The Cystic Fibrosis Gene: A Molecular Genetic Perspective

Lap-Chee Tsui¹ and Ruslan Dorfman²

¹The University of Hong Kong, Hong Kong, Special Administrative Region, China

²Geneyouin Inc., Maple, Ontario, Canada

Correspondence: tsuilc@hku.hk; Ruslan.dorfman@geneyouin.ca

The positional cloning of the gene responsible for cystic fibrosis (CF) was the important first step in understanding the basic defect and pathophysiology of the disease. This study aims to provide a historical account of key developments as well as factors that contributed to the cystic fibrosis transmembrane conductance regulator (CFTR) gene identification work. A redefined gene structure based on the full sequence of the gene derived from the Human Genome Project is presented, along with brief reviews of the transcription regulatory sequences for the CFTR gene, the role of mRNA splicing in gene regulation and CF disease, and, various related sequences in the human genome and other species. Because CF mutations and genotype-phenotype correlations are covered by our colleagues (Ferec C, Cutting GR. 2012. Assessing the disease-liability of mutations in CFTR. *Cold Spring Harb Perspect Med* doi: 10.1101/cshperspect.a009480), we only attempt to provide an introduction of the CF mutation database here for reference purposes.

The cloning of the gene responsible for cystic fibrosis (CF) is a classic example of disease gene identification based on genetic linkage analysis. The first genetic analysis for CF could be traced to 1946, when Andersen and Hodges proposed that the disease was caused by a recessive mutation (Andersen and Hodges 1946), but linkage analyses were only briefly applied by Morton and Allen for cystic fibrosis of the pancreas in 1956 (Allen et al. 1956; Morton and Steinberg 1956). In fact, genetic linkage for disease gene identification was not broadly attempted before the discovery of polymorphic DNA, termed “restriction fragment length polymorphisms” markers in the early 1980s, and the

suggestion of their use as markers for genetic mapping (Botstein et al. 1980). Hence, early CF linkage and association studies were conducted with polymorphic biochemical markers, such as immunoglobulins and protein markers (Tsui et al. 1985b; Schmiegelow et al. 1986). Ironically, however, the first CF linkage was detected with a serum enzyme marker, paraoxonase (PON) (Eiberg et al. 1985).

IDENTIFICATION OF CFTR GENE

There have been previous reviews on the identification of the cystic fibrosis gene (Tsui and Buchwald 1991; Tsui 1995). To avoid repeating

Editors: John R. Riordan, Richard C. Boucher, and Paul M. Quinton

Additional Perspectives on Cystic Fibrosis available at www.perspectivesinmedicine.org

Copyright © 2013 Cold Spring Harbor Laboratory Press; all rights reserved; doi: 10.1101/cshperspect.a009472

Cite this article as *Cold Spring Harb Perspect Med* 2013;3:a009472

L.-C. Tsui and R. Dorfman

all that has been said, we have decided to provide only a historical account of key developments as well as factors that contributed to the CF gene identification.

1. CF was first found linked to PON, but this linkage did not directly result in identification of the CF gene because the chromosome location of PON was not known. Nevertheless, the finding confirmed CF to be a relatively homogeneous genetic disease and suggested that it would be possible to identify the CF gene by genetic linkage analysis with a large number of nuclear (two-generation) families.
2. The first DNA marker found linked to CF was an anonymous marker known as CRI-917 (Tsui et al. 1985a), which was subsequently designated D7S15.
3. Before the chromosome location of D7S15 was eventually published (Knowlton et al. 1985), rumor spread that the CF gene was localized to chromosome 7 (Newmark 1985). This information quickly led to the detection of linkage between CF and several other DNA markers on the long arm of chromosome 7 (Wainwright et al. 1985; White et al. 1985). In other words, the localization of CF to the long arm of chromosome 7 was almost instantly confirmed. However, further gene mapping was limited by the lack of human sequence data (Botstein and Risch 2003; Cardon and Abecasis 2003; Flint et al. 2005) because, essentially for each candidate region, a physical map had to be created, overlapping DNA segments isolated and analyzed for the presence of coding sequences (known as “chromosome walking”), and each gene assessed for its candidacy as the sought-after gene. This general approach was subsequently dubbed “positional cloning” (Collins 1990, 1992). The technique of gene hopping or jumping techniques (Collins et al. 1987) was also used in an attempt to reduce the number of DNA fragments required to be isolated while moving along the chromosome (Kerem et al. 1989a,b; Rommens et al. 1989a,b).
4. The identification of the CF gene was published in a series of three papers in 1989 (Kerem et al. 1989a; Riordan et al. 1989; Rommens et al. 1989a). The gene was named *cystic fibrosis transmembrane conductance regulator* (*CFTR* for short). The supporting evidence for the gene was the discovery of the most common mutation in CF, named $\Delta F508$ (now renamed p.F508del). Because this 3-bp deletion was found on a relatively rare extended chromosome haplotype, it was possible to apply allelic association (commonly known as “linkage disequilibrium”) in the fine mapping of the CF gene, and haplotype mapping (ancestral recombination) was proposed as a means to map disease genes by Cox and Chakravarti (Cox et al. 1989).
5. The strong allelic association led to the suggestion that the p.F508del mutation arose from one single event (Cutting et al. 1989; Kerem et al. 1989a). Furthermore, the occurrence of this common mutation in a region with extended linkage disequilibrium was used to form part of the argument for the initial identification of the *CFTR* gene (Kerem et al. 1989a).
6. It is of interest to note, however, that haplotype analysis was used to assess the genetic heterogeneity of the disease before the cloning of the gene (Tsui and Buchwald 1988). For example, it was used to show that pancreatic-sufficient CF patients had more variable and different haplotypes than pancreatic-insufficient patients (Kerem et al. 1989b).
7. It is also of interest to note that a great deal of effort was used to argue for the 3-bp p.F508del deletion as a bona fide mutation because it was not immediately apparent. Its status was somewhat questionable until the results of subsequent functional analysis of the p.F508del-*CFTR* protein (Drumm et al. 1991). The status of *CFTR* as the CF-causing gene was confirmed earlier, however, by the discovery of additional disease-causing mutations (Cutting et al. 1990).
8. Although CF is a relatively homogeneous, monogenic disease with distinct clinical

features, it is possible that a very small number of clinically diagnosed CF patients do not have mutations in the *CFTR* gene. In other words, some patients may be “phenocopies” of CF; despite their classical CF appearance, the disease in these patients may be caused by mutations in other genes. In support of this possibility, it was reported that mouse models with ENaC channel overexpression showed a phenotype reminiscent of CF (Donaldson and Boucher 2007). In fact, some patients with CF-like disease but mutations in only one copy of their *CFTR* genes were found to carry gain-of-function mutations in ENaC-encoding subunits (Sheridan et al. 2005; Azad et al. 2009; Fajac et al. 2009).

CDNA CLONING, MUTATION ANALYSIS, AND *CFTR* GENE STRUCTURE

The initial characterization of the *CFTR* gene was based on the isolation and alignment of genomic clones with the cDNA clones derived from the T84 human colonic adenocarcinoma cell line and those containing the p.Phe508del mutation isolated from a CF sweat gland cDNA library. Some of corresponding genomic DNA segments were missed because either the exons therein were small and thus refractory to isolation by hybridization or certain exons were skipped in the cDNA clones used for the construction of consensus sequence. Therefore, the *CFTR* gene was thought to contain 24 exons (Riordan et al. 1989). Subsequently, it became apparent that the gene contains 27 exons (Table 1); exons 6b, 14b, and 17b were missed in the original publication.

It is now well established that the full-length *CFTR* mRNA contains 6128 nucleotides. Alignment of this sequence with the genomic DNA sequence derived from the Human Genome Project has provided the precise exon–intron structure and intronic sequences of the *CFTR* gene (Table 1).

Based on the recent sequencing and functional data, the transcription module of *CFTR* was established to be >216 kb in size (see Fig. 1). The *CFTR* gene itself spans only 189.36 kb; however, the immediate promoter can be extended

as far as 20.9 kb upstream, where the CTCF-dependent insulator element is located—the expanded promoter region includes the regulatory binding element required for proper gene expression (Blackledge et al. 2007). An insulator element 6.8 kb downstream from *CFTR* includes a DHS site with a CTCF binding site followed by additional DHS sites, at +6.8 kb, +7.0 kb, and +15.6 kb, which appear to contribute to DNA looping, limiting the boundary of the *CFTR* transcriptional unit (Blackledge et al. 2009). Thus, the total size of the *CFTR* transcription unit between these two insulating elements is ~216.7 kb.

SEGMENTAL DUPLICATION OF THE EXON 9 REGION

One of complications in the *CFTR* gene structure analysis stemmed from the segmental duplication of the ~30-kb region including exon 9 (Rozmahel et al. 1997). These multiple segmental duplications containing exon 9 and its flanking intron sequences (Fig. 2) have been found in multiple locations across the human genome (Liu et al. 2004). Although how these sequences became amplified with two flanking LINE1 elements is not exactly clear, multiple copies of exon 9-related sequences become problematic when the exon 9 sequence is being examined for mutation analysis, especially when PCR amplification is used (El-Seedy et al. 2009). The pyrimidine track immediately upstream of exon 9 also appears to influence mRNA splicing, causing exon 9 skipping (see below) (Chu et al. 1991, 1993).

REGULATION OF *CFTR* TRANSCRIPTION

To understand the pathophysiology of CF, one of the approaches is to discover the *CFTR* expression pattern in different tissue surveys (Gregory et al. 1990; Trezise and Buchwald 1991; Ward et al. 1991). *CFTR* is found to be expressed in the epithelial cells of a variety of tissues and organs, whose functions are significantly affected in CF patients: lung and trachea, pancreas, liver, intestines, and sweat glands. Low levels of *CFTR* transcripts can be found in kidney, uterus, ovary,

Table 1. Summary of *CFTR* coding regions according to historical and current nomenclature

Exon	Historical exon name	Start	End	Length	Exon	Start	End	Length
Exon 1	Exon 1	117,120,017	117,120,201	185	Intron 1-2	117,120,202	117,144,306	24,105
Exon 2	Exon 2	117,144,307	117,144,417	111	Intron 2-3	117,144,418	117,149,087	4670
Exon 3	Exon 3	117,149,088	117,149,196	109	Intron 3-4	117,149,197	117,170,952	21,756
Exon 3	Exon 3	117,170,953	117,171,168	216	Intron 4-5	117,171,169	117,174,329	3161
Exon 5	Exon 5	117,174,330	117,174,419	90	Intron 5-6	117,174,420	117,175,301	882
Exon 6	Exon 6a	117,175,302	117,175,465	164	Intron 6-7	117,175,466	117,176,601	1136
Exon 7	Exon 6b	117,176,602	117,176,727	126	Intron 7-8	117,176,728	117,180,153	3426
Exon 8	Exon 7	117,180,154	117,180,400	247	Intron 8-9	117,180,401	117,182,069	1669
Exon 9	Exon 8	117,182,070	117,182,162	93	Intron 9-10	117,182,163	117,188,694	6532
Exon 10	Exon 9	117,188,695	117,188,877	183	Intron 10-11	117,188,878	117,199,517	10,640
Exon 11	Exon 10	117,199,518	117,199,709	192	Intron 11-12	117,199,710	117,227,792	28,083
Exon 12	Exon 11	117,227,793	117,227,887	95	Intron 12-13	117,227,888	117,230,406	2519
Exon 13	Exon 12	117,230,407	117,230,493	87	Intron 13-14	117,230,494	117,231,987	1494
Exon 14	Exon 13	117,231,988	117,232,711	724	Intron 14-15	117,232,712	117,234,983	2272
Exon 15	Exon 14a	117,234,984	117,235,112	129	Intron 15-16	117,235,113	117,242,879	7767
Exon 16	Exon 14b	117,242,880	117,242,917	38	Intron 16-17	117,242,918	117,243,585	668
Exon 17	Exon 15	117,243,586	117,243,836	251	Intron 17-18	117,243,837	117,246,727	2891
Exon 18	Exon 16	117,246,728	117,246,807	80	Intron 18-19	117,246,808	117,250,572	3765
Exon 19	Exon 17a	117,250,573	117,250,723	151	Intron 19-20	117,250,724	117,251,634	911
Exon 20	Exon 17b	117,251,635	117,251,862	228	Intron 20-21	117,251,863	117,254,666	2804
Exon 21	Exon 18	117,254,667	117,254,767	101	Intron 21-22	117,254,768	117,267,575	12,808
Exon 22	Exon 19	117,267,576	117,267,824	249	Intron 22-23	117,267,825	117,282,491	14,667
Exon 23	Exon 20	117,282,492	117,282,647	156	Intron 23-24	117,282,648	117,292,895	10,248
Exon 24	Exon 21	117,292,896	117,292,985	90	Intron 24-25	117,292,986	117,304,741	11,756
Exon 25	Exon 22	117,304,742	117,304,914	173	Intron 25-26	117,304,915	117,305,512	598
Exon 26	Exon 23	117,305,513	117,305,618	106	Intron 26-27	117,305,619	117,306,961	1343
Exon 27	Exon 24	117,306,962	117,308,715	1754				

Data are based on the Ensembl release 61–Feb 2011 (http://www.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000001626;r=7:117119358-117308719;t=ENST000000003084).

Figure 1. Scheme of *CFTR* gene, transcript, and mutation distribution. (A) Scaled schematic of exons and introns followed by current and historical exon numbering. (B) A rescaled exon scheme with minimized introns, which are drawn not to scale. (C) Distribution of known mutations and polymorphisms as vertical bars.

thyroid, and even higher levels in salivary gland and bladder, but the epithelial cell function is not seriously compromised in tissues and organs of CF patients. It is possible that there is sufficient compensation of the missing function by other ion transporters. It is of interest to note that the low levels of CFTR expression in these tissues are driven off an alternative promoter (McCarthy and Harris 2005). These studies have also led to the identification of several tissue-specific splicing isoforms and splicing elements (see below).

The promoter region of the *CFTR* gene lacks a TATA box and is quite GC rich, although generally it is not methylated.³ As such, the CFTR promoter has not been well defined. The immediate promoter region contains an important *cis*-acting element within 632 nucleotides upstream of the translation start site (Lewandowska et al. 2010). The definition of the key promoter regions is also suggested by several mutations that appear to disrupt transcription initiation; for example, the c.-234T>A (-102T>A) muta-

tion was first reported by Claustres and coworkers (Romey et al. 1999). Mutations in more distant regions have been shown to affect the efficiency of transcription: The c.-1750A>G polymorphism reduces the transcription efficiency by >50% (Lewandowska et al. 2010).

The predominant CFTR transcription start site is as described in the original study by Rommens et al. (1989a), meaning that the majority of CFTR transcripts are driven from the key promoter, described in the previous section. However, although the canonical transcripts are found in cells with high CFTR expression, alternative transcription start sites are apparently used in cell lines with low expression levels (McCarthy and Harris 2005). CFTR transcripts from even more distant transcription start sites between -868 and -794 can be found in CFPAC and T84 cell lines (McCarthy and Harris 2005).

The immediate promoter region has also been characterized by consensus binding sites for several transcription factors: CTCF, AP-1, SP1, GRE, CRE, C/EBP, and Y-box proteins (McCarthy and Harris 2005). DNase I hypersensitive site (DHS) mapping has been used to map various putative enhancer sequences

³http://www.ensembl.org/Homo_sapiens/Location/View?g=ENSG0000001626;r=7:117078291-117109965#r=7:117094129-117220824.

L.-C. Tsui and R. Dorfman

Figure 2. Scheme reconstructed based on data from the Rozmahel et al. (1997) publication showing the schematic diagram of proposed exon 9 duplication by retrotransposition.

within the CFTR intragenic regions (Table 2). Presumably, multiple transcription factors can bind to chromatin at these sequences, opening the DNA and extending the physical interactions with the promoter, thereby affecting transcription. HNF1 α binding sites, indicative of putative enhancer elements, can be found in multiple locations inside introns 10, 17a, and 20 (Mouchel et al. 2004; McCarthy et al. 2009); it has been shown that RNAi-mediated inhibition of the HNF1 α could lead to reduction of CFTR expression. Additional enhancer elements have been located in introns 1 and 11, and HNF1 α and p300 are involved in the regulation of CFTR expression (Ott et al. 2009a). However, understanding of the transcriptional control of *CFTR* gene expression is far from comprehensive.

SPLICING

As described in the previous section, the majority of CFTR transcripts start from around the same site. There are, however, transcripts with alternative splicing, resulting in skipping of different exons. Some alternatively spliced species can produce truncated CFTR protein with partial function; for example, an alternatively spliced CFTR transcript missing exon 5 found in heart may produce an active channel albeit with significantly reduced function (Xie et al. 1995).

Many of the alternatively spliced transcripts are expected to introduce translation reading frameshifts, thereby introducing premature translation termination that can induce nonsense-mediated RNA decay, removing the aberrant mRNA species. For some mutations,

Table 2. List of genomic regulatory regions in the human *CFTR* gene

Regulatory region	Analysis	Location	Length (bp)
ENSR00000633238	Ensembl Regulatory Build	7:117110774–117111212	481
ENSR00000633239	Ensembl Regulatory Build	7:117116334–117117806	1617
ENSR00000633240	Ensembl Regulatory Build	7:117119693–117120138	488
ENSR00000633241	Ensembl Regulatory Build	7:117121119–117121448	360
ENSR00000633242	Ensembl Regulatory Build	7:117125053–117125742	756
ENSR00000633243	Ensembl Regulatory Build	7:117141802–117141975	186
ENSR00000633244	Ensembl Regulatory Build	7:117145373–117145667	319
ENSR00000633245	Ensembl Regulatory Build	7:117199627–117199834	226
ENSR00001050415	Ensembl Regulatory Build	7:117215123–117215415	317
ENSR00000633246	Ensembl Regulatory Build	7:117220696–117220951	280
ENSR00000633247	Ensembl Regulatory Build	7:117221137–117221424	312
ENSR00000633248	Ensembl Regulatory Build	7:117222043–117223014	1068
ENSR00000633249	Ensembl Regulatory Build	7:117227963–117229004	1144
ENSR00000683157	Ensembl Regulatory Build	7:117236602–117237299	764
ENSR00000633250	Ensembl Regulatory Build	7:117237932–117238146	233
ENSR00000633251	Ensembl Regulatory Build	7:117240434–117240790	387
ENSR00000633252	Ensembl Regulatory Build	7:117240865–117241512	708
ENSR00000633253	Ensembl Regulatory Build	7:117259273–117259612	370
ENSR00000633254	Ensembl Regulatory Build	7:117279990–117280443	496
ENSR00000633255	Ensembl Regulatory Build	7:117285455–117286319	949
ENSR00000633256	Ensembl Regulatory Build	7:117288545–117288840	320
ENSR00000633257	Ensembl Regulatory Build	7:117293467–117294015	603
ENSR00000633258	Ensembl Regulatory Build	7:117294880–117295179	330
ENSR00000633259	Ensembl Regulatory Build	7:117299551–117299849	323
ENSR00000633260	Ensembl Regulatory Build	7:117305402–117305885	532
ENSR00000633261	Ensembl Regulatory Build	7:117306006–117306648	703
ENST00000003084	miRanda microRNA target predictions: hsa-miR-28-3p	7:117307497–117307518	22
ENSR00001050416	Ensembl Regulatory Build	7:117307574–117307827	278
ENST00000003084	miRanda microRNA target predictions: hsa-miR-28-3p	7:117307750–117307771	22
ENST00000003084	miRanda microRNA target predictions: hsa-miR-433	7:117308205–117308229	25
ENSR00000633262	Ensembl Regulatory Build	7:117309937–117310180	268

however, the resulting mRNA species with exon skipping can persist and produce a dysfunctional protein (Hull et al. 1994). In the case of the c.2491G>T (p.Glu831X, also known as 2623G>T, E831X) variant, the presence of an alternative spliced mutant transcript may explain a mild disease presentation and pancreatic sufficiency in patients carrying this nonsense mutation (Hinzpeter et al. 2010).

Based on studies with minigene constructs, Hinzpeter et al. (2010) have proposed that the c.2491G>T mutation not only introduces a premature stop codon, but also creates a binding

site (CAGTAG) for the splicing factor that recognizes the consensus sequence NAGNAG. The study shows that the mutation can lead to the production of three different mRNAs: CFTR_831-873del, CFTR_E831X, and CFTR_E831del, and, unexpectedly, the latter transcript could bypass the premature termination codon and produce a functional p.E831del protein.

ROLE OF SPLICING REGULATION IN CF DISEASE

Mutations occurring at the intron–exon boundaries with highly conserved sequences, that is,

L.-C. Tsui and R. Dorfman

splicing junctions at -1 , -2 , -3 , and $+1$, $+2$, $+3$ positions are expected to affect splicing of the immediately adjacent exons. The consequence of such mutations, for example, c.1117-1G>A and c.1209+1G>A, are generally considered to be severe, whereas mutations occurring at more distant positions, for example, $+5$, $+6$, or -5 and -6 , are mild, typically associated with mild CF disease. Although the sequences between $+6$ and $+30$ are to some degree conserved in evolution, suggestive of a possible role of them in splicing (Aznarez et al. 2008), there is not much information regarding the severity of mutations found beyond the ± 6 positions, except for c.1117-18G>T, c.1209+18A>C, and c.2909-15T>G, which are frequently found in infertile males.

Some of the mutations are found to affect splicing efficiency, presumably because they alter the binding sites of splicing factors (Aznarez et al. 2003). The most noted example in the *CFTR* gene is the polypyrimidine track located in intron 8 in front of exon 9; the length of the T-track varies from five to nine Ts (Chu et al. 1993). Although the 5T variant is associated with low efficiency of splicing, the effect by itself is insufficient to cause any known disease. However, if the 5T allele is present in combination with another mild mutation, p.Arg117His (R117H) (Gervais et al. 1993), the combined effect of R117H-5T (reduced protein function and lower abundance of correctly spliced transcript) is apparently sufficient to cause CF, albeit the pancreatic sufficient form (Chu et al. 1993; Massie et al. 2001; Zuccato et al. 2004). Other examples of mutations affecting splicing efficiency include several missense (D648V and T665S) and nonsense (E664X) mutations, presumably due to the disruption of the ESE elements within exon 13 (Aznarez et al. 2003).

It is of interest to note that synonymous variants (or polymorphisms) are much less common in the *CFTR* gene than in other disease-related genes. The reason for this is unknown, but it may be a property of genomic regions with strong allelic association (or linkage disequilibrium). However, some of these rare synonymous variants, for example, c.2679G>T (p.Gly893Gly), c.2898G>A, c.3204C>T, c.3285

A>T, and c.3897A>G (rs1800131), could have an effect on splicing because they occur in the consensus sequences for ESEs (exon splicing enhancers) or ESRs (exon splicing repressors). The exon 12 variant c.2679G>T (p.Gly893Gly), for example, is associated with a mild form of CF; the mutation affects a highly conserved sequence and causes skipping of exon 12 (Pagani et al. 2005; Raponi et al. 2007). In fact, the sequence alteration in exon 12 has been found coupled with a new 5' splice site in exon 15, resulting in a transcript lacking 76-amino-acid codons (Faa et al. 2010). This and other observations bring about a suggestion that some of the synonymous changes can affect *CFTR* mRNA folding and, in turn, protein synthesis (Bartoszewski et al. 2010). Supporting this view, synonymous mutations have been shown to affect posttranslational protein modification and protein stability and were found to be overrepresented among the "top hits" identified in several Genome Wide Association Studies of common diseases (Chen et al. 2010; Li et al. 2010; Plotkin and Kudla 2010), again showing a crucial, but frequently underestimated, role of synonymous variations in disease pathophysiology.

TRANS-SPlicing

Trans-splicing has been described for the *CFTR* gene, making the initial characterization of the gene structure difficult (Rommens et al. 1989a). It is unclear how often it happens for other genes, but *trans*-splicing has been suggested as a possible strategy for the correction of *CFTR* mutations (Wardle and Harris 1995; Mansfield et al. 2000; Buratti and Baralle 2001; Liu et al. 2002). For example, it has been feasible to reconstruct a full-length *CFTR* transcript with recombinant DNA containing exons 10–24, provided in *trans*, in transfected cells (Mansfield et al. 2000, 2003; Liu et al. 2002). It remains to be determined, however, whether it is possible to engineer gene replacement therapy by introducing a large *CFTR* cDNA construct into patient cells. Nevertheless, this approach may be useful for gene repair for rare missense and frameshift mutations, which are

unlikely to benefit from drugs with molecular chaperone properties (see Thomas 2013).

STUDY OF CFTR IN MICE AND OTHER SPECIES

Because of the difficulty in studying the *CFTR* gene in human tissues, extensive studies have been devoted to the isolation and characterization of genes homologous to *CFTR* in other species. CFTR (ABCC7) is a member of the C subfamily of ABC transporters, which includes sulfonylurea receptors (SURs) and multidrug resistance-associated proteins (MRPs). The ABCC subfamily proteins have two transmembrane membrane-spanning domains (MSDs), each consisting of six transmembrane (TM) segments, although some the MRP channels have an additional MSD₀ (Deeley et al. 2006; Toyoda et al. 2008). All ABC proteins also have two nucleotide-binding domains (NBDs), but only CFTR has the novel R domain (CFTR protein structure and function) (for review, see Hunt et al. 2013; Hwang and Kirk 2013). Better understanding of CFTR function may derive from studies of orthologous genes in other species. One line of such investigation has been attempts to search for naturally occurring *CFTR* gene mutations in various animal species, including sheep, guinea pigs, ferrets, dogs, and even chickens (Tebbutt 1995). Unfortunately, although some sequence variations have been identified, none of them appear to correlate with any CF-like disease.

Because mouse is the best-studied experimental animal for human diseases, several CFTR-deficient murine lines were created with the use of recombinant DNA technologies (Snouwaert et al. 1992). These include simple knockout (KO) and specific gene replacements on various genetic strains of mice, providing a spectrum of clinical phenotypes (Davidson and Rolfe 2001; Guilbault et al. 2007; Wilke et al. 2011). Despite undetectable CFTR transcripts in all the complete KO strains—*Cftr*^{tm1Unc}, *Cftr*^{tm1Cam}, *Cftr*^{tm1Hsc}, *Cftr*^{tm3Bay}, *Cftr*^{tm3Uth}—the survival varied from <5% to 40%. Furthermore, survival in the hypomorphic strains

Cftr^{tm1Hgu} and *Cftr*^{tm1Bay} was largely unaffected, reminiscent of the residual CFTR function for some CF mutations and disease severity in humans, particularly with respect to pancreatic status and extent of CF lung disease (Wilke et al. 2011). The *Cftr*^{tm1Kth}, *Cftr*^{tm2Cam}, and *Cftr*^{tm1Eur} mice were created to study the p.Phe508del mutation, whereas other lines, such as *Cftr*^{tm1G551D}, for other missense mutations found in CF patients. Despite intestinal obstruction and reduced weight, the symptoms observed in CFTR-deficient mice appear to be very different from those of the human disease, particularly in terms of pulmonary involvement.

The availability of CFTR-deficient mice has, however, allowed the study of possible modifying genes for CF, because the same CFTR mutation(s) can result in different disease presentations among different CF patients. Variability in phenotypic manifestations was also observed in mice, and it appeared to be largely dependent on the genetic background of the knockout mice (Rozmahel et al. 1996). Indeed, these differences were used to map the *CFM1* and several other genetic modifier loci associated with varying “disease severity” in CFTR-deficient mice (Rozmahel et al. 1996; Haston et al. 2002a,b; Haston and Tsui 2003).

The CFTR-deficient pig (Rogers et al. 2008 a,b) appears to be a very promising animal model for CF because the mutant animals essentially recapitulate all of the CF-specific gastrointestinal abnormalities and the critical aspects of CF lung disease (Meyerholz et al. 2010). For example, meconium ileus is fully penetrant in the mutant animals: All of the newborn CFTR-deficient piglets suffer from intestinal obstruction and defects in gallbladder development. However, there are differences in disease severity between the completely null CF pigs (Rogers et al. 2008a,b) and the p.Phe508del pigs (Ostedgaard et al. 2011).

Ferrets with *CFTR* gene knockout may provide yet another promising animal model for CF. The CFTR-deficient animals have been found to show many of the characteristics of human CF disease, including defective airway chloride transport and submucosal gland fluid secretion;

L.-C. Tsui and R. Dorfman

meconium ileus with variable penetrance (75% MI); pancreatic, liver, and vas deferens diseases; and a predisposition to lung infection in early life (Sun et al. 2010).

CROSS-SPECIES STUDIES OF CFTR SEQUENCES

Study of sequences conserved across multiple species may identify potential regulatory sequences important for *CFTR* gene expression and splicing. Such information may be obtained from the ENCODE project (ENCyclopedia Of DNA Elements), which has been developed to identify conserved regions and putative gene expression regulatory regions; *CFTR* is one of the several candidate genes included in the study (ENCODE Project Consortium 2004; Birney et al. 2007). The ChIP-seq method that involves DNA digestion and chromatin immunoprecipitation followed by sequencing was used to identify the transcription factor and histone binding sites. RNA sequencing was used for more accurate gene and exon mapping. DNA methylation and other DNA modifications were assessed in multiple cell types. The conserved regions and functional DNA modification and protein–DNA interaction sites identified by ENCODE (<http://encodeproject.org>) are compared with the *CFTR* variants recorded in the Cystic Fibrosis Mutation Database in the PhenCode browser (<http://globin.bx.psu.edu/phencode/>) (Giardine et al. 2007).

However, not many *CFTR* mutations have been found associated with these putative regulatory regions thus far (Ott et al. 2009b). There are several possible explanations for the apparent paucity of mutations detected in these regions. First, the biological significance of these sequences has not been investigated. Second, mutation screening studies in CF patients have been primarily concentrated in the coding regions, including at most the immediate flanking intronic regions. Third, mutations in regulatory regions may only affect the expression levels of *CFTR*, resulting in very mild disease, not recognized as CF.

Identification of crucial regulatory sequences that can direct proper tissue specificity may be

critically useful in the construction of *CFTR* expression vectors for CF gene therapy because, although *CFTR* deficiency could be rescued by overexpression of wild-type cDNA (Drumm et al. 1990), expression of the gene in atopic tissues could lead to adverse side effects. One study showed that atopic expression of *CFTR* in the heart could cause arrhythmia and sudden death of transgenic animals (Ye et al. 2010). Therefore, most of the CF gene therapy studies have only used heterologous promoters that are well characterized. Among these, the KRT18 promoter appears to be promising in targeting *CFTR* expression in the lung epithelial cells (Koehler et al. 2001).

CF MUTATION DATABASE AND MUTATION NOMENCLATURE

The Cystic Fibrosis Genetic Analysis Consortium was formed to facilitate *CFTR* mutation analysis, at a time before the gene structure was fully characterized. A set of general agreements and guidelines was agreed upon at the North American CF Conference in Florida on October 13, 1989. Briefly, members of the Consortium were encouraged to report data to the Consortium before submission for publication. Because it would usually take at least three months for any publication to come to print, early communication within the group would allow other members to use the information and screen their respective populations in a coordinated and concerted effort. The Consortium also helped disseminate information and ideas regarding novel methods and technical tips that would improve genetic testing. Members agreed that the group that reported the information would hold the priority to publish the primary data. Other members of the Consortium might use the information for their own research before the publication of the original report but might not publish any secondary observations without the consent of the original group.

In addition, a standardized mutation nomenclature system was proposed (Beaudet and Tsui 1993). This was the first time a systematic nomenclature was used to describe human disease mutations. These guidelines were adopted

Table 3. Summary statistics for classification of *CFTR* mutations recorded in CFMDB (January 2011)

Mutation type	Count	Frequency (%)
Missense	765	40.41
Frameshift	303	16.01
Splicing	225	11.89
Nonsense	160	8.45
In frame in/del	37	1.95
Large in/del	49	2.59
Promoter	15	0.79
Sequence variation	269	14.21
Unknown	70	3.7
Total	1893	

by the genetic community and applied in the naming of mutations in other genes, and laid the basis for subsequent versions of mutation nomenclature developed by the Human Genome Variation Society.⁴

After the launching of the Consortium in 1989, its membership quickly grew to include 130 groups of CF laboratories from more than 30 countries. The Consortium exchanged mutation reports and facilitated the discussions and research on genotype–phenotype correlations, confirmed clinical data, and enabled development of genetic screening tools. With the development of Internet technologies, because the data held by the Consortium were essentially open to the entire scientific community, a Web-based Cystic Fibrosis Mutation Database (CFMDB) was launched in 1995. Even to date, it is one of the most cited genetic databases and is visited by more than 500 unique users per day. The Database contains information for more than 1890 disease-causing mutations and variants of different types⁵ (Tables 3 and 4: mutation distribution summary). Although the changes in HGV nomenclature rules have necessitated a revision (April 2010) of CFMDB, all of the previous (legacy) names for CF mutations are retained as a display option along with new maps and up-to-date numbering systems according to the current knowledge of the gene structure.

All new mutation submissions, however, would only appear in the revised HGV format.

CONCLUDING REMARKS

It has often been said that, had the full human genome been mapped and sequenced, the task of finding the cystic fibrosis gene would have taken only a fraction of the time and cost. The identification of the *CFTR* gene has, however, provided an entry point to understand the basic defect underlying this multisystemic disease, whereby better detection and treatment can be devised for the sake of all of the affected individuals.

In addition to providing the first example of positional cloning, the CF gene cloning work also provided the concept of haplotype mapping, which was illustrated on the cover of the September 8, 1989 issue of *Science* (see Fig. 3). Although it was not properly introduced in the report of Kerem et al. (1989), the solution for fine-mapping of human traits by haplotype analysis was initially attempted by Cox and Chakravarti (Cox et al. 1989) and later properly introduced by Lander and Kruglyak (Kruglyak 1997, 2005, 2008).

The spirit of sharing and cooperation was exemplified in CF genetic research as early as the CF gene mapping stage; the primary DNA marker data set used in the CF gene mapping study was shared with the human gene mapping community, so that different algorithms on multipoint mapping could be tested (Spence et al. 1989).

After identification of the *CFTR* gene, crucial information for the detection of additional mutations was shared among members of the CF Genetic Analysis Consortium before the gene structure was entirely elucidated and published (Zielenski et al. 1991). New mutation data were also shared before formal publication as describe in the preceding section.

Most remarkable, however, was the honorable competition and cooperative spirit that could be found even during the race to the cloning of the CF gene. In 1987, the group led by Robert Williamson had used a rather elegant technique and identified a DNA fragment very

⁴<http://www.hgvs.org/mutnomen/history.html>.

⁵<http://www.genet.sickkids.on.ca/StatisticsPage.html>.

L.-C. Tsui and R. Dorfman

Table 4. Distribution of mutations in CFTR protein domains

Domain	Exon number	Missense	Frameshift/stop	In frame in/del	Splicing	Total
MSD1	Exon 4	18	12	2	0	32
MSD2	Exon 4-5	13	5	2	1	21
MSD3	Exon 6a-6b	10	3	0	0	13
MSD4	Exon 6b	10	5	1	0	16
MSD5	Exon 7-8	13	7	2	0	22
MSD6	Exon 8	9	9	0	0	18
NBD1	Exon 10-13	147	37	3	1	188
R	Exon 13-14a	58	87	1	1	147
MSD7	Exon 15	13	6	0	0	19
MSD8	Exon 16-17a	12	10	2	1	25
MSD9	Exon 17b	25	2	0	0	27
MSD10	Exon 17b	12	9	0	0	21
MSD11	Exon 19	2	7	0	0	9
MSD12	Exon 20	6	7	0	0	13
NBD2	Exon 20-24	74	65	0	0	139
Total		422	271	13	4	710 ^a

^aPlease note that the total number of mutations is smaller because synonymous variants are not shown in the summary; also, different cDNA mutations that lead to identical missense changes and mutations outside of the functionally defined domains are now shown in the summary. Therefore, the total number of missense mutations is substantially smaller than indicated in Table 3.

Figure 3. Schematic diagram illustrating the concept of haplotype mapping. (A) The color bars represent chromosomes from different patients with cystic fibrosis; this diagram was used as the basis for the cover design of the September 8, 1989 issue of *Science*. (B) The DNA marker haplotypes associated with the CF chromosomes carrying ΔF508; the “1s” and “2s” inside the colored bars on the *left* represent schematic alleles of DNA markers around the CFTR gene; the original data appeared in Table 3 on p. 1076 of Kerem et al. (1989); the numbers on the *right* denote the counts of CF and normal (N) chromosomes with the corresponding DNA marker haplotypes shown on the *left*.

close to the CF gene (Scambler et al. 1986; Estivill et al. 1987). In fact, their data misled them and many others to believe that they had cloned the gene. Due credit must be given to Williamson and his colleagues, however, because they were quick to admit to the scientific world that the candidate gene they had isolated was not the right gene (Wainwright et al. 1988). Their timely sharing of this valuable piece of information reversed the funding decision at the National Institutes of Health (of the United States), so that the CF gene cloning work in several laboratories could be continued without any interruption of funding.

We would like to end this work with a quote from a previous review article (Tsui and Estivill 1991):

Finally, there has been tremendous competition in the field of disease gene cloning. . . . It is difficult to describe the feelings of performing repeated searches for clone after clone without an open reading frame, constructing and screening genomic and cDNA libraries one after another, watching the come and go of a candidate gene, and working under the fear that the gene has perhaps been identified by another group . . . although being the first to report appears to be the only consolation of this hard work, one should keep in mind that gene identification is in fact only the first step . . . in understanding a disease. . . .

REFERENCES

*Reference is also in this collection.

- Allen FH Jr, Dooley RR, Shwachman H, Steinberg AG. 1956. Linkage studies with cystic fibrosis of the pancreas. *Am J Hum Genet* **8**: 162–176.
- Andersen DH, Hodges RG. 1946. Celiac syndrome; genetics of cystic fibrosis of the pancreas, with a consideration of etiology. *Am J Dis Child* **72**: 62–80.
- Azad AK, Rauh R, Vermeulen F, Jaspers M, Korbmacher J, Boissier B, Bassinet L, Fichou Y, des Georges M, Stanke F, et al. 2009. Mutations in the amiloride-sensitive epithelial sodium channel in patients with cystic fibrosis-like disease. *Hum Mutat* **30**: 1093–1103.
- Aznarez I, Chan EM, Zielenski J, Blencowe BJ, Tsui LC. 2003. Characterization of disease-associated mutations affecting an exonic splicing enhancer and two cryptic splice sites in exon 13 of the cystic fibrosis transmembrane conductance regulator gene. *Hum Mol Genet* **12**: 2031–2040.
- Aznarez I, Barash Y, Shai O, He D, Zielenski J, Tsui LC, Parkinson J, Frey BJ, Rommens JM, Blencowe BJ. 2008. A systematic analysis of intronic sequences downstream of 5' splice sites reveals a widespread role for U-rich motifs and TIA1/TIAL1 proteins in alternative splicing regulation. *Genome Res* **18**: 1247–1258.
- Bartoszewski RA, Jablonsky M, Bartoszewska S, Stevenson L, Dai Q, Kappes J, Collawn JE, Bebek Z. 2010. A synonymous single nucleotide polymorphism in $\Delta F508$ CFTR alters the secondary structure of the mRNA and the expression of the mutant protein. *J Biol Chem* **285**: 28741–28748.
- Beaudet AL, Tsui LC. 1993. A suggested nomenclature for designating mutations. *Hum Mutat* **2**: 245–248.
- Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**: 799–816.
- Blackledge NP, Carter EJ, Evans JR, Lawson V, Rowntree RK, Harris A. 2007. CTCF mediates insulator function at the CFTR locus. *Biochem J* **408**: 267–275.
- Blackledge NP, Ott CJ, Gillen AE, Harris A. 2009. An insulator element 3' to the CFTR gene binds CTCF and reveals an active chromatin hub in primary cells. *Nucleic Acids Res* **37**: 1086–1094.
- Botstein D, Risch N. 2003. Discovering genotypes underlying human phenotypes: Past successes for Mendelian disease, future approaches for complex disease. *Nat Genet* **33**: 228–237.
- Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* **32**: 314–331.
- Buratti E, Baralle FE. 2001. Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J Biol Chem* **276**: 36337–36343.
- Cardon LR, Abecasis GR. 2003. Using haplotype blocks to map human complex trait loci. *Trends Genet* **19**: 135–140.
- Chen R, Davydov EV, Sirota M, Butte AJ. 2010. Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. *PLoS ONE* **5**: e13574.
- Chu CS, Trapnell BC, Murtagh JJ Jr, Moss J, Dalemans W, Jallat S, Mercenier A, Pavirani A, Lecocq JP, Cutting GR, et al. 1991. Variable deletion of exon 9 coding sequences in cystic fibrosis transmembrane conductance regulator gene mRNA transcripts in normal bronchial epithelium. *EMBO J* **10**: 1355–1363.
- Chu CS, Trapnell BC, Curristin S, Cutting GR, Crystal RG. 1993. Genetic basis of variable exon 9 skipping in cystic fibrosis transmembrane conductance regulator mRNA. *Nat Genet* **3**: 151–156.
- Collins FS. 1990. Identifying human disease genes by positional cloning. *Harvey Lect* **86**: 149–164.
- Collins FS. 1992. Positional cloning: Let's not call it reverse anymore. *Nat Genet* **1**: 3–6.
- Collins FS, Drumm ML, Cole JL, Lockwood WK, Vande Woude GF, Iannuzzi MC. 1987. Construction of a general human chromosome jumping library, with application to cystic fibrosis. *Science* **235**: 1046–1049.

L.-C. Tsui and R. Dorfman

- Cox TK, Kerem B, Rommens J, Iannuzzi MC, Drumm M, Collins FS, Dean M, Tsui L-C, Chakravarti A. 1989. Mapping of the cystic fibrosis gene using putative ancestral recombinants. *Am J Hum Genet* **45**: A136.
- Cutting GR, Antonarakis SE, Buetow KH, Kasch LM, Rosenstein BJ, Kazazian HH Jr. 1989. Analysis of DNA polymorphism haplotypes linked to the cystic fibrosis locus in North American black and Caucasian families supports the existence of multiple mutations of the cystic fibrosis gene. *Am J Hum Genet* **44**: 307–318.
- Cutting GR, Kasch LM, Rosenstein BJ, Zielenski J, Tsui LC, Antonarakis SE, Kazazian HH Jr. 1990. A cluster of cystic fibrosis mutations in the first nucleotide-binding fold of the cystic fibrosis conductance regulator protein. *Nature* **346**: 366–369.
- Davidson DJ, Rolfe M. 2001. Mouse models of cystic fibrosis. *Trends Genet* **17**: S29–S37.
- Deeley RG, Westlake C, Cole SP. 2006. Transmembrane transport of endo- and xenobiotics by mammalian ATP-binding cassette multidrug resistance proteins. *Physiol Rev* **86**: 849–899.
- Donaldson SH, Boucher RC. 2007. Sodium channels and cystic fibrosis. *Chest* **132**: 1631–1636.
- Drumm ML, Pope HA, Cliff WH, Rommens JM, Marvin SA, Tsui LC, Collins FS, Frizzell RA, Wilson JM. 1990. Correction of the cystic fibrosis defect in vitro by retrovirus-mediated gene transfer. *Cell* **62**: 1227–1233.
- Drumm ML, Wilkinson DJ, Smit LS, Worrell RT, Strong TV, Frizzell RA, Dawson DC, Collins FS. 1991. Chloride conductance expressed by $\Delta F508$ and other mutant CFTRs in *Xenopus* oocytes. *Science* **254**: 1797–1799.
- Eiberg H, Mohr J, Schmiegelow K, Nielsen LS, Williamson R. 1985. Linkage relationships of paraoxonase (PON) with other markers: Indication of PON–cystic fibrosis syndrome. *Clin Genet* **28**: 265–271.
- El-Seedy A, Dudognon T, Bilan F, Pasquet MC, Reboul MP, Iron A, Kitzis A, Ladeveze V. 2009. Influence of the duplication of CFTR exon 9 and its flanking sequences on diagnosis of cystic fibrosis mutations. *J Mol Diagn* **11**: 488–493.
- ENCODE Project Consortium. 2004. The ENCODE (ENCyclopedia Of DNA Elements.) Project. *Science* **306**: 636–640.
- Estivill X, Farrall M, Scambler PJ, Bell GM, Hawley KM, Lench NJ, Bates GP, Kruyer HC, Frederick PA, Stanier P, et al. 1987. A candidate for the cystic fibrosis locus isolated by selection for methylation-free islands. *Nature* **326**: 840–845.
- Faa V, Coiana A, Incani F, Costantino L, Cao A, Rosatelli MC. 2010. A synonymous mutation in the CFTR gene causes aberrant splicing in an Italian patient affected by a mild form of cystic fibrosis. *J Mol Diagn* **12**: 380–383.
- Fajac I, Viel M, Gaitch N, Hubert D, Bienvenu T. 2009. Combination of ENaC and CFTR mutations may predispose to cystic fibrosis–like disease. *Eur Respir J* **34**: 772–773.
- Flint J, Valdar W, Shifman S, Mott R. 2005. Strategies for mapping and cloning quantitative trait genes in rodents. *Nat Rev Genet* **6**: 271–286.
- Gervais R, Dumur V, Rigot JM, Lafitte JJ, Roussel P, Claustres M, Demaille J. 1993. High frequency of the R117H cystic fibrosis mutation in patients with congenital absence of the vas deferens. *N Engl J Med* **328**: 446–447.
- Giardine B, Riemer C, Hefferon T, Thomas D, Hsu F, Zielenski J, Sang Y, Elnitski L, Cutting G, Trumbower H, et al. 2007. PhenCode: Connecting ENCODE data with mutations and phenotype. *Hum Mutat* **28**: 554–562.
- Gregory RJ, Cheng SH, Rich DP, Marshall J, Paul S, Hehir K, Ostedgaard L, Klinger KW, Welsh MJ, Smith AE. 1990. Expression and characterization of the cystic fibrosis transmembrane conductance regulator. *Nature* **347**: 382–386.
- Guilbault C, Saeed Z, Downey GP, Radzioch D. 2007. Cystic fibrosis mouse models. *Am J Respir Cell Mol Biol* **36**: 1–7.
- Haston CK, Tsui LC. 2003. Loci of intestinal distress in cystic fibrosis knockout mice. *Physiol Genomics* **12**: 79–84.
- Haston CK, Corey M, Tsui LC. 2002a. Mapping of genetic factors influencing the weight of cystic fibrosis knockout mice. *Mamm Genome* **13**: 614–618.
- Haston CK, McKerlie C, Newbigging S, Corey M, Rozmahel R, Tsui LC. 2002b. Detection of modifier loci influencing the lung phenotype of cystic fibrosis knockout mice. *Mamm Genome* **13**: 605–613.
- Hinzpeter A, Aissat A, Sondo E, Costa C, Arous N, Gammeiro C, Martin N, Tarze A, Weiss L, de Becdelievre A, et al. 2010. Alternative splicing at a NAGNAG acceptor site as a novel phenotype modifier. *PLoS Genet* **6**: e1001153.
- Hull J, Shackleton S, Harris A. 1994. The stop mutation R553X in the CFTR gene results in exon skipping. *Genomics* **19**: 362–364.
- * Hunt JF, Wang C, Ford RC. 2013. Cystic fibrosis transmembrane conductance regulator (ABCC7) structure. *Cold Spring Harb Perspect Med* doi: 10.1101/cshperspect.a009514.
- * Hwang T-C, Kirk KL. 2013. The CFTR ion channel: Gating, regulation, and anion permeation. *Cold Spring Harb Perspect Med* doi: 10.1101/cshperspect.a009498.
- Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC. 1989a. Identification of the cystic fibrosis gene: Genetic analysis. *Science* **245**: 1073–1080.
- Kerem BS, Buchanan JA, Durie P, Corey ML, Levison H, Rommens JM, Buchwald M, Tsui LC. 1989b. DNA marker haplotype association with pancreatic sufficiency in cystic fibrosis. *Am J Hum Genet* **44**: 827–834.
- Knowlton RG, Cohen-Haguenaer O, Van Cong N, Frezal J, Brown VA, Barker D, Braman JC, Schumm JW, Tsui LC, Buchwald M, et al. 1985. A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. *Nature* **318**: 380–382.
- Koehler DR, Hannam V, Belcastro R, Steer B, Wen Y, Post M, Downey G, Tanswell AK, Hu J. 2001. Targeting transgene expression for cystic fibrosis gene therapy. *Mol Ther* **4**: 58–65.
- Kruglyak L. 1997. What is significant in whole-genome linkage disequilibrium studies? *Am J Hum Genet* **61**: 810–812.
- Kruglyak L. 2005. Power tools for human genetics. *Net Genet* **37**: 1299–1300.

- Kruglyak L. 2008. The road to genome-wide association studies. *Nat Rev Genet* **9**: 314–318.
- Lewandowska MA, Costa FF, Bischof JM, Williams SH, Soares MB, Harris A. 2010. Multiple mechanisms influence regulation of the cystic fibrosis transmembrane conductance regulator gene promoter. *Am J Respir Cell Mol Biol* **43**: 334–341.
- Li Y, Vinckenbosch N, Tian G, Huerta-Sanchez E, Jiang T, Jiang H, Albrechtsen A, Andersen G, Cao H, Korneliussen T, et al. 2010. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet* **42**: 969–972.
- Liu X, Jiang Q, Mansfield SG, Puttaraju M, Zhang Y, Zhou W, Cohn JA, Garcia-Blanco MA, Mitchell LG, Engelhardt JE. 2002. Partial correction of endogenous $\Delta F508$ CFTR in human cystic fibrosis airway epithelia by spliceosome-mediated RNA trans-splicing. *Nat Biotechnol* **20**: 47–52.
- Liu X, Li X, Li M, Acimovic YJ, Li Z, Scherer SW, Estivill X, Tsui LC. 2004. Characterization of the segmental duplication LCR7–20 in the human genome. *Genomics* **83**: 262–269.
- Mansfield SG, Kole J, Puttaraju M, Yang CC, Garcia-Blanco MA, Cohn JA, Mitchell LG. 2000. Repair of CFTR mRNA by spliceosome-mediated RNA trans-splicing. *Gene Ther* **7**: 1885–1895.
- Mansfield SG, Clark RH, Puttaraju M, Kole J, Cohn JA, Mitchell LG, Garcia-Blanco MA. 2003. 5' exon replacement and repair by spliceosome-mediated RNA trans-splicing. *RNA* **9**: 1290–1297.
- Massie RJ, Poplawski N, Wilcken B, Goldblatt J, Byrnes C, Robertson C. 2001. Intron-8 polythymidine sequence in Australasian individuals with CF mutations R117H and R117C. *Eur Respir J* **17**: 1195–1200.
- McCarthy VA, Harris A. 2005. The CFTR gene and regulation of its expression. *Pediatr Pulmonol* **40**: 1–8.
- McCarthy VA, Ott CJ, Phylactides M, Harris A. 2009. Interaction of intestinal and pancreatic transcription factors in the regulation of CFTR gene expression. *Biochim Biophys Acta* **1789**: 709–718.
- Meyerholz DK, Stoltz DA, Pezzulo AA, Welsh MJ. 2010. Pathology of gastrointestinal organs in a porcine model of cystic fibrosis. *Am J Pathol* **176**: 1377–1389.
- Morton NE, Steinberg AG. 1956. Sequential test for linkage between cystic fibrosis of the pancreas and the MNS locus. *Am J Hum Genet* **8**: 177–189.
- Mouchel N, Henstra SA, McCarthy VA, Williams SH, Phylactides M, Harris A. 2004. HNF1 α is involved in tissue-specific regulation of CFTR gene expression. *Biochem J* **378**: 909–918.
- Newmark P. 1985. Testing for cystic fibrosis. *Nature* **318**: 309.
- Ostedgaard LS, Meyerholz DK, Chen JH, Pezzulo AA, Karp PH, Rokhlina T, Ernst SE, Hanfland RA, Reznikov LR, Ludwig PS, et al. 2011. The $\Delta F508$ mutation causes CFTR misprocessing and cystic fibrosis-like disease in pigs. *Sci Transl Med* **3**: 74ra24.
- Ott CJ, Blackledge NP, Kerschner JL, Leir SH, Crawford GE, Cotton CU, Harris A. 2009a. Intronic enhancers coordinate epithelial-specific looping of the active CFTR locus. *Proc Natl Acad Sci* **106**: 19934–19939.
- Ott CJ, Suszko M, Blackledge NP, Wright JE, Crawford GE, Harris A. 2009b. A complex intronic enhancer regulates expression of the CFTR gene by direct interaction with the promoter. *J Cell Mol Med* **13**: 680–692.
- Pagani F, Raponi M, Baralle FE. 2005. Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proc Natl Acad Sci* **102**: 6368–6372.
- Plotkin JB, Kudla G. 2010. Synonymous but not the same: The causes and consequences of codon bias. *Nat Rev Genet* **12**: 32–42.
- Raponi M, Baralle FE, Pagani F. 2007. Reduced splicing efficiency induced by synonymous substitutions may generate a substrate for natural selection of new splicing isoforms: The case of CFTR exon 12. *Nucleic Acids Res* **35**: 606–613.
- Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL, et al. 1989. Identification of the cystic fibrosis gene: Cloning and characterization of complementary DNA. *Science* **245**: 1066–1073.
- Rogers CS, Hao Y, Rokhlina T, Samuel M, Stoltz DA, Li Y, Petroff E, Vermeer DW, Kabel AC, Yan Z, et al. 2008a. Production of CFTR-null and CFTR- $\Delta F508$ heterozygous pigs by adeno-associated virus-mediated gene targeting and somatic cell nuclear transfer. *J Clin Invest* **118**: 1571–1577.
- Rogers CS, Stoltz DA, Meyerholz DK, Ostedgaard LS, Rokhlina T, Taft PJ, Rogan MP, Pezzulo AA, Karp PH, Itani OA, et al. 2008b. Disruption of the CFTR gene produces a model of cystic fibrosis in newborn pigs. *Science* **321**: 1837–1841.
- Romey MC, Guittard C, Carles S, Demaille J, Claustres M, Ramsay M. 1999. First putative sequence alterations in the minimal CFTR promoter region. *J Med Genet* **36**: 263–264.
- Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N, et al. 1989a. Identification of the cystic fibrosis gene: Chromosome walking and jumping. *Science* **245**: 1059–1065.
- Rommens JM, Zengerling-Lentes S, Kerem B, Melmer G, Buchwald M, Tsui LC. 1989b. Physical localization of two DNA markers closely linked to the cystic fibrosis locus by pulsed-field gel electrophoresis. *Am J Hum Genet* **45**: 932–941.
- Rozmahel R, Wilschanski M, Matin A, Plyte S, Oliver M, Auerbach W, Moore A, Forstner J, Durie P, Nadeau J, et al. 1996. Modulation of disease severity in cystic fibrosis transmembrane conductance regulator deficient mice by a secondary genetic factor. *Nat Genet* **12**: 280–287.
- Rozmahel R, Heng HH, Duncan AM, Shi XM, Rommens JM, Tsui LC. 1997. Amplification of CFTR exon 9 sequences to multiple locations in the human genome. *Genomics* **45**: 554–561.
- Scambler PJ, Law HY, Williamson R, Cooper CS. 1986. Chromosome mediated gene transfer of six DNA markers linked to the cystic fibrosis locus on human chromosome seven. *Nucleic Acids Res* **14**: 7159–7174.
- Schmiegelow K, Eiberg H, Tsui LC, Buchwald M, Phelan PD, Williamson R, Warwick W, Niebuhr E, Mohr J, Schwartz M, et al. 1986. Linkage between the loci for cystic fibrosis and paraoxonase. *Clin Genet* **29**: 374–377.

L.-C. Tsui and R. Dorfman

- Sheridan MB, Fong P, Groman JD, Conrad C, Flume P, Diaz R, Harris C, Knowles M, Cutting GR. 2005. Mutations in the β -subunit of the epithelial Na^+ channel in patients with a cystic fibrosis-like syndrome. *Hum Mol Genet* **14**: 3493–3498.
- Snouwaert JN, Brigman KK, Latour AM, Malouf NN, Boucher RC, Smithies O, Koller BH. 1992. An animal model for cystic fibrosis made by gene targeting. *Science* **257**: 1083–1088.
- Spence MA, Tsui LC, Hodge SE, MacCluer JW, Elston RC. 1989. Genetic Analysis Workshop 6: Multipoint mapping of loci in the region of cystic fibrosis. *Prog Clin Biol Res* **329**: 1–9.
- Sun X, Sui H, Fisher JT, Yan Z, Liu X, Cho HJ, Joo NS, Zhang Y, Zhou W, Yi Y, et al. 2010. Disease phenotype of a ferret CFTR-knockout model of cystic fibrosis. *J Clin Invest* **120**: 3149–3160.
- Tebbutt SJ. 1995. Animal studies of cystic fibrosis. *Mol Med Today* **1**: 336–342.
- * Thomas PJ. 2013. CFTR folding and cellular processing. *Cold Spring Harb Perspect Med* doi: 10.1101/cshperspect.a009530.
- Toyoda Y, Hagiya Y, Adachi T, Hoshijima K, Kuo MT, Ishikawa T. 2008. MRP class of human ATP binding cassette (ABC) transporters: Historical background and new research directions. *Xenobiotica* **38**: 833–862.
- Trezise AE, Buchwald M. 1991. In vivo cell-specific expression of the cystic fibrosis transmembrane conductance regulator. *Nature* **353**: 434–437.
- Tsui LC. 1995. The cystic fibrosis transmembrane conductance regulator gene. *Am J Respir Crit Care Med* **151**: S47–S53.
- Tsui LC, Buchwald M. 1988. No evidence for genetic heterogeneity in cystic fibrosis. *Am J Hum Genet* **42**: 184.
- Tsui LC, Buchwald M. 1991. Biochemical and molecular genetics of cystic fibrosis. *Adv Hum Genet* **20**: 311–152.
- Tsui L-C, Estivill X. 1991. Identification of disease genes on the basis of chromosomal localization. In *Genome analysis. Vol. 3: Genes and phenotypes* (ed. Davies KE, Tilghman S), pp. 1–36. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Tsui LC, Buchwald M, Barker D, Braman JC, Knowlton R, Schumm JW, Eiberg H, Mohr J, Kennedy D, Plavsic N, et al. 1985a. Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science* **230**: 1054–1057.
- Tsui LC, Cox DW, McAlpine PJ, Buchwald M. 1985b. Cystic fibrosis: Analysis of linkage of the disease locus to red cell and plasma protein markers. *Cytogenet Cell Genet* **39**: 238–239.
- Wainwright BJ, Scambler PJ, Schmidtke J, Watson EA, Law HY, Farrall M, Cooke HJ, Eiberg H, Williamson R. 1985. Localization of cystic fibrosis locus to human chromosome 7cen-q22. *Nature* **318**: 384–385.
- Wainwright BJ, Scambler PJ, Stanier P, Watson EK, Bell G, Wicking C, Estivill X, Courtney M, Boue A, Pedersen PS. 1988. Isolation of a human gene with protein sequence similarity to human and murine int-1 and the *Drosophila* segment polarity mutant wingless. *EMBO J* **7**: 1743–1748.
- Ward CL, Krouse ME, Gruenert DC, Kopito RR, Wine JJ. 1991. Cystic fibrosis gene expression is not correlated with rectifying Cl^- channels. *Proc Natl Acad Sci* **88**: 5277–5281.
- Wardle CJ, Harris A. 1995. Down-regulation of CFTR expression in an HT29-derived cell line by stable antisense RNA production. *Biochim Biophys Acta* **1261**: 417–423.
- White R, Woodward S, Leppert M, O'Connell P, Hoff M, Herbst J, Lalouel JM, Dean M, Vande Woude G. 1985. A closely linked genetic marker for cystic fibrosis. *Nature* **318**: 382–384.
- Wille M, Buijs-Offerman RM, Aarbiou J, Colledge WH, Sheppard DN, Touqui L, Bot A, Jorna H, de Jonge HR, Scholte BJ. 2011. Mouse models of cystic fibrosis: Phenotypic analysis and research applications. *J Cyst Fibros* **10**: S152–S171.
- Xie J, Drumm ML, Ma J, Davis PB. 1995. Intracellular loop between transmembrane segments IV and V of cystic fibrosis transmembrane conductance regulator is involved in regulation of chloride channel conductance state. *J Biol Chem* **270**: 28084–28091.
- Ye L, Zhu W, Backx PH, Cortez MA, Wu J, Chow YH, McKerlie C, Wang A, Tsui LC, Gross GJ, et al. 2010. Arrhythmia and sudden death associated with elevated cardiac chloride channel activity. *J Cell Mol Med* **15**: 2307–2316.
- Zielenski J, Rozmahel R, Bozon D, Kerem B, Grzelczak Z, Riordan JR, Rommens J, Tsui LC. 1991. Genomic DNA sequence of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. *Genomics* **10**: 214–228.
- Zuccato E, Buratti E, Stuaní C, Baralle FE, Pagani F. 2004. An intronic polypyrimidine-rich element downstream of the donor site modulates cystic fibrosis transmembrane conductance regulator exon 9 alternative splicing. *J Biol Chem* **279**: 16980–16988.

The Cystic Fibrosis Gene: A Molecular Genetic Perspective

Lap-Chee Tsui and Ruslan Dorfman

Cold Spring Harb Perspect Med 2013; doi: 10.1101/cshperspect.a009472

Subject Collection [Cystic Fibrosis](#)

Antibiotic and Anti-Inflammatory Therapies for Cystic Fibrosis

James F. Chmiel, Michael W. Konstan and J. Stuart Elborn

Structure and Function of the Mucus Clearance System of the Lung

Brenda M. Button and Brian Button

New Pulmonary Therapies Directed at Targets Other than CFTR

Scott H. Donaldson and Luis Galiotta

The Cystic Fibrosis Airway Microbiome

Susan V. Lynch and Kenneth D. Bruce

Cystic Fibrosis Transmembrane Conductance Regulator (ABCC7) Structure

John F. Hunt, Chi Wang and Robert C. Ford

Status of Fluid and Electrolyte Absorption in Cystic Fibrosis

M.M. Reddy and M. Jackson Stutts

The Influence of Genetics on Cystic Fibrosis Phenotypes

Michael R. Knowles and Mitchell Drumm

Perspectives on Mucus Properties and Formation—Lessons from the Biochemical World

Daniel Ambort, Malin E.V. Johansson, Jenny K. Gustafsson, et al.

The Cystic Fibrosis Intestine

Robert C. De Lisle and Drucy Borowitz

Cystic Fibrosis Transmembrane Regulator Correctors and Potentiators

Steven M. Rowe and Alan S. Verkman

The Cystic Fibrosis of Exocrine Pancreas

Michael Wilschanski and Ivana Novak

Dynamics Intrinsic to Cystic Fibrosis Transmembrane Conductance Regulator Function and Stability

P. Andrew Chong, Pradeep Kota, Nikolay V. Dokholyan, et al.

The Cystic Fibrosis Gene: A Molecular Genetic Perspective

Lap-Chee Tsui and Ruslan Dorfman

The CFTR Ion Channel: Gating, Regulation, and Anion Permeation

Tzyh-Chang Hwang and Kevin L. Kirk

Assessing the Disease-Liability of Mutations in CFTR

Claude Ferec and Garry R. Cutting

Supramolecular Dynamics of Mucus

Pedro Verdugo

For additional articles in this collection, see <http://perspectivesinmedicine.cshlp.org/cgi/collection/>