

## Testing for cystic fibrosis using allelic association

A J IVINSON\*, A P READ\*, R HARRIS\*, M SUPER†, M SCHWARZ†, J CLAYTON SMITH\*, AND R ELLES\*

From \*the Department of Medical Genetics, St Mary's Hospital, Manchester M13 0JH; and †Clinical Genetics Unit, Royal Manchester Children's Hospital, Pendlebury, Manchester M27 1HA.

**SUMMARY** A particular haplotype defined by probes XV2c, KM19, and CS-7 at the *D7S23* locus was found on 90% of chromosomes which carry cystic fibrosis (CF), but on only 11% of normal chromosomes in a UK sample of CF carriers. We show how such data can be used to calculate carrier risks for people with and without a family history of CF, and give examples of clinical applications. For parents or sibs of dead CF patients, phase and genotypes can often be assigned with only 1 to 2% error. However, this method is not suitable for prenatal testing where there is no history of CF; for couples with no family history, no fetus can be shown to be at more than 2% risk of being affected.

The first prenatal diagnoses of cystic fibrosis (CF) used the microvillar enzyme test of Brock *et al.*<sup>1</sup> This test can be applied to any pregnancy by amniocentesis at about 18 weeks of gestation, but its specificity is low unless the prior risk is 1 in 4 or greater. The discovery of DNA polymorphisms linked to the CF locus<sup>2-4</sup> allowed prenatal diagnosis to be done earlier in pregnancy using chorion villus biopsy, with a lower error rate. We have described the principle of using these DNA markers elsewhere<sup>5</sup> and summarised our experience with our first 30 cases.<sup>6</sup> The main limitation of this test is that it is applicable only to a minority of families, that is, those who have a surviving affected child.

DNA amplification using the polymerase chain reaction (PCR)<sup>7</sup> has extended the scope of gene tracking. It is now possible to use stored Guthrie blood spot cards<sup>8</sup> or histological specimens<sup>9</sup> to obtain DNA from a long dead child and to score polymorphisms more easily and rapidly than previously. Nevertheless, gene tracking still requires a family history of the disease and material from an affected patient.

The original DNA markers met and pJ3-11<sup>2-4</sup> show less than 1% recombination with the CF locus,<sup>10</sup> but there is little allelic association with CF (that is, normal and CF chromosomes carry much the same distribution of marker types). A second generation of probes, XV2c, KM19, and CS-7,<sup>11 12</sup> defines the locus *D7S23* which lies much closer to the CF locus. The clinical importance of these new markers lies not so much in the reduced recombina-

tion error compared to the earlier markers, but in the strong population association between particular *D7S23* haplotypes and the CF disease allele.

This allelic association presumably exists because most of the CF disease genes in the European population are descended from a single ancestral mutation, and so CF chromosomes still tend to carry a particular *D7S23* haplotype. By good fortune, the original mutation evidently happened on a chromosome carrying an unusual haplotype. As a result, most CF chromosomes in the UK carry a different *D7S23* haplotype from most normal chromosomes. Thus, it now becomes possible to identify putative CF carrying chromosomes even in families where there is no material from an affected patient. We report here our experience of using these allelic associations to extend the scope of prenatal diagnosis and carrier detection, and show a simple way of calculating risks for any haplotype or genotype.

### Materials and methods

We analysed 114 informative Caucasian families, mainly from the north western region of England. The polymorphisms studied are shown in table 1. Conventional Southern blotting<sup>13</sup> was used for the XV2c and some KM19 analyses. Some KM19 and all CS-7 analyses were done by polymerase chain reaction,<sup>7</sup> using oligonucleotide primer sequences kindly communicated by Professor Bob Williamson. Starting with less than 1 µg DNA, 30 rounds of amplification were carried out using a Peerless Systems programmable mechanical arm. One-third of each sample was removed, digested with the

TABLE 1 RFLPs used and allele definitions.

Conventional analysis				PCR analysis			
Probe	Enzyme	Allele	Size (kbp)	Marker	Enzyme	Allele	Size (bp)
KM19	PstI	K1	7.8	KM19	PstI	K1	950
		K2	6.6			K2	650+300
XV2c	TaqI	X1	2.1	CS-7	HhaI	CS1	330
		X2	1.4			CS2	165

appropriate enzyme, and electrophoresed on 2% Nusieve agarose gels. RFLP types were read by UV fluorescence from gels stained with ethidium bromide (fig 1).

#### RISK CALCULATIONS

All risk estimates are based on simply counting numbers of CF and non-CF (NCF) chromosomes with each D7S23 haplotype and applying Bayes's theorem<sup>14</sup> to the numbers.

Let  $p$  be the prior risk that a chromosome carries CF (1 in 2 for the sib of a patient and 1 in 45 for a randomly selected person in the UK). If haplotype A is seen in  $c$  out of  $C$  CF chromosomes and  $n$  out of  $N$  NCF chromosomes, then the risk that an unknown chromosome with this haplotype carries CF is

$$R_A = (p \cdot c/C) / [p \cdot c/C + (1-p) \cdot n/N].$$

If a phenotypically normal person has haplotypes A and B with risk  $R_A$  and  $R_B$ , the risk that he is a CF carrier is

$$[R_A(1-R_B) + R_B(1-R_A)] / (1-R_A R_B).$$

A small BASIC program, available on request,

has been written to do these sums. From standard statistics, there is a 95% chance that the true value of  $c/C$  in the population lies within 1.96 SD of the value observed in our sample, and the SD is  $\sqrt{[c(C-c)]/C}$ . Substituting  $c/C$  (1.96 SD) in the calculation of carrier risk gave the confidence intervals shown in table 2.

#### Results

Table 3 shows the distribution of marker alleles and table 4 the distribution of haplotypes among normal (NCF) and CF chromosomes. These results are generally similar to data on XV2c-KM19-CS-7 haplotypes reported<sup>12</sup> from the UK, Denmark, Spain, and Finland, and to XV2c-KM19 haplotypes from Holland<sup>15</sup> and Germany.<sup>16</sup> The distribution of haplotypes among both normal and CF chromosomes was not significantly different in our results and in the UK series of Estivill *et al.*,<sup>12</sup> and therefore these two series were combined for calculating risks. Tables 2 and 5 show the risks for each haplotype and genotype calculated by the method described. Inspection of table 2 shows that haplotype D is associated with a high risk of carrying CF, and haplotypes A, B, E, and H with a low risk.

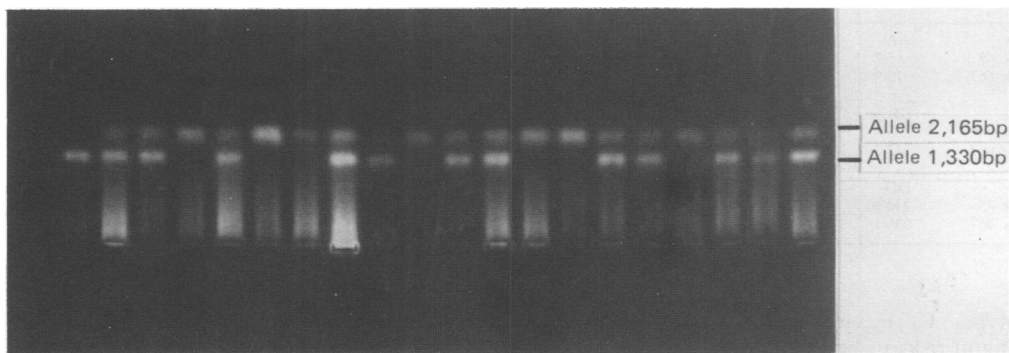


FIG 1 CS-7/HhaI polymorphism scored by polymerase chain reaction. 1  $\mu$ g samples of genomic DNA were amplified for 30 cycles, digested with HhaI, and electrophoresed in a 2% Nusieve agarose gel. The fragments were visualised by UV fluorescence after staining with ethidium bromide. When the HhaI site is present the 330 bp amplified sequence is cleaved into two 165 bp fragments.

TABLE 2 Risk that a chromosome of each common haplotype carries the CF gene, calculated using the combined data in table 4.

Haplotype	Combined Nos		1000×risk±95% CI	
	CF	NCF	Prior=1/2	Prior=1/45
A X1,K1,CS1	1	24	36.7±6	0.9±0.1
B X1,K1,CS2	1	11	76.8±12	1.9±0.3
D X1,K2,CS2	118	15	878±5	140±5
E X2,K1,CS1	7	51	112±6	2.8±0.2
H X2,K2,CS2	3	15	155±12	4.1±0.1
Others (C,F,G)	0	3	—	—

TABLE 3 Distribution of alleles at the D7S23 locus in normal (NCF) and CF chromosomes in parents of CF children.

Allele	X1	X2	K1	K2	CS1	CS2
CF chromosomes	205	23	19	200	8	75
NCF chromosomes	94	132	164	55	50	33

TABLE 4 D7S23 haplotypes in parents of CF children: our data compared to the data of Estivill *et al.*<sup>12</sup> Not all chromosomes in our series were typed for CS-7.

Haplotype	Estivill <i>et al.</i> <sup>12</sup> (UK)		Our data		Our X-K haplotypes	
	CF	NCF	CF	NCF	CF	NCF
A X1,K1,CS1	0	12	1	12	3	49
B X1,K1,CS2	1	3	0	8		
C X1,K2,CS1	0	1	0	0	147	19
D X1,K2,CS2	51	8	67	7		
E X2,K1,CS1	5	24	2	27	8	70
F X2,K1,CS2	0	0	0	2		
G X2,K2,CS1	0	0	0	0	8	25
H X2,K2,CS2	1	4	2	11		
Total	58	52	72	67	166	163

TABLE 5 Risk (×1000) that a phenotypically normal person with the given haplotypes (defined in table 4) is a CF carrier. Prior risks of 45 and 667 per thousand are the 1/22 population risk and the 2/3 risk for a sib of a CF patient, respectively.

Genotype	AA	AB	AD	AE	AH
Prior risk 45	1.7	2.8	141.3	3.7	5.0
Prior risk 667	70.9	108.2	878.6	140.7	181.1
	BB	BD	BE	BH	HH
Prior risk 45	3.8	142.0	4.7	6.0	8.3
Prior risk 667	142.7	879.3	172.8	210.3	268.0
	DD	DE	DH	EE	EH
Prior risk 45	246.6	142.7	143.7	5.7	7.0
Prior risk 667	935.1	879.9	880.7	200.8	235.9

Haplotypes C, F, and G are still too rare for meaningful risks to be calculated.

### Case reports

The following cases illustrate the value of these

associations in circumstances where gene tracking was not possible.

#### CASE 1

A 30 year old man reported a history of persistent sinusitis, malabsorption, azoospermia, and recurrent bronchopulmonary infection, with evidence of pulmonary tuberculosis and pancreatic steatorrhoea. In childhood a diagnosis of CF had been overturned; however, the course of his illness suggested that CF was now a strong possibility. DNA analysis showed that he was homozygous for haplotype D. From table 2, 82% of people with CF but only 1.6% of normal homozygotes (and 23% of CF carriers) are expected to have the DD genotype. The odds in favour of his having CF, given these results, are

$$0.82 : (0.016 + 1/23 \times 0.23) = 31:1.$$

Thus the DNA studies supported the diagnosis of CF, and provided anecdotal support for the view that unusually mild CF may result from the same mutation as the classical disease.

#### CASE 2 (FIG 2)

A 29 year old nulliparous woman presented at 11 weeks' gestation. Her sister had died many years before of CF, confirmed by necropsy. She and her non-consanguineous husband were given a risk for CF of 1/120 in the present pregnancy, based on the assumptions that she had a 2/3 and he a 1/20 risk of being a carrier. The couple were unwilling to accept this 1/120 risk and remained extremely anxious. DNA studies (fig 2) showed that both partners were homozygous for the low risk haplotype E, while the woman's parents (both presumably carriers) each had one low risk (E) and one high risk (D) haplotype. The risk that the husband is a carrier, from table 5, is 0.57%. The wife's risk is lower than the figure of 20% in table 5 because of the results on her parents. For each parent, the phase could be

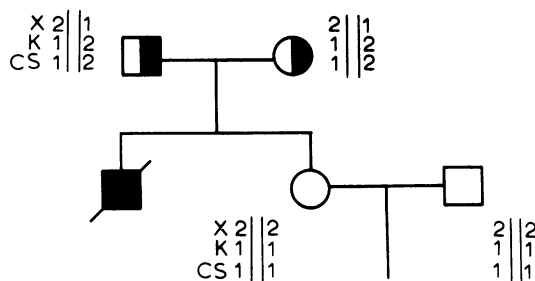


FIG 2 D7S23 haplotypes in case 2. X=XV2c/TaqI, K=KM19/PstI, CS=CS-7/HhaI.

$D^{CF}E^{NCF}$  or  $D^{NCF}E^{CF}$  with equal prior probability. Each chromosome is at 1 in 2 prior risk of carrying CF, therefore from table 2 the relative likelihoods of the two phases are

$$0.878 \times (1 - 0.112) : (1 - 0.878) \times 0.112 = 57:1.$$

The chance that one or other of the woman's E haplotypes carries CF is 2/58 or 3.5%. Thus, the risk that both partners are carriers is 0.57%  $\times$  3.5%, and the risk of an affected child is 1 in 20 000.

CASE 3 (FIG 3)

A 41 year old woman presented at 10 weeks' gestation and gave a history of two CF children who died early in life followed by two healthy children. DNA analysis (fig 3) showed that the father was homozygous for haplotype D, the mother was BD, and the fetus was also BD. Its odds of inheriting a CF gene from the father were 1:1. Calculation of the mother's phase follows the method used in case 2 above. She could be  $B^{NCF}D^{CF}$  or  $B^{CF}D^{NCF}$  with equal prior probability, and each chromosome is at 1 in 2 prior risk. From table 2 the relative likelihoods of the two phases are

$$(1 - 0.077) \times 0.878 : 0.077 \times (1 - 0.878) = 86:1.$$

Since the fetus inherited the B haplotype, it has a 1 in 87 chance of inheriting the CF gene from the mother, and an overall risk of 1 in 174 of being affected.

In this family we were able to confirm the low risk using pJ3-11, one of the less closely linked markers.<sup>10</sup> All five family members were heterozygous 2-1. In isolation this result means that the two surviving unaffected boys must share either two or no haplotypes. Since one boy is 2-1 and the other 2-2 with KM19 (fig 3), they cannot share two haplotypes, so they must share none. Given that neither boy is

affected, they must both be carriers: one received CF from his father and NCF from his mother, the other vice versa. The fetus has the same genotype for all four markers as the older boy and therefore, barring recombination, it too is a carrier.

Discussion

We believe that the main clinical application of these results should be for prenatal diagnosis where couples are both known to be carriers but have no surviving affected child. The problem in these cases is to decide which chromosome in each parent carries CF; 90% of CF chromosomes carry haplotype D while 85% of NCF chromosomes carry haplotypes A, B, E, or H. From the figures in table 2 we predict that over 75% of carriers will be AD, BD, ED, or HD, and the chance of identifying the CF chromosome wrongly in such a carrier is only 1 to 1.5%, as shown for cases 2 and 3 above. This compares with the 0.5% risk of error when the CF chromosome is identified by studying an affected child with a single marker from the met or pJ3-11 series.<sup>10</sup> A second useful application is for carrier couples who have requested donor insemination. About half of all sperm donors carry only haplotypes A, B, or E, and by using them the risk of a CF child is negligibly low.

For a person with no family history of CF, carrying haplotype D raises the carrier risk from 1/23 to 1/7. Even if the population were screened for couples both carrying haplotype D, their DD fetuses would be at only 1/7  $\times$  1/7 = 2% risk of CF. Thus, there is no case for offering these tests to people without a family history of CF: population screening must await the definition of the CF mutation itself.

DNA testing is often requested by couples where one partner has no family history and the other has a dead CF sib, as in our case 2. Those who persist with the request after counselling are presumably those who are unwilling to take a 1 in 120 risk (the risk before any DNA results). Some laboratories might refuse to do tests because of the likelihood that they will lead to terminations of pregnancies at relatively low risk. These problems cannot be entirely avoided, but they can be minimised by trying hard to establish haplotypes before a pregnancy, using other family members as necessary, and by haplotyping the high risk partner before doing any tests on the low risk partner. A possible protocol would be as follows.

- (1) Establish haplotypes for the high risk partner. Unless this person has at least one D (high risk) haplotype, useful prenatal testing will not be possible. Given that he or she does not have a D haplotype, the next highest risk would be seen if the fetus inherits an H haplotype from the high

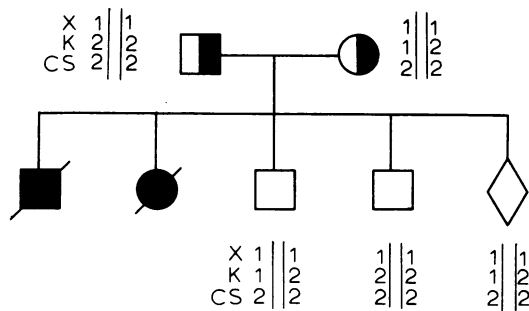


FIG 3 D7S23 haplotypes in case 3. X=XV2c/TaqI, K=KM19/PstI, CS=CS-7/HhaI.

risk and a D haplotype from the low risk parent. From table 2, the risk that the fetus is affected is  $0.155 \times 0.14 = 2\%$ . This is too low for the microvillar enzyme test to be useful.

- (2) If the high risk partner has at least one D haplotype, the haplotypes of the low risk partner are established next. In about half of all cases only the low risk haplotypes A, B, or E will be present. In these cases the couple can be reassured. Even if a fetus inherits haplotype D from the high risk partner, if it inherits A, B, or E from the low risk parent it is at no more than 0.3% risk of being affected (table 2).
- (3) A fetus which inherits haplotype D from both parents is at about 1/7 risk, and perhaps the microvillar enzyme test would be appropriate later in the pregnancy.

All the results given here depend on being able to count the numbers of CF and non-CF chromosomes with each haplotype; hence, indeterminate haplotypes cause problems. These arise when the parents and the affected child are all heterozygous for two or more markers so that it is not possible to identify the CF and non-CF haplotypes. Sometimes analysis of grandparents or unaffected sibs will help, but some indeterminate cases will remain. As has been pointed out,<sup>16</sup> the missing haplotypes are not a random selection of those in the population, and ignoring them could bias the estimates of haplotype frequencies. Only a minority of haplotypes are indeterminate (16/155 in our data), so any bias will not be serious for common haplotypes, but the frequency of some rare haplotypes might be underestimated. We used the Hapmax program<sup>16</sup> (kindly made available by Dr M Krawczak) to make a maximum likelihood estimate of the true haplotype frequencies from our data in table 4. The only significant change was an increase in the frequency of haplotype E on CF chromosomes from 2.8% (2/72, table 4) to 6.9%. The risks used throughout this paper are derived from combined data, ours plus those of Estivill *et al*,<sup>12</sup> and in these combined data the E frequency is 7/130 = 5.4%. We conclude that indeterminate haplotypes are not seriously biasing our risk estimates.

We thank Professor Bob Williamson and Dr Carolyn Williams for the probes XV2c, KM19, and CS-7 and the primer sequences used in the PCR,

and Dr Dian Donnai for referring patients. This work was supported by the North Western Regional Health Authority, the Department of Health, and the Cystic Fibrosis Research Trust.

#### References

- 1 Brock DJH, Bedgood D, Barron L, Hayward C. Prospective prenatal diagnosis of cystic fibrosis. *Lancet* 1985;i:1175-8.
- 2 Knowlton RG, Cohen-Hagenauer O, Nguyen VC, *et al*. A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. *Nature* 1985;318:380-2.
- 3 White R, Woodward S, Leppert M, *et al*. A closely linked genetic marker for cystic fibrosis. *Nature* 1985;318:382-4.
- 4 Wainwright BJ, Scambler PJ, Schmidtke J, *et al*. Localisation of cystic fibrosis locus to human chromosome 7cen-7q22. *Nature* 1985;318:384-5.
- 5 Read AP. Principles of DNA-based prenatal diagnosis of cystic fibrosis: an introduction for the clinician. *J R Soc Med* 1987;80 (suppl 15):9-12.
- 6 Super M, Ivinson A, Schwarz M, *et al*. Clinical experience of prenatal diagnosis of cystic fibrosis by use of linked DNA probes. *Lancet* 1987;ii:782-4.
- 7 Saiki RK, Scharf S, Faloona F, *et al*. Enzymatic amplification of  $\beta$ -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 1985;230:1350-4.
- 8 Williams C, Weber L, Williamson R, Hjelm M. Guthrie spots for DNA-based carrier testing in cystic fibrosis. *Lancet* 1988;ii:693.
- 9 Imprain CC, Saiki RK, Ehrlich HA, Teplitz RL. Analysis of DNA extracted from formalin-fixed paraffin-embedded tissues by enzymatic amplification and hybridisation with sequence-specific oligonucleotides. *Biochem Biophys Res Commun* 1987;142:710-6.
- 10 Beaudet A, Bowcock M, Buchwald M, *et al*. Linkage of cystic fibrosis to two tightly linked DNA markers. Joint report from a collaborative study. *Am J Hum Genet* 1986;39:681-93.
- 11 Estivill X, Farrall M, Scambler PJ, *et al*. A candidate for the cystic fibrosis locus isolated by selection for methylation-free islands. *Nature* 1987;326:840-5.
- 12 Estivill X, Scambler PJ, Wainwright BJ, *et al*. Patterns of polymorphism and linkage disequilibrium for cystic fibrosis. *Genomics* 1988;1:257-63.
- 13 Maniatis T, Fritsch EF, Sambrook J, eds. *Molecular cloning: a laboratory manual*. New York: Cold Spring Harbor Laboratory, 1982.
- 14 Vogel F, Motulsky AG. *Human genetics*. Berlin: Springer-Verlag, 1979:594.
- 15 te Meerman GJ, ten Kate LP, Cobben JM, von Essen T, Buys CHCM, Halley D. Prenatal diagnosis of cystic fibrosis where single affected child has died. *Lancet* 1988;ii:745.
- 16 Krawczak M, Konecki D, Schmidtke J, *et al*. Allelic association of the cystic fibrosis locus and two DNA markers, XV2-c and KM19, in 55 German families. *Hum Genet* 1988;80:78-80.

Correspondence to Dr A P Read, Department of Medical Genetics, St Mary's Hospital, Hathersage Road, Manchester M13 0JH.