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.

a 26% difference in amino acid sequence, the newly discovered theta gene is at least 200 million years old.

In addition to the apparently functional theta gene in the alpha-globin locus, there appear to be at least two other theta-globins in the human genome, at least one of which is on a different chromosome from the alpha-globin locus. Errant globin genes of this sort are rare in humans.

Shen and his colleagues find theta-globin genes in the orangutan and the baboon, which, they suspect, are also functional. However, in the galago, a prosimian, the rabbit and horse, the theta gene has been inactivated. Why the gene apparently remains functional in higher primates—if that truly is the case—remains to be established.

c-*myc* Implicated in RNA Processing

In a series of colorful and dramatic slides, David Spector of the Cold Spring Harbor Laboratory presented evidence that the human c-*myc* protein might be associated with RNA processing.

Previously, the c-*myc* protein has been suggested to be a DNA-binding protein, but Spector and his colleagues, N. F. Sullivan and R. A. Watt, find this unconvincing. As support they adduce the observation that treatment with DNase I enzyme fails to release the protein from nuclear preparations whereas RNase treatment does.

More specifically, however, double-label immunofluorescence techniques show that the c-*myc* protein colocalizes in the nuclei of virally transformed quail cells with snRNP complexes, which are considered to be important components in the RNA processing machinery. Moreover, c-*myc* protein injected into cells migrates into the nucleus and again colocalizes with snRNP complexes.

Several other experimental techniques, including immunoelectron microscopy and immunoblotting of fractionated extracts, confirm the close association of c-*myc* protein and snRNP complexes. Antibodies directed against c-*myc* protein bring down a series of small RNA's, which appear to be snRNA's of the U1, U2, U4, and U6 classes that are associated with RNA processing.

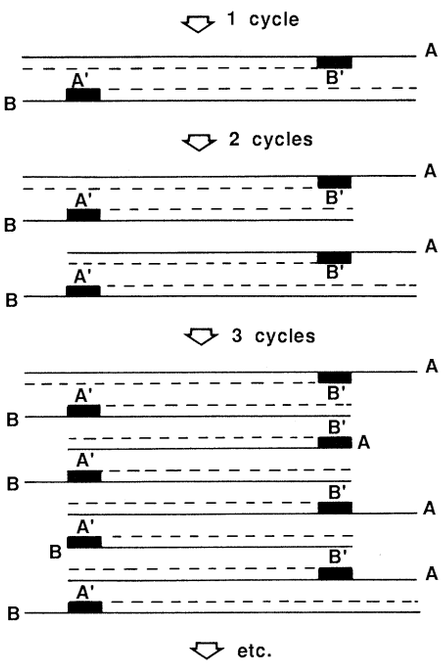
Although these several lines of evidence all point in the same general direction, they are only indirect indications of an RNA processing role for c-*myc* protein. A better test, which Spector and his colleagues are currently preparing, will be to see if antibodies against c-*myc* protein will block RNA processing in efficient in vitro systems. ■

Amplifying DNA By the Magic of Numbers

Sometimes the simplest of techniques yield results so dramatic they are difficult to believe. Just so with Polymerase Chain Reaction (PCR), which in the right circumstances allows the amplification of a desired DNA sequence from minute quantities of messily impure starting material.

Developed by Kary Mullis of Cetus Corporation, PCR has one prerequisite: that the sequences flanking the DNA region of interest should be known. Given that, the method works as follows.

Oligonucleotide primers that flank the required region, and complementary to opposite strands, are synthesized. These are added in large excess to the starting material, which in theory could be a single human cell



lysed in aqueous medium. The double-stranded chains are melted by heating and the primers are extended enzymically by DNA polymerase.

This first cycle gives a doubling of the region of interest but includes extension beyond it. The next cycle produces equally long extensions but also a small number of discrete strands that contain the region of interest plus the primer. From then on the number of chosen strands grows exponentially while the longer strands remain in linear growth: 10 cycles give 1000 short strands, 20 cycles a million, 30 cycles a billion, and so on.

The technique, which runs at about 5 minutes per cycle, has of course been automated. So, given the required oligonucleotides, it is possible effectively to select out a piece of target DNA in workable quantities during an afternoon's operation.

Because the repeated cycles of heating and cooling necessitated the repeated addition of polymerase, Mullis and his colleagues decided to try using an enzyme from a thermophilic organism, which survives undamaged through the heating part of the cycle.

So far the longest stretch of DNA that has been "amped" in this way is 1 kb. The reason for this is that at about this length, DNA chains start to assume tight secondary structure conformations, which can block the enzyme during primer extension.

The smallest quantity of starting material that Mullis has as yet successfully worked with is whole genome DNA from 150 human cells. There is no theoretical lower limit, he says, and it should be possible to start with a single cell. The production of target DNA also has no theoretical limit, beyond the kinetic barriers imposed by increasing concentration.

In addition to plucking a target gene out of an overwhelming background of unwanted DNA, the PCR technique can be used to construct inserts for cloning and further manipulation. For instance, restriction sites, promoters, and other sequences can be tacked on to the end of the primers, which then becomes part of the amplified chain.

Although it was first developed almost 2 years ago, PCR has been slow to attract attention, perhaps because it is so simple. At the Cold Spring Harbor meeting, however, it seemed to have passed a threshold and was described by some participants as the hottest thing going. ■

First Success with Reverse Genetics

In applying molecular biology to clinical medicine, one great hope has been to track down the cause of genetic diseases by going from the gene to the defective function, the so-called bottom-up approach or reverse genetics. Stuart Orkin and his colleagues at Harvard Medical School believe they have done this with chronic granulomatous disease (CGD), one of the classic inherited diseases.

In identifying what they believe to be the defective protein in CGD, Orkin and his colleagues will have chalked up the first successful search for the cause of an inherited disease in which there was no initial presumption about the nature of the protein. It was clear from other presentations that similar claims for cystic fibrosis, Huntington's disease, and Duchenne muscular

dystrophy won't be long in coming.

Patients with CGD suffer chronic and recurrent bacterial infections, largely due to a metabolic defect in the phagocytic cells, polymorphonuclear leukocytes and monocytes. The most obvious metabolic problem is a breakdown in the NADPH oxidase system, and a defect in one of the b-type cytochromes has been implicated as the cause of it all.

The disease is X-linked, and the locus has been pinpointed to the short arm of the chromosome. Orkin and his colleagues had an opportunity to zero in on the disease gene with the recent identification of a boy suffering not only from CGD but also from Duchenne muscular dystropy, and retinitis pigmentosa. The boy had a large deletion—in the order of 3 to 5 megabases—within Xp21, which is somewhat more proximal on the arm of the chromosome than had been supposed for the location of the CGD gene.

Then, capitalizing on a series of clones developed for the region by Louis Kunkel, and a good deal of blind luck, Orkin and his colleagues proceeded to search for their quarry. They made cDNAs from a normal phagocytic cell and did a subtraction with cDNA's from a granulomatous cell, which in theory should leave the DNA that is missing in CGD. This time theory worked out in practice, and a piece of cDNA lit up on a Southern blot when scanned with Kunkel's clones.

The putative disease protein is in the order of 54 kilodaltons, or 468 amino acids in length, and is clearly not a cyctochrome. What role—direct or indirect—it plays in the oxidase system remains to be established.

The Harvard team is now beginning the search for this protein in normal cells. And, says Orkin, it will not be too long before he and his colleagues will be in a position to test the effect of reintroducing the protein into defective cells. ■

AIDS Virus Entry Pinpointed in Brain

The virus that causes acquired immune deficiency syndrome (AIDS) apparently uses a simple method of entry into its target cells. All that is necessary and sufficient is the presence of the T4 receptor, reports Richard Axel and P. Maddon of Columbia University College of Physicians and Surgeons.

T4 is a surface glycoprotein that is a member of the immunoglobulin supergene family. In normal circumstances the molecule serves as a receptor that enables the T helper cell to recognize its target cell in mediating an immune response. It now appears that the AIDS virus is able to bind to T4 and subsequently gain entry to the cell by endocytosis. A related molecule, T8, which characterizes another set of T cells, does not offer a binding site to the AIDS virus.

Axel and Maddon made cDNA and genomic clones of T4, and in collaboration with British researchers Steven McDougal and Robin Weiss were able to get the antigen expressed on the surface of B lymphocytes and epithelial cells. They found that whereever T4 was located, the AIDS virus could bind and enter, no matter what kind of cell was involved.

Given this important link between T4 and AIDS virus entry into cells, the Columbia team turned its attention to the known, extensive involvement of brain tissue in the disease. Brain cells, after all, were not known to bear T4 antigen. It turns out, however, that some do, particularly those in the forebrain, where AIDS neuropathology is most marked.

Axel speculated about the possible therapeutic benefits of soluble T4, perhaps modified in some manner, which might mop up AIDS virus before it could reach its normal target areas. There are many obvious potential immunological problems with this approach. ■

Chimeric Receptors Give Clues to Oncogene Action

One strong theme that came through the meeting was the very striking domain assembly of transmembrane receptor molecules, a feature that lends itself to experimental manipulation. And as cellular growth control resides in some molecules of this class, an understanding of their function gives an insight into the action of some oncogenes.

One particularly striking manipulation, reported by Axel Ullrich and his colleagues at Genentech Inc., involved the assembly by recombinant DNA technology of a chimeric receptor made up of the extramembrane domain of the insulin receptor and the trans- and intramembrane domains of epidermal growth factor.

The extramembrane section of the insulin receptor is a hexamer, made up of two alpha and two beta subunits, whereas the equivalent section of the epidermal growth factor molecule is a monomer. It was therefore of some interest to see if the binding of insulin would trigger the kinase activity of the epidermal growth factor. It did, with considerable efficiency.

William Rutter of the University of California, San Francisco, reported various modifications of the insulin receptor. One of these involved removal of much of the extramembrane portion, which had much the same effect as insulin binding, showing that the normal binding releases some kind of constraint.

More subtle manipulations, such as modifying a key pair of tyrosine residues, revealed how the functions—direct and indirect—of the receptor could be dissected. And, by empirical comparison, the effects of certain oncogenes may be dissected too. ■

Important Advance In Gene Therapy

The ability securely and accurately to insert new genes into target cells, and the expression of those genes once they are in place, continue to be the great stumbling blocks to true advancement of gene therapy.

Oliver Smithies and his colleagues at the University of Wisconsin continue to develop methods for targeting specific chromosomal regions for transfected DNA. Aiming at insertion between the zeta- and beta-globin genes in the beta-globin locus, they report a success rate of about 1 in 1000, which he admits is not high. Nevertheless, progress is being made, involving both this and other systems.

Meanwhile, Stuart Orkin and his colleagues at Harvard Medical School and Thomas Caskey and his colleagues at Baylor College of Medicine have made some headway with in vivo expression. They have been working independently for some time with the insertion of the human adenosine deaminase (ADA) gene into mouse hematopoietic cells. This system is an experimental model for the repair by gene therapy of severe combined immunodeficiency.

Using the SV40 promoter as part of their insert, Orkin and his colleagues, like every other group, had until recently been able to obtain good expression of ADA in cell culture but not in live animals. But, having stripped the virus down to its skeletal minimum, Orkin was able to report high level in vivo expression of the human gene. Using a similar approach, Caskey and his group were able to report equal success.

For reasons that are unclear, the degree of expression varies, ranging from trace amounts in some instances to levels comparable with the endogenous enzyme in others. The result is clearly an important step forward, but, as Orkin admits, it is something of a step in the dark. Caskey characterizes the results as "encouraging." ■