
D7S15-D7S8	21	−17.01	−4.04	0.15	1.73	2.25	2.21	1.85	1.31	0.71	0.22	2.28	...
COL1A2-PON	21	−1.90	1.47	2.25	2.29	2.05	1.67	1.23	0.79	0.39	0.13	2.33	...
COL1A2-TCRB	22	−14.80	−4.68	−1.20	0.24	0.86	1.03	0.92	0.66	0.35	0.25	1.03	...
COL1A2-MET	29	−9.38	−0.33	2.38	3.2	3.25	2.88	2.27	1.54	0.81	0.18	3.29	.10–.30
COL1A2-D7S8	32	−14.16	−3.05	0.5	1.78	2.15	2.04	1.66	1.13	0.59	0.21	2.16	...
TCRB-PON	8	−7.08	−3.15	−1.66	−0.84	−0.52	−0.28	−0.13	−0.06	−0.02	0.50	0.00	...
TCRB-MET	12	−3.11	0.49	1.54	1.81	1.76	1.53	1.19	0.8	0.41	0.16	1.82	...
TCRB-D7S8	17	−3.04	1.10	2.25	2.5	2.36	2.02	1.55	1.03	0.53	0.15	2.5	...
MET-PON	16	−8.13	−1.85	0.25	1.07	1.37	1.36	1.17	0.86	0.49	0.22	1.39	...
MET-D7S8	27	23.65	21.42	18.57	15.68	12.78	9.9	7.09	4.46	2.19	0.00	24.21	.00–.02
D7S8-PON	15	−9.24	−2.91	−0.73	0.18	0.57	0.67	0.6	0.43	0.23	0.25	0.67	...

* D7S15, PON, COL1A2, MET, and D7S8 data derived from a cohort of 47 Canadian and 13 Danish CF families;
† Including three from HGCMR, four from London, five from Salt Lake City, and nine from Stanford-Berkeley.





COL1A2- EcoR I:	A A	A B	A B	A A	A B
RSA I:	A B	B B	A B	A B	B B
Msp I:	B B	B B	B B	B B	B B
D7S15- Hind III:	A A	A B	A B	A A	A B
HINC II:	B A	A A	B A	B B	B A
PON-	1 1	1 2	1 1	1 1	1 2
MET-(H) Taq I:	A B	B B	A B	A B	B A
(D) Taq I:	A A	A A	A A	A A	A B
Ban I:	A B	B B	A B	A B	B B
D7S8- Msp I:	A A	A B	A B	A B	B A
Taq I:	A A	A A	A A	A A	A A
TCRB- Bgl II:	B B	B A	B A	B A	A A

A A	A A	A A	A A
A B	B B	B B	B B
B B	B B	B B	B B
B B	B A	B A	A B
B B	B A	B A	A B
1 1	1 1	1 1	1 1
A A	A A	A A	A A
A B	B B	A B	B B
A B	B B	A B	B B
A B	B B	A A	A B
A A	A A	A A	A A
A B	A B	B B	A B

FIG. 1.—Chromosome 7 marker segregation in Toronto family 17 and 19

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 2

HOMOGENEITY TEST

	Hypotheses*	Max. lod	α	θ	Components of χ^2 Source		d.f.	χ^2	*P*
CF vs. D7S8	H2	52.171	0.97	.01	H2 vs. H1	Heterogeneity	1	4.267	.0194
	H1	50.037	(1)	.01	H1 vs. H0	Linkage	1	100.075	.0000
	H0	(0)	(0)	(.5)	H2 vs. H0	Total	2	104.342	.0000
CF vs. MET	H2	52.784	0.98	.01	H2 vs. H1	Heterogeneity	1	0.105	.3727
	H1	52.732	(1)	.01	H1 vs. H0	Linkage	1	105.463	.0000
	H0	(0)	(0)	(.5)	H2 vs. H0	Total	2	105.568	.0000

* 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 HOMOG program [22].

TABLE 3

LOD SCORES FOR CF VS. MET AND CF VS. D7S8 IN FAMILIES WITH NO APPARENT MALABSORPTION

MARKER LOCI	FAMILY	LOD SCORES AT θ OF				
		.00	.10	.20	.30	.40
CF vs. MET	3	2.31	1.81	1.29	0.74	0.23
	14	2.43	1.9	1.34	0.76	0.23
	15	0.30	0.21	0.13	0.06	0.02
	19	$-\infty$	−0.44	−0.19	−0.08	−0.02
	27	0.30	0.21	0.13	0.06	0.02
	40	0.68	0.52	0.35	0.19	0.05
Total		$-\infty$	4.21	3.05	1.73	0.53
CF vs. D7S8	3	2.31	1.81	1.29	0.74	0.23
	14	2.43	1.9	1.34	0.76	0.23
	15	0.60	0.43	0.27	0.13	0.03
	19	$-\infty$	−0.89	−0.39	−0.15	−0.04
	27	0.30	0.21	0.13	0.06	0.02
	40	0.68	0.52	0.35	0.19	0.05
Total		$-\infty$	3.98	2.99	1.72	0.53

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 ($\chi^2 = 4.27$, 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 4

HAPLOTYPES OF NORMAL AND CF CHROMOSOMES

HAPLOTYPE			No. CHROMOSOMES							
			TOTAL		FRENCH-CANADIAN		OTHER CANADIAN		DANISH	
pmetH *TaqI*	pmetD *TaqI*	D7S8 *MspI*	Normal	CF	Normal	CF	Normal	CF	Normal	CF
–	–	–	8	36	4	10	3	20	1	6
–	–	+	13	20	1	2	12	13	0	5
+	–	+	25	14	6	3	17	10	2	1
+	–	–	11	8	2	0	9	7	0	1
–	+	+	11	5	1	1	10	4	0	0
–	+	–	2	2	0	0	2	1	0	1
Total			70	85	14	16	53	55	3	14

families, including two large families each with four affected and multiple unaffected 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, we next 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 (*Taq*I for both p*met*H and p*met*D, which are greater than 10 kilobases (kb) apart; *Msp*I 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 "– – –" is found to be represented over four times more frequently in the CF chromosome pool than in the normal (42% in CF vs. 11% in normals). Conversely, the high frequency normal chromosome "+ – +" (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 χ^2 is divided by the sample size and degrees of freedom to obtain a standardized measure of linkage disequilibrium, r^2. The

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 (χ^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 Charkravati and Jeffrey C. Murray for helpful discussions on linkage disequilibrium; Natasa Plavsic, Danuta Markiewicz, Dara Kennedy, Martha Zsiga, and Stefanie Zengerling for DNA analysis; Richard Rozmahel for assistance in computing; Ray White for providing the p*met*H and p*met*D 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

1. Eiberg H, Schmiegelow K, Tsui L-C, et al.: Cystic fibrosis, linkage with PON. Eighth International Workshop on Human Gene Mapping. *Cytogenet Cell Genet* 40:623, 1985
2. Eiberg H, Mohr J, Schmiegelow K, Nielsen LS, Williamson R: Linkage relationships of paraoxonase (PON) with other markers: indication of PON-cystic fibrosis synteny. *Clin Genet* 28:265–271, 1985
3. Schmiegelow K, Eiberg H, Tsui L-C, et al.: Linkage between the loci for cystic fibrosis and paraoxonase. *Clin Genet* 29:374–377, 1986
4. Tsui L-C, Buchwald M, Barker D, et al.: Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science* 230:1054–1057, 1985
5. Knowlton RG, Cohen-Haguenauer O, Nguyen VC, et al.: A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. *Nature* 318:380–382, 1985
6. White R, Woodward S, Leppert M, et al.: A closely linked genetic marker for cystic fibrosis. *Nature* 318:382–384, 1985
7. Wainwright BJ, Scambler PJ, Schmidtke J, et al.: Localization of cystic fibrosis locus to human chromosome 7cen-q22. *Nature* 318:384–385, 1985
8. Scambler PJ, Wainwright BJ, Farrall M, et al.: Linkage of COL1A2 collagen gene to cystic fibrosis and its clinical implications. *Lancet* ii:1241–1242, 1985
9. Buchwald M, Zsiga M, Markiewicz D, et al.: Linkage of cystic fibrosis to the proα2 (1) collagen gene, COL1A2, on chromosome 7. *Cytogenet Cell Genet* 41:234–239, 1986
10. Scambler PJ, Wainwright BJ, Watson E, et al.: Isolation of a further anonymous informative DNA sequence from chromosome seven closely linked to cystic fibrosis. *Nucleic Acids Res* 14:1951–1961, 1986
11. Tsui L-C, Zengerling S, Willard HF, Buchwald M: Mapping of the cystic fibrosis locus on chromosome 7. *Cold Spring Harbor Symp Quant Biol.* In press, 1986

12. BEAUDET A, BOWCOCK A, BUCHWALD M: Linkage of cystic fibrosis to two tightly linked DNA markers: joint report from a collaborative study. *Am J Hum Genet* 39:681–693, 1986
13. TSIPOURAS P, BØRRESEN AL, DICKSON LA, BERG K, PROCKOP DJ, RAMIREZ F: Molecular heterogeneity in the mild autosomal dominant forms of osteogenesis imperfecta. *Am J Hum Genet* 36:1172–1179, 1984
14. MEYERS JC, CHU M-L, FARO SH, CLARK WJ, PROCKOP DJ, RAMIREZ F: Cloning a cDNA for the proα2 chain of human type I collagen. *Proc Natl Acad Sci USA* 78:3516–3520, 1981
15. YANAGI Y, YOSHIKAI Y, LEGGETT K, CLARK SP, ALEKSANDER I, MAK TW: A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. *Nature* 308:145–149, 1984
16. DEAN M, PARK M, LE BEAU MM, ET AL.: The human met oncogene is related to the tyrosine kinase oncogenes. *Nature* 318:385–388, 1985
17. BARTELS I, GRZESCHIK K-H, COOPER DN, SCHMIDTKE J: Regional mapping of six cloned DNA sequences on human chromosome 7. *Am J Hum Genet* 38:280–287, 1986
18. MORTON NE: Sequential tests for the detection of linkage. *Am J Hum Genet* 7:277–318, 1955
19. OTT J: Estimation of the recombinant fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. *Am J Hum Genet* 26:588–597, 1974
20. LATHROP GM, LALOUEL J-M, JULIER C, OTT J: Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 81:3443–3446, 1984
21. OTT J: Linkage analysis and family classification under heterogeneity. *Ann Hum Genet* 47:311–320, 1983
22. CONNEALLY PM, EDWARDS JH, KIDD KK, ET AL.: Report of the Committee on Methods of Linkage Analysis and Reporting. Eighth International Workshop on Human Gene Mapping. *Cytogenet Cell Genet* 40:356–359, 1985
23. THOMPSON MW: Genetics of cystic fibrosis, in *Perspectives in Cystic Fibrosis,* edited by STURGESS JM, Toronto, Canadian Cystic Fibrosis Foundation, 1980, pp 281–291
24. SING CF, RISSER DR, HOWATT WF, ERICKSON RP: Phenotypic heterogeneity in cystic fibrosis. *Am J Med Genet* 13:179–195, 1982
25. HEDRICK PW, THOMSON G: A two-locus neutrality test: applications to humans, *E. coli* and lodgepole pine. *Genetics* 112:135–156, 1986