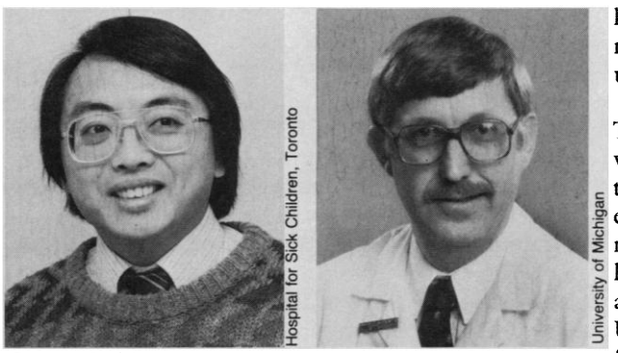
News & Comment

The Cystic Fibrosis Gene Is Found

Researchers have identified the major gene defect that causes cystic fibrosis. The discovery should lead to better diagnosis and perhaps improved therapies for the now fatal disease

THE RACE TO FIND the cystic fibrosis gene is over. In three papers to be published in the 8 September issue of *Science*, researchers from Toronto and Ann Arbor report that they have cloned the gene and pinpointed the gene defect that causes most cystic fibrosis cases. "The data are virtually irrefutable that they have the right gene," says Louis Kunkel of Children's Hospital Medical Center in Boston, a cloning expert who led the successful search for the gene causing Duchenne muscular dystrophy.



Gene sleuths. *Lap-Chee Tsui (left), Francis Collins, and their colleagues tracked down the cystic fibrosis gene.*

Cystic fibrosis researchers have looked long and hard for their gene—and with good reason. The disease is the most common genetic disorder of Caucasians. In the United States, it strikes one child in every 2000. An estimated 30,000 people have the disease today, and their prospects are grim. Most will die before their thirtieth birthday. Perhaps not surprisingly then, news of the gene discovery began to leak out before the scheduled publication of the papers describing the research, and this in turn prompted the editors of *Science* to drop their normal embargo policy (also see box on p. 924).

The discovery means that scientists can improve cystic fibrosis diagnosis, including prenatal diagnosis, and also devise better screening tests for people who carry a defective copy of the gene and run the risk of having children with disease. It also raises hopes for better cystic fibrosis treatments, perhaps new drugs or even gene therapy to replace the defective gene itself.

None of this could have even been considered until scientists could get a handle on the basic protein defect that causes cystic fibrosis. "Now we can really study what the basic defect is and we may be able to treat the defect directly, not just the symptoms," says Lap-Chee Tsui, the leader of one of the groups that cloned the gene. No one can now predict, however, how long it might take to do this or even if it will prove to be possible.

The search for the cystic fibrosis gene has been highly competitive, if not out-and-out contentious at times (*Science*, 8 April 1988, p. 141, and 15 April 1988, p. 282). But in the end, a collaborative effort by the groups of Tsui and John Riordan at Toronto's Hospital for Sick Children, together with Francis Collins at the Howard Hughes Medical Institute at the University of Michigan, bagged the gene.

The researchers appear to have a clear victory. "We have a lot of papers in press, but we don't have the gene," says chief competitor Robert Williamson of Saint Mary's Hospital Medical School in London, who has also been rumored to be close to cloning the cystic fibrosis gene. "If we couldn't get it, we're very pleased that Francis and Lap-Chee were the ones to do it."

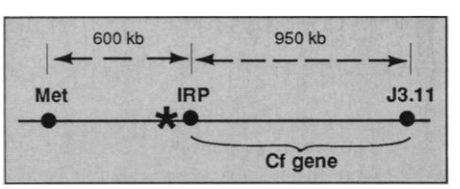
The collaboration between Tsui and Collins began in the fall of 1987, when the two researchers, who had previously been working independently, got together in San Diego at the annual meeting of the American Society for Human Genetics. "It was clear by then that this was a very hard problem that was not going to be solved without a great deal of labor," Collins says.

The cystic fibrosis gene was such a tough nut to crack because, in the absence of information about the protein it encodes, researchers did not know what they were looking for among the estimated 100,000 genes in the human genome. Researchers have managed to clone a few other genes without knowing what their products were—the Duchenne muscular dystrophy gene is one of them—but some of the defects responsible for the malfunction of these genes were large rearrangements that made them relatively easy to spot once their approximate locations in the genome were known. The cystic fibrosis gene did not carry any such convenient tag, unfortunately.

In 1985, however, 2 years before Tsui and his colleagues joined forces with Collins and his team, the Toronto group had provided a big boost to efforts to find the gene when they mapped it to chromosome 7. Williamson and Ray White of the Howard Hughes Medical Institute at the University of Utah in Salt Lake City further narrowed its location by identifying two "markers," the *met* oncogene and a DNA sequence designated J3.11 that flanked the gene (see diagram).

Once flanking markers have been identified, they can serve as starting points for zeroing in on a target gene. But *met* and J3.11 are almost 1600 kilobases apart—too great a distance to be traversed in a reasonable time by standard chromosome "walking" techniques. (A researcher walks a chromosome by identifying overlapping cloned fragments of DNA until the final destination is reached.)

Tsui consequently decided to use a brute force approach known as saturation mapping to find new marker sequences that were closer to the cystic fibrosis gene than either *met* or J3.11. To do this, the Toronto workers had to look for identifiable DNA sequences that are inherited along with the cystic fibrosis gene in the members of families with cystic fibrosis. The frequency with which a particular marker is transmitted along with the target gene gives an estimate of how close together they are. "Lo and behold, after screening 250 markers," Tsui says, "we found two that happened to be between *met* and J3.11." The genetic studies



At the start of the search. *The cystic fibrosis gene was known to be somewhere between the IRP gene and J3.11, but closer to IRP. The asterisk marks the location at which the successful hunt for the gene began.*

indicated that both were closer to the cystic fibrosis gene than either *met* or J3.11.

It was at that point that Tsui joined forces with Collins, who had developed a technique called chromosome jumping that can skip over lengthy segments of DNA. Not only is jumping faster than walking, but it also has the advantage of being able to move over unclonable DNA sequences. The human genome is studded with such sequences and they can stop a walk in its tracks.

It would have been ideal if the two new markers identified by Tsui had flanked the cystic fibrosis gene. But they didn't. They were located close together between *met* and another gene that the Williamson group had originally identified in the spring of 1987.

At the time, the London workers thought that they had the cystic fibrosis gene itself and said as much in an article published in *Nature*. By the fall of that year, however, their hopes were cruelly dashed when additional work showed that it was not. But the gene, which became known as the IRP (for *int*-related protein) gene because the protein it encodes resembles the product of the *int* oncogene, did prove to be the closest marker yet for the cystic fibrosis gene.

Despite some initial disappointment at the location of the new markers, Tsui, Collins, and their colleagues decided to plunge ahead. They began walking and jumping at one marker, moving at first in both directions. "You have to move in both directions until you cross a landmark that tells you which way you are going," Collins explains. The IRP gene was one of the first landmarks crossed, and the researchers knew that they were heading in the right direction.

The researchers had to jump and walk across 280 kilobases of DNA before encountering the beginning of the cystic fibrosis gene. Along the way they searched for potential genes by comparing the DNA sequences they were traversing with DNAs from other organisms. If structurally related sequences could be found in other organisms, that would mean that the sequence had been conserved during evolution, a good indication that it has an essential function. They found three conserved DNA sequences but were quickly able to eliminate two of them as candidates for the cystic fibrosis gene.

The third proved to be the key to the prize—but not without some initial anxiety. Genetic studies in cystic fibrosis families indicated that the potential gene segment was in the right location, but when the researchers looked for signs that it might be actively expressed, they could not find any after an extensive search. "That was quite disappointing," Tsui says. It looked as if the DNA sequence was not part of any active gene.

And this is the point where Tsui's Toronto colleague Riordan made an essential contribution. The Riordan group had made "libraries" of DNAs copied from the messenger RNAs present in sweat gland cells, which is one of the cell types in which the cystic fibrosis gene is supposed to be expressed. Each DNA copy corresponds to an active gene, and one of those from the sweat gland library proved to contain a segment matching the conserved sequence that Tsui, Collins, and their colleagues had found.

The extent of the match-up was quite small. The two DNAs shared only 113 base pairs of sequence, a circumstance that may explain why the researchers originally had so much trouble showing that the conserved sequence was part of an active gene. Further analysis showed, however, that those 113

The CF Gene Hits the News

When scientists attack problems cooperatively, they often hasten solutions. But there's at least one downside to this. It may fatally wound the hallowed tradition of the publishing embargo. Some secrets are apparently just too good to keep.

That certainly was the case with the discovery of the cystic fibrosis gene. With some two dozen researchers at two institutions and perhaps a half-dozen funding agencies involved in two countries, it may come as no surprise that someone leaked the news to the press. Reports began appearing more than 2 weeks before the papers describing the achievement were scheduled to be published, prompting *Science* editor Daniel Koshland to take the highly unusual step of lifting the embargo that is normally imposed on data in press at the journal. And that's not all the leaks did.

The papers are to appear in the 8 September issue, a near record 5 weeks after they were first received. In the normal course of events, a press conference would have been held on 7 September. The patent agents for the two institutions where the work was done wanted to file the patent applications before then.

But those plans changed on 22 August when Reuters News Service put out a story on its wire that said—correctly—that Francis Collins of the University of Michigan and Lap-Chee Tsui of the Hospital for Sick Children in Toronto had discovered the cystic fibrosis gene. "Once Reuters had broken the story we thought it was unfair to the rest of the press to withhold the information," Koshland says.

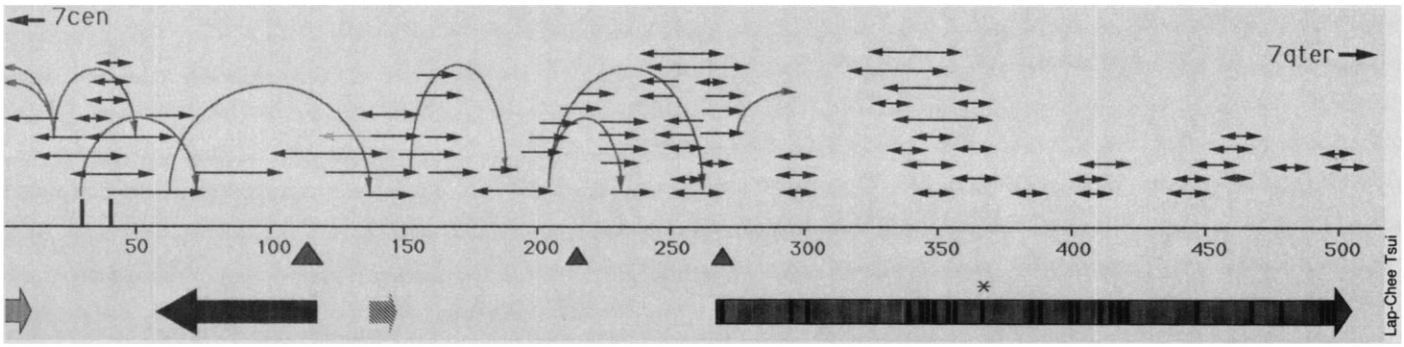
The researchers then held two press conferences, one in Toronto and one in Washington, on one exhausting day: 24 August. Howard Hughes Medical Institute, which supports Collins' work, provided private jets that made this possible. *Science*'s own press office was under siege by reporters who were not pleased to learn that the actual papers, still grinding their way through the editorial mill, would not be available for another week. At least one veteran reporter grumped that it was just like cold fusion all over again. It wasn't, of course; the three papers had been accepted.

Meanwhile, on learning that the press conference was going to be moved up, patent officials at the University of Michigan became concerned because the patents on the gene discovery and its applications had not yet been filed, and they were afraid that public disclosure of the research might jeopardize the awarding of patents, especially in Europe and Japan.

According to Robert Gavin of Michigan's Office of Intellectual Properties, the patent rights will be jointly held by the Hospital for Sick Children and the University of Michigan, with Howard Hughes receiving a share of Michigan's royalties. The discovery could be worth a lot of money because it should make it possible for the first time to screen members of the general population for defective cystic fibrosis genes. Since an estimated 1 person in 20 is a carrier, the potential testing market is large.

So with the press conference date moved up, the patent agent for the Hospital for Sick Children had to move quickly, filing the patent at the U.S. patent office at 11:45 p.m. on 22 August. Gavin was still concerned that the descriptions of the applications were not sufficiently detailed, however. James Friesen, the director of the Research Institute at the Hospital for Sick Children, concedes, "We knew that what we put in was not very polished. We thought we had another week and a half." They didn't, but patent applications can be amended.

Can the decorum of science publishing be enforced in the future? Koshland certainly prefers this. "Basically we try to keep embargoes because we firmly believe in the principle that people should be able to see all the data at the same time," he avers. But in the new age of high-stakes science and large collaborative groups, that may be more easily said than done. ■ **J.L.M.**



The long march to the cystic fibrosis gene. *The trek began at a site, shown here at the left of the diagram, that Lap-Chee Tsui's group had identified as close to the cystic fibrosis gene on human chromosome 7. The 280 kilobases of DNA between the start site and the beginning of the gene were covered by a combination of chromosome "walking" and "jumping." The straight arrows above the line represent the DNA segments cloned during the walk and the curved arrows represent the jumps taken. The long arrow on the lower right depicts the cystic fibrosis gene, which spans about 250 kilobases. The dark bars are the 24 exons that specify the amino acid sequence of the protein encoded by the gene. These are separated by the noncoding sequences known as introns. The exon marked with the asterisk contains the mutation found in 70% of defective cystic fibrosis genes. The middle arrow on the lower left denotes the IRP gene that was identified by Robert Williamson's group during their search for the cystic fibrosis gene. The other two right arrows also mark sequences with protein-coding capabilities and the three triangles point to sequences of a type frequently found near gene start sites.*

base pairs were likely to come from the starting end of the cystic fibrosis gene. This provided the probe they needed to clone the whole gene. But even that did not come easily. "To finish the cloning we spent night and day with lots of people," Tsui remarks. "We could only get bits and pieces and then had to fit everything together."

The gene proved to be quite large, extending across nearly 250 kilobases of genomic DNA. Like other genes of higher organisms, it consists of a mosaic of protein-coding exons—24 in this case—separated by nonprotein-coding introns.

Sequence analysis revealed that the protein encoded by the gene contains 1480 amino acids and that it has all the earmarks of a membrane protein, possibly of an ion channel. The protein sequence resembles those of several other proteins known to be involved in transporting substances across membranes. And that, says Robert Beall of the Cystic Fibrosis Foundation, "is very compatible with our current hypothesis of what causes cystic fibrosis."

The patients' main problem is the abnormally thick mucus that they produce, especially in the lungs. As a result, they fall prey to repeated infections that destroy the lung tissue, eventually leading to the patients' deaths. Although researchers had not been able to identify the protein defect causing the excessively thick mucus, recent evidence has indicated that the fault may lie in the inability of the lung cells to secrete chloride ions, and therefore water, into the mucus.

The structure of the cystic fibrosis protein now suggests that it may be a membrane channel for chloride ions. Moreover, Tsui, Riordan, Collins, and their colleagues have found that the gene encoding the protein is altered in cystic fibrosis patients, a change that might well cause the protein to malfunction.

Approximately 70% of the gene mutations are caused by the loss of a single specific trinucleotide codon. As a result, the corresponding protein is lacking just one amino acid, the phenylalanine at position 508. The researchers never see this change in the gene on normal chromosomes. This observation provides the proof that they have the correct gene, Tsui says. They are now looking for the remaining 30% of the mutations that alter this gene.

The site of the phenylalanine deletion may provide some clues as to how the protein malfunctions in cystic fibrosis patients. It affects what may be an important region regulating the protein's activity. The region contains an apparent binding site for adenosine triphosphate (ATP), a compound that provides energy for many cell functions. A nearby region also contains several target sequences for phosphate addition by the protein kinases A and C, both important protein regulators.

The loss of the phenylalanine may therefore interfere with chloride ion transport by preventing ATP binding to the cystic fibrosis protein and depriving it of the energy it needs or by rendering it unresponsive to activation by the protein kinases.

Currently, clinicians can only treat cystic fibrosis patients by attempting to control their infections and other symptoms. But now that the cystic fibrosis gene and protein are in hand it may at last be possible to design more rational therapies aimed at the specific defect itself. One possibility is to develop drugs that can act through the protein to restore normal chloride transport. "This will require a long period of research and development," Tsui says, "but we have at least reached a starting point."

Ultimately, it may even be possible to use gene therapy to correct the defect. Introducing the normal gene into lung cells should be sufficient to help patients, Beall suggests. They have other symptoms, but these can be controlled. It is the lung defect that kills. Until a way can be found to deliver a functioning cystic fibrosis gene into lung cells, gene therapy will remain something of a long shot, however.

Improved detection of carriers of defective cystic fibrosis genes is a much more immediate prospect. A person has to inherit two bad genes to get the disease. A carrier has only one defective copy and does not have any symptoms by which he or she might be identified. But if two carriers have a baby, their child has a 25% chance of being affected.

Before the cystic fibrosis gene was discovered, carriers could only be detected in families already known to carry the defective gene because some of their members had the disease. Genetic counselors were forced to look for markers known to be inherited with the cystic fibrosis gene, rather than the gene itself, and this procedure requires a knowledge of family genetics.

But the new work should make it possible to identify defective cystic fibrosis genes in anyone. This will require, Tsui points out, that the remaining 30% of the mutations in the cystic fibrosis gene be identified. But the researchers are hard at work on this project, and it may be completed in a year or two.

The long march to the cystic fibrosis gene was obviously arduous. But the successful procedures worked out by Tsui, Collins, and their colleagues for isolating the gene should also be applicable to the identification of the genes causing other genetic diseases. "It was a long task," Kunkel says, "but it shows that it can be done. It can be done again."

■ JEAN L. MARX