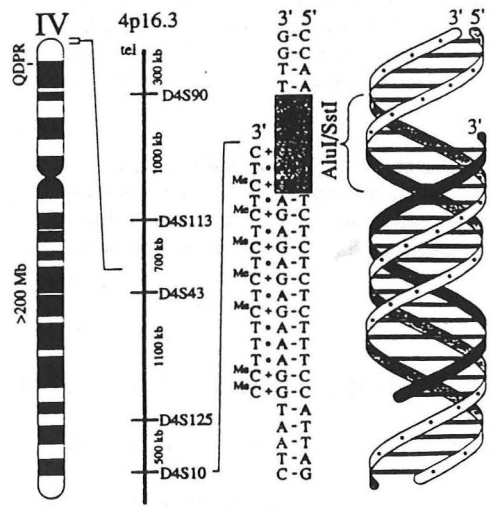
THE GENOME PROJECT - AN UPDATE

AND THE MEDICAL, SOCIAL, ETHICAL,

AND LEGAL IMPLICATIONS



Rody P. Cox, M.D.

Internal Medicine Grand Rounds

February 11, 1993

This review of the genome project will be from the standpoint of a general internist. It is essential that we as physicians understand in a general sense the strategy of mapping the human genome and the enormous benefits that it promises for medical practice. However, this information also has the potential for significant abuse. The results of the human genome project are already trickling into medical practice and influencing it significantly. Soon this information will become a deluge. We as general internists must examine the ways this information can be used for the benefit of our patients. This requires us to prepare ourselves to intelligently guide and counsel them in using this important new information. We and our professional organizations should also be involved in planning safeguards to prevent the very real possibility that some of the information may be misused to the detriment of society.

The objective of the human genome project is to provide a source book for human biology and medicine that will be studied for many decades to come. Spinoffs from the human genome project are already evident from the ever-accelerating pace of discoveries in disorders such as muscular dystrophy(1), cystic fibrosis(2), Marfan's syndrome(3), familial polyposis of the colon(4), and many others. Unlike some other mega-projects - for example, the Superconducting Super-Collider - it is not necessary to have the finished product before payoff can be realized.

The first phase of the human genome project is: 1) to provide a physical map of every human chromosome with adequate markers along the chromosomes so that undiscovered genes can be readily localized. The second phase about which there is some controversy is sequencing the entire human genome. All scientists involved in the human

TABLE 1.

GOALS OF THE HUMAN GENOME PROJECT

I. Provide a physical map of each human chromosome

II. Sequence the human genome

III. Elucidate the function of all 50,000 to 100,000 human genes

. . . Map the genomes of well studied model organisms; yeast, C. elegan, Drosophila and mouse.

genome project agree with physical mapping as being of the utmost importance. However, there is "soul-searching" as to whether sequencing the entire genome is cost effective and necessary(5). Apparently only about 5% of the DNA in the human genome appears to

contain information. The remainder of the DNA is intronic sequences, repetitive DNA, and other so-called "junk" DNA. Some mappers feel that sequencing should be confined to the biologically interesting DNA, that is, promoter-regulatory regions and genes that are transcribed. Others feel that sequencing will have evolutionary significance, and that some of the DNA that is now believed to be non-informational may turn out to have important biological significance. The last phase will take decades and that is to understand the functions of each of the genes and how they determine the developmental plan of humans and the molecular basis of normal cellular physiology (6,7).

The human genome project is an international effort to map each of the human chromosomes as well as to map the genomes of several other organisms that are well studied, e.g. yeast, the nematode ,C elegans, the fly Drosophila, and the mouse. These organisms are model systems where detailed genetic information is already available. Their genomes are simpler than humans, therefore they provide a means of interpreting human gene functions as new genes are discovered.

Genomes of Model Organisms

The relative simplicity of the genomes of simpler organisms and the homology and conservation of function at the level of proteins provides a useful model for linking gene structure with protein function. Without gene-to-protein function correlations, the mapping of the 50,000 to 100,000 human genes would have limited value. For example, the yeast genome is 1/200th the size of the human genome(8). Human genes are interrupted by long stretches of uninformative DNA that are called introns. Figure 1 compares the size of the HMG-CoA reductase gene of humans and yeast. It is obvious that mapping and sequencing the yeast genome is much simpler.

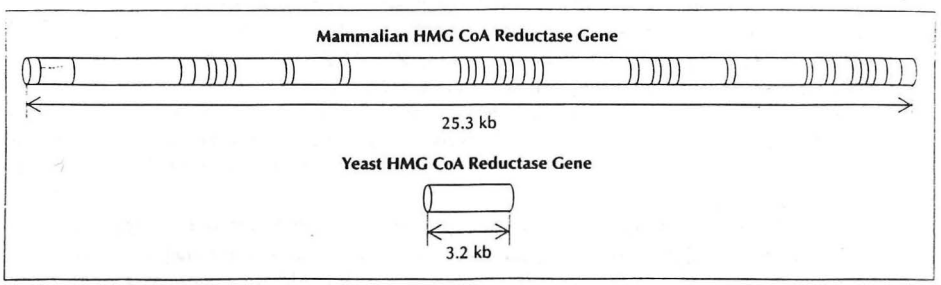


Figure 1. **Because it is more densely packed with coding information, the yeast genome represents a miniature version of genomes in higher eukaryotes. An example is provided by the mammalian and yeast genes that encode an essentially homologous protein, HMG-CoA reductase. The mammalian version extends to 25.3 kb, most of it taken up by noncoding introns (yellow), compared with a mere 3.2 kb in yeast, all of it consisting of exons.**

A basic biological principle is gene conservation during evolution. Complex organisms are elaborations of the successful design of primitive species(9). The genes are similar, with respect to information and the encoded proteins have a high degree of sequence identity (Fig. 2). For example, a human genome mapper may isolate a human gene whose

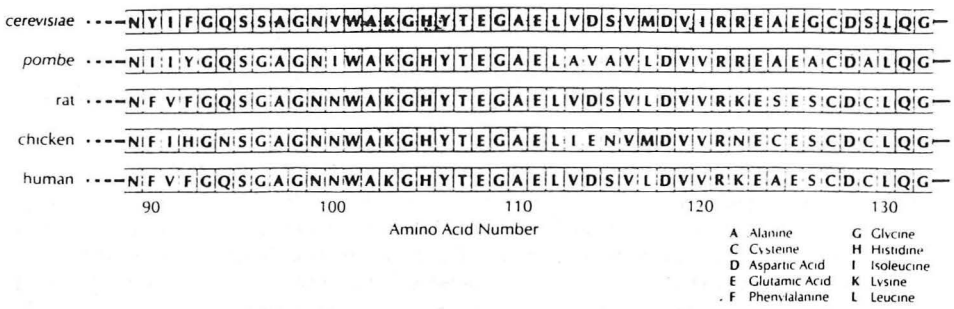


Figure 2. **Gene conservation is a major reason why yeasts serve as an important model organism for studies of the human genome. This is illustrated by the genes coding for the protein tubulin in yeasts, chickens, rats, and humans..**

function is not known. In collaboration with other Genome Project scientists a homologous gene can be located and cloned in a simpler organism, such as yeast. In many instances the yeast gene will be known and its function understood. Therefore, the function of the human gene can be inferred from that of its homologue. On the other hand, if the yeast gene is not known it can be cloned, expressed and the protein characterized. The gene can be mutated or deleted and its effect on the growth and physiology of the yeast studied. The human gene can be inserted into yeast to determine if it corrects the mutant phenotype. The function of the human gene is then clearly defined. A physical map of the yeast genome already is available at 110kb resolution(10).

C. elegans (a simple nematode) has a genome (six chromosomes) with 100 million base pairs. This genome is much larger than other simple model organisms. However, sequencing its genome probably will be complete by the year 2000 (11). The advantage of studying C. elegans is in the area of developmental biology because the exact lineage and fate of all 959 cells in this small metazoan (1mm in length) is known (12). A complete wiring diagram of the nervous system in all the 302 neurons and the connections among them are known.

The eminence and detailed knowledge of the gentics of Drosophila and mice strains is familiar to all of us. A dense map of the mouse genome would be a valuable source for studying many human genetic problems. More than 3,600 loci have been identified in the mouse and more than 2,700 are located on specific chromosomes(13). Major applications of mouse genetics to the human genome include analysis of mutant phenotypes, the genetic dissection of complex traits and the development of comparative maps for mice and humans (14). The importance of mapping and studying the genomes of these model organisms is critical to the central aim of the Human Genome Project understanding the function of all 50,000 to 100,000 genes.

STRATEGIES FOR CONSTRUCTING A CHROMOSOMAL MAP

As I discussed in the introduction, my approach will be that of a general internist rather than a gene mapper. For those who would like more detail about the technical aspects of gene mapping, I would refer you to Joe Goldstein's Grand Rounds on September 28, 1989, on sequencing the human genome. Many of the predictions that Joe made concerning advances in technology and increases in the efficiency of mapping have been achieved. It is estimated that the human genome contains between 50,000 and 100,000 genes. By 1993 3,521 genes had been mapped and more than 13,000 loci were defined by DNA markers (courtesy of Dr. A. Bowcock and P. Ceverha, 1993).

GENOMIC MAPPING

The mapping of the human genome requires a number of different sophisticated techniques. Therefore, there is not a single genomic map but rather several kinds of maps which correspond in different ways to the human genome. As McKusick has pointed out, these genetic maps are different just as a highway map, a topographic map, and a surveyor's map may all correspond to the same geographic location but they are different representations(15). The genomic mapping techniques also provide different kinds of information which I will review briefly. There are two basic categories of genomic maps: 1) the genetic or meiotic linkage map and 2) the physical map of each human chromosome.

I. Genetic Linkage Map

Genetic linkage maps show the relative frequency with which a chromosomal elements (that is genes or gene markers) are co-inherited, that is their degree of linkage which corresponds to the physical distance between them on a chromosome. The tighter the linkage

the less often they are separated during meiotic recombination and the more often they are inherited together. Figure 3 shows the pedigrees of African-American families carrying both the gene for color blindness and G6PD deficiencies (16). These genes are located on the X

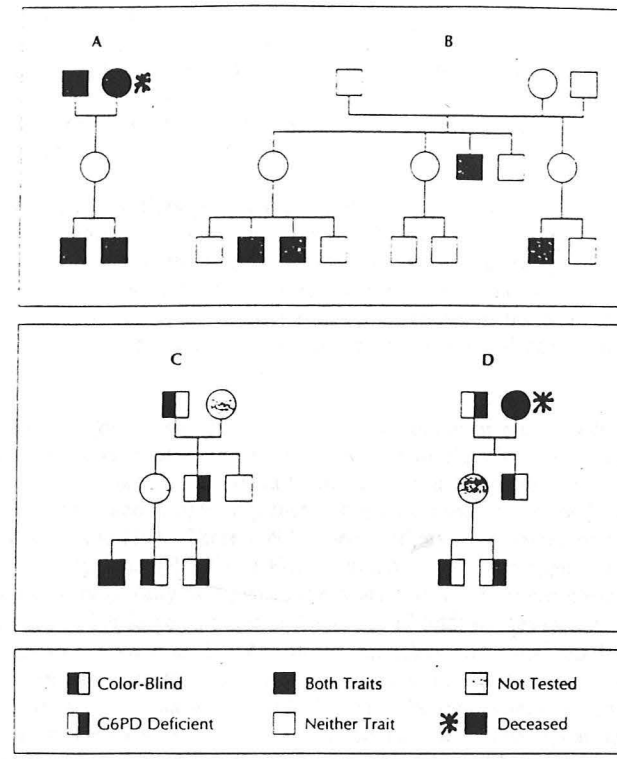


Figure 3. **Initial evidence that genes responsible for color blindness and for deficiency in glucose-6-phosphate dehydrogenase (G6PD) are close neighbors on the X chromosome was derived from three-generation pedigrees of black families in Baltimore. In each of the four examples shown, the presence of both X-linked traits among the brothers in a family identified their mother as doubly heterozygous. If her father—the boys' maternal grandfather—had both traits (A) or neither (B), the traits were in the coupling phase in the mother (i.e., both traits were on the same X chromosome). If the maternal grandfather had either trait—that is, color blindness (C) or G6PD deficiency (D)—the traits were in repulsion (on different X chromosomes). Analysis of the complete set of pedigrees yielded evidence of close linkage: The chance of recombination disrupting the coinheritance of color blindness and G6PD deficiency was less than 5%. The investigation was conducted at Johns Hopkins University in the early 1960s by Ian H. Porter, Jane Schulze, and Victor A. McKusick.**

chromosome and this study established their linkage. If mothers who are carriers of both genes (doubly heterozygous) passed both genes to their sons frequently, these genes are

linked. The "grandfather method" was used to determine if color-blindness and G6PD deficiency genes were on the same X chromosome - that is coupled, or if they were on different X chromosomes - repulsion. A and B show that the traits were in the coupling phase while in C and D the genes were in the repulsion phase. Both genes are now known to be located at the end of the long arm of the X chromosome at band Xq28 (15). The units used to measure the closeness of two genes on the same chromosome is called a centiMorgan (cM), and was named after the geneticist, Thomas Hunt Morgan. cMs measure the frequency of recombination between two genes. One cM is equal to a recombination frequency of 1% and corresponds roughly to a physical interval of about 1,000,000 bases(15,17).

Linkage is established through statistical analysis of co-inheritance patterns in multi-generational pedigrees from which it is possible to calculate the probability that two loci will be inherited together versus the probability that they will separate during meiotic recombination. The reliability of estimates of distance between genes is expressed mathematically as a lod score(18). The lod score is a logarithm of the odds of linkage at a given distance, that is, the higher the lod score the greater the confidence of the estimate of linkage.

The ideal linkage for a meiotic map would be composed of highly polymorphic markers spread evenly throughout the genome. This map would allow the mapping of any genotypic or disease trait. Initial steps toward construction of a meiotic map were largely based on Southern blotting and simple sequence variation caused by loss or gain of a restriction endonuclease site (19). These DNA-based markers were named restriction fragment length polymorphisms (RFLP). RFLP's remain useful for mapping, but they have an index of heterozygosity of only 50%, since the restriction site is either present or absent. They therefore require rather large families and analysis of multiple restriction sites. Helen Hobbs, Grand Rounds on November 8, 1990, has an excellent and more detailed description of RFLP's. More recently, investigations have concentrated on DNA markers based on highly polymorphic mini- and micro- satellites (20-22). The mini-satellites consist of variable numbers of tandem repeats (VNTRs) of between 8 and 12 bases. The repeat unit in a micro-satellite appears to be mainly mono-di, tri- or tetra- nucleotides. While mini-satellites tend to cluster toward the telomeric end of chromosomes, micro-satellites appear to be more generally distributed and are currently the markers of choice when looking for polymorphisms with restriction enzymes. A further advantage of the micro-satellite is that they can be scored by using the polymerase chain reaction (PCR). Figure 4 shows a pedigree and the inheritance of short tandem repeats (STRs) polymorphism (23). The STRs are very useful for localizing

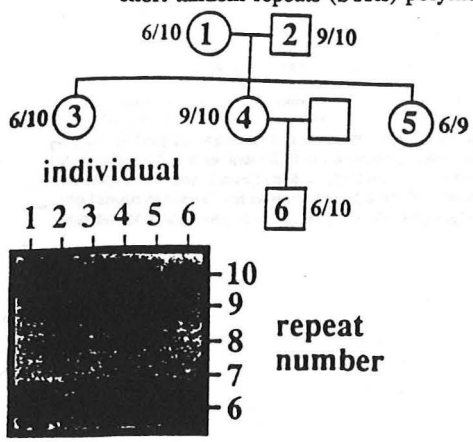


Figure 4 Inheritance of STR polymorphisms. In the pedigree shown above, circles indicate females and squares indicate males; the numbers within the symbols relate to the individuals shown in the panel below. The analysis shown here is PCR amplification of an autosomal $(CATT)_{1-10}$ repeat, the length of the radiolabeled amplification product varying according to the number of repeated units within it. The numbers shown alongside the pedigree symbols indicate the alleles detected in that individual. Figure supplied by A. Edwards.

genes because of their distribution and polymorphism. The best study class of micro-satellites are CA di-nucleotide repeats. These have an estimated 12,000 polymorphic $(dC\text{-}dA)_n$, $(dG\text{-}dT)_n$ loci in the human genome(24). Using these markers alone it would be estimated that one could create a map with a resolution of around 0.5cM.

Obviously, genetic linkage maps or meiotic maps require large families. In France, the Centre d'Etude du Polymorphisme Humaine (CEPH) maintains and distributes DNA from a large number of families where there are all four grandparents, parents and at least 10-12 children (25). This DNA is distributed internationally by CEPH and provides a resource for genomic mapping using the genetic linkage method. Figure 5 demonstrates the use of

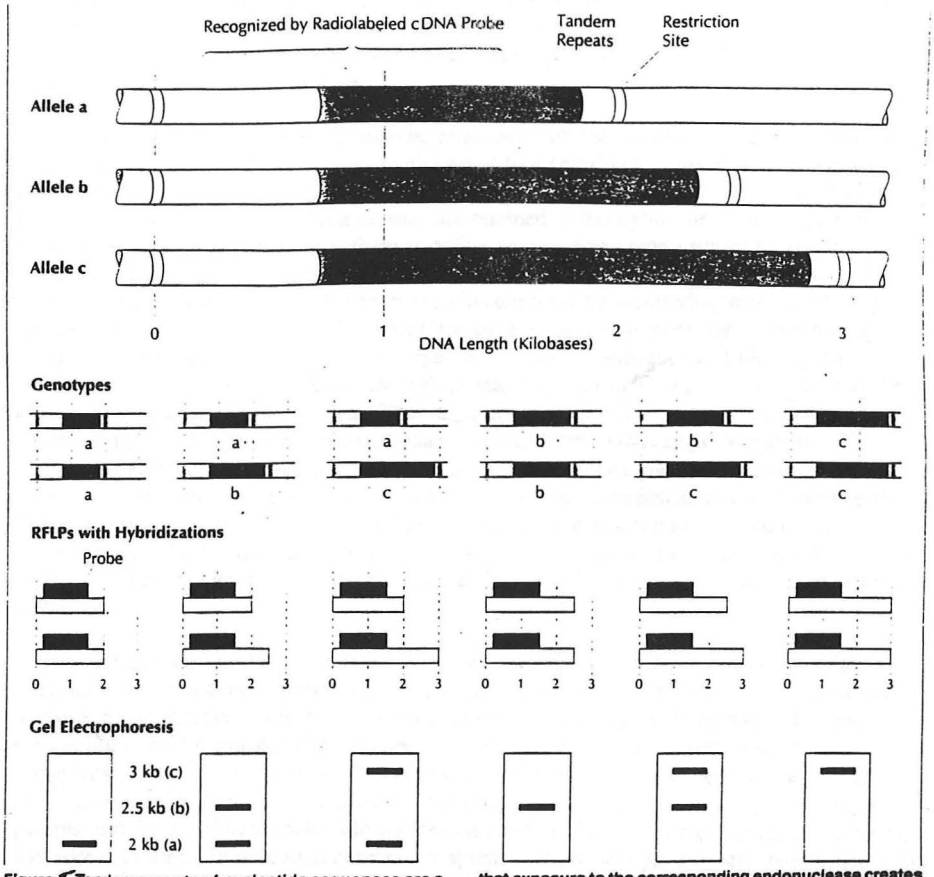


Figure 5. **Tandem repeats of nucleotide sequences are a type of genetic marker identified more recently than restriction endonuclease cleavage sites. In this example, increasing length of the repeat region, called a VNTR (variable number of tandem repeats), "pushes" a cleavage site, so that exposure to the corresponding endonuclease creates RFLPs of increasing length from allele *a* to *b* to *c*. Electrophoretic analysis of radiolabeled hybridizations identifies genotype, and coinheritance of a particular VNTR with a disease serves to localize the disease-related gene.**

variable tandem repeats (VNTR's) which are recognized by a radioactively labelled cDNA probe, to map three alleles that have different lengths of VNTRs (15). The genotype after restriction endonuclease cleavage of sites is shown in the panel below the three different length VNTR's. The VNTR pushes or extends the cleavage site so that exposure to the specific endonuclease creates restriction fragment length polymorphisms of increasing length from allele A to B to C. One then uses a probe complementary to the VNTR and hybridizes it. The alleles are separated by electrophoresis identifying the various genotypes. This method can be used to localize a disease-related gene or it can be used to map the chromosome in which this DNA resides by using a single copy (eg., cDNA) marker that is known to be associated with a specific chromosome and with a specific region on that chromosome. Helen Hobbs presented a more detailed discussion of VNTR and DNA fingerprinting in her Grand Rounds.

Linkage between genes can also be established on the likelihood of their separating when a standard dose of radiation is administered to a cell (26a). This method is known as radiation mapping. The preferred technique is to use a hybrid cell, that is, a fused human and rodent cell line. The rodent chromosomes are retained in the hybrid cell but the human chromosomes are eliminated until there is only 1 or 2 human chromosomes remaining, or even just a fragment of a human chromosome. The human chromosome retained in the rodent-human hybrid can be established and characterized by identifying specific expressed human proteins that differ from the rodent homologues on electrophoresis. Alternatively, specific DNA probes that hybridize to a specific human chromosome will identify the chromosome and the region. Radiation hybrid mapping was exploited by David Cox at the University of California in San Francisco. This approach provides a general method for constructing long-range maps of mammalian chromosomes. DNA regions spanning 20 megabase pairs were resolved into a high resolution contiguous map by radiation mapping (26a). This statistical method depends on x-ray breakage of chromosomes to determine the distance between DNA markers as well as their order on chromosomes. In addition, the method allows the relative likelihood of alternative marker orders to be determined. Thus, radiation hybrid mapping is a method for genetic linkage that does not depend on meiotic recombination.

Another genetic-linkage method that does not require family or pedigree studies is analysis of DNA sequence polymorphism in individual gametes (26b-d). Sperms are used to measure recombination frequencies between genetic markers by PCR methods. In brief, by comparing a male's somatic DNA sequences from any single diploid cell, with the DNA sequences present in the haploid meiotic products themselves, namely sperm, one can by direct analysis determine genetic linkage. Instead of comparing somatic DNA sequences of parents and their children, each child representing the product of a single meiosis, a mapper has access to virtually unlimited numbers of sperms. Individual gametes are determined to be parental or recombinant in genotype for individual loci using PCR amplification (on the order of 10^9-fold) of those loci; followed by allele specific oligonucleotide (ASO) probing gel electrophoresis of the PCR products. Many sperm can be individually typed and compared with similar analysis of the male's diploid cell. Dividing the number of recombinant sperm

examined by the total number gives an estimate of the frequency of recombination. Studying thousands of human sperm provides a fine structure linkage map at resolution far greater than using pedigree analysis of large families. At present, it is possible to type more than 500 meiotic products in one week with sperm mapping by PCR (26b-d). This method appears to have the potential for measuring genetic distances to less than 1cM.

A Common Language for Physical Mapping

As described above the emphasis on physical mapping of the human genome and that of model organisms has precedence over the sequencing of the genome. However, the different mapping strategies, some of which are described above and others which follow this section, require a common language to merge the mapping data gathered by diverse methods. The advent of the polymerase chain reaction (PCR) appears to provide a common path towards consolidating mapping data gathered in different laboratories by different methods. Moreover, it avoids the logistics and expense of managing a huge collection of cloned segments of DNA that would be required to be exchanged among investigators in order to facilitate mapping. Using short DNA sequences which can be detected with high specificity and sensitivity with PCR as landmarks along a chromosome, , the DNA sequence itself provides the basic unit for physical maps. In a seminal paper Olson, Hood, Cantor and Botstein (27) suggested the use of short tracks of single-copy DNA sequences (that is sequences that occur only once in the genome) that can be recovered at any time by PCR as the landmarks that define the position on the physical map. These landmarks were called sequence-tagged sites or STSs. Construction of a physical map would be determined by the order in spacing of DNA segments (STSs) each of which is identified uniquely by their sequence. This apparently simple idea solves the problem of merging data from many sources, and at the same time eliminates the need for large clone archives. Thus, the physical map can evolve naturally towards having the complete DNA sequence of the human genome. Since physical mapping and genetic linkage mapping is carried out by a variety of different techniques it would permit a common reference to these sequence-tagged sites (STSs). Thus, STSs will enhance mapping strategy by permitting all of the various techniques (that is restriction maps, contig maps, radiation maps and results of pulse-fields gel electrophoresis) to be expressed in a common language. In most instances the STS will encompass 200 to 500 base pairs of sequence each of which would be operationally unique in the human genome. The PCR assay for STS could be implemented by synthesizing short (approximately 20 nucleotides) oligodeoxynucleotide primers chosen to be complementary to opposite strands in opposite ends of the sequence tract. A DNA sample would be tested for the presence of the sequence by testing its capacity to serve as a template for the *in vitro* synthesis of the tract in the presence of these two primers. The procedure involves many automated cycles of DNA synthesis in a PCR apparatus, consequently when the assay is positive large amounts of product are made and it can be detected without radioactivite labelling.

The advantage of STSs for mapping landmarks is that the STSs can be completely

described as information in a database. It eliminates the need for biological materials. STSs also facilitates collection and storage of mapping data from diverse sources (28). It also allows the physical mapping of data to evolve in a straightforward way and facilitates comparisons between findings in different laboratories. The current goal for a five year period is to develop a map with an STS resolution at an average spacing of 100kb (27). This would require the mapping of approximately 30,000 STSs. Moreover, converting the existing sets of mapped DNA probes to STSs will impart a major momentum to the human genome project. Once the database was compiled it would lead to a direct ability to extend the physical mapping of chromosomes to a higher and higher resolutions.

Cytogenetic Maps

For many years physical maps of the human genome were rooted in cytogenetics, however, the resolution of a banded chromosome is only about 5 million bases per band (29). With Giemsa staining it is believed that most of the genes occur in the lighter bands while the dark bands reflect heterochromatin or heavily condensed chromatin. Recently fluorescent *in situ* hybridization (FISH) has revolutionized cytogenetic mapping by enabling the simultaneous sub-localization and ordering of several DNA sequences. This procedure is both rapid and sensitive with resolution being dependent on the degree of condensation of the target chromatin. Since the condensation of chromatin limits the distances that can be mapped, metaphase chromosome being tightly packed have a resolution of only 5 to 3 million bases. Prophase chromosomes are less condensed than metaphase chromosomes but they are only marginally better in resolving FISH localization. The chromatin in interphase nuclei are much less condensed and resolution of fluorescent probes to somatic interphase cells is said to be from about 2 million bases to perhaps 50,000 bases (29). With fluorescent probes different colors can be used to mark the site of hybridization in an interphase nucleus. Recently, Barbara Trask and her associates fluorescently labelled 13 well-characterized probes from the distal 4 megabase region of chromosome 4p (30). That is the distal short arm which contains the Huntington's disease region. Using pairs of fluorescent probes of different colors, her group was able to map the order of these probes along the chromosomes with the remarkable resolution of 100 to 2000 kilobases. Figure 6 shows the amazing accuracy of this

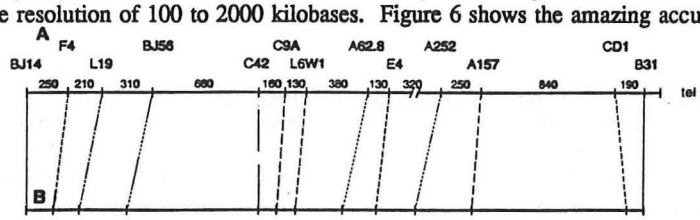


Fig. 6 Comparison of published map with map constructed from interphase distance measurements. (**A**) The published map of the terminal 4 Mbp of 4p16.3 showing the most likely locations of 13 cosmid probes used in this study (*11–15*). The numbers are the published estimates of genomic distance between the probes in kilobase pairs (*11–15*). Literature values vary by 10 to 30%. There is a gap in the map between E4 and A62.8. Cosmid probes used were BJ14 (locus D4S126), BJ56 (D4S127), C42 (42RB1.8, at D4S43), C9A (at D4S43), L6W1 (at J107, D4S166), A62.8 (D4S113), E4 (D4S168), A252 (D4S115), A157 (D4S111), CD1 (D5. D4S90), and B31 (D4S182) (*11–15*). (**B**) Map derived from relative physical distance measured among the cosmids in interphase chromatin. Distance estimates are the squares of the mean interphase distances between nearest neighbors.

fluorescent *in situ* hybridization when compared to a cosmid and contig physical map of the terminal 4 Mbp of chromosome 4p. The A portion of the figure shows the distances as calculated from hybridization with pairs of fluorescent probes. As can be seen the correlation is very close.

To provide even more extended chromatin, to improve the accuracy and resolution of fluorescent probe localization by *in situ* hybridization, a technique for mapping interphase pronuclei has been devised (31). The essence of the technique is to fuse human sperm to hamster eggs. The hamster egg cytoplasm processes the tightly packaged sperm DNA into large diffuse networks of chromatin fiber bundles providing hybridization targets more extended than those available in somatic interphase cell nuclei. The physical distance between hybridization signals in these pronuclei were measured in micrometers and correlated to genomic distances determined by restriction fragment analysis of the human factor VIII and color vision region of the X-chromosome q28. It was estimated that the mean pronuclear distances between hybridization sites were about three times larger than those found in somatic interphase cells. The relationship between the physical and genomic distance was linear from less than 50kb to at least 800kb. These results using a sperm-egg system promise to extend mapping range obtainable by FISH analysis in pronuclei to provide a high resolution physical map.

In addition to their usefulness in assigning a particular gene to a chromosomal region and in mapping chromosome regions FISH has also been useful in explaining cytological observations. For example, as we know the number of Barr bodies per cell nucleus is always 1 less than the number of X-chromosomes. The Barr body is believed to represent the inactivated X-chromosome. Recently Barbara Migeon and her collaborators using FISH markers have discovered that the Barr body represents a looped X-chromosome with the telomeres attached to the nuclear membrane (32). This finding explains the location and the nature of the Barr body. The use of FISH can be applied not only to metaphase and prophase chromosomes and the interphase and pronuclei but can also be used with somatic hybrids and also with the yeast artificial chromosome (YAC). It is noteworthy that somatic hybrids and *in situ* hybridization have contributed most to the chromosomal localization of genes (Table 2).

TABLE 2.
NUMBERS OF GENES OR DNA MARKERS LOCALIZED TO DATE

Techniques	Genes	Marker DNA (Anonymous)
In situ hybridization	1,218	1,948
Somatic cell hybridization	1,878	6,770
Restriction endonuclease analysis of DNA	1,260	498
Linkage/family studies	595	2,685
Total single copy DNA	3,521	13,067

(Courtesy of Drs. A. Bowcock and P. Ceverha, 1993)

The Contig Map

A higher resolution physical map can be constructed for each human chromosome by using overlapping cloned segments of DNA, these are called contig because they are contiguous overlapping regions of DNA. The contigs may range in size from 20 to several thousand kilobases depending on whether they are cloned in plasmids or by using a newer technique, the yeast artificial chromosome (YAC) (33). YAC's are constructed by inserting human DNA between a yeast chromosome centromere and telomere. Contig maps have the potential to provide more information than other forms of physical maps, since the contigs can be assigned to specific chromosomal bands and encompass single copy genes or STSs. They may also be used to produce a restriction map, that is a physical map of contigs that are sub-divided by nearby restriction endonuclease sites.

Physically mapping a chromosome by the contig method requires a technique where one can selectively cleave a relatively large segment of DNA from a specific chromosomal site. Recently the development of the "triple helix strategy" has enabled investigators to target a particular region of the chromosome for excision and later amplification(34,35). The triple helix method is shown in Fig. 7. A specific oligonucleotide containing as few as 15 bases can be targeted to a known homologous region near the segment of chromosome to be cleaved. An enzyme, called a recombinase, is used to insure specificity of pairing with the homologous duplex of DNA to form a three strand complex. This allows specific targeting of

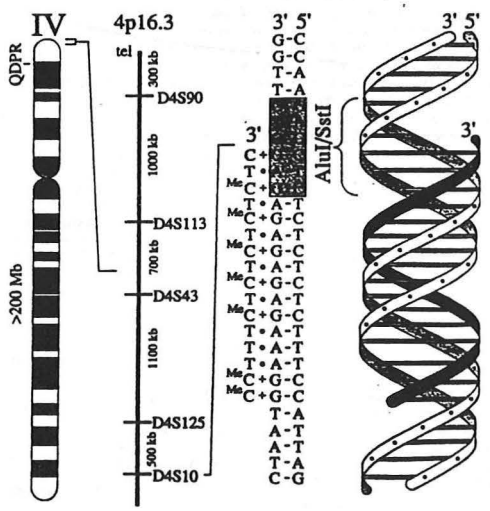


Fig. 7. (**Left**) Schematic diagram of human chromosome 4 showing the cytogenetic location of the QDPR locus at 4p15.3. (**Left-Center**) A physical map of 4p16.3 to 4p16ter showing the location and distance between the loci described. On the basis of pulsed field gel mapping, the size estimate from D4S10 to the telomere is approximately 4 Mb (*17*). The distances indicated between the loci are approximations from the physical map (*17*). (**Right-Center**) The sequence and relative orientation of the Alu I–Sst I target site identified by the degenerate pyrimidine oligonucleotide search. The site is recognized by a 16-base ^{Me}CT oligonucleotide with ^{Me}C + GC and TAT base triplets (Alu-16). (**Right**) Schematic diagram of the triple-helix complex. The pyrimidine oligonucleotide is bound in the major groove, parallel to the purine strand of the DNA duplex and covers the purine half of the Alu I methylation site.

any sequence in the DNA for cleavage. The oligonucleotide forming the central strand in the triple helix is constructed to overlay a restriction endonuclease site. An appropriate methylase for the particular consensus sequence of the restriction endonuclease is added, and all of the restriction sites except the one that is covered by the oligonucleotide are methylated. Methylation sites are resistant to cleavage by the specific restriction enzyme. The methylase enzyme is inactivated and the oligonucleotide is then disassociated from the DNA helix. The addition of the specific restriction enzyme then cleaves only at the predetermined site that was covered by the oligonucleotide (34).

A recent example of the power of this technique was carried out by Maynard Olson's laboratory where they were able to clone and assemble the cystic fibrosis gene (CFTR gene) located on chromosome 7(36). This achievement was made possible by the use of YAC cloning for isolating and analyzing this large block of human DNA. As described above in the general discussion of mapping Olson used sequence-tagged sites (STS), a class of landmarks that are designated as the natural language of the physical mapping project(27). A given region of DNA can be mapped by determining the order of a series of STS and measuring the distance between them. Fig. 8A shows the YAC-based STS contig mapping. The use of the presence or absence of additional STSs define the overlapping relationships between different segments of chromosomes carried in the YAC (Fig. 8B). The isolated

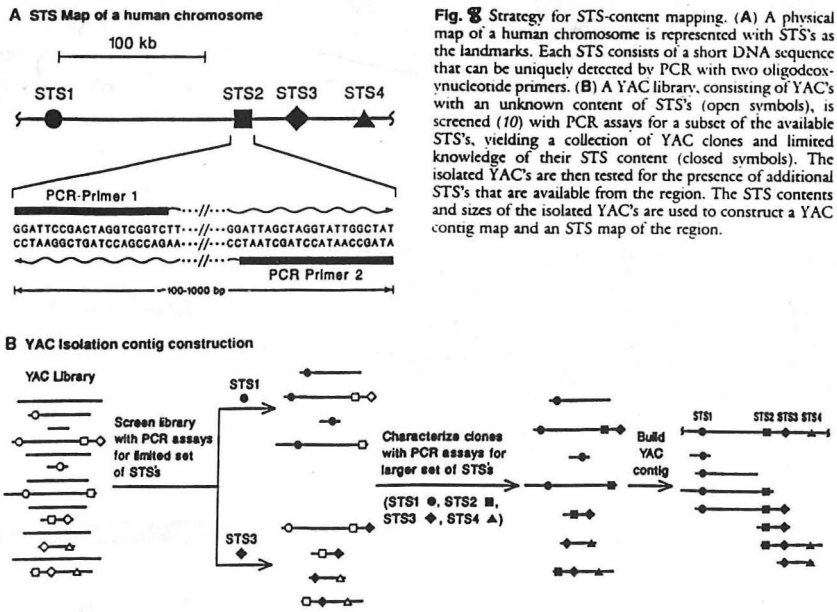


Fig. 8 Strategy for STS-content mapping. (**A**) A physical map of a human chromosome is represented with STS's as the landmarks. Each STS consists of a short DNA sequence that can be uniquely detected by PCR with two oligodeoxynucleotide primers. (**B**) A YAC library, consisting of YAC's with an unknown content of STS's (open symbols), is screened (*10*) with PCR assays for a subset of the available STS's, yielding a collection of YAC clones and limited knowledge of their STS content (closed symbols). The isolated YAC's are then tested for the presence of additional STS's that are available from the region. The STS contents and sizes of the isolated YAC's are used to construct a YAC contig map and an STS map of the region.

YAC's are tested for the presence of additional STSs that are known to be present in this region. The STSs content and size of the isolated YAC's are used to construct a contig map and an STS map of the region around the cystic fibrosis gene. The size of the human DNA in the yeast artificial chromosome can be determined by pulse-field gel electrophoresis. Olson's study showed that several of the YAC isolates contained significant portions of the CFTR gene but none of them contained the entire genomic sequence of more than 250 kilobases (Fig. 9). They therefore carried out an experiment, selecting two YAC's containing

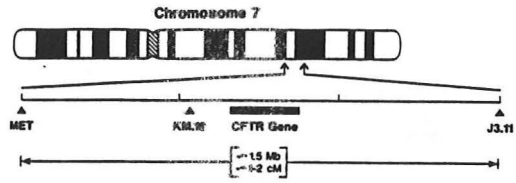


Fig. 9. The CF region of human chromosome 7q. The CFTR gene spans more than 250 kb within the interval defined by the genetic markers KM.19 and J3.11 (*14*).

large segments of the CFTR gene that overlap one another (Fig. 10). Homologous recombination between the pair of overlapping YAC's appeared to be a promising method to generate a single YAC that would span the entire gene. In yeast the average ratio of physical to genetic distance is only 3kb/cM. A high frequency of meiotic recombination is therefore expected. Shown in Fig. 10 are the two YAC chromosomes that together span the entire CFTR gene. The hatched area in Fig. 10 indicates the areas of overlap. These two YAC

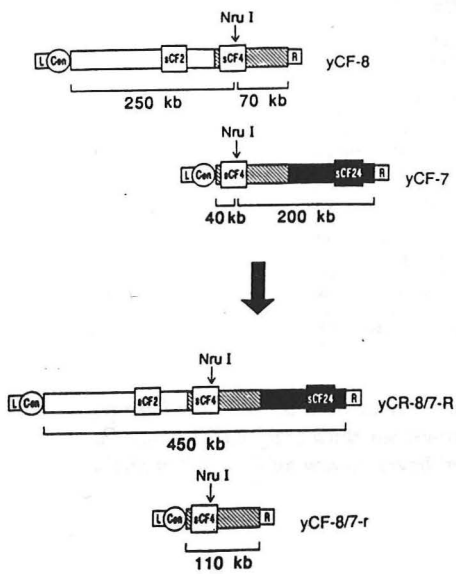


Fig. 10 Schematic model for the recombination of two YAC's, yCF-8 and yCF-7. Mapping of these YAC's with Nru I revealed the relative alignment of the two clones. The region common to both parental YAC's (hatched area) is approximately 110 kb in size and contains sCF4. The sizes and STS content of the regions distinctive for each YAC (open and closed areas in yCF-8 and yCF-7, respectively) were also determined. Homologous recombination between yCF-8 and yCF-7 within the common segment of human DNA would yield two products: a large YAC (yCF-8/7-R) containing the common and two distinctive regions, and a smaller YAC (yCF-8/7-r) consisting only of the 110-kb common region. Since the two parental YAC's are in the same orientation with respect to the vector arms, each of the reciprocal products obtained after recombination would be expected to contain a single centromere.

chromosomes were present in haploid yeast of opposite mating types. Following mating the diploid strains containing both YAC's underwent meiosis and sporulation. It was hoped that homologous recombination would occur between the two chromosomes with overlapping regions and that one recombinant would represent the 450 kilobase CFTR gene and that the other recombinant would be 110 kilobases of the overlapping regions. The spores in the resultant tetrad were separated and analyzed. PCR-based assessment of STS content and size measures by pulse-field gel electrophoresis were used to identify recombinants. Figure 11 shows the results of this experiment. Each of the spores analyzed showed that two were the

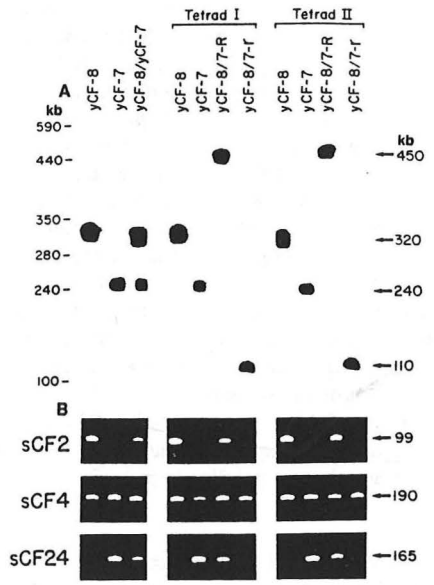


Fig. 11 Characterization of recombination products from the yCF-8 × yCF-7 mating. Yeast strains containing yCF-8 and yCF-7 were mated, and the resulting spores were analyzed, yielding two tetrads containing YAC's that appeared to be the products of reciprocal recombination. (**A**) Intact chromosomes from each of the parental YAC clones, the diploid strain yCF-8/yCF-7, and the indicated spores from each tetrad were separated by pulsed-field gel electrophoresis and analyzed by gel-transfer hybridization with ^{32}P-labeled total human DNA as the probe. For both tetrads, designated I and II, the parental YAC's, yCF-8 (320 kb) and yCF-7 (240 kb), as well as the recombinant products, yCF-8/7-R and yCF-8/7-r, were each found in a single spore. The two recombinant YAC's were of the predicted sizes (450 kb and 110 kb, respectively), as shown in Fig. 6. (**B**) Each of the yeast strains analyzed in (A) was tested for the presence of sCF2, sCF4, and sCF24 by PCR as described in the legend to Fig. 4.

parental type and the third showed the recombinant 450 kilobase chromosome that contained the entire CFTR gene while the fourth recombinant in the tetrad showed the expected 110 kilobase DNA. This was observed in both tetrad 1 and tetrad 2 that were analyzed. This

experiment demonstrates the power of the YAC-amplified human DNA and the ability of yeast to force recombination and to isolate specific genes that span regions that are so large they could not be isolated by any other means. This study also emphasizes the importance of the STS strategy in physical genomic mapping. These experiments were successfully used in assembling a YAC contig map containing more than 1.5 million bases of DNA from the region of chromosome 7. Moreover, they were able to assemble the entire cystic fibrosis gene in a single 450kb segment of DNA (36).

Although the YAC method of amplifying DNA has greatly improved our ability to recover large chunks of genetic material there are some disadvantages. For example, YAC's may form chimeras, these are DNA fragments containing two different human chromosomes linked together(33). This tends to occur because overlapping ends of two chromosomes may have repetitive sequences that presumably cause them to associate(). To resolve this problem and mitigate it, YAC libraries that are made from human-rodent hybrid cells that contain only a fragment of the human genome and show lower levels of co-cloning of human DNA than those made from total human DNA. It should be noted that human DNA from human-rodent hybrid cell lines will co-clone with mouse or hamster DNA whichever was the host. However these co-cloned YAC's can be easily distinguished by their hybridization to both human and hamster DNA probes. Moreover, one can select for study only those YACs that have small fragments of rodent DNA with large fragments of human DNA for further study.

Many innovative techniques are being used to physically map human chromosomes. Among them are chromosome micro-dissection in which a specific area of a chromosome is physically dissected and isolated (37). The deproteinized chromosome segment is then amplified by PCR and sub-cloned into plasmids or into YACs for further characterization. There have also been several new techniques for trapping exon from genomic DNA (38,39,40). These techniques are aimed at identifying areas in the genome that are transcribed and expressed.

Our time is limited and I want to consider the implications of the genome project to society and to medicine. We will have an opportunity to hear the latest in the human genome mapping advances. Michael Brown is organizing a symposium with the experts from the human genome project which will be held here at Southwestern in Gooch Auditorium on May 6, 1993, at 2:30 p.m. This symposium will obviously be an authoritative review of the current status of the Genome Project. We will have an opportunity to really update the strategies for mapping the human genome.

MEDICAL IMPLICATIONS OF THE GENOME PROJECT

As primary care physicians we can participate in the human genome project by being alert to patients who exhibit features of a genetic disease. In particular we should be alert to those individuals who appear to have concurrently two different genetic diseases. These patients must be refered to genetic centers for detailed cytogenetic analysis. The cloning of

the muscular dystrophy gene and the gene causing chronic granulomatous disease were facilitated by a physician recognizing these two diseases in a single patient. The patient was referred to a genetic center, and cytogenetic analysis found a gross chromosomal deletion in the X chromosome of this male child. The deletion provided the focus for positional cloning studies (41). The human genome project will greatly enhance our ability to clone genes whose protein products are unknown (Table 3). As the physical map of the human

TABLE 3.

IMPACT OF THE GENOME PROJECT ON MEDICINE

I.	Enhance efficiency of gene localization and cloning
II.	Isolation of disease-associated genes
III.	Understand the pathophysiology of genetic diseases
IV.	Lead to innovative treatment of genetic diseases
V.	Recognition of predisposition to specific diseases
VI.	Recognition of susceptibility to environmental hazards
VII.	Selective prevention of diseases
VIII.	Presymptomatic genetic diagnosis

chromosome achieves higher and higher resolution, it will be possible to detect disease-causing genes with increasing efficiency and speed. This unparalleled opportunity to discover disease producing genes and to begin to study the pathophysiology by which these genes produce disease will revolutionize medicine. The long-term achievements of the human genome project will provide enormous benefits to mankind by allowing more effective treatment of disease, by permitting a recognition of predisposition to particular diseases or to environmental hazards, and by enhancing our ability to selectively employ preventive measures to minimize genotypic effects on the phenotypes.

Presymptomatic Genetic Diagnosis

Currently medical diagnostic tests reveal diseases, for example diabetes, that can be treated by replacing a deficiency, or by detecting a cancer at a stage when it can be surgically removed or controlled with chemotherapy. However, in the new era of DNA-based diagnostic testing our prospective outcome of such diagnoses must be examined critically. DNA-based diagnostics can clearly identify Duchenne muscular dystrophy. Prior to the discovery of the gene for this disease, a young boy with muscle weakness would be subjected to a battery of tests including the measurement of muscle enzymes, electromyography, and muscle biopsy. Combined results of these tests might have a high diagnostic accuracy if Duchenne muscular dystrophy was present, particularly if there was a good family history. Now we can unequivocally identify an abnormal gene for dystrophin, the muscle protein deficient in Duchenne dystrophy. At birth we can predict with certainty that a child at risk will eventually develop Duchenne muscular dystrophy. Female carriers of the Duchenne

mutation can be detected unequivocally and pre-natal DNA-based diagnosis is readily available for her fetus.

DNA-based diagnostic methods for other diseases can be more ambiguous as discussed in a provocative chapter by Tom Caskey entitled "Presymptomatic Genetic Diagnosis - A Worry for the United States". Adult polycystic kidney disease (APKD) illustrates the issues and challenges which DNA-based testing may cause. APKD is an autosomal dominantly-inherited disorder which occurs at a frequency of 1 to 5 per 1,000 live births and is characterized by the early onset of hypertension and renal failure (43,44). Each child of an affected individual has a 1 in 2 chance of inheriting the disease. It is estimated that in the United States there are 500,000 people at risk and approximately 7,000 new patients are recognized every year. As with other dominant diseases, which have their onset in adult life (for example, Huntington's disease, myotonic dystrophy, and Alzheimer's disease) the age at which the symptoms occur varies even within a given family. The reason for differences in expression and the variability of onset is not known. However, to a patient with APKD it is critical, because it influences the age at which dietary therapy is advised, medical treatment is initiated and dialysis and renal transplantation are required. Prior to DNA-based testing, the diagnosis of APKD was entertained if more than 3 renal cysts were documented by ultrasonography, or if there was a family history of the disease (45). Figure 12 illustrates the typical course of APKD disease and presents a number of the problems and challenges of pre-symptomatic diagnosis when there is no specific treatment. The ability to detect renal cysts

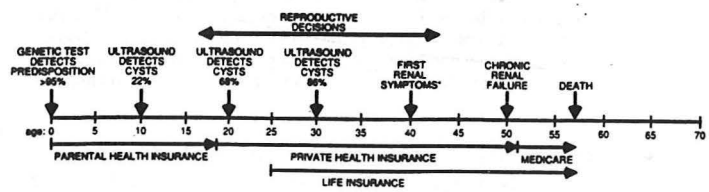


Figure 12. Events in the life of an APKD patient. A typical course of APKD is shown. *Hematuria, renal stones, and/or hypertension. The ability to detect renal cysts by ultrasound increases with age,[3] but genetic tests for predisposition to APKD can be performed even before birth. The APKD patient has many years of symptom-free health care and life insurance coverage before medical treatment is required. After eighteen months of chronic renal failure, the Medicare system takes over payment of medical costs from the private health carrier until the time of death.

by ultrasonography increases with age, but the genetic tests for the predisposition to APKD can be preformed even before birth. As illustrated in the figure, the APKD patient has many years of symptom-free health and potential insurance coverage before medical treatment is required. The DNA-based testing can accurately determine the probability of having the APKD gene at an age before reproductive decisions are made.

The APKD gene was localized to the short arm of chromosome 16 by its linkage to the 3' hypervariable region of the alpha-globin gene (46). Figure 13 shows a pedigree of a family affected with APKD. There are abundant genetic markers in the region of

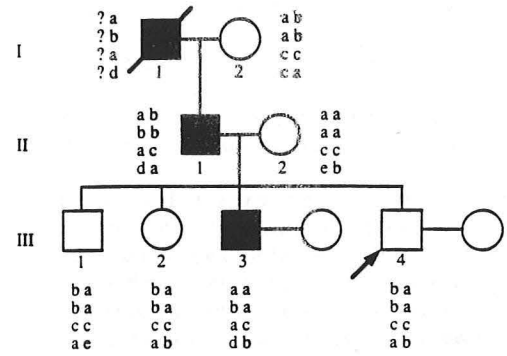


Figure 13. Pedigree of a family affected with APKD. Squares represent males, circles represent females, filled symbols indicate individuals with APKD. I-1. Renal failure, death at age forty-one. II-1. Age fifty, undergoing chronic kidney dialysis. III-1. Age thirty-two, without renal cysts. III-2. Age thirty, no ultrasound studies. III-3. Age twenty-eight, has four renal cysts. III-4. Consultand, age twenty-three, no ultrasound studies. Each person inherits one chromosome from each parent, and the alleles present at four loci close to the APKD locus (two on either side) are arranged vertically for each chromosome. DNA was not available from individual I-1, but since he was affected and the APKD gene is associated with the "a,b,a,d" group of alleles in this family, he was assumed to carry the "a,b,a,d" haplotype.

chromosome 16 in which the APKD gene resides. In most families there are multiple polymorphic alleles close to the APKD gene (46). Observing which of the alleles are inherited in conjunction with the APKD gene in most families enables prediction of the disease with an accuracy approaching 95%. As is seen in the pedigree those individuals that have the a,b,a,d haplotype are those destined to have the disease. It should be recognized that the cooperation of family members is necessary in order to ascertain the haplotype responsible for the disease. This obviously raises complications if certain family members are unwilling to provide specimens of blood. If DNA is not available from enough family members or if the genetic markers within the family are not appropriately distributed then DNA studies may be uninformative. In this circumstance the risk to the patient remains at 50%. If the DNA studies are informative with the family's cooperation, then the patient will know his risk of APKD is less than 5% or greater than 95%.

Once a patient is identified with a high risk of APKD his physicians would be able to follow renal function and blood pressure because azotemia and hypertension are mitigated by medical and dietary therapy. Routine medical surveillance should be the rule and the cost of these observations is low. Vigorous treatment of hypertension and renal dysfunction may

delay the need for dialysis. However, most patients require dialysis 6 to 8 years after developing significant hypertension or azotemia. National costs for patients with renal failure is currently estimated at over $400 million dollars per year and is covered by the government under the Medicare End-Stage Renal Disease (ESRD) program. Recently the government is reducing funding of Medicare for renal disease and has displaced some of the cost to private insurance carriers. Private insurers must now bear the expense for 18 months before Medicare funds may be used for dialysis. It is likely that this saving trend will continue and perhaps become even more stringent. Since private health care insurers are sharing the costs for caring for APKD patients, their attitudes concerning presymptomatic DNA testing will influence the available funding for care of patients predicted to develop the disease.

If a patient at risk for APKD elects to be screened it may have serious consequences. Such testing would place the patient in either a normal or high insurance risk category depending if the test revealed a normal result or the presence of the mutant gene. These results would markedly influence the cost of insurance (47). An employer either for insurance pool risks or prediction of long-term impact on job performance might insist that DNA studies clarify the potential employee's genetic risk. The patient's wife may request information since reproduction is a joint decision. How far can we go to protect the patients confidentiality?

At present we accept insurance carriers' right to determine whether an individual has a prior existing medical condition. It is presently accepted that these prior existing conditions do influence eligibility or rate of insurance. Many of these potential risks may become moot, if Congress enacts a universal insurance plan that would cover chronic renal disease, including APKD. Indeed most physicians would insist that pre-symptomatic diagnosis is not a disease and therefore should not be considered a prior existing condition. There is likely to be a great deal of controversy concerning this issue. Unfortunately, courts have recently decided litigation that restricts an individual's rights to privacy. It is clear from this example, that we as general internists and specialists in internal medicine have a major stake in the human genome project. Many genes will be detected that cause diseases in adult life or even at older ages. It is likely that we will be able to diagnose these diseases before there is any effective therapy. The knowledge that an individual is an asymptomatic carrier of such a gene may have adverse psychological effects. In a recent study from Canada, 136 participants had DNA-based testing to determine their risk for Huntington's disease (48). The subjects at risk were divided into three groups based on the test results: an increased-risk group, a decreased-risk group, and a group with no change in risk because of uninformative family studies. After counselling a battery of tests were administered for psychological distress. The decreased-risk group, as expected, had lower scores for distress than baseline results before testing. The increased-risk group showed no significant change in distress from baseline. The group with no change in risk scored lower with higher distress than their baseline results. Follow-up at 12 months gave similar results. Predictive testing for Huntington's disease improves the psychological well-being of those at decreased-risk without apparently worsening the distress of subjects at increased-risk. However, a longer follow-up of psychological effects of positive-testing is needed before firm conclusions are made.

IMPLICATIONS OF THE GENOME PROJECT FOR SOCIETY, LAW, AND ETHICS

The planners of the human genome project have earmarked 3% of the genome budget funds to explore ethical, legal and social concerns (ELSI Project). This amount is estimated to be about 90 million dollars over fifteen years (49). This is a wise policy since there are serious social issues that arise through the potential application of gene mapping. There is a need to assure that the benefits of the project are maximized and that the potential risks are minimized. It is important that we as physicians and our organizations are in the forefront planning policy and advocating for the welfare of our patients. Our profession has the tradition and experience to counsel and advise our patients and their families. Therefore, it is critical that we understand the Genome Project. There are a number of important issues emanating from human genome research: these include misinterpretation of the social importance of certain genomic findings, prevention of discrimination in employment and education, inappropriate political uses, eugenic interpretations, protection of confidentiality and the databank and legal issues (Table 4).

TABLE 4.
IMPACT OF THE GENOME PROJECT ON SOCIETY

I. Misinterpretation of genomic finding with respect to individual behavior and quantitative traits

II. Prevent discrimination in employment and education

III. Inappropriate use for political purposes

IV. Genomics and eugenics

V. Protection of the confidentiality of information - "the slippery slope"

VI. Protection of the databanking of the Genome Project

VII. Legality of patenting DNA sequences

In general, society expects the fruits of our research to create a better world. The proponents of the genome project have raised expectations, and I believe they can achieve them, providing the social and ethical pitfalls can be avoided and/or minimized. As the number of genes mapped increases, the gap between what we can diagnose and what we can treat is widened. Moreover, genes that contribute to the determination of certain quantitative traits have been shown to set limits not absolute levels, and can be modified by the environment (50). A too exact interpretation of the importance of the genetic constitution is not only wrong, but can lead to dangerous assumptions and policies. For example, one can predict that weak associations will be found between a particular genomic constitution and

certain forms of behavior or quantitative traits. On further analysis some of these associations will be spurious. Others, however, will be statistically significant but their importance may be exaggerated out of all proportion to their genetic importance for society.

One common ethical and legal question is to what extent is a person responsible for himself. Society may ask if an individual who commits a violent act was responsible for his behavior. A legal argument concerning the XYY genotype was used to defend a mass murderer in Chicago. There may be those who try to use genetics as an explanation or excuse for individual behavior traits, as well as for emphasizing differences between groups. The genome project may enhance the tendency to explain differences if genetic correlates suggest some ethically, legally or economically consequential outcome. There may be a temptation to assume that genetic correlates are fundamental (that is outside the individual's responsibility). The findings of genetic connections to important human traits and behavior may be used to explain mental illness, as well as addiction, occupational and environmental illness, and the educational and occupational attainments of different racial and ethnic groups. Certain of these genomic constitutions may be valid and may be helpful in finding more effective treatments and means of prevention or correction. However, history is rich with examples of scientific perspectives used inappropriately for political purposes.

I would refer you to Tyler Miller's Grand Rounds of September 17, 1992, where he examines the eugenics movement with respect to "genes and intellect". The eugenic movement exploits convenient explanations for intractable social problems. In 1988 a school physician wrote in the New York Times Magazine (August 14, 1988) that the tendency for people to do physical harm to others was "intrinsic, fundamental, and natural". Genes have been blamed for nearly every conceivable aberration and vice of human life. However, there is a disturbing continuity between genomics and eugenics (51). Many people believe that human talent and disabilities are inherited and indeed there is some truth in this. However, human biological differences are modifiable since the genes provide a range of performance rather than an absolute level of attainment (50). We must be on our guard, because in the late 20th century, in America, as in mid-century and late 20th century in Germany, it is dangerous to assume that the genes determine our destiny.

We as physicians must take a leadership role in preserving the confidentiality of information that relates to patients, their diagnosis, prognosis, and other medical finding, including genetic testing. The control of this information is difficult, as courts have decided in the past, that information access should be limited to a "need-to-know basis". The difficulty of achieving this goal is exemplified by HIV testing (52-55). Clinical information on HIV status has led to loss of housing, employment, and insurance. These are among the documented consequences of disclosure. The physician-patient relationship is built on trust and ethical and legal protections of that trust must be firmly and unequivocally established. Genetic information can properly be considered private, requests or demands for information concerning a patient's genetic constitution is an unwarranted invasion of privacy. However, in the past the courts have developed a concept of "duty to disclose". An example is the Tarasoff case decided in California in 1976 (56). That decision held a psychotherapist liable

for the murder of a female student. The therapist's patient had fantasized about his intention to kill Ms. Tarasoff if his love was not returned. The psychotherapist contacted the campus police, who detained the patient but then released him because he seemed rational. However, later the patient did carry out the murder. The California Supreme Court ruled that it was the psychotherapist's duty to warn the potential victim and to take other preventive steps. It is a short step from this decision to the breaching of confidentiality in other situations. If a physician or counsellor discovers genetic information about a patient that could substantially affect the interests or health of the patient's relatives, do they have a "duty to disclose"? This situation constitutes a genuine dilemma in which the physician or counsellor has two obligations that cannot be simultaneously met. One is the duty to the patient to preserve confidentiality versus a duty to disclose medically relevant information to a third party who could benefit substantially from that information.

The databanking of findings in the genome project will be computerized information resources. If this data is not appropriately coded or anonymous, the ability of someone with access to a number of different databases can pose a series of queries requiring yes or no answers that can yield highly specific information about an individual. The FBI has already proposed a national DNA database for the purpose of law enforcement (57,58). Currently that database is limited to profiles of persons convicted of violent crimes or those involved in unsolved cases. Moreover, DNA samples from certain of these individuals are stored in a repository controlled by the FBI. We are indeed on a slippery slope!

CONCLUSION

The human genome project provides an unparalleled opportunity to learn about human biology and to begin to understand at the molecular level human development and physiology. It also provides the opportunity to understand disease and to develop innovative treatments. The dangers for the misuse of this information have been detailed, and we as physicians individually and through our organizations must work for the benefit of our patients, and for the retention of confidentiality and trust of the patient-doctor relationship. In the end, the human genome project will probably be the crowning achievement to the last decade of the 20th century and to the first few decades of the 21st century.

REFERENCES

1. Hoffman EP, Brown RH Jr., Kunkel LM. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. Cell 51:1363; 1987.

2. Rommens JH, et al. Identification of the cystic fibrosis gene: Chromosome walking and jumping. Science 245:1059; 1989.

3. Lee B, et al. Linkage of Marfan syndrome and a phenotypically related disorder to two different fibrillin genes. Nature 352:330; 1991.

4. Kinzler KW, et al. Identification of a gene located at chromosome 5q21 that is mutated in colorectal carcinoma. Science 251:1366; 1991.

5. Weinberg RA. There are two large questions. FASEB Journal 5:78; 1991.

6. Mapping and sequencing the human genome. National Research Council, National Academy Press, Washington, DC, 1988.

7. Understanding our genetic inheritance: The U.S. human genome project: The first five years, FY 1991-1995. Dept. HHS, Dept. DOE, National Academy Press, Washington, DC, 1988.

8. Botstein D, Fink GR. Yeast: An experimental organism for modern biology. Science 240:1439; 1988.

9. Botstein D. Why yeast. Hosp Pract Oct. 15, 1991, p. 121-128.

10. Link AJ, Olson MV. Physical map of the Saccharomyces cerevisiae genome at 110-kilobase resolution. Genetics 127:681; 1991.

11. Coulson A, Kozono Y, Lutterbach B, Shownkeen R, Sulston J, Waterston R. YACs and the C. elegans genome. Bioessays 13:413; 1991.

12. Roberts L. The worm project. Science 248:1310; 1990.

13. Chapman VM, Klein J, Nadeau JH, Silver LM (Eds): Encyclopedia of the Mouse Genome I. *Special Issue Mammalian Genome.* 1:1-523; 1991.

14. Copeland NG, Jenkins NA. Development and application of a molecular genetic linkage map of the mouse genome. Trends Genet 7:113; 1991.

15. McKusick VA. Genomic mapping and how it has progressed. Hosp Pract Oct. 15, 1991, p. 50-62.

16. Porter IH, Schulze J, McKusick VA. Genetical linkage between the loci for glucose-6-phosphate dehydrogenase deficiency and colour-blindness in American negroes. Ann Hum Genet 26:107; 1962.

17. McKusick VA. Current trends in mapping human genes. FASEB J 5:12; 1991.

18. Stephens JC, et al. Mapping the human genome: Current status. Science 250:237, 1990.

19. Botstein D, White RL, Skolnick MH, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms (RFLPs). Am J Hum Genet 32:314; 1980.

20. Jeffreys AJ, Wilson V, Thein SL. Hypervariable 'minisatellite' regions in human DNA. Nature 314:67; 1985.

21. Royle NJ, Clarkson RE, Wong Z, Jeffreys AJ. Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. Genomics 3:352; 1990.

22. Weber JL. Informativeness of human $(dC\text{-}dA)_n$, $(dG\text{-}dT)_n$ polymorphisms. Genomics, 7:524; 1990.

23. Rossiter BJF, Caskey CT. Molecular studies of human disease. FASEB J 5:21; 1991.

24. Armour JAL, Jeffreys AJ. Biology and applications of human minisatellite loci. Curr Opin Genet Develop 2:850; 1992.

25. Cann HM. CEPH maps. Curr Opin Genet Develop 2:393; 1992.

26a. Cox DR, et al. Radiation hybrid mapping: A somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science 250:245; 1990.

26b. Li H, Gyllensten UB, Cuy X, Saiki RK, Erlich HA, Arnheim N. Amplification analysis of DNA sequences in single human sperm and diploid cells. Nature 335:414; 1988.

26c. Goradia TM, Stanton VP Jr., Cui XF, Aburatani H, Li HH, Lange K, Housman DE, Arnheim N. Ordering three DNA polymorphisms on human chromosome 3 by sperm typing. Genomics 10:748; 1991.

26d. Hubert R, Stanton VP Jr., Aburatani H, Warren J, Li H, Housman DE, Arnheim N. Sperm typing allows accurate measurement of the recombination fraction between D3S2 and D3S3 on the short arm of human chromosome 3. Genomics 12:683; 1992.

26e. Zhang L, Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N. Whole genome amplification from a single cell: Implications for genetic analysis. Proc Natl Acad Sci US 89:5847; 1992.

27. Olson M, Hood L, Cantor C, Botstein D. A common language for physical mapping of the human genome. Science 245:1434; 1989.

28. Pearson ML, Soll D. Human genome project: A paradigm for information management in the life sciences. FASEB J 5:35; 1991.

29. Trask BJ, Massa H, Kenwrick S, Gitschier T. Mapping of human chromosome Xq28 by two color fluorescence *in situ* hybridization of DNA sequences in interphase cell nuclei. Am J Hum Genet 48:1; 1991.

30. van den Engh G, Sachs R, Trasks BJ. Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model. Science 257:1410; 1992.

31. Brandriff B, Gordon L, Trask BL. A new system for high resolution DNA sequence mapping in interphase pronuclei. Genomics 10:72; 1991.

32. Walker CL, Cargile CB, Floy KM, DeLannoy M, Migeon BR. The Barr body is a looped X chromosome formed by telomere association. Proc Natl Acad Sci USA 88:6191; 1991.

33. Schlessinger D, et al. Yeast artificial chromosome-based genome mapping: Some lessons from Xq24-q28. Genomics 11:783; 1991.

34. Ferrin LJ, Camerini-Otero RD. Selective cleavage of human DNA: RecA-assisted restriction endonuclease (RARE) cleavage. Science 254:1494; 1991.

35. Strobel SA, et al. Site-specific cleavage of human chromosome 4 mediated by triple-helix formation. Science 254:1639; 1991.

36. Green ED, Olson MV. Chromosomal region of the cystic fibrosis gene in yeast artificial chromosomes: A model for human genome mapping. Science 250:94; 1990.

37. Kao FT, Yu JW. Chromosome microdissection and cloning in human genome and genetic disease analysis. Proc Natl Acad Sci USA 88:1844; 1991.

38. Southern EM. Genome mapping: cDNA approaches. Curr Opin Genet Develop 2:412; 1992.

39. Duyk GM, Kim S, Myers RM, Cox DR. Exon trapping: A genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. Proc Natl Acad Sci USA 87:8995; 1990.

40. Buckler AJ, et al. Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. Proc Natl Acad Sci USA 88:4005; 1991.

41. Orkin SH. Reverse genetics and human disease. Cell 47:845; 1986.

42. Caskey CT. "Presymptomatic genetic diagnosis - A worry for the United States" in Gene Mapping: Using Law and Ethics as Guides. GJ Annes, S Elias (eds). Oxford Univ Press, 1992; p. 177-185.

43. Torres VE, Holley KE, Offord KP. "General features of autosomal dominant polycystic kidney disease" in Problems in Diagnosis and Management of Polycystic Kidney Disease. JJ Grantham, KD Gardner (eds.) Polycystic Kidney Research Foundation, Kansas City, MO 1985; p. 49-69.

44. Dalgaard OZ, Norby S. Autosomal dominant polycystic kidney disease in the 1980's. Clinical Genetics 36:320; 1989.

45. Bear JC, McManamon P, Morgan J, Payne RH, Lewis H, Gault MH, Churchill DN. Age at clinical onset and at ultrasonographic detection of adult polycystic kidney disease: Data for genetic counselling. Am J Med Genetics 18:45; 1984.

46. Reeders ST, Breuning MH, Davies KE, Nicholls DR, Jarman AP, Higgs DR, Pearson PL, Weatherall DJ. A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. Nature 317:542; 1985.

47. Genetic Testing Committee to the Medical Section of the American Council of Life Insurance (1989). The potential role of genetic testing in risk classification. American Council of Life Insurance, Washington, DC.

48. Wiggins S, et al. The psychological consequences of predictive testing for Huntington's disease. NEJM 327:1401; 1992.

49. Touchette N. Letting the genome out of the bottle. J NIH Reas 2:17; 1990.

50. King JC. The Biology of Race, Harcourt Brace Jovanovich, Inc, NY, 1971.

51. Kevles DJ. In the Name of Eugenics: Genetics and Uses of Human Heredity. Alfred A. Knopg Publ, New York; 1985.

52. Edgar H. Sandomire H. Medical privacy issues in the age of AIDS: Legislative options. Am J Law and Med 16:160; 1990.

53. Annas GJ. The Rights of Patients, 2nd edition. Southern Illinois University Press, Carbondale, IL. 1989: p. 177.

54. Field MA. Testing for AIDS: Uses and abuses. Am J Law and Med 16:49; 1990.

55. Rennert S. AIDS/HIV and Confidentiality: Model Policy and Procedures. Am Bar Association, Washington, DC, 1991.

56. The discussion of the Tarasoff case is adapted from the discussion in Ruth Macklin, "HIV-infected psychiatric patients: Beyond confidentiality." Ethics and Behavior 1:3-20; 1991.

57. Technial Working Group on DNA Analysis Methods (TWGDAM). The combined DNA index system (CODIS): A theoretical model. FBI Laboratory, Quantico, VA, October 15, 1989, p. 284-286: and see National Research Council, DNA Technology in Forensice Science, National Academy Press, Washington, DC; 1992.

58. U.S. Congress, Office of Technology Assessment. Genetic Witness: Forensic Uses of DNA Tests. OTA-BA-438 U.S. Government Printing Office, Washington, DC, July 1990, p. 21.