Molecular Biology of *Homo sapiens*

In a landmark symposium at Cold Spring Harbor, human genetics was seen to be on the threshold of a new era, one in which humans would become like experimental genetic systems

No one could accuse James Watson of modesty in his selection of the title for the 51st Cold Spring Harbor Symposium on Quantitative Biology, namely, "Molecular Biology of *Homo sapiens*." But, with his usual perspicacity, Watson had contrived to pull together a gathering that was at least a match for the grandeur of the title, and more so. What had been clear to Watson in initiating the meeting, and quickly became evident to those who attended, was that the human animal is on the threshold of being a truly accessible genetic system for study.

Already on display at the symposium were the fruits of the techniques and analysis of molecular biology, which ranged from the identification of oncogenes to speculations on the first members of *Homo sapiens*, and from the commercial use of human gene products to the first experimental steps of gene therapy.

And in the not too distant future clearly lies the promise of a time when the entire human genome will be charted and mapped in both genetic and physical dimensions. Whether this is in 2 years at best or 5 at worst, it will usher in the opportunity to study human genetics in its most subtle shades, namely, in dealing with complex polygenic systems, be they predispositions to diseases, anatomical characteristics, or behaviors.

It is just a little more than two decades since the Cold Spring Harbor symposium last turned its attention to human genetics. "The contrast with the present gathering is striking," noted Walter Bodmer, who is director of the Imperial Cancer Research Fund, London, and one of the handful of people who had attended both meetings. "Only very few of the papers in 1964 mentioned DNA. Here, there wasn't a single paper that didn't."

Victor McKusick of Johns Hopkins University, who also had attended the earlier meeting, recalls that the 1964 program included three sections: population genetics, somatic cell genetics, and proteins. "All deductions about mutations were made through proteins at that time," he says. "Diseases like Huntington's chorea and cystic fibrosis were mentioned only in passing." At Watson's meeting these conditions were notable for the speed with which researchers are closing in on the disease genes. They represent paradigm cases of modern clinical genetics. As Bodmer said, "The real challenge is to explain these diseases at the molecular level, to go from the bottom up, from the genes to the defective phenotype."

***"Humans deserve a genetic linkage map. It is part of the description of* Homo sapiens."**

This bottom-up, or reverse genetics, approach is conceptually the same whether it involves single genes, such as some of the classic inherited diseases, or many, which might include the more common disorders, such as hypertension and probably some psychiatric conditions. This involves identifying the stretch of DNA that segregates with the condition through affected families; zeroing in on the gene(s) involved; identifying the protein(s) coded for; and finally elucidating the role of the protein(s) in normal metabolism. The difference is that as the number of variables in the initial genetic analysis increases, the complexity of solving it—of pinpointing the relevant locus or loci—rises exponentially.

Not surprisingly, the recent meeting was much more sharply focused than its predecessor, and was divided into seven sections, each of which could have been a legitimate meeting of its own. The seven topics were as follows: human gene mapping; human cancer genes; genetic diagnosis; human evolution; drugs made off human genes, including clotting/anticlotting factors, anticancer agents, and growth factors; receptors; and, of course, gene therapy.

Note that population genetics as a topic, which had represented one third of the 1964 meeting, figured minimally in 1986. Variation, which is the stuff of population genetics, is now being scrutinized at the individual, molecular level, not so much in populations.

Strikingly evident through all this, as Thomas Caskey of Baylor College of Medicine noted, was the tremendous impact of the many new techniques of molecular biology, whose targets range from single nucleotide bases to whole genomes. With DNA sequencing now being a routine—and recently automated—methodology, and techniques for separating large segments of DNA and even whole chromosomes becoming generally used, "the ability to dissect the human genome is now tremendous," says Caskey.

So powerful are these various techniques that it is now possible to contemplate mapping, and even sequencing, the entire human genome. Such a prospect proved to be a heady thought for many participants, who became intoxicated with the idea of so much information that would become available to them. Not surprisingly, there was a good deal of enthusiasm for scooping up this "grail of human genetics," as Walter Gilbert has called it. "To wait the 100 years that total sequencing would take at current rates is clearly too long," says Caskey. "But deciding how to achieve that goal in 10 years, and deciding on its relative priority, that's difficult." (See *Science*, 27 June, page 1598.)

Caskey, who had been allotted the impossible job of summarizing the meeting, also noted the important and growing input from the commercial sector. "Twenty-five of the 123 papers were from the corporate sector," he said. In fact, in the session on growth factors all seven papers were from researchers with biotechnology companies. It goes without saying that this was not so in 1964. But it is notable that neither was it the case just 5 years ago.

Clinically oriented researchers were there in force too, and for good reason. Even aside from the more classical areas of human genetics, such as inherited diseases and genetic diagnosis, the rate of production of information in medically important areas, such as oncogenes, growth factors, and receptors, is quite staggering. "The impact of all this is already changing patient care," says Caskey. "Clinical and genetic diagnosis programs are being redesigned." So fast is the development of new understanding and new techniques, however, that the clinical world is in danger of being left behind. "The number of people in medical genetics who know about these new areas is very few," comments Caskey.

But perhaps the aspect of human genetics that eventually will have the greatest impact is the construction of genetic and physical maps of the genome. This is what will make

Homo sapiens a genetic system for detailed study. A good example of the initial approach, and an illustration of the requirements, comes from the current race to find the cystic fibrosis gene.

At the end of last yeast four separate groups announced more or less simultaneously that they were closing in on the gene. Each had found at least one genetic marker that was near the putative disease gene. By near is meant something like 1 to 3 million bases, which is still frustratingly distant in terms of physically locating the gene. Even the fastest method of chromosome walking would take a very long time, with steps being in the order of 10 kilobases at a time.

What is needed here, for instance, is a series of overlapping cosmids, which might cover as much as 40 kilobases each. These are still pretty small steps, but if their order on the chromosome is known, then searching through unknown territory (in terms of the genes contained therein) becomes a much simpler task, provided of course there are some signposts to help orient the search. Hence the need for a physical map, for which there are several potential methods for their construction.

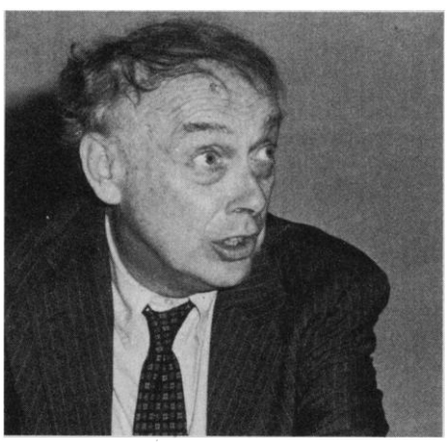
But even a complete physical map will not tell you where the cystic fibrosis gene is, or any other unknown gene for that matter. Here is where the genetic map is required, and here is where the most oft-repeated acronym of the meeting resides: RFLP, roughly pronounced "ruflup." Restriction fragment length polymorphisms, or ruflups, are discrete sites of variation in the genome, of which several hundred have been discovered. Variation is the stuff of geneticists' analysis, providing there is a sufficient number of alleles at a sufficient frequency. Some ruflups fulfill these requirements and are useful for genetic analysis because they are easily detectable with standard laboratory techniques.

The idea is to find a ruflup that is inherited with the same pattern as the genetic condition one is tracing. Given such concordance, it is then possible to pinpoint the disease locus to within a few million base pairs of DNA, as described for cystic fibrosis. This approach, of using single markers, has also brought the genes for Huntington's disease and Duchenne muscular dystrophy within grasp.

However, the search technique becomes substantially more powerful when a series of markers is used, preferably scattered evenly throughout the genome. David Botstein of the Massachusetts Institute of Technology calculates that one reasonably polymorphic marker located every 20 million bases would provide an adequate map.

Using a technique called simultaneous search, developed by Eric Lander at the Whitehead Institute in Cambridge, Massachusetts, the amount of raw data—that is, family histories—required for tracking down candidate loci is reduced by an order of magnitude, simply by comparing many markers rather than just one. The real benefit of a complete ruflup map, however, would be analyzing those conditions in which several genes are involved, perhaps being expressed to different degrees. The mathematics and computing power required for such analyses is currently horrendous, but Lander and Botstein are developing methods for reducing it to tractable dimensions.

This shift from single gene to multigene conditions, from "simple" genetics to more subtle influences of many genes and uncertain regulatory effects, is likely to become the high ground of human genetics. As Joseph Goldstein of the University of Texas puts it: "Now that we are learning about major gene effects at the molecular level we are gaining the tools to begin to open up the black box of the common diseases." The classic inherited diseases, though often clinically devastating, are numerically rather rare. The more common diseases to which Goldstein refers and which might have subtle, multigene effects, include hypertension, diabetes, and some psychiatric disorders.



James D. Watson: *An eye on the genetics of* Homo sapiens.

Although Helen Donis-Keller and her colleagues at Collaborative Research, Inc. have peppered chromosome 7 with various types of genetic markers (in their search for the cystic fibrosis gene), so far the remainder of the genome is only sparsely signposted. However, it surely cannot be very long before these remaining markers are in place. "Humans deserve a genetic linkage map," says Ray White of the Howard Hughes Medical Institute, University of Utah. "It is part of the description of *Homo sapiens*." ■

ROGER LEWIN

Cold Spring Harbor Briefings: By Roger Lewin

New Alpha-Globin Gene Discovered

It is not often that a new human gene is discovered, particularly so in such well-charted territory as the alpha-globin locus. But Che-Kun James Shen and his colleagues at the University of California, Davis, were able to report such a discovery, which is a gene they name theta-globin.

The new gene, which is located way down at the 3′ end of the 30-kilobase locus, has every appearance of being functional, says Shen which poses the question of when the gene is expressed.

What is currently understood about the expression of the alpha genes is the following. The zeta-globin gene, which is located at the extreme 5′ end of the locus, is functional very early in embryonic life. The alpha 1 and 2 genes, which are located toward the 3′ end of the locus, then kick in around 6 weeks and continue through maturity. Located in between the zeta and alpha genes are three inactive pseudogenes. In other words, it has been assumed that just one alpha-globin transition—zeta to alpha—occurs throughout life. The discovery of a potentially functional third gene in the alpha group raises the possibility of a third transition.

Three transitions in the alpha-globin group would bring the locus into accord with the beta-globin genes, in which there is a shift from embryonic (epsilon), to fetal (gamma G and A) to postnatal-through-adult globin (beta). In the alpha and beta loci as they are currently known, the direction of temporal expression of the genes is consistently 5′ to 3′. But if Shen and his colleagues, Jon Marks and Jeng-Pyng Shaw, are correct, the theta gene will break that neat arrangement. According to the admittedly rather thin evidence available so far, it seems that the theta gene might be expressed very early in embryonic life, perhaps some time prior to 5 weeks.

There is evidence for the appearance of unidentified globins at this period, which Shen believes might include theta-globin. Two other possibilities—that theta-globin might be expressed during stress or old age—are apparently not supported. So, if the new gene is active very early in life it is at the "wrong" end of the alpha locus.

Judging by the sequence divergence between the alpha and theta genes, which gives

"Molecular Biology of *Homo sapiens*," Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 28 May to 4 June.