New techniques for physical mapping of the human genome

PAUL R. BILLINGS,*,†,1 CASSANDRA L. SMITH,*,‡ AND CHARLES R. CANTOR*,‡

*Human Genome Center, Lawrence Berkeley Laboratory, Berkeley, California 94720, USA; †Department of Medicine, Pacific Presbyterian Medical Center, San Francisco, California 94115, USA; and ‡Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA

ABSTRACT **We describe improvements in techniques and strategies used for making maps of the human genome. The methods currently used are changing and evolving rapidly. Today's techniques can produce ordered arrays of DNA fragments and overlapping sets of DNA clones covering extensive genomic regions, but they are relatively slow and tedious. Methods under development will speed the process considerably. New developments include a range of applications of the polymerase chain reaction, enhanced procedures for high resolution in situ hybridization, and improved methods for generating, manipulating, and cloning large DNA fragments. More detailed genetic and physical maps will be useful for finding genes, including those associated with human diseases, long before the complete DNA sequence of the human genome is available.—Billings, P. R., Smith, C. L., and Cantor, C. R. New techniques for physical mapping of the human genome. *FASEB J.* 5: 28-34; 1991.**

Key Words: genetic analysis • mapping strategy • physical map • genomic DNA • phenotype • chromosomal region

THE ULTIMATE GOAL OF THE Human Genome Project is to determine the complete sequence of the DNA that comprises the human genome (1). During the next 15 years, this effort will require sequencing 3×10^9 base pairs (bp)[2] of DNA to finish the physical map of the human genome and obtain the genetic information contained in the 23 pairs of chromosomes found in normal nucleated human cells. The completed physical map can be expected to be useful in: *1*) understanding the organization and regulation of human genes, and elucidating their protein (or nonprotein) products; *2*) understanding human biology, especially the immensely complex development and physiology of the brain; *3*) locating the approximately 10^5 genes present in the human genome, of which many may be involved in clinical disorders; *4*) tracing the evolutionary origins of human sequences; and *5*) facilitating the interspecies comparative analysis of DNA sequences and the development of animal models for human conditions.

The completed physical map will simplify the genetic analysis of human diseases and provoke scientific projects considering the products of newly located genes and their role in biology and medicine. The processes required for completing the map of the human genome should lead to useful new technologies with far-reaching effects on biological sciences. Finally, in a project the size and scope of the Human Genome Project, important but unexpected or serendipitous benefits are likely to arise.

Although the complete DNA sequence of the human mitochondrial genome is available (2) and the number of cloned sequenced human DNAs is growing exponentially, a variety of factors (including cost, instrumentation, and computational considerations) make extensive random sequencing of the human genome unfeasible at present. Sequencing of mRNAs and genomic regions proximate to or at known disease-associated genes or sites of biological interest will proceed by using current methods and should become increasingly efficient. However, until dramatic advances in sequencing methods make genomic sequence analysis easier, refinements in chromosomal, genetic, and physical mapping will provide many of the early rewards of the Human Genome Project. **Figure 1** depicts different types of maps and their interrelationship.

GENERAL MAPPING STRATEGIES

Until recently, two general approaches termed top down and bottom up were used to describe the strategies used in mapping projects (3). Outlines of these two basic methods are illustrated in **Fig. 2.** Top down strategies involve the identification and ordering of progressively smaller fragments of DNA from a chromosome of interest (4). It is a satisfying approach in that it usually yields complete maps, as demonstrated in several relatively small genomes and large regions of mammalian genomes (5, 6). Unfortunately, top down-generated arrays of fragments or clones are not always useful for finding specific genes or generating long stretches of linear sequence.

Bottom up mapping builds up contigs (contiguous regions) of overlapping sets of DNA clones by identifying them in ways that allow one mapped site to be aligned with others. This approach can be enhanced with different walking or jumping protocols (techniques for characterizing DNA nearby an identified site) that are also used in top down schemes (3). Because this basic method always produces an ordered set of DNA clones, it is extremely useful for finding a gene localized to a small region [fewer than 2 megabases (Mb)] of DNA. Bottom up strategies are difficult to complete, especially over large regions of chromosomes, and are therefore not useful by themselves for making finished maps of genomes.

As the two major mapping strategies both have significant drawbacks, most investigators ultimately combine or blend these different approaches. For instance, a combination of top down and bottom up methods form the basis of positional or reverse genetics strategies for characterizing disease loci (7). Positional approaches have been used to clone the genes associated with Duchenne muscular dystrophy, cystic fibrosis, chronic granulomatous disease, neurofibromatosis, and others.

---

[1]To whom reprint requests should be sent.

[2]Abbreviations: bp, base pairs; Mb, megabase; PFG, pulsed-field gel electrophoresis; kb, kilobases; YAC, yeast artificial chromosomes; PCR, polymerase chain reaction; STS, sequenced tagged sites.

0892-6638/91/0005-28/$01.50.

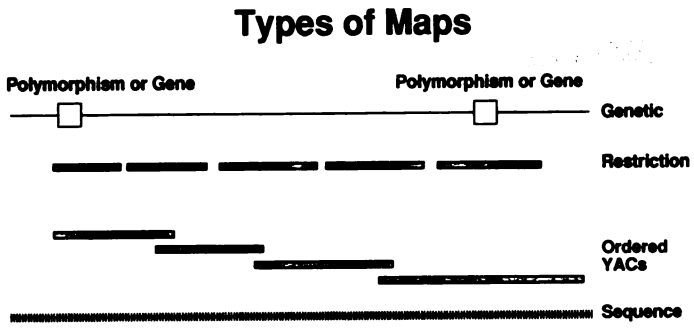

Figure 1. Four types of maps and their interrelationship. The ultimate physical map is bp sequence. Libraries of YAC containing clones can have average insert sizes of 200–400 kb. Restriction fragments 1–2 Mb in length can routinely be separated by PFG. A 1% recombination frequency in genetic mapping is denoted as 1 centimorgan (cm) and measures, on the average, 1 Mb. Not shown is chromosomal mapping which directly places genetic sites within or outside of bands (5 Mb) or as close as 100 kb apart using in situ techniques. This direct form of mapping links genetic, restriction, YAC based methods, and large cosmid strategies.

These general mapping strategies are now being augmented by rapid advances in direct chromosomal localization of DNA fragments by in situ hybridization. In addition, increasingly efficient DNA sequencing strategies (including automation) will spur production of detailed human physical genetic maps and allow more complete regional DNA sequencing.

Before discussing recent advances in the construction of genetic and physical maps in humans, several new but already well accepted and widely used techniques will be reviewed.

NEW TECHNOLOGY FOR MAKING PHYSICAL MAPS

Restriction enzymes are used to generate DNA fragments in all mapping, cloning, and sequencing protocols. Unfortunately, most naturally occurring restriction enzymes cut DNA so frequently that the number of resulting DNA fragments is too large to order readily for mapping. Furthermore, a variety of factors, particularly DNA methylation, makes restriction enzymes at times unusable. Recently, naturally occurring restriction enzymes that recognize 8-bp sequences and therefore cut genomic DNA less frequently have been identified (3, 8, 9). These enzymes along with certain others with 6-bp recognition sites share the fact that they recognize DNA containing one or two occurrences of the dinucleotide sequence CpG. This DNA motif is underrepresented in the human genome (10). It often occurs near expressed genes and is usually methylated. Examples of these restriction enzymes include *Not* I and *Mlu* I, which recognize the sequences GGCGGCCC and ACGCGT, respectively. Experimentally, these enzymes produce large fragments of cut human genomic DNA that are useful for mapping studies. Recently, new techniques have been described for cutting genomic DNA into large fragments (see below).

Once cut, DNA fragments are usually separated for analytic and preparative purposes. The ability to fractionate large DNA fragments by pulsed-field gel electrophoresis (PFG) has already improved mapping and cloning of large DNA molecules significantly (6, 11). PFG allowed research on fragments in the resolution gap between conventional electrophoretic separation methods [DNA fragments <40 kilobases (kb)] and cytogenetic positional studies (3–5 Mb of DNA). PFG can be extended to fractionate DNA molecules up to about 10 Mb in size (12). The development of PFG has meant that conventional and novel mapping strategies can be applied to large genomic regions. Although developments in direct visualization of DNA may provide useful alternatives and adjuncts to large DNA separation methods (see below), PFG separation will remain a technical mainstay in mapping approaches.

Special clones (termed linking clones) contain single rare cutting restriction enzyme recognition sites (e.g., for *Not* I). The use of these clones as probes for hybridization on chromosomal DNA digested with the same enzyme allows rapid ordering of large fragments separated by PFG methods. Since the first description of linking clones, a number of different protocols have been described for isolating this DNA (13, 14). Recently an oligonucleotide probe for the *Not* I recognition site was used for rapid screening of genomic libraries (15). A novel and simple enrichment protocol for direct selection of linking clones from a circular library has also been described (16).

Strategies employing partial as well as complete digestion of genomic DNA by restriction enzymes have also been used for ordering DNA fragments. A technique blending partial digestion by *Not* I with a linking clone method has been described (3). Improvements in the control of partial restriction enzyme digestion have been reported (17).

In addition, arrays of clones containing smaller DNA inserts (cosmids) have routinely been mapped by specifically identifying (fingerprinting) clones chosen at random or by hybridization to a known sequence. A number of different fingerprinting methods have been shown to be useful. The classical methods have used the pattern of restriction fragments or the occurrence of restriction sites (18, 19). These strategies have been applied successfully to the human genome (20). Other investigators have taken advantage of the occurrence of repetitive sequences that appear to be widely dispersed in human genomic DNA to fingerprint DNA present in genomic clones and align them (21). Although the presence of repetitive DNA sequences can be helpful in mapping, it often confounds cloning, walking, and sequencing experiments.

Aside from these tested approaches for cutting, separating, and ordering DNA for mapping purposes, another important development has been the improvement in methods for cloning large DNA segments. Until recently, the largest conventional cloning vectors were cosmids, which can accommodate DNA fragments up to 20–40 kb in length. Sub-

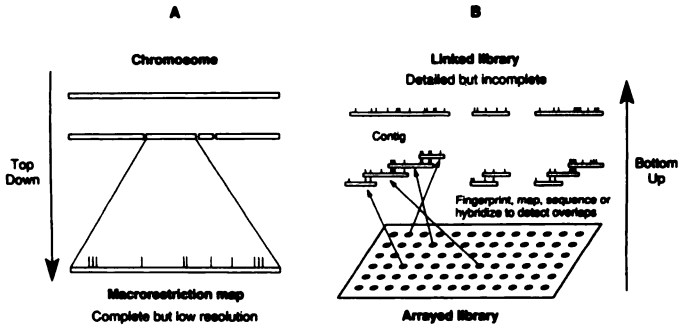

Figure 2. Two general schemes for physical mapping (3). Top down approaches yield complete maps but may fail to isolate pertinent cloned genes (*A*). Bottom up schemes generate sequence and clones, but are hard to finish (*B*). Strategies using a blend of both techniques are attractive.

sequently, it was noted that yeast could be transformed, albeit less efficiently than bacterial hosts, with artificially constructed chromosome vectors containing DNA fragments as large as 1 Mb (22). These vectors are maintained as independently segregating yeast artificial chromosomes (termed YACs). They allow cloning of the large DNA segments such as partial genomic digests with ordinary restriction enzymes or total digests generated by restriction enzymes that cut genomic DNA infrequently. Recently large fragment cloning systems have been tested using bacterial hosts (see below).

Finally, the polymerase chain reaction (PCR) has now been fully incorporated into mapping and recombinant DNA research. This method allows for the amplification of very small amounts of DNA (23). In the short time since its development, this technique has simplified protocols for probe production, cloning, DNA sequencing, and the identification of mapped sites. Some newer PCR-based methods are reviewed in subsequent sections of this article. Although problems with PCR methods have been noted (24) and other amplification methods have been described (25), the ability of PCR to produce large amounts of sequence for further analysis (even by nonradioactive techniques) is a major technical advance, the benefits of which are still accumulating.

IMPROVED CHROMOSOMAL AND GENETIC MAPPING

In this section, we will review recent advances in the construction of chromosomal and genetic maps. Chromosomal maps are produced by visually localizing a gene, DNA sequence, or phenotype to a particular chromosomal region. Genetic maps evolve from studying the segregation at meiosis of known genetic sites. Chromosomal mapping can utilize similar statistical methods to those used in classic genetic linkage studies (e.g., when human phenotypes are linked with specific inherited chromosomal variations or syndromes correlated with aneuploidy). Both chromosomal and genetic mapping may be used to locate genetic loci defined only by a phenotype.

Early in this century, chromosomal mapping began with the recognition of the coinheritence of hemophilia and colorblindness with gender. This observation indicated that these clinical phenotypes were genetic and localized the genes to chromosome X. Seventy years later, two completed chromosomal maps of the human genome were published with markers an average of 15–20 Mb apart (26, 27). These sites have been superimposed on the increasing detailed banding profiles of the 24 unique human chromosomes (28). At least a 10-fold increase in the density of mapped sites will be required for the completed human chromosomal maps to be useful for random DNA sequence localization.

Advances in in situ hybridization are rapidly enhancing the utility of direct visualization as a mode of chromosomal mapping. In situ hybridization to metaphase chromosomes has been feasible for several years. This technique allows orientation of DNA sequences which lie as close as 2–5 Mb apart (29, 30). Unfortunately, the compaction of metaphase chromatin usually prevents resolution of sites separated by smaller distances. However, in situ hybridization mapping information can be supplemented with information derived from chromosomal banding patterns. Since the average high-resolution band contains approximately 5 Mb of DNA, in situ hybridization improves the resolution of maps derived from chromosomal banding patterns. A four- to fivefold increase in banding refinement might allow chromosomal bands to be associated with specific cloned DNA segments such as YACs in the future. Similarly, improvements in in situ hybridization might allow placement of markers flanking a genetic site of interest on a single YAC. These developments could prove helpful in the analysis of syndromes and phenotypes associated with chromosomal aberrations.

Recently, modifications of in situ hybridization methods using interphase chromosome preparations have been reported (31). Interphase chromatin is not as compacted as metaphase chromatin. Thus, using interphase chromosomes can increase the resolution of two or more closely spaced sites. Using cosmid-sized DNA probes, experienced practitioners of interphase in situ hybridization can discriminate DNA sequences as close as 100 kb apart. Data from these direct observations can then be confirmed by physical or genetic mapping methods. This technique can quickly yield high-density maps using markers spaced every 3 Mb, and it is especially useful near the centromeres or telomeres of chromosomes. It should facilitate mapping YAC libraries and moving these analyses to cosmid-based direct DNA sequencing.

Given the ever enlarging catalog of cytogenetically defined sites that are associated with disorders, the direct analyses of DNA sequences from clinically implicated chromosomal regions should prove interesting. Several laboratories have reported the use of DNA subtraction and chromosomal microdissection techniques for generating DNA probes and sequences directly from specific chromosome regions (32–35). For instance, analysis of the region of chromosome 8 reported to be associated with Langer-Giedion syndrome has been shown to be facilitated by strategies using a regionally localized DNA probe production protocol (36). A number of refinements in DNA subtraction and microdissection will hasten the analysis of chromosomal syndromes involving insertions, deletions, and translocations.

Improvements in chromosomal mapping will be matched by genetic approaches. The fact that PCR can be used on single sperm or DNA molecules will have a major effect on genetic mapping experiments (37, 38). The reliability and power of genetic mapping will always depend on the number of meioses assessed. Because the results of millions of meioses are present in a single sperm sample, a semen collection can be used after PCR amplification to establish the order of genetic sites with recombination frequencies far less than 1%. A 1% recombination frequency corresponds to an average physical distance of about 1 Mb. Hence, this method will be particularly useful for confirming the order of probes mapped by in situ hybridization methods and for establishing a fine set of genetic markers in genomic regions with low recombination frequencies.

For instance, this approach might be used to estimate the distance between two DNA markers that flank a genetic site assayed only by a disease phenotype. Note that this method only samples male meioses, and the relationship between recombination frequencies and physical distance can exhibit sexual dimorphism. Also, single sperm PCR can only be used to analyze identified DNA markers, unlike other genetic mapping strategies. It is not helpful for directly studying gene loci known only by a disease phenotype. A similar limitation applies to the otherwise novel and powerful method for fine genetic mapping by mitotic recombination (39).

Finally, although the use of radiation-based techniques for genetic analysis has a long tradition, Cox and collaborators (40) have pioneered methods using radiation-induced chromosomal fragmentation and cell fusions to create panels of cells with varying human chromosomal components. In this form of genetic mapping the frequency with which genetic

sites remain together after fragmentation is a measure of the distance between them. This method is extremely useful for establishing the order and distance between DNA markers but does not allow the mapping of phenotypic traits, unless they can be assayed in the hybrid cell lines.

DISEASE GENE MAPPING

The advances in chromosomal and genetic mapping just reviewed should result in the rapid localization of most of the Mendelian traits and disorders of humans. A variety of computational, epidemiological approaches have been proposed to use maps and linkage strategies for locating more complex human traits (41, 42). In general, these methods compare the coinheritance of traits or markers in pairs of relatives (for instance, siblings) with that found in a random population sample. A variety of factors, including the occurrence of phenocopies and etiologic heterogeneity, may confound these new methods. In the past, genetic linkage strategies for dissecting complex human traits have only yielded a C3 polymorphism linkage in hypercholesterolemia and a chromosome 12 marker for familial osteoarthrosis (43, 44). The application of new approaches with the aid of more complete genomic maps should facilitate the mapping of complex human phenotypes and polygenic traits.

Disease gene mapping may also be propelled by the recognition that groups of families of genetic loci and chromosomal regions are conserved during evolution (45, 46). These conserved linkages define syntenic genomic regions. For instance, it appears that more than 80% of the linkage groups are syntenic between the laboratory mouse and humans (47). Thus, the localization of one member of a linkage group in both species will likely lead to mapping the entire family. Aside from expanding the mapping-derived information from the localization of a single marker, synteny may allow the identification of human disease-associated genetic sites in animals and may hasten the development of animal models of human disorders. Unfortunately, the phenotypes produced by identical genes in different species may not be the same (48). This could limit interspecies analysis of gene function for linked but uncharacterized genetic loci. It may also restrict transgenic approaches for mapping and analyzing sites defined only by species-specific phenotypic expression.

The extent of preservation of syntenic regions and linked gene families during biological evolution has not yet been fully determined. However, the inspection of genetic and physical maps for *Escherichia coli* (49), *Saccharomyces* (12), *Drosophilia* (50), and murine species (47), as well as for less traditional genetic model systems (51), should soon allow this question to be assessed. The result is likely to reinforce our relatedness to other coinhabitants of this planet and further call into question creationist explanations of the origins of human life.

NEWER TECHNIQUES FOR PHYSICAL MAPPING

As expected, the challenge posed by the immense task of mapping the human genome is generating novel methods and techniques for physical mapping. These will enhance and replace established methods for constructing physical maps and ordered sets of clones. These advances include better methods for generating and cloning large DNA fragments, improved fingerprinting strategies for ordering overlapping clones into contigs, and attempts at constructing totally new mapping paradigms.

Several new restriction endonucleases have been discovered in the past 2 years (8, 9, 52). It remains likely, however, that such enzymes will be rather rare and will never provide the same kind of powerful resource that four to six base-specific enzymes have offered. For this reason, alternative methods for specifically cutting DNA into large fragments have been studied. Several investigators report success in creating artificial restriction enzymes which can be designed to cut DNA into large fragments with ends of known sequence. These novel enzymes use oligonucleotides, oligopeptides, proteins, or DNA binding ligands to recognize DNA and nucleases, metal chelates or photoactive groups to cleave DNA after binding (53).

Another approach involves altering the specificity of existing restriction nucleases by modifying their recognition sites. Some success has been reported using methylases that partially overlap the sequence recognized by a particular restriction nuclease (54). A much more powerful technique utilizes methylation of all but pre-selected restriction enzyme cleavage sites by the cognate methylase—a methylating enzyme specific for the same DNA sequence as the particular restriction enzyme. Recognition site protection from methylation can be accomplished either by using triple helix formation to shield specific sites against the methylase or by using a DNA binding protein which binds at sequences that overlap a particular subset of restriction enzyme cleavage sites (55–57). The generation of large DNA fragments by new enzymes or digestion techniques may be a key development in linking mapping and direct sequencing strategies.

Since the introduction of YACs in cloning strategies, it has become clear that it is far easier to order a small number of large clones than a much larger number of small clones. An additional great advantage of YAC cloning over cosmid-based methods is that many YACs are large enough to span entire human genes. Recent developments in YAC cloning have resulted in libraries with average DNA inserts of more than 400 kb; this is considerably larger than the first YACs (23, 58, 59). A difficulty in using yeast cells for cloning is that the amount of cloned DNA per cell is relatively small because YACs exist as single copies in host cells. New methods for amplifying YACs up to 10 to 20 copies per cell promise to make procedures for screening yeast libraries much easier in the future (60).

Despite the success of YAC utilizing methods, it would also be useful to have available a large capacity bacterial cloning vector. Some progress has been made in developing F plasmids and P1 bacteriophage vectors in *E. coli* for this purpose (61, 62). The potential capacitiy of F vectors, judged by naturally occurring F episomes, is more than 2.0 Mb. It remains to be determined whether transformation procedures can be developed that will actually allow such large pieces of DNA to be introduced into *E. coli*.

Several methods have been reported which could improve conventional cloning methods of single copy sequences. Genomic subtraction (63) or genomic difference cloning (64) allows unique DNA sequences between two different DNA samples to be identified and cloned. In both of these protocols, biotin-streptavidin subtraction steps are combined with PCR amplification of the remaining DNAs. These strategies have the potential for more general application to problems including the analysis of genetic events such as chromosomal deletions, insertions, and translocations.

Aside from new restriction enzymes and more efficient cloning of large DNA fragments, other methods that show

promise for making mapping more efficient include improvements in fluorescence-activated sorting of chromosomes (65) and an extension of DNA sequence information that can be read from a single template. PCR can further improve the efficiency of DNA analysis (66). Not only can this method assist strategies using STS or STAR (see below), it can also be used as noted in a previous section to generate probes from microdissected chromosomal DNA. PCR-based strategies can speed chromosome walking, jumping, and linkage strategies (67, 68). New approaches may use biotinylated PCR primers which allow the capture of DNA by streptavidin bound to solid supports including beads. Although these methods are early in their development, they may speed the sequence analysis of large regions of DNA and could be automated.

A different approach for the production of ordered libraries has been suggested by techniques using hybridization with relatively short oligonucleotides (69). In principle, there is sufficient information in the fingerprint of oligonucleotide hybridizations to link up an entire genomic library de novo. The success of these methods in practice will depend on the degree of false negatives or false positives in such hybridizations. Preliminary results are encouraging but it is too early to tell whether this fingerprinting method will be able to supplant more traditional ordering approaches. A variation of this method has been proposed as a means to completely sequence genomic DNA (70).

In addition, new fingerprinting approaches for the construction of overlapping libraries have been attempted. They utilize DNA probes generated from recurrent DNA encoding protein motifs or gene regulatory regions. Repetitive DNA sequences can serve as PCR primer binding sites for amplifying single copy sequences. For example, probes have been generated between *Alu* and *Line* repeats, and these have been used for fingerprinting of clones and large DNA fragments (21, 35, 71–73).

The development of reliable semi-automated DNA sequencing instruments has prompted suggestions of how to replace DNA mapping by sequencing, or at least how to use these instruments to facilitate mapping. Automated fluorescent gel readers are already in use to assist clone fingerprinting in cosmid contig construction (74). One widely discussed general protocol which would lend itself to automation proposes determining a small section of DNA sequence on every DNA probe of interest, and then use of this sequence to construct PCR primers which can be used to replace hybridization for all subsequent mapping procedures (75). This approach, termed sequenced tagged sites (STS), has appeal but it is not yet clear how to use these bits of sequence information optimally for map construction.

One potentially important modification of the STS approach is to sequence just the overlapping areas of two otherwise complementary clone sets. For example, one could sequence a set of linking clones for an enzyme like *Not* I (a group in which all the clones contain a *Not* I cleavage site) and a corresponding set of jumping clones (a group of clones whose inserts consist of the two ends of individual DNAs produced from *Not* I cleaved, PFG separated fragments which then are ligated together with intervening sequence removed) to generate sequence tagged rare restriction sites (STAR) (36). Overlap between the sequence data would allow the two sets of clones to be placed in order; each linking clone inserted between two jumping clones and vice versa. Variants of STS-like strategies have been discussed. Which of these methods will be most useful needs further experimentation.

CONCLUSIONS

Although the physical mapping and DNA sequencing of the human genome will not be completed for many years, it is already clear that the task will generate a wide variety of novel approaches and techniques as well as refinements of older methods. An endeavor which demands the identification of 10^5 or more genes, the precise mapping and sequencing of clones containing 3×10^9 bp of DNA, and possibly the analysis of all human DNA differences which could be more than 1.5×10^6 (about 3×10^6 DNA variations per person $\times$ 5×10^9 people) is a daunting and potentially tedious undertaking. Yet the initial efforts at physical mapping have been characterized by creativity, increasing sophistication, and recognition of the importance of simplification for purposes of eventual automation. But most significantly, they also demonstrate that mapping is a healthy science, rapidly evolving and changing to better generate useful data.

It is likely that the conduct of physical mapping 5 or 10 years from now will resemble current methods only in that the importance of the data produced will require the considered analyses of investigators and review by the scientific community. The effect of an ever more detailed map and the technologies its pursuit fosters is already evident. Techniques and strategies reviewed here and others not yet conceived will simplify physical mapping in the future. It is hoped that the explosion in interest and creativity which characterizes the current scene in the mapping and sequencing of human DNA will in turn generate an equally impressive burst of insight from those investigators who must show how genetic information modifies biological processes and is best applied in human societies. FJ

This work was supported in part by a grant from the U.S. Department of Energy (DE-F602-87ER-GD582).

REFERENCES

1. Understanding our genetic inheritance: the U.S. Human Genome Project: the first five years FY 1991–1995 (1990) *U.S. Dept. of Energy, Ofc. of Energy Research,* p. 5
2. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R., and Young, I. G. (1981) Sequence and organization of the human mitochondrial genome. *Nature (London)* **290,** 457–465
3. Oliva, R., Lawrance, S. K., Wu, Y., and Smith, C. L. (1990) Human chromosomes: molecular studies. In *Encyclopedia of Human Biology* (Dulbecco, R., et al., eds) Academic, San Diego. In press
4. Cantor, C. R., and Smith, C. L. (1989) Construction and applications of physical maps of the human genome. *Jikken Igaku* **8,** 15–22
5. Smith, C. L., and Condemine, G. (1990) New approaches for physical mapping in small genomes. *J. Bacteriol.* **172,** 1167–1172
6. Oliva, R., Fong, H., and Smith, C. L. (1990) Physical maps: pulsed field gel electrophoresis. In *Mapping the Human Genome: Methodology and Application to Human Disease* (Naylor, S., ed) Academic, San Diego. In press
7. Orkin, S. (1986) Reverse genetics in human disease. *Cell* **47,** 847–850
8. Nelson, J. M., Miceli, S. M., Lechevalier, M. P., and Roberts, R. J. (1990) *FseI,* a new type II restriction endonuclease that recognizes the octanucleotide sequence 5' GGCCGGCC 3'. *Nucleic Acids Res.* **18,** 2061–2064
9. Tautz, N., Kaluza, K. L., Frey, B., Jarsch, M., Schmitz, G. G.,

and Kessler, C. (1990) *Sgr*AI, a novel class-II restriction endonuclease from *Streptomyces griseus* recognizing the octanonucleotide sequence 5′-CR/CCGGYA-3′. *Nucleic Acids Res.* **18,** 3087
10. Sved, J., and Bird, A. (1990) The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. *Proc. Natl. Acad. Sci. USA* **87,** 4692-4696
11. Cantor, C. R., Smith, C. L., and Mathew, M. K. (1988) Pulsed-field gel electrophoresis of very large DNA molecules. In *Annual Review of Biophysics and Biophysical Chemistry* (Engelman, D. M., Cantor, C. R., and Pollard, T. D., eds) p. 287, Annual Reviews Inc., Palo Alto
12. Fan, J.-B., Chikashige, Y., Smith, C. L., Niwa, O., Yanagida, M., and Cantor, C. R. (1989) Construction of a *Not* I restriction map of the fission yeast *Schizosaccharomyces pombe* genome. *Nucleic Acids Res.* **12,** 2801-2817
13. Smith, C. L., Warburton, P., Gaal, A., and Cantor, C. R. (1986) Analysis of genome organization and rearrangements by pulsed field gradient gel electrophoresis. In *Genetic Engineering* (Setlow, J. K., and Hollaender, A., eds) Vol. 8, pp. 45-70, Plenum, New York
14. Smith, C. L., Lawrence, S. K., Gillespie, G. A., Cantor, C. R., Weissman, S. M., and Collins, F. S. (1987) Strategies for mapping and cloning macro-regions of mammalian genomes. In *Methods in Enzymology, Molecular Genetics of Mammalian Cell* (Gottesman, M., ed) Vol. 151, pp. 461-489, Academic, New York
15. Melmer, G., and Buchwald, M. (1990) Laboratory methods: use of short oligonucleotides to screen cosmid libraries for clones containing G/C-rich sequences. *DNA and Cell Biol.* **9,** 377-385
16. Ito, T., and Sakaki, Y. (1988) A novel procedure for selective cloning of *Not* I linking fragments from mammalian genomes. *Nucleic Acids Res.* **16,** 9177-9184
17. Hanish, J., and McClelland, M. (1990) Methylase-limited partial *Not* I cleavage for physical mapping of genomic DNA. *Nucleic Acids Res.* **18,** 3287-3291
18. Coulson, A., Sulston, J., Brenner, S., and Karn, J. (1986) Toward a physical map of the genome of the nematode *Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA* **83,** 7821-7825
19. Kohara, Y., Akiyama, K., and Isono, K. (1987) The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50,** 495-508
20. Evans, G. A., and Lewis, K. A. (1989) Physical mapping of complex genomes by cosmid multiplex analysis. *Proc. Natl. Acad. Sci. USA* **86,** 5030-5034
21. Stallings, R. L., Torney, D. C., Hildebrand, C. E., Longmire, J. L., Deaven, L. L., Jett, J. H., Doggett, N. A., and Moyzis, R. K. (1990) Physical mapping of human chromosomes by repetitive sequence fingerprinting. *Proc. Natl. Acad. Sci. USA* **87,** 6218-6222
22. Burke, D. T., Carle, G. F., and Olson, M. V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236,** 806-812
23. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharff, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239,** 487-492
24. Keohavong, P., and Thilly, W. G. (1989) Fidelity of DNA polymerases in DNA amplification. *Proc. Natl. Acad. Sci. USA* **86,** 9253-9257
25. Van Brunt, J. (1990) Amplifying genes: PCR and its alternatives. *Biotechnology* **8,** 291-301
26. Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T. P., Bowden, D. W., Smith, D. R., Lauder, E. S., et al. (1987) A genetic linkage map of the human genome. *Cell* **51,** 319-337
27. White, R., and Lalouel, J.-M. (1988) Sets of linked genetic markers for human chromosomes. *Annu. Rev. Genet.* **22,** 259-279
28. Yunis, J. J. (1981) Mid-prophase human chromosomes. The attainment of 2000 bands. *Hum. Genet.* **56,** 293-298
29. Albertson, D. B., Fishpool, R., Sherrington, P., Nacheva, E., and Milstein, C. (1988) Sensitive and high resolution in situ hybridization to human chromosomes using biotin labelled probes: assignment of the human thymocyte CD 1 antigen genes to chromosome 1. *EMBO J.* **7,** 2801-2806
30. Bhatt, B., Burns, J., Flannery, D., and McGee, J. O'D. (1988) Direct visualization of single copy genes on banded metaphase chromosomes by nonisotopic in situ hybridization. *Nucleic Acids Res.* **16,** 3951-3961
31. Lichter, P., Cremer, T., Tang, C. J., Watkins, P. C., Manuelides, L., and Ward, D. C. (1988) Rapid detection of human chromosome 21 aberrations by in situ hybridization. *Proc. Natl. Acad. Sci. USA* **85,** 9664-9668
32. Saunders, R. D. C., Glover, D. M., Ashburner, M., Siden-Klamos, I., Louis, C., Monastirioti, M., Savakis, C., and Kafatos, F. (1989) PCR amplification of DNA microdissected from a single polytene chromosome band: a comparison with conventional microcloning. *Nucleic Acids Res.* **17,** 9027-9037
33. Johnson, D. H. (1990) Molecular cloning of DNA from specific chromosomal regions by microdissection and sequence-independent amplification of DNA. *Genomics* **6,** 243-251
34. Wesley, C. S., Ben, M., Kreitman, M., Hagag, N., and Easnes, W. F. (1990) Cloning regions of the *Drosophila* genome by microdissection of polytene chromosomes DNA and PCR with nonspecific primer. *Nucleic Acids Res.* **18,** 599-603
35. Ledbetter, S. A., Nelson, D. L., Warren, S. T., and Ledbetter, D. H. (1990) Rapid isolation of DNA probes within specific chromosome regions by interspersed repetitive sequence polymerase chain reaction. *Genomics* **6,** 475-481
36. Ludecke, H.-J., Senger, G., Calussen, U., and Horsthemke, B. (1989) Cloning defined regions of the human genome by microdissection of banded chromosomes and enzymatic amplification. *Nature (London)* **338,** 348-350
37. Li, H., Cui, X., and Arnheim, N. (1990) Direct electrophoresis detection of the allelic state of single DNA molecules in human sperm by using the polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* **87,** 4580-4584
38. Ruano, G., Kidd, K. K., and Stephens, J. C. (1990) Haplotype of multiple polymorphisms resolved by enzymatic amplification of single DNA molecules. *Proc. Natl. Acad. Sci. USA* **87,** 6296-6300
39. Morley, A. A., Grist, S. A., Turner, D. R., Kutlaca, A., and Bennett, G. (1990) Molecular nature of in vivo mutations in human cells at the autosomal HLA-A. *Cancer Res.* **50,** 4584-4587
40. Cox, D. R., Burmeister, M., Price, E. R., Kim, S., and Myers, R. M. (1990) Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* **250,** 245-250
41. Lander, E. S., and Botstein, D. (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121,** 185-199
42. Ott, J. (1990) Invited editorial: cutting a Gordian knot in the linkage analysis of complex human traits. *Am. J. Hum. Genet.* **46,** 219-221
43. Ott, J., Schrott, H. G., Goldstein, J. L. j., Hazzard, W. R., Allen, F. H., Falk, C. T., and Motulsky, A. G. (1974) Linkage study in a large kindred with familial hypercholesterolemia. *Am. J. Hum. Genet.* **26,** 598-603
44. Palotie, A., Väisänen, P., Ott, J., Ryhänen, L., Elima, K., Vikkula, M., Cheah, K., Vuorio, E., and Peltonen, L. (1989) Predisposition to familial osteoarthrosis linked to type II collagen gene. *Lancet* **1,** 924-927
45. Nadeau, J. H. (1989) Maps of linkage and synteny homologies between mouse and man. *Trends Genet.* **5,** 82-86
46. Sawyer, J. R., and Hozier, J. C. (1986) High resolution of mouse chromosomes: banding conservation between man and mouse. *Science* **232,** 1632-1635
47. Searle, A. G., Peters, J., Lyon, M. F., Evans, E. P., Edwards, J. H., and Buckle, V. J. (1987) Chromosome maps of man and mouse, III. *Genomics* **1,** 3-18
48. Erickson, R. P. (1989) Why isn't a mouse more like a man? *Trends Genet.* **5,** 1-3
49. Smith, C. L., Econome, J., Schutt, A., Klco, S., and Cantor, C. R. (1987) A physical map of the *E. coli* K12 genome. *Science* **236,** 1448-1453
50. Garza, D., Ajioka, J. W., Burke, D. T., and Hartl, D. L. (1989) Mapping the *Drosophila* genome with yeast artificial chromosomes. *Science* **246,** 641-646
51. Kuspa, A., Vollrath, D., Cheng, Y., and Kaiser, D. (1989) Physi-

cal mapping of the *Myxococcus xanthus* genome by random cloning in yeast artificial chromosomes. *Proc. Natl. Acad. Sci. USA* **86**, 8917-8921
52. Waterbury, P. G., Rehfuss, R. P., Carroll, W. T., Smardon, A. M., Faldadz, B. D., Huckaby, C. S., and Lane, M. J. (1989) Specific cleavage of the yeast genome at 4'-ATCGATCGAT-3'. *Nucleic Acids Res.* **17**, 9493
53. Hélène, C., Thuong, N. T., Saison-Behmoaras, T., and François, J.-C. (1989) Sequence-specific artificial endonucleases. *Tibtech* **7**, 310-315
54. Patel, Y., Van Cott, E., Wilson, G. G., and McClelland, M. (1989) Cleavage at the twelve-base-pair sequence 5'-TCTAGATCTADA-3' using M · *Xba*I (TCTAG^{m6}A) methylation and *Dpn*I (G^{m6}A/TC) cleavage. *Nucleic Acids Res.* **18**, 1603-1607
55. Hanvey, J. C., Shimizu, M., and Wells, R. D. (1989) Site-specific inhibition of EcoRI restriction/modification enzymes by a DNA triple helix. *Nucleic Acids Res.* **18**, 157-161
56. Maher, L. J., Wold, B., and Dervan, P. B. (1989) Inhibition of DNA binding proteins by oligonucleotide-directed triple helix formation. *Science* **245**, 725-730
57. Koob, M., and Szybalski, W. (1990) Cleaving yeast and *Escherichia coli* genomes at a single site. *Science* **250**, 271-273
58. Imai, T., and Olson, M. V. (1990) Second-generation approach to the construction of yeast artificial-chromosome libraries. *Genomics* **8**, 297-303
59. Albertsen, H. M., Abderrahim, H., Cann, H. M., Dausset, J., Le Paliser, D., and Cohen, D. (1990) Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proc. Natl. Acad. Sci. USA* **87**, 4256-4260
60. Smith, D. R., Smyth, A. P., and Moir, D. T. (1990) Amplification of large artificial chromosomes. *Proc. Natl. Acad. Sci. USA* **87**, 8242-8246
61. O'Connor, M., Pfeifer, M., and Bender, W. (1989) Construction of large DNA segments in *Escherichia coli*. *Science* **244**, 1307-1312
62. Sternberg, N., Ruether, J., and deRiel, K. (1990) Generation of a 50,000-member human DNA library with an average DNA insert size of 75-100 kbp in a bacteriophage P1 cloning vector. *New Biologist* **2**, 151-162
63. Straus, D., and Ausubel, F. M. (1990) Genomic subtraction for cloning DNA corresponding to deletion mutations. *Proc. Natl. Acad. Sci. USA* **87**, 1889-1893
64. Wieland, I., Bolger, G., Asouline, G., and Wigler, M. (1990) A method for difference cloning: gene amplification following subtractive hybridization. *Proc. Natl. Acad. Sci USA* **87**, 2720-2724
65. Minoshima, S., Kawasaki, K., Fukuyama, R., Maehawa, M., Kudoh, J., and Shimizu, N. (1990) Isolation of giant DNA fragments from flow-sorted human chromosomes. *Cytometry* **11**, 539-546
66. Schwarz, K., Hansen-Hagge, T., and Bartram, C. (1990) Improved yields of long PCR products using gene 32 protein. *Nucleic Acids Res.* **18**, 1079
67. Rosenthal, A., and Jones, D. S. C. (1990) Genomic walking and sequencing by oligo-cassette mediated polymerase chain reaction. *Nucleic Acids Res.* **18**, 3095-3096
68. Kandpal, R. P., Shukla, H., Ward, D. C., and Weissman, S. M. (1990) A polymerase chain reaction approach for constructing jumping and linking libraries. *Nucleic Acids Res.* **18**, 3081
69. Craig, A. G., Nizetic, D., Hoheisel, J. D., Zehetner, G., and Lehrach, H. (1990) Ordering of cosmid clones covering the Herpes simplex virus type I (HVS-I) genome: a test case for fingerprinting by hybridization. *Nucleic Acids Res.* **18**, 2653-2660
70. Drmanac, R., Labat, I., Bruckner, I., and Crkvenjakov, R. (1989) Sequencing of megabase plus DNA by hybridization: theory of the method. *Genomics* **4**, 114-128
71. Cotter, F. E., Hampton, G. M., Nasipuri, S., Bodmer, W. F., and Young, B. D. (1990) Rapid isolation of human chromosome-specific DNA probes from a somatic cell hybrid. *Genomics* **7**, 257-263
72. Zischler, H., Nanda, I., Schafer, R., Schmid, M., and Epplen, J. T. (1989) Digoxigenated oligonucleotide probes specific for simple repeats in DNA fingerprinting and hybridization in situ. *Hum. Genet.* **82**, 227-233
73. Lichter, P., Ledbetter, S. A., Ledbetter, D. H., and Ward, D. C. (1990) Fluorescence in situ hybridization with *Alu* and L1 polymerase chain reaction probes for rapid characterization of human chromosomes in hybrid cell lines. *Proc. Natl. Acad. Sci. USA* **87**, 6634-6638
74. Carrano, A. V., Lamerdin, J., Ashworth, L. K., Watkins, B., Branscomb, E., Slezak, T., Raff, M., De Jong, P. J., Keith, D., McBride, L., Meister, S., and Krinock, M. (1989) A high-resolution, fluorescence-based, semiautomated method for DNA fingerprinting. *Genomics* **4**, 129-136
75. Olson, M., Hood, L., Cantor, C., and Botstein, D. (1989) A common language for physical mapping of the human genome. *Science* **245**, 1434