The gene defect in cystic fibrosis and clinical applications of the knowledge

Maurice Super MD FRCPE *Paediatric Genetics Unit, Royal Manchester Children's Hospital, Manchester M27 1HA*

Keywords: cystic fibrosis; genetics

This article describes developments since the location of the cystic fibrosis (CF) gene, linkage disequilibrium of closely linked markers and CF, the discovery of the CF gene—CF transmembrane conductance regulator (CFTR), its putative structure and its major pathological mutation, ΔF_{508}; the distribution in Europe and in the world of this mutation and theories about heterozygote advantage and meiotic drive. The discovery of further mutations, their ethnic distribution and studies of genotype phenotype correlation are discussed, including the discovery of pathological effects of the gene in apparent heterozygotes. Finally clinical applications of the new knowledge are described.

From the locus on chromosome 7 to discovery of the gene itself

The discovery of the locus of the CF gene in 1986[1] brought the immediate ability to offer accurate prenatal diagnosis[2] and carrier detection in siblings of affected people. As the search for the gene narrowed, polymorphic markers in strong linkage disequilibrium with CF were discovered which brought the option of increasing or decreasing the likelihood of carrier status in those with no family history. Notable amongst these were the cDNA probes, XV2c and Km19. When the CF gene itself was discovered, a particular haplotype pattern of these probes was found with a number of separate CF mutations, including the most frequent mutation. A heterozygous advantage is generally taken to be operating to maintain the high frequency of carriers of CF. The finding of this haplotype association outside the CF gene itself and associated with different mutations, raises the intriguing possibility that this advantage has nothing to do with the CF gene itself but that a 'hitchhiking' effect has operated, the advantage arising from a property outside the gene[3]. Though there are many polymorphic markers providing information about this region, the XV2c, Km19 haplotype is most frequently quoted with the 'B' haplotype showing the powerful association with CF (XV2c type 1, Km19 type 2).

The CF gene

In September 1989 the CF gene itself was described, together with its full complementary DNA sequence. The methods of discovery[4], the characterization of the functional protein termed cystic fibrosis transmembrane conductance regulator (CFTR)[5] and clinical and haplotype details of the major mutation, ΔF_{508}[6] were described in three seminal articles, models of their kind for any serious student of biological science. The three major scientists, Drs Lap-Chee Tsui, John Riordan and Francis Collins received the first Paul Di Sant'Agnese medals from the Cystic Fibrosis Foundation (Di Sant'Agnese first described the sweat defect in CF).

The claims made in the three articles have largely stood the test of time over the past 2 years, though it is still admitted that the proposed appearance of CFTR remains the best fit of a computer simulation, rather than being the result of scientific observation. The protein is said to contain two nucleotide (ATP) binding and a large regulatory domain of positively and negatively charged aminoacids intracytoplasmically and two hydrophobic intramural domains. There is also a glycosylated extracellular portion.

As expected from the autosomal recessive nature of the disease CF, it occurs when there are abnormalities of both CF genes—significant homozygous and compound heterozygous alterations both result in disease. They probably result in incorrect folding, inadequate glycosylation and finally abnormal trafficking, eg of chloride ions[7].

Table 1 gives details of abbreviations used in describing gene mutations. The commonest mutation discovered, ΔF_{508} designates a missing phenylalanine at position 508 of the 1480 aminoacids of CFTR. It arises from a codon deletion in exon 10 in the first putative nucleotide binding fold. In the largely Canadian population studied first ΔF_{508} comprised 70% of the CF genes with the remainder not identified. A genetic analysis consortium for the purpose of sharing population data and information on new mutations was set up by Dr Lap-Chee Tsui and has proved a model of international co-operation. The distribution of ΔF_{508} worldwide, and more particularly in Europe has proved interesting. A cline from south-east to north-west in Europe exists, with Turkey having the lowest described frequency (27%) and Denmark the highest at 88%[3]. The overall frequency in Britain is 76%, but in our population of the North West of England the frequency based on analysis of a very large number of chromosomes is 81%[8].

Table 1. Amino-acid abbreviations

A=Alanine	C=Cysteine	D=Aspartic Acid
E=Glutamic Acid	F=Phenylalanine	G=Glycine
H=Histidine	I=Isocleucine	K=Lysine
L=Leucine	M=Methionine	N=Asparagine
P=Proline	Q=Glutamine	R=Arginine
S=Serine	T=Threonine	V=Valine
W=Tryptophan	Y=Tyrosine	

X=Stop Codon Delta=Deletion Ins=Insertion
Number followed by + or − denotes intron/exon boundary alteration

In the third of the original articles describing the gene[6], the authors described clinical features and included invariable pancreatic insufficiency associated with the 49% of patients demonstrating homozygosity for ΔF_{508}. Subsequently pancreatic sufficiency has been described a few times with this homozygous state, though the original authors claim that even in these rare cases pancreatic insufficiency will occur, in time. The ΔF_{508} mutation occurs usually with haplotpye B, but it was immediately noted that haplotype B also occurs with a number of non-ΔF_{508} CF chromosomes.

ΔF_{508} is situated in exon 10 in the first putative nucleotide binding fold. Soon after its description we were fortunate to detect an adjacent mutation, ΔI_{507}[9] on the basis of the so-called allele specific oligonucleotide test in a family study showing a much weaker than normal band indicative of imperfect annealing of the DNA strands.

Most mutations associated with significant CF have involved the putative nucleotide binding folds, especially the first fold. Numerically exon 11 has proved the site of most mutations; nevertheless ΔF_{508} remains by far the commonest mutation discovered in all studied populations except in Ashkenazi Jews where, interestingly, an exon 20 mutation, W1282X, comprises 60%. Other ethnic differences have been observed with the exon 11 mutation, G551D being fairly prevalent in those of Celtic origin. An interesting history of Europe could be written through its most prevalent autosomal recessive disorder, CF, and a number of us are making a start in this study. It should prove possible to date the occurrence of various mutations. To date 107 separate CF mutations have been described, though only 28 have been found in any appreciable numbers and in more than one population group.

Mutations discovered have consisted of in-frame deletions (such as ΔF_{508} and ΔI_{507} implying a deleted phenlylalanine and iso-leucine), point mutations resulting in amino-acid substitutions (eg G551D, glycine substituted by aspartic acid), in stop-codon point mutations (eg R553X-arginine substituted by a stop codon or G542X or W1282X), insertions at intron-exon boundaries, eg 621+1, G to T (ie guanine to thymine). A few insertions resulting in frame shift have been reported too, though none in any significant numbers.

The clinician wishes to know whether there are genotype, phenotype correlations. The original describers of the ΔF_{508} separated groups into serious and mild on the basis of pancreatic insufficiency and sufficiency and by and large such a classification is justified, though pancreatic insufficient individuals can sometimes have mild lung disease. Factors outside the gene seem to be associated with differing levels of severity and, of course, one must remember differing environmental influences such as time of diagnosis, vigour of treatment and social class which may play a role. Compound heterozygotes for ΔF_{508} and G551D or W1282X are no better off than homozygotes for ΔF_{508}. On the other hand homozygotes for G551D have milder disease but are pancreatic insufficient. Compound heterozygotes for ΔF_{508} and R117H however do have milder disease. This supports a crucial role for the nucleotide binding folds in determining severity—R117H is in exon 3, part of the intramural anchoring protein presumably not playing a major role in ion transport. It has long been recognized that there is a specific predisposition for meconium ileus to recur in affected individuals within families. One of the stop mutations, G542X occurring in compound heterozygote form with ΔF_{508} appears to have an increased association with meconium ileus and information is being gathered in from the genetic analysis consortium to confirm this association. Moves are afoot to ensure uniform gathering of clinical information among members of the consortium which is advised by a number of experienced clinicians. This should improve the power and accuracy of the analysis of genotype phenotype association.

Heterozygote observations

An increase in heterozygotes for ΔF_{508} has been claimed in patients with bronchorrhoea, in infants with raised immune reactive trypsin and in sterile males with absence of the vas deferens; analysis of the genotype in DNA extracted from nasal polyps removed at operation showed an excess of the mutation G551D. None of the subjects in any of these groups has accepted diagnostic features of CF. The question is whether the heterozygote with the other CFTR gene apparently normal may result in these features or whether a minor mutation, eg in a biologically relatively inactive part of the gene, for instance in the intramural portion may explain the phenomenon. Parents of children with CF are clearly both fertile and productive bronchitis is not recognized in them either. This is an extremely interesting area for further research and advances in CF treatment

Table 2. Results of genetic analysis of 504 individuals from north-west England with established cystic fibrosis

Genotype	*Patients*
$\Delta F_{508}/\Delta F_{508}$	337
ΔF_{508}	61
AA	6 (includes 3 homozygotes, G551D)
ΔF_{508}/?	83
A/?	11
??	6 n=504

A=Identified non-ΔF_{508} mutation
?=CF mutation as yet not identified

Table 3. Distribution of CF mutations in north-west England

Mutation	*CF chromosomes screened*	*CF chromosomes with mutation*	
ΔF_{508}	1008	818	81.15%
G551D	215*	34	2.98%
R553X	215*	7	0.61%
G542X	203*	11	1.02%
R560T	113*	6	1.00%
N1303K	118*	6	0.96%
DI507	117*	5	0.81%
R117H	116*	4	0.65%
621+1	159*	10	1.19%
W1282X	159*	1	0.24%
V520F	63*	3	0.90%
G85E	40*	2	0.48%
Total			91.99%

*non-ΔF_{508}chromosomes

Table 4. Applications of the discovery of the CF gene

		Genotype/phenotype study
C	Genotyping in known CF	Confirmation of diagnosis Questioning of diagnosis Appropriate screening of relatives Prenatal diagnosis Geographical origins
CD	Early (eg before sweat test) diagnosis	
CSF	Carrier screening where there is a family story, with spouses	
CS	Carrier screening—sperm and ovum donors and recipients	
	—opportunistic—preconception clinic	
	general genetic clinic	
	—'population'—pregnant	
	—non-pregnant	

may have implications for symptomatic heterozygotes. Genetic analysis in these disparate groups seems worthwile and family studies should rapidly identify whether there must be a genetic explanation for the existence of symptomatic heterozygotes. Of course there are also patients in whom a diagnosis of CF has been seriously considered but features have not been absolutely typical—analysis sometimes reveals one CF mutation and it becomes difficult to decide whether a diagnosis of CF applies or whether these are heterozygotes who have been detected by chance.

Clinical applications

In well-studied populations like that in Britain there are a number of very useful applications of the new knowledge (see Tables 2 and 3). In our own North-Western population we have identified 92% of the genes associated with CF and have reported on 12 separate mutations found[8] and have recently found a 13th (see Table 2): this implies that a molecular diagnosis is possible in 84% of instances.

Table 4 shows the many ways in which this new knowledge may be applied right now. Thus there is a role in diagnosis, either preclinically or prenatally, in the investigation of unusually presenting patients and in the questioning of the diagnosis; there are some genotype-phenotype correlations which are possible; relatives and their spouses can be screened for carrier status. The existence of a family story of CF ought to be enquired about as part of general medical history taking and carrier screening offered to those who reply positively. Since 90% of children with CF are born into families with no known family history a more important exercise is to offer screening to the known relatives (and spouses) of those with CF and then to offer screening to the relatives of those detected in this way. Those who attend assisted reproduction clinics involving donated eggs or sperm should also be screened which allows low CF risks to be planned.

Population carrier screening pilot projects are being assessed currently in three United Kingdom studies. The degree of uptake and public interest in being screened will be intriguing to learn.

Hurry on the day when the ultimate application of the new knowledge is achieved, a cure or simple treatment for CF.

References

1 Tsui L-C, Buchwald M, Barker D, *et al.* Cystic fibrosis locus defined by a genetically linked polymorphic marker. *Science* 1985;**230**:1054-7

2 Super M, Ivinson A, Schwarz MJ, *et al.* Clinic experience; prenatal diagnosis of cystic fibrosis. *Lancet* 1987;**ii**:782-4

3 European working group on cystic fibrosis genetics (EWGCFG) Gradient of distribution in Europe of the major CF mutation and of its associated haplotype. *Hum Genet* 1990;**85**:436-41

4 Rommens JM, Ianuzzi MC, Kerem BS, *et al.* Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* 1989;**245**:1059-65

5 Riordan JR, Rommens JM, Kerem B-S, *et al.* Identification of the cystic fibrosis gene: cloning and characterisation of complementary DNA. *Science* 1989;**245**:1066-73

6 Kerem B-S, Rommens JM, Buchanan JA, *et al.* Identification of the cystic fibrosis gene: genetic analysis. *Science* 1989;**245**:1073-80

7 Gregory RJ, Rich DP, Cheng SH, *et al.* Maturation and function of CFTR variants bearing mutations in putative nucleotide binding domains 1 & 2. *Mol Cell Biol* 1991;**11**:3886-93

8 Super M, Schwarz MJ. Cystic fibrosis mutations in the north-west of England. *Eur J Pediatr* 1992;**151**:108-11

9 Schwarz MJ, Summers C, Heptinstall I, *et al.* A deletion mutation of the cystic fibrosis transmembrane conductance regulator (CFTR) locus: delta I507. In: Tsui L-C, ed. *The identification of the CF (cystic fibrosis) gene.* New York: Plenum Press, 1991