

Linkage of cystic fibrosis to the pro α 2(I) collagen gene, COL1A2, on chromosome 7

M. Buchwald,^{1,2} M. Zsiga,¹ D. Markiewicz,¹ N. Plavsic,¹ D. Kennedy,¹ S. Zengerling,¹ H.F. Willard,^{1,2} P. Tsipouras,³ K. Schmiegelow,⁴ M. Schwartz,⁴ H. Eiberg,⁵ J. Mohr,⁵ D. Barker,⁶ H. Donis-Keller,⁶ and L.-C. Tsui^{1,2}

¹ Department of Genetics, The Hospital for Sick Children, Toronto, ² Departments of Medical Genetics and Medical Biophysics, University of Toronto, ³ Departments of Pediatrics, UMDNJ-Rutgers Medical School, Piscataway, NJ, ⁴ Department of Clinical Genetics, University of Copenhagen, ⁵ University Institute of Medical Genetics, Copenhagen, and ⁶ Department of Human Genetics, Collaborative Research, Inc., Lexington, MA

Abstract. A linkage has been detected between the locus for cystic fibrosis (CF) and the pro α 2(I) collagen gene (COL1A2) which is located in the region q21.3→q22.1 of chromosome 7. Based on the combined linkage data derived from 50 informative two-generation nuclear families collected in Canada and Denmark, the distance between COL1A2 and CF is estimated to be 19 centiMorgans. Close linkage has also been detected between COL1A2 and the DNA marker D7S15 (formerly D0CRI-917) and the serum enzyme activity marker paraoxonase (PON), both of which have previously been found linked to CF. The results of the two-point and three-point linkage analyses indicate that the most probable order of these four genetic loci is COL1A2 - D7S15 - PON - CF.

Cystic fibrosis (CF) is the most common severe autosomal recessive disorder in Caucasians (reviewed in Talamo et al., 1983). While the gene defect in most human genetic diseases has been identified through biochemical studies of patient tissues or cells, this approach has not been successful in the case of CF. An alternative approach is to locate the CF gene by identification of a closely linked chromosomal marker. A large number of markers have been studied over the past years, covering a significant portion of the human genome (Steinberg et al., 1956; Steinberg and Morton, 1956; Goodchild et al., 1976; Beaudet et al., 1985; Scambler et al., 1985a, 1986; Tsui et al., 1985b,c; Watkins et al., 1985). Recently, Eiberg and co-workers (Eiberg et al., 1985; Schmiegelow et al., 1986) have discovered a linkage between the disease locus,

CF, and a polymorphic genetic determinant for serum paraoxonase activity, PON. The estimated genetic distance between CF and PON is approximately 10 centiMorgans (cM); however, the chromosomal location of PON has yet to be determined. We have also discovered a linkage between CF and a DNA locus defined by two restriction fragment length polymorphisms (RFLPs) detected with a randomly isolated DNA probe Lam4-917 (Tsui et al., 1985a). The estimated distance between this RFLP locus, D0CRI-917, and CF is 15 cM and that between D0CRI-917 and PON is 5 cM. Three point linkage analysis data indicate the order of the three genetic loci to be D0CRI-917 - PON - CF.

Subsequently, by hybridization analysis using DNA of rodent-human cell hybrids, Knowlton et al. (1985) have assigned D0CRI-917 to chromosome 7. Therefore the locus D0CRI-917 has been renamed as D7S15. Since CF is genetically linked to D7S15 (Tsui et al., 1985a), it is reasoned that CF is also on chromosome 7 (Knowlton et al., 1985). To corroborate this finding and to further define the location of CF, we have initiated CF linkage analysis with other markers known to be on chromosome 7. Here we

Supported by grants from the Canadian Cystic Fibrosis Foundation (CCFF); the National Institutes of Health, USA; the Medical Research Council of Canada, and the Sellers Fund from the Hospital for Sick Children in Toronto. L.-C. T. is a Research Scholar of CCFF.

Request reprints from: Dr. M. Buchwald, Department of Genetics, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, M5G 1X8 (Canada).

report the discovery of a linkage between *CF* and the pro α 2(I) collagen gene (*COL1A2*) which has been previously localized to the long arm of chromosome 7, region q21→q22 (Junien et al., 1982), and more recently, to region q21.3→q22.1 (Retief et al., 1985). Based on the linkage data derived from 50 informative *CF* families, the distance between *COL1A2* and *CF* is estimated to be 19 centiMorgans (cM). Furthermore, a close linkage has also been detected between *D7S15* and *COL1A2* based on data from 28 of these families.

Materials and methods

Families. The *CF* families used in this study have been described previously (Eiberg et al., 1985; Tsui et al., 1985a). They were all two-generation nuclear families each with two or more affected children. DNA samples from each individual were prepared either directly from peripheral blood samples or from Epstein-Barr virus-transformed lymphoblast cultures.

DNA probes. A number of RFLPs associated with *COL1A2* have been described (Tsipouras et al., 1984; Brebner et al., 1985; Grobler-Rabie et al., 1985). The DNA probes used in this study are NJ-3, NJ-1 (Tsipouras et al., 1984) and Hf-32 (Myers et al., 1981) which detect three different RFLPs, *EcoRI*, *MspI* and *RsaI*, respectively. The polymorphic locus *D7S15* has been described previously (Tsui et al., 1985a); both RFLPs, *HindIII* and *HincII*, are detected with the probe Lam4-917.

DNA analysis. The methods of DNA analysis are essentially as described by Tsui et al. (1985a) except the following. The digested DNA samples were size-fractionated by electrophoresis on agarose gels (0.4% for *EcoRI*-digested samples, 0.7% for *MspI*, *HindIII* and *HincII*, and 1% for *RsaI*). ³²P-labeled probes were synthesized using the random primer procedure of Feinberg and Vogelstein (1983, 1984) to a specific activity of approximately 5×10^8 cpm/ μ g.

Methods of linkage analysis. To obtain the maximal likelihood estimate of the recombination fraction between genetic markers, the method of lod score analysis (Morton, 1955) was employed. The calculations for two point and three point linkage analyses were performed with the computer programs, LIPED (Ott, 1974) and LINKAGE (Lathrop et al., 1984), respectively. The confidence interval for the maximal likelihood estimate of the recombinant fraction was derived by taking the θ values which correspond to a lod score of ($z - 1$) (Conneally et al., 1985). The genetic distance (d) was estimated from the recombination frequency (r) using Haldane's function (Haldane, 1919): $d = -0.5 \ln(1 - 2r)$. The value for r was taken from the maximal likelihood estimate for θ in this study.

Results

Study of the *EcoRI*, *RsaI* and *MspI* RFLPs in the *CF* population

A panel of 53 *CF* families was examined for the inheritance of the three RFLPs associated with the

pro α 2(I) collagen gene. Among them, 50 (37 Canadian and 13 Danish families) were found to be informative for *CF* linkage analysis, as either one or both of the parents in these families were heterozygous for one to three RFLPs. Examples of RFLP analysis are shown in Fig. 1. The allele frequencies for the three RFLPs in these families (data not shown) are close to those previously reported (Tsipouras et al., 1984; Falk et al., 1985; Grobler-Rabie et al., 1985).

Linkage relationship between *CF* and *COL1A2*

To examine the linkage relationship between *CF* and *COL1A2*, a standard lod (z) score analysis (Morton, 1955) was performed with the LIPED program based on the segregation patterns of the RFLPs and the disease in each of the 50 families. The combined result is summarized in Table I. The maximal likelihood estimate for recombination fraction (θ) between *CF* and *COL1A2* is at 0.16 with a lod score of 4.36 (odds ratio: 23,000:1), assuming the recombination frequencies in males and females are equal. A total of 115 *CF* children were included in this analysis. By counting the number of informative meioses, the apparent recombination fraction among these children is 16/146 (11%). Although this number is slightly lower than that derived from the maximal likelihood estimate, it is within the confidence interval for θ which is between 0.10 and 0.26. A maximal lod score (z) of 4.51 (odds ratio of 32,000:1) can be obtained at male recombination fraction (θ_M) of 0.14 and female recombination fraction (θ_F) of 0.20. The score which is far above 3 (odds ratio of 1,000:1), the value generally accepted for proof of linkage, indicates linkage between *CF* and *COL1A2*.

Linkage relationship among *COL1A2* and *D7S15* and *PON*

We next examined the linkage relationship between *COL1A2* and the other two *CF*-linked genetic markers, the DNA marker *D7S15* (Tsui et al., 1985a) and the serum enzyme marker *PON* (Eiberg et al., 1985; Schmiegelow et al., 1986). In our previous analysis (Tsui et al., 1985a), the most probable order for *D7S15*, *PON*, and *CF* was determined to be *D7S15* - *PON* - *CF*. The distance between *D7S15* and *CF* was estimated to be 15 cM, that between *D7S15* and *PON*, 5 cM (Tsui et al., 1985a) and that between *PON* and *CF*, 10 cM (Eiberg et al., 1985; Schmiegelow et al., 1986). In the present study, a total of 29 families were found to be informative for linkage analysis between

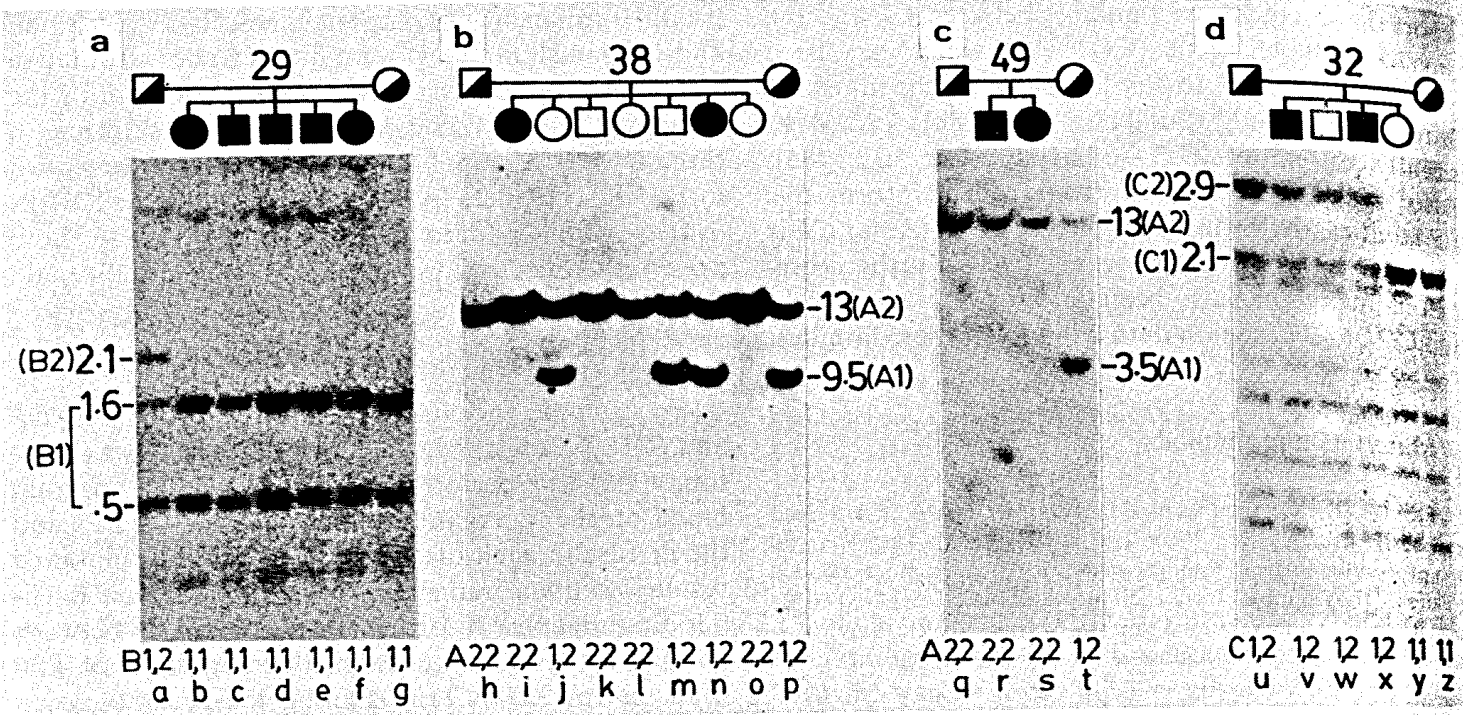


Fig. 1. Segregation of *MspI*, *EcoRI*, and *RsaI* RFLPs of *COLIA2* in CF families. DNA samples were (a) (lanes a-g) digested with *MspI* and hybridized with probe NJ-1; (b) (lanes h-p) digested with *EcoRI* and hybridized with probe NJ-3 3.2; (c) (lanes q-t) digested with *EcoRI* and hybridized with probe NJ-3 3.5; (d) (lanes u-z) digested with *RsaI* and hybridized with probe Hf-32. The family pedigree is shown above each panel: affected children are represented by filled squares and circles; unaffected children by open squares and circles. The RFLP allele assignment of each individual is indicated below each lane. The sizes of hybridizing fragments are in kilobases.

Table I. Linkage analysis of *COLIA2*, *CF*, *D7S15* and *PON*

Loci	Number of informative families	Lod (z) scores at recombinant fractions (θ) of								
		0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
<i>CF - COLIA2</i>	50	-14.07	-0.47	3.35	4.32	4.19	3.51	2.59	1.63	0.79
<i>COLIA2 - D7S15</i>	29	11.21	13.45	12.77	11.31	9.52	7.59	5.58	3.62	1.84
<i>COLIA2 - PON</i>	21	-0.68	2.06	2.60	2.51	2.19	1.76	1.29	0.81	0.40

$\hat{\theta} = 0.16$, $z = 4.36$; confidence interval: 0.10 - 0.26, $\hat{z} = 4.51$, $\theta_M = 0.14$, $\theta_F = 0.20$.
 $\hat{\theta} = 0.05$, $z = 13.45$; confidence interval: 0.02 - 0.11, $\hat{z} = 13.70$, $\theta_M = 0.08$, $\theta_F = 0.03$.
 $\hat{\theta} = 0.11$, $z = 2.61$; confidence interval: 0.04 - 0.26, $\hat{z} = 3.28$, $\theta_M = 0.22$, $\theta_F = 0.04$.

COLIA2 and *D7S15*. As shown in Table I, a maximal likelihood estimate for θ between *COLIA2* and *D7S15* is 0.05 with a lod score of 13.45. The confidence interval is between 0.02 and 0.11. The maximal lod (z) score is 13.70 at θ_M of 0.08 and θ_F of 0.03. These data establish that *COLIA2* and *D7S15* are linked.

Twenty-one families were also informative for linkage between *COLIA2* and the other CF-linked marker *PON*. A maximal likelihood estimate of θ for the linkage between *PON* and *COLIA2* is observed at

0.11, with a lod score of 2.61 (Table I), consistent with the assumption that these two markers are linked.

Multi-point linkage analysis

Based on all the available two point linkage data (this study and Tsui et al., 1985a), the apparent order for the four genetic markers is *COLIA2 - D7S15 - PON - CF*. To further examine the order of these four genetic loci, the multipoint LINKAGE program was used. First, the proper order for *COLIA2*, *D7S15*, and

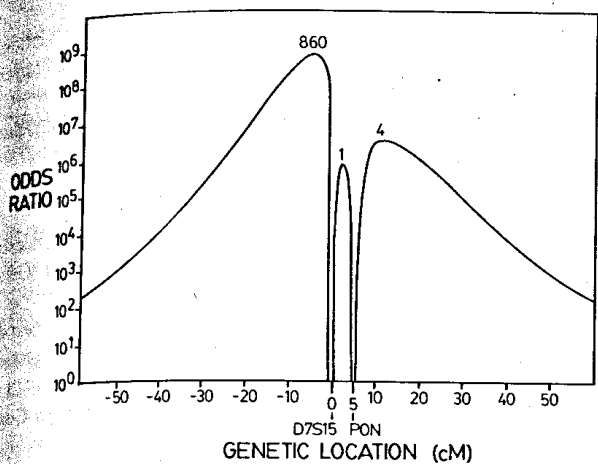


Fig. 2. Likelihood of the map location of *COLIA2* with respect to *D7S15* and *PON*. Data are selected from 22 CF families informative for all 3 genetic loci. Horizontal axis: genetic distance (w) from *D7S15*; vertical axis: odds ratio for location of *COLIA2* at wcf versus *COLIA2* at infinite distance (i.e., no linkage). The distance between *D7S15* and *PON* is assumed to be 5 cM. The relative odds for the orders *COLIA2* - *D7S15* - *PON*: *D7S15* - *PON* - *COLIA2*: *D7S15* - *COLIA2* - *PON* is 860 : 4 : 1.

PON was investigated. By fixing the genetic distance between *D7S15* and *PON* as 5 cM, the most likely location for *COLIA2* was determined. The relative odds calculation for the orders *COLIA2* - *D7S15* - *PON*, *COLIA2* - *PON* - *D7S15*, *PON* - *COLIA2* - *D7S15* was performed using data from 22 CF families in which all of these markers were found to be informative. As shown in Fig. 2, the most probable order for these three markers is *COLIA2* - *D7S15* - *PON*. Then, by fixing the distance between *COLIA2* and *D7S15* and between *D7S15* and *PON*, the most likely location for *CF* was determined. The latter analysis was based on information derived from 50 CF families, each of which showed segregation of at least three of these markers. The result in Fig. 3 suggests that the most likely order is *COLIA2* - *D7S15* - *PON* - *CF*.

Discussion

Based on linkage data from 50 informative families, the maximal likelihood estimate for the recombination fraction between *COLIA2* and *CF* is 0.16, assuming the recombination frequencies are equal in male and female. Using Haldane's function (Haldane, 1919), the genetic distance between *COLIA2* and *CF* is estimated to be 19 cM. In addition, our best estimate for the distance between *COLIA2* and *D7S15* is 5 cM and that between *COLIA2* and *PON*, 11 cM. These results, together with those from previous studies (Eiberg et al., 1985; Tsui et al., 1985a, Schmieglow et al., 1986) allow us to construct a linkage map for *COLIA2*, *D7S15*, *PON* and *CF*, as shown in Fig. 4.

Since *COLIA2* has been localized to the long arm of chromosome 7q21.3→q22.1 (Retief et al., 1985), it is likely that the closely linked DNA marker *D7S15* is also on the proximal half of this long arm. Con-

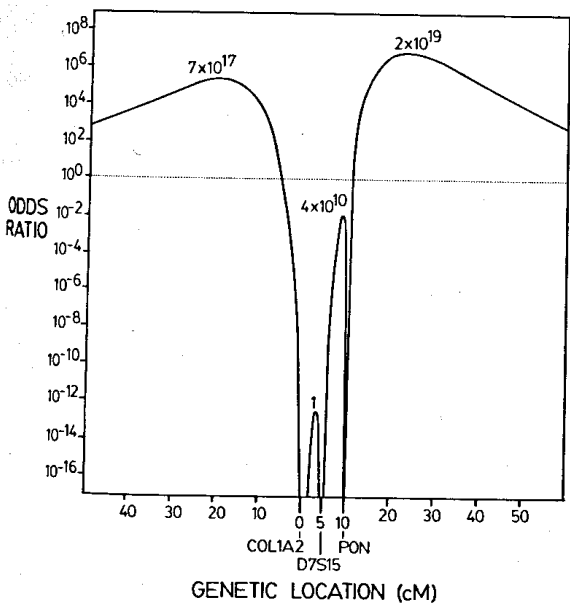


Fig. 3. Likelihood of the map location of *CF* with respect to *COLIA2*, *D7S15*, and *PON*. Data are selected from 50 CF families informative for at least 3 of the markers. Horizontal axis: genetic distance (w) from *COLIA2*; vertical axis: odds ratio for location of *CF* at wcf versus *CF* at infinite distance (i.e., no linkage). The genetic distances between *COLIA2* and *D7S15* and between *D7S15* and *PON* are both assumed to be 5 cM. The relative odds for the orders *COLIA2* - *D7S15* - *PON* - *CF*: *CF* - *COLIA2* - *CF* - *D7S15* - *PON*: *COLIA2* - *D7S15* - *CF* - *PON*: *COLIA2* - *CF* - *D7S15* - *PON* is 2×10^{19} : 7×10^{17} : 4×10^{10} : 1.

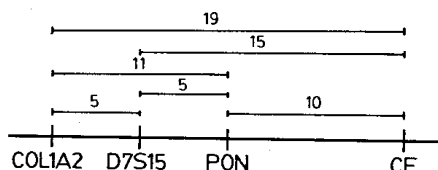


Fig. 4. Proposed linear order of *COLIA2*, *D7S15*, *PON*, and *CF*. The numbers represent estimated distances, assuming equal recombination frequencies between male and female. The distance between *PON* and *CF* is taken from Schmieglow et al. (1986); the distance between *D7S15* and *CF* and between *D7S15* and *PON* are from Tsui et al. (1985c).

sistent with this assumption, a rodent/human hybrid cell line which has lost the segment distal to band 7q31.1 was found to retain D7S15 sequences (unpublished results). Based on the linkage distance from *COL1A2*, it is probable that *CF* is also on the long arm of chromosome 7, either close to the centromere or the middle of the long arm. However, since the correlation between recombination and physical distances is unknown for this chromosome region, it is possible that *CF* is on the short arm of chromosome 7, close to the centromere. A more accurate localization of *CF* awaits the identification of a closer marker or one flanking the locus.

Note added in proof: After submission of this manuscript, the authors noted that Scambler et al. (1985b) have also discovered linkage between *CF* and *COL1A2*. The localization of *CF* to the long arm of chromosome 7 is further supported by the results of White et al. (1985) and Wainwright et al. (1985).

Acknowledgements

The authors wish to thank Dr. Francesco Ramirez for providing all the collagen probes.

References

- Beaudet AL, Rosenbloom CL, Spence JE, O'Brien WE, Cole S, Ferrell RE: Gene mapping for cystic fibrosis. Eighth International Workshop on Human Gene Mapping. *Cytogenet Cell Genet* 40:579 (1985).
- Brebner D, Grobler-Rabie A, Bester A, Mathews C, Boyd C: Two new polymorphic markers in the human pro α 2(I) collagen gene. *Hum Genet* 70:25-27 (1985).
- Conneally PM, Edwards JH, Kidd KK, Lalouel J-M, Morton NE, Ott J, White R: 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).
- 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).
- Falk CT, Schwartz RC, Ramirez F, Tsipouras P: Use of molecular haplotypes specific for the human pro α 2(I) collagen gene in linkage analysis of the mild autosomal dominant forms of osteogenesis imperfecta. *Am J hum Genet* (in press, 1985).
- Feinberg AP, Vogelstein B: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analyt Biochem* 132:6-13 (1983).
- Feinberg AP, Vogelstein B: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity (Addendum). *Analyt Biochem* 137:266-267 (1984).
- Goodchild MC, Edwards JH, Glenn KP, Grindey C, Harris R, Mackintosh P, Wentzel J: A search for linkage in cystic fibrosis. *J med Genet* 7:417-419 (1976).
- Grobler-Rabie A, Brebner D, Vandenplas S, Wallis G, Dalgleisch R, Kaufman R, Bester A, Mathew C, Boyd CD: Polymorphism of DNA sequence in the human pro α 2(I) collagen gene. *J med Genet* (in press, 1985).
- Haldane JBS: The combination of linkage values, and the calculation of distances between loci of linked factors. *J Genet* 8:299-309 (1919).
- Junien C, Weil D, Myers JC, Cong NY, Chu M-L, Foubert C, Gross M-S, Prockop DJ, Kaplan JC, Ramirez F: Assignment of the human pro α 2(I) collagen structural gene (*COL1A2*) to chromosome 7 by molecular hybridization. *Am J hum Genet* 34:381-387 (1982).
- Knowlton RG, Cohen-Haguenaer O, Ngeyen Van Cong, Frezal J, Brown V, Barker D, Braman JC, Schumm JW, Tsui L-C, Buchwald M, Donis-Keller H: A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. *Nature* 318:380-382 (1985).
- Lathrop GM, Lalouel JM, Julier C, Ott J: Strategies for multilocus linkage analysis in humans. *Proc natn Acad Sci USA* 81:3443-3446 (1984).
- Morton NE: Sequential tests for the detection of linkage. *Am J hum Genet* 7:277-318 (1955).
- Myers JC, Chu M-L, Faro SH, Clark WJ, Prockop DJ, Ramirez F: Cloning a cDNA for the pro α 2 chain of human type I collagen. *Proc natn Acad Sci USA* 78:3516-3520 (1981).
- Ott J: Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human studies. *Am J hum Genet* 26:588-597 (1974).
- Retief E, Parker MI, Retief AE: Regional chromosome mapping of human collagen genes alpha 2(I) and alpha 1(I) (*COL1A2* and *COL1A1*). *Hum Genet* 69:304-308 (1985).
- Scambler PJ, Bell G, Watson E, Farrall M, Bates G, Davies K, Lench N, Ashworth A, Williamson R, Tippet P, Wainwright B: Cystic fibrosis linkage exclusion data. *Cytogenet Cell Genet* 41:62-63 (1986).
- Scambler P, Robbins T, Gilliam C, Boylston A, Tippet P, Williamson R, Davies KE: Linkage studies between polymorphic markers on chromosome 4 and cystic fibrosis. *Hum Genet* 69:250-254 (1985a).
- Scambler PJ, Wainwright BJ, Farrall M, Bell J, Stanier P, Lench NJ, Bell G, Kruyer H, Ramirez F, Williamson R: Linkage of *COL1A2* collagen gene to cystic fibrosis, and its clinical implications. *Lancet* ii:1241-1242 (1985b).
- Schmiegelow K, Eiberg H, Tsui L-C, Buchwald M, Phelan PD, Williamson R, Warwick W, Niebuhr E, Mohr J, Schwartz M, Koch C: Linkage between the loci for cystic fibrosis and paraoxonase. *Clin Genet* (in press, 1986).
- Steinberg AG, Morton NE: Sequential test for linkage between cystic fibrosis of the pancreas and the MNS locus. *Am J hum Genet* 8:177-189 (1956).
- Steinberg AG, Schwachman H, Allen FH, Dooley RR: Linkage studies with cystic fibrosis of the pancreas. *Am J hum Genet* 8:162-176 (1956).
- Talamo RC, Rosenstein BJ, Berninger RW: Cystic fibrosis, in Stanbury JB, Wyngaarden JB, Fredrickson DS, Goldstein JL, Brown MS (eds): *The Metabolic Basis of Inherited Disease*, 5th edition, pp 1889-1917 (McGraw-Hill, New York 1983).

- Tsipouras P, Borresen 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).
- Tsui L-C, Buchwald M, Barker D, Braman JC, Knowlton RG, Schumm J, Eiberg H, Mohr J, Kennedy D, Plavsic N, Zsiga M, Markiewicz D, Akots G, Brown V, Helms C, Gravius T, Parker C, Rediker K, Donis-Keller H: Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science* 230:1054-1057 (1985a).
- Tsui L-C, Cox DW, McAlpine PJ, Buchwald M: Cystic fibrosis: progress on mapping of the disease locus by red cell and plasma protein markers. *Cytogenet Cell Genet* 39:238-239 (1985b).
- Tsui L-C, Zsiga M, Kennedy D, Plavsic N, Markiewicz D, Buchwald M: Cystic fibrosis: progress in mapping the disease locus using polymorphic DNA markers. I. *Cytogenet Cell Genet* 39:299-301 (1985c).
- Wainwright BJ, Scambler PJ, Schmidtke J, Watson EA, Law H-Y, Farrall M, Cooke HJ, Eiberg H, Williamson R: Localization of cystic fibrosis to human chromosome 7cen \rightarrow q22. *Nature* 318:384-385 (1985).
- Watkins P, Stanislovitis P, Hoffman N, Klinger K, Schwartz R, Doherty R, Kidd K, Tanzi R, Gusella J: Exclusion of the locus for cystic fibrosis from a linked set of DNA markers on chromosome 21. *Am J hum Genet* 37:A181(S37) (1985).
- White R, Woodward S, Leppert M, O'Connell P, Hoff M, Herbst J, Lalouel J-M, Dean M, Vande Woude G: A closely linked genetic marker for cystic fibrosis. *Nature* 318:382-384 (1985).

Received: 19 November 1985

Accepted: 2 December 1985