Human Chromosomes, Physical Maps of

Nikolay Bukanov, Denan Wang, and Cassandra L. Smith

1 Introduction

2 Genomic Mapping Methods and Terms

3 Types of Physical Maps

4 Genomic Restriction Map Construction

5 Construction of Overlapping Libraries and Genomic Restriction Maps

6 Prospectus

Key Words

***Alu* Repeat** A 300 base pair (bp) DNA interspersed repeat sequence that occurs about once every 5000 bp. The most common repeat element in the human genome, it is CG rich and appears to occur preferentially in light Giemsa bands on condensed chromosomes.

Comparative Genomic Hybridization (CGH) A molecular cytogenetic method capable of detecting and locating relative genomic sequence copy number differences between pairs of DNA samples. Equal concentrations of differentially labeled DNAs are hybridized simultaneously to metaphase chromosome spreads. Regions with deleted or amplified DNA sequences are seen as changes in the ratio of the intensities of these two labels along the target chromosome.

Expressed Sequence Tag (EST) A segment of a sequence from cDNA clone that corresponds to an mRNA. The mapping ESTs helps make maps and points directly to expressed genes.

Fluorescence In Situ Hybridization (FISH) Hybridization of nucleic acid probes to chromosomes immobilized on microscope slides or filters.

Inter-*Alu* PCR A polymerase chain reaction method that uses primers contained in *Alu* repeat element to amplify single-copy sequences between adjacent *Alu* elements.

***Kpn* Repeat** The second most commonly occurring repeat sequence in the human genome, occurring on average once every 50,000 bp. It is a relatively AT-rich, LINE repeat and occurs preferentially in dark Giemsa bands on condensed chromosomes.

Linking Clone A clone containing DNA sequences that span a restriction enzyme recognition site.

Long Interspersed Repeat Elements (LINEs) Long interspersed repeats that appear to be similar to retroposons.

P1-Derived Artificial Chromosomes (PAC) An *E. coli* bacteriophage cloning system used for cloning of fragments up to 150 kilobases (kb) in size.

Polymerase Chain Reaction (PCR) A method for amplifying DNA by alternatively denaturing double-stranded DNA, annealing pairs of primers located near each other on complementary strands, and synthesizing the DNA between the primers using DNA polymerase.

Polymorphism Link-up A mapping approach that establishes continuity between restriction fragments by taking advantage of the naturally occurring polymorphism in different DNAs. In some cell lines hybridized probes appear to identify different fragments, whereas in others they appear to identify the same fragment. The pattern of occurrence of these fragments can be used to assess whether the probes identify the same or adjacent fragments.

Pulsed Field Gel (PFG) Electrophoresis A method of electrophoresis that exposes nucleic acids to alternating electrical fields. Fractionation is based on the speed at which the molecules can change directions.

RARE (RecA-Assisted Restriction Endonuclease) Cleavage A site-specific DNA cleavage method based on the ability of

RecA protein from *E. coli* to pair an oligonucleotide to its homologous sequence in duplex DNA. This three-stranded DNA complex is protected from methylase. After methylation and removal of RecA, restriction endonuclease cleavage is limited to the site previously protected from methylation. If pairs of oligonucleotides are used, a specific fragment can be cleaved out of genomes.

Retroposons A group of DNA sequence elements that appear to transpose through an RNA intermediate. Retroposon elements do not code for reverse transcriptase, do not have terminally redundant sequences, and do have a 3′-poly(A)$_n$ stretch. A variable-sized target duplication occurs at the site of integration.

Sequence-Tagged Restriction Site (STAR) A short DNA sequence used to identify the DNA surrounding a restriction nuclease cleavage site.

Sequence-Tagged Site (STS) A short DNA sequence used to identify a DNA segment.

Short Interspersed Repeat Elements (SINEs) Repeated sequences less than 500 bp long and present in a high copy number (10^5 per human genome). The *Alu* repeat element consists of SINEs.

Yeast Artificial Chromosome (YAC) An artificial yeast chromosome constructed by cloning genomic fragments into vectors that can replicate in yeast. YACs have the following characteristics: a yeast centromere, two telomeric sequences, and a selectable marker.

This entry reviews the types of physical map that can be constructed and the methods used to construct them. In particular, the emphasis is on the construction of genomic restriction maps and ordering overlapping libraries using top-down mapping approaches. The same approach allows the use of genomic DNA in functional studies, as well, and this area is explored also.

1 INTRODUCTION

In the past most genome studies were limited to organisms having well-developed genetic systems. Thus, a few model systems were studied very intensely (e.g., bacteriophage lambda, *E. coli, Drosophila,* mice) in a large number of different ways. The accumulation of diverse knowledge on these model organisms further promoted use of the few model systems. Now a number of molecular techniques have been developed that allow the construction of physical maps for virtually any chromosome. This means that biological analysis of an uncharacterized organism can begin with a physical dissection. Knowledge gained from this initial foray provides the foundation for functional studies. Hence, a new synergism in biology is provided by bottom-up genome studies.

Genomic physical mapping experiments were facilitated by the development of pulsed field gel (PFG) electrophoresis and accompanying techniques. These techniques allow the isolation, characterization, and manipulation of large pieces of DNA, including intact chromosomes from virtually any organism. For instance, intact megabase (Mb) chromosomes may be analyzed directly; larger chromosomes are analyzed after cleavage into specific megabase

pieces, or after cloning as megabase segments. Thus far, the size fractionation limit of this technique is about 10 Mb.

Bacterial genomes are small and range from approximately 1 to 15 Mb in size, reflecting, in part, the relative ability of these organisms to be free-living. These genomes are ideal sizes for PFG analysis, as evidenced by the large number of complete, low resolution, genomic restriction maps for microorganisms that continue to be published. Overlapping clone libraries have also been constructed for several small genomes. One bacterial genome (1.8 Mg in size) has now been completely sequenced, as well as a number of the smaller *Saccharomyces cerevisiae* chromosomes.

The study of some lower eukaryotic genome such as protozoa was hampered not only by the lack of well-developed genetic systems, but also by the failure of the chromosomes to condense during cell division. Thus, it has been particularly useful to apply PFG electrophoresis to the analysis of protozoan genomes, which range from a size overlapping that of large bacterial genomes, to about 100 Mb. Surprisingly, one parasitic protozoan genome, the *Giardia lamblia* genome, was found to be only about 12 Mb in size (i.e., smaller than the *Saccharomyces* genome).

Large, complex genomes, like the human genome (estimated to be 3000 Mb in size) have, until recently, been particularly recalcitrant to molecular dissection. Mammalian chromosomes (estimated to be 50–300 Mb in size) condense; hence, they can be seen in the microscope. A method that allows a finer division of these large genomes is the differential staining of the condensed chromosomes into smaller regions (estimated at 5–10 Mb in size) that are presumed to reflect regional differences in GC content. The division of the genome into chromosomes and chromosomal bands provides convenient pieces for top-down mapping approaches.

In top-down mapping, the genome is divided into units to facilitate study. For instance, the natural division of the genome into chromosomes could represent one such division and chromosomal banding another. Further division would depend on the method of analysis. Conventional recombinant DNA approaches to genome analysis allowed the characterization of molecules up to about only 0.05 Mb. The "resolution gap," 0.05–10 Mg, was exactly the size range that was most amenable to study by PFG techniques. Thus, the entire size range of DNA molecules can now be analyzed. This allowed the extension of recombinant DNA methods to large cloning system such as yeast artificial chromosomes (YACs) and P1 artificial chromosomes (PACs).

The application of PFG analysis to the human genome has been particularly useful for positional cloning of disease genes. Here, genetic analysis allows the region of the genome containing a gene of interest to be narrowed to 1–10 Mb. Then, PFG analysis can determine the size of the region and the number of putative gene candidates it contains. The goal of physical mapping experiments is to facilitate the isolation of more genetic markers and to isolate candidate genes.

2 GENOMIC MAPPING METHODS AND TERMS

Many molecular techniques may be applied with varying degrees of accuracy to the characterization of chromosomes. Hence, there is no unifying concept of what constitutes a chromosomal physical map. In some cases this lack has led to ambiguities in the literature with respect to the definition of a genetic map.

Any map will consist of markers or objects. The order, and per-

haps the distance, between pairs of objects will be known (some might argue that a map must consist of both order and distance). Object order along a chromosome should be maintained irrespective of the method used to construct the map. However, map distances will be method dependent. For instance, the amount of recombination along a chromosome is not constant. Hence, it is not surprising that a comparison of the physical and genetic distances along the long arm of chromosome 21 revealed at least a sixfold variation in the distances. It is quite clear that the ultimate map is the entire sequence of a chromosomal DNA. All maps and objects will be anchored to the DNA sequence once it is available. However, many biological map objects (e.g., function) are imprecisely defined or have imprecisely defined locations by their very nature.

Classically, a genetic map was composed of chromosomal loci and the amount of recombination between the loci was the genetic distance. For eukaryotic organisms, genetic distance, expressed in centimorgans (cM), is a measure of the coinheritance of genetic markers. This type of analysis requires the examination of several family generations. In contrast to bacterial like *E. coli,* recombination is measured as the time of transfer and integration of DNA from one cell into the chromosome of another cell.

In some instances, the placement of genes on various physical maps has been referred to as a genetic map (i.e., "a gene map"), although only physical distances and locations were known. Cytogenetic maps lead to further confusion in terminology. In these maps the presence or absence of a gene or map object is correlated with an observable genomic location (i.e., a chromosome band).

Here, a physical map is considered to be any map consisting of objects that have been located by physical rather than by genetic methods. Thus, a physical map can consist of objects located along the chromosome, such as a chromosomal band, a breakpoint, a genomic restriction fragment, and the location of a clone on a chromosome. We are concerned only with the maps made by ordering restriction fragments and clones. These maps are constructed using molecular methods.

Another concept whose use results in some confusion is "locus." Classically this term defined genes (i.e., "genetic locus"). More generally, however, a chromosomal locus, representing a location on a chromosome, can consist of objects other than genes. Examples include DNA sequences sequenced tagged sites (STS), expressed-sequenced tags (ESTs), probe sites, restriction enzyme sites, clone sites, centromeres, telomeres, and chromosomal breakpoints (such as those that occur naturally or are induced by ultraviolet breakage). To add to the confusion, new genetic markers based on anonymous DNA sequences now define genetic loci because they are used in genetic mapping experiments. In this entry, a *chromosomal locus* is any location on a chromosome that can be identified in a distinctive manner.

3 TYPES OF PHYSICAL MAPS

The term "physical maps" has been used to describe both genomic restriction maps and overlapping libraries. For instance, low resolution, genomic restriction maps may be created by hybridization experiments using cloned sequences as probes to order large restriction fragments generated by enzymes that cleave genomic DNA infrequently. Genomic restriction maps may also be created by analyzing restriction sites contained on overlapping clones. The resulting restriction map is of low, or high, resolution depending on the restriction enzyme used and the frequency at which its recogni-

tion site occurs within the genomic DNA. In some cases this information is obtained during the ordering of the library, or it may be generated afterward (see Section 5). Furthermore, an overlapping library may in itself represent a map consisting of ordered objects (clones) whose size can be approximated but not stated with certainty.

4 GENOMIC RESTRICTION MAP CONSTRUCTION

The first step in creating a genomic restriction map is choosing the DNA source. For many organisms or chromosomes this is obvious, since there is a well-characterized isolate or a cell line that may be useful. For the human genome, for instance, some complications associated with the analysis of polymorphic diploid DNA can be avoided by using DNA from hybrid cell lines containing haploid human chromosomes.

Usually, the genomic DNA is extracted and purified intact, in agarose, to prevent shear damage. Small ($<$ 10 Mb) chromosomal DNAs may be sized directly by PFG techniques. Other chromosomal DNAs must be cleaved with a restriction enzyme before they are subjected to PFG analysis. The largest size standard for PFG electrophoresis is 6 Mb. Thus, analytical methods use this limit, although enhanced versions of the technique allow the fractionation of molecules greater than 6 Mb.

Usually maps are constructed using restriction enzymes that have large recognition sequences or a recognition site that occurs infrequently in the genome of interest. In some cases, it has been useful to test a battery of enzymes. The usefulness of a particular enzyme may be estimated roughly from the size of the site or the GC content of the test organism. Even so, these predictions are somewhat inaccurate because the genomic DNA sequences do not occur at random, whereas calculated occurrences assume randomness. Furthermore, the frequency of occurrence in cloned sequences may not be representative, inasmuch as most past molecular studies have focused on genes.

Genomic map construction is similar to putting a puzzle together. It is much easier if all the pieces are known in advance. For small genomes this usually involves finding convenient restriction enzymes that cleave the genome into a reasonable number of fragments that can be resolved by PFG analysis and visualized by simple ethidium bromide staining. A method was also developed for visualizing all the megabase restriction fragments for individual human chromosomes. This approach analyzes PFG-fractionated restriction fragments from genomic DNA obtained from monosomic hybrid cell lines. Here, human-specific interspersed repetitive hybridization probes are used to identify the human megabase restriction fragments (Figure 1). This approach works as long as the restriction enzyme digestion goes to completion. Otherwise, partial digestion products obscure the complete digestion products. For the human genome the restriction enzymes *Not*I, *Sgr*AI, *Fse*I, and *Asc*I have this desired characteristic and cut the genome into fragments that appear to average 1 Mb in size. In mammalian genomes, partial cleavage is observed, and many restriction enzymes cleave infrequently because partial CpG methylation in their recognition sites inhibits total cleavage.

The most commonly occurring interspersed repeat in the human genome is *Alu,* a short interspersed repeating element (SINE) that is estimated to occur about every 5000 base pairs (bp). The second most commonly occurring interspersed human repeat is the *Kpn* re-

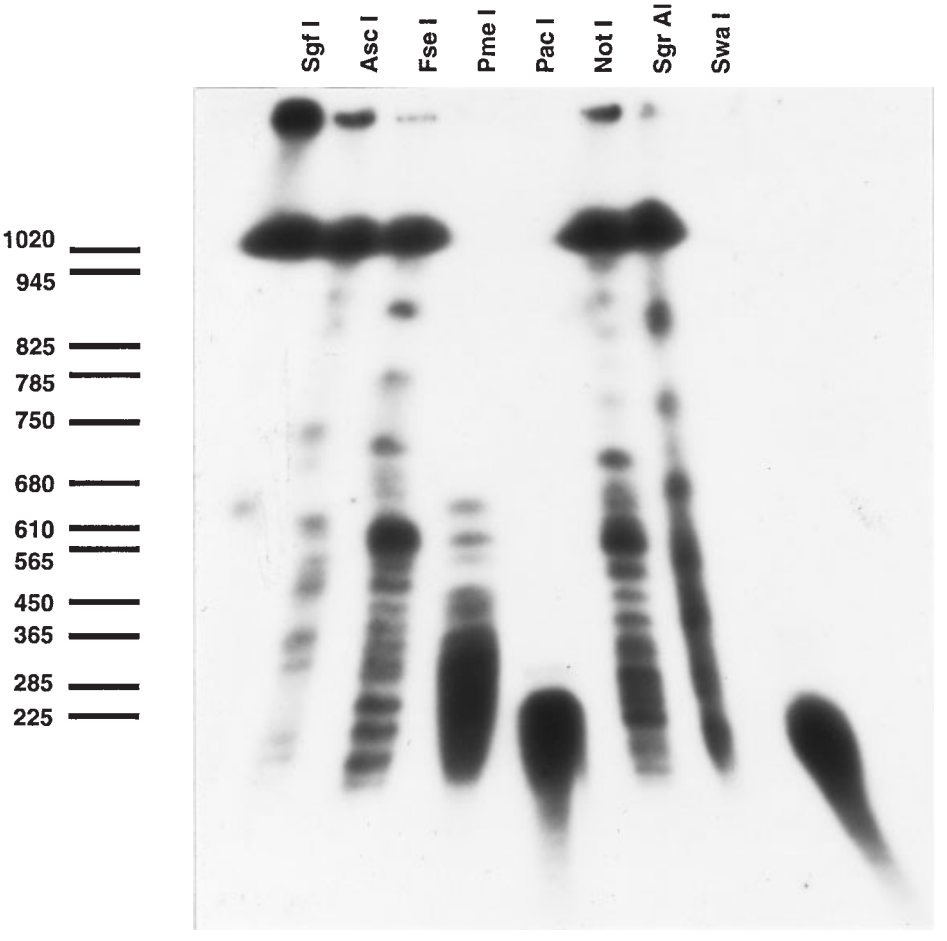

Figure 1. Detection of PFG fractionated restriction fragments of human chromosomes. DNA from a monosomic hybrid cell line containing chromosome 20 was digested to completion with different enzymes, PFG-fractionated, and hybridized to the human-specific *Alu* probe. The recognition sites for the enzymes used are: *Sgf* I (GCGATCGC), *Asc* I (GGCGCGCC), *Fse* I (GGCCGG/CC), *Pme* I (GTTTAAAC), *Pac* I (TTAATTAA), *Not* I (GCGGCCGC), *Sgr*AI (CrCCGGyG), and *Swa* I (ATTT/AAAT).

peat. The *Kpn* element, a LINE (long interspersed) repeat, is estimated to occur at a frequency tenfold less than that of the *Alu* repeat. Theoretically the use of these two repeats as hybridization probes to hybrid cell line DNA cleaved with the appropriate enzyme should reveal all the human megabase fragments. Since, however, this approach will not distinguish two fragments of the same size, it is best used with very high resolution PFG fractionations. Restriction fragments identified by SINE and/or LINE repeats will be differently distributed on a chromosome in a manner that reflects the apparent differential distribution of these repeat elements on chromosomes.

The use of repeat sequences as hybridization probes can reveal information about the size and distribution of restriction fragments, but nothing about order. For monosomic hybrid cell line DNA, human-specific repetitive telomeric and centromeric sequences may be used to identify the ends of the physical map as well as to provide an important anchor for regions of condensed chromosomes visualized microscopically. Although these sequences occur on all, or most, human chromosomes, they are limited to the human component of hybrid cell lines Hence, they act like single-copy probes for these samples.

In conventional mapping experiments interstitial megabase frag-

ments are linked to single-copy sequences that are used as hybridization probes. Single-copy sequences that have been located on genetic maps can be used for the regional assignment of restriction fragments and can serve as anchors between genetic and the physical maps. The accuracy of fragment location will reflect the accuracy provided by the genetic map. In this approach, the catalogued megabase restriction fragments containing interspersed repeats are linked to single-copy sequences. This allows for the identification of unassigned megabase fragments, which then become candidates to fill map gaps, as discussed toward the end of this section.

The construction of a genomic restriction map requires that neighboring fragments be linked. At least one neighbor can be identified unambiguously by hybridizing a single-copy sequence to partially digested genomic DNA. The difficulty of interpreting the partial digestion data increases dramatically with the number of partial digestion products. For instance, it is important to consider that both the neighboring fragments may be of the same size. That is, the possibility that a single partial product band could represent two different products of the same size must always be borne in mind. Thus, it is important to confirm neighboring fragments in multiple experiments using probes from different fragments. Many times, the confusion associated with partial digest data can be sorted out by

obtaining partial digest data from probes located on adjacent, or nearly adjacent, fragments. In these cases different sets of partial digest fragments are identified by the individual probes. The correct map interpretation is one that is consistent with all the data obtained in the different hybridization experiments.

The analysis of partial products is simplified if the probe is from a telomere region or is situated on a small fragment located next to a very large fragment. In both cases, the partial digest information reflects the order of fragments in only one direction, as opposed to the usual bidirectional information obtained from the use of interstitial probes. This approach also eliminates the problem of comigration of same-size products associated with bidirectional partial mapping data.

There are several other approaches to proving that two restriction fragments are adjacent. By far the most powerful is the use of linking libraries—small insert libraries that contain DNA segments from two adjacent megabase restriction fragments. Hence, a *Not*I linking clone used as a hybridization probe to genomic DNA digested with *Not* I will identify adjacent *Not*I fragments. A complete linking library would suffice to construct, in the most efficient manner, a complete genomic restriction map. Without a complete linking library, partial digest strategies combined with polymorphism link-up and double restriction enzyme strategies can provide information on neighboring fragments. In both approaches, parallel analysis of a region provides a regional signature.

The same polymorphism link-up refers to the results obtained when one probe is used to analyze restriction enzyme digested DNA from different cells. For instance, a restriction site may be missing in one cell line, either because there is a mutation at the restriction enzyme site or because methylation is interfering with restriction enzyme cleavage. Hence, when a number of DNA sources are examined, it is often possible to fingerprint the polymorphism of a restriction site (i.e., assuming that the site is polymorphic). Thus, probes that are usually located on separate, but adjacent, fragments can, in some DNAs, be found on the same fragment, which will equal in size the total size of the two smaller fragments seen in other cell lines. In such a case, the pattern of occurrence of the two distinct smaller fragments, detected by the different probes, is self-consistent. This is very similar to the more familiar approach of fingerprinting a region using many different restriction enzymes.

Megabase restriction fragments themselves may also be fingerprinted by cleavage with a second enzyme. The products are analyzed by hybridization experiments using probes that are believed to be located on the original megabase fragment. The partial digest product data will be different for the two probes and will depend on their location on the original megabase fragment. The interpretation of the distribution of the cleavage sites of the second enzyme must be consistent with data obtained using the two interstitial (or terminal) probes. This approach is particularly useful for analyzing megabase fragments that are at the limit of the PFG technique, because the products of the second enzyme will be within the PFG limit.

Besides restriction enzymes, a number of other enzymes have been adapted for cutting genomic DNA into specific large pieces. Most recently, methods have been developed using protection assays to cleave DNA at specific DNA sequences. These methods usually depend on the formation of some structure (e.g., a triplex structure or D-loop structure) having an associated single-stranded region. D loops are formed with the aid of recA protein at targeted sites. In one strategy, termed the "Achilles' heel" strategy, or RARE

(for recA-assisted restriction endonuclease) cleavage, the genomic DNA containing a D loop is treated with DNA methylase, which modifies DNA at all its recognition sites except one that is located in the D loop. Methylation makes the DNA resistant to cleavage by the cognate restriction enzyme. The methylase is removed, and the DNA is then subjected to digestion with the cognate restriction enzyme. Only the restriction enzyme recognition site in the previously formed D loop is now susceptible to cleavage by the restriction enzyme.

This type of approach allows for site-directed cleavage of genomic DNA at specific locations. The combination of two RARE reactions from nearby sites would produce a fragment whose length would be equivalent to the distance between the two sites. Furthermore, fragments smaller than 10 Mb in size could be purified from the remaining genomic DNA using PFG electrophoresis (Figure 2). Such a technique could be extended to mapping an entire chromosome or genome. Here, adjacent pairs of STSs would be used to generate megabase fragments. The comprehensive set of polymorphic genetic markers that span the entire human genome would be a particularly useful set of such STSs. This material would allow the precise physical distances between each of the markers to be determined and would provide a mean of purifying the genomic DNA contained in this region between the markers. We have developed a number of procedures that allow the reliable use of DNA contained in gel slices as templates in polymerase chain reactions (PCRs). For instance, the DNA can be isolated and amplified using a PCR reaction, either nonspecific, randomly primed or specifically primed. Furthermore, the distribution of probes along such a stretch of DNA could be analyzed using some of the new fluorescent hybridization methods [e.g., fiber fluorescence in situ hybridization (FISH)] that order probes along stretched DNA.

It is very easy to start maps but very difficult to finish them. Each gap in each map presents a unique problem. It is important to treat

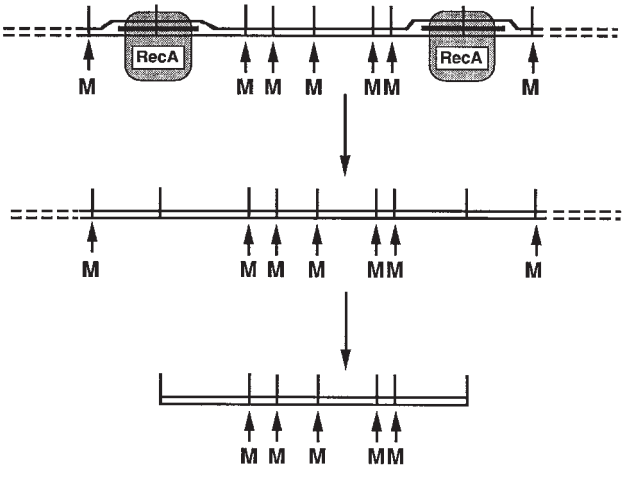

Figure 2. The RARE strategy for generating megabase fragments. Vertical bars indicate the location of recognition site for a specific restriction–methylation enzyme system. Horizontal heavy bars represent of STSs targeted by RecA protein mediated D-loop formation. (A) After D-loop formation, enzyme recognition sites except sites that are located in the D loops are modified by incubation with methylase. (B) RecA protein and methylase are removed and sample is incubated with restriction enzyme. (C) The use of two RecA targets will generate a fragment whose length is equal to the distance between the STSs.

each gap individually and design the best strategy for dealing with the gap. The best strategy will depend on the size of the restriction fragment, the amount of polymorphism in the region, the number of cloned sequences from the region, and the presence or absence of known unassigned megabase fragments. In some cases, the megabase restriction fragment themselves are useful hybridization probes to deploy, using partial digests, in the identification of neighboring fragments. For the human restriction fragments from monosomic hybrid cell lines, this can be done by amplifying single-copy human sequences contained between *Alu* elements using inter-*Alu* PCR amplification.

The development of methods that permit the use of DNA in gel slices also can provide a basis for functional studies. For instance, a recently development method termed comparative genome hybridization (CGH) uses differentially labeled cDNA, or genomic DNA, from two samples. The labeled samples are mixed together in equal amounts and used as a hybridization probe. The target samples are metaphase chromosomes. Quantitative analysis of the hybridization results provides information on the relative amount of DNA from specific regions on the chromosome in each of the probe samples. CGH experiments provide positional information of differences between samples. The resolution of this method is low (10–30 Mb), and the method requires sophisticated image construction. The method, similar to conventional genetic positional cloning approaches, provides positional information about the location in the genome of important DNA sequences.

An alternative to conventional CGH would utilize genome DNA arrays that have been generated from size-fractionated genomic restriction fragments. An array consisting of 1500 elements would allow 1–3 Mb resolution. Here the analysis is much easier because the samples are in an orderly array. This approach would also provide a DNA of interest in hand as one of the array elements. It would not give positional genomic information directly.

The positional information can be generated in a number of other ways. For instance, each array element could be subjected to PCR analysis, hence linked to the set of genomic STSs or ESTs that are serving as anchors between various genetic and physical maps. The density of genetic markers at this time would not allow assignment of all the DNA in such samples. Alternatively, the DNA in each array element could be independently assigned to genome positions using FISH experiments. In FISH, positional information is obtained when DNA probes hybridize to a particular location on a metaphase chromosome.

5 CONSTRUCTION OF OVERLAPPING LIBRARIES AND GENOMIC RESTRICTION MAPS

The construction of an overlapping library begins with the selection of the genomic DNA sample. The next decision entails choice of the type of vector that will be used. The most distinguishing feature of vector possibilities is the size of the DNA that can be cloned into them. It is helpful to begin whole-genome mapping projects using the largest DNA fragments possible, thus minimizing the number of clones to order. In the past, the largest cloning vectors, cosmids, contained cloned sequences about 40 kb in size. Recently, several systems that allow the cloning of larger segments have been developed. For instance, in bacteria, fragments up to 150 kb may be cloned in PACs. Less developed is the use of the *E. coli* F^+ plasmid or P1 phage plasmid cloning system, which has the potential to con-

tain very large fragments. For instance, it has been known for some time that *E. coli* can maintain F^+ plasmids carrying up to 2.5 Mb of *E. coli* sequences. Megabase fragments also can be cloned into YACs. Recent progress in this area has shown that libraries containing clones with more than 1 Mb may be isolated, although not on a routine basis.

In conventional chromosome walking (or "crawling") experiments, individual clones are hybridized to an entire library to detect overlapping clones. This method, though guaranteed to detect overlapping clones, is slow and laborious. The efficiency of the method may be enhanced by taking a multiplex approach, using intelligent pools of samples to minimize the work. For instance, clones may be hybridized to a minimum set of clone pools. Then each clone is present in more than one pool, and the pattern of hybridization to the entire set of pools will reveal the identity of the overlapping clones. Most other more efficient methods of clone ordering have been unable by themselves to produce finished maps. Thus, the random approaches described in the remainder of this section often must be abandoned and replaced by focused efforts to fill in map gaps.

The ordering method will regulate how the cloning is done as well as which vector is used. For instance, most bottom-up, fingerprinting methods require that libraries be constructed using partial restriction enzyme digests to ensure that overlaps exist. There are a number of ways to fingerprint clones to identify overlaps. Bottom-up strategies for library ordering usually involve testing individual clones to search for an overlapping restriction fragment or restriction site pattern. This type of fingerprinting is easy to automate. Clones may also be fingerprinted using interspersed repetitive sequences. Here, clones, or restriction fragments of the clones, are hybridized to a set of short oligonucleotides or longer interspersed repetitive sequences.

A variation on this approach that has proven to be quite efficient is the use of genomic DNA restriction fragments as hybridization probes to order clone libraries. Megabase restriction fragment probes allow the assignment of clones to particular fragments. If the fragments are mapped, the clones can then be regionally assigned in the genome. Clones in each region are ordered simultaneously, using pools of small genomic restriction fragments. The small restriction fragments in the pools come from different regions of the genome. Each pool generally consists of genomic restriction fragments of a particular size range. Since the clones were assigned earlier to megabase restriction fragments, two clones on different megabase fragments can be differentiated even though they hybridize to the same pool of small restriction fragments. This method is a multiplex approach because each hybridization experiment order, in parallel, clones from different portions of the genome.

Another ordering approach involves detecting overlaps by DNA sequencing. Here, the sequences may be chosen at random sites (STSs) or the sequences can be collected at specific locations originally called sequence-tagged restriction sites (STARs). The STAR approach may be used with both partial and complete digest libraries. DNA sequence information may be collected at specific restriction sites, including the ends of the cloned sequences. If the library is generated by a complete restriction enzyme digestion, clone linking is done in conjunction with another library (i.e., a linking clone library that spans the restriction enzyme site). If the library is generated by a partial digest, the ends of enough clones must be sequenced to ensure that overlaps will be detected. Recently, this approach was used to sequence an entire 1.8 Mb chromosome. The method eliminates the need for up front clone ordering, but a highly

representative library is required, and sequence read lengths must be long enough to span most small repeats.

6 PROSPECTUS

Genomic restriction maps and overlapping libraries are complementary resources. It is important to have genomic restriction maps because they provide accurate distances and can guard against the use of incorrect information generated from cloning artifacts that may occur during library construction of maintenance. Genomic restriction maps are relatively easy to make, and they can be used to compare many samples. For instance, many disease gene hunts have been aided by gross chromosomal rearrangements such as translocations, insertions, or deletions. Overlapping libraries are comparatively difficult to construct and maintain; however, they provide genomic DNA in a convenient and useful form.

See also DNA Sequencing in Ultrathin Gels, High Speed; Gene Order by FISH and FACS; Nucleic Acid Sequencing Techniques; Partial Denaturation Mapping; Restriction Endonucleases and DNA Modification Methyltransferases for the Manipulation of DNA; Robotics and Automation in Genome Research; Whole Chromosome Complementary Probe Fluorescence Staining.

Bibliography

Billings, P. R., Smith, C. L., and Cantor, C. R. (1991) New techniques for the physical mapping of the human genome. *FASEB J*. 5:28–34.

Burmeister, M., and Ulanovsky, L., Eds. (1992) *Methods in Molecular Biology, Pulsed-Field Gel Electrophoresis. Protocols, Methods, and Theories*. Humana Press, Totowa, NJ.

Cantor, C. R., Smith, C. L., and Matthew, M. K. (1988) *Annu. Rev. Biophys. Biophys. Chem*. 17:287–304.

Collins, F. S. (1995) Positional cloning moves from perditional to traditional. *Nature Genet*. 9:347–350.

Ferrin, L. J., and Camerini-Otero, R. D. (1964) Long-range mapping of gaps and telomeres with *RecA*-assisted restriction endonuclease (RARE) cleavage. *Nature Genet*. 6:379–383.

Fleischmann, R. D. et al., (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science,* 269:496–512.

Florijn, R. J., Bonden, L. A., Vrolijk, H., Wiegant, J., Vaandrager, J. W., Baas, F., den Dunnen, J. T., Tanke, H. J., van Ommen, G. J., and Raap, A. K. (1995) High-resolution DNA fiber-FISH for genomic DNA mapping and colour bar-coding of large genes. *Hum. Mol. Genet*. 4:831–836.

Fonstein, M., and Haselkorn, R. (1995) Physical mapping of bacterial genomes. *J. Bacteriol*. 117:3361–3369.

Gyapay, G., Morissette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, M., and Weissenbach, J. (1994) The 1993–94 Généthon human genetic linkage map [see comments]. *Nature Genet*. 7:246–339.

Heiskanen, M., Kallioniemi, O., and Palotie, A. (1995). Fiber-FISH: Experiences and a refined protocol. *Genet. Anal. (Biomol. Eng.),* in press.

Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F., and Pinkel, D. (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science,* 258:818–821.

Lawrence, B. J., Carter, K. C,. and Gerdes, M. J. (1992) Extending the capabilities of interphase chromatin mapping. *Nature Genet*. 2:171–172.

Ledbetter, S. A., and Nelson, D. L. (1991) Genome amplification using primers directed to interspersed repetitive sequences (IRS-PCR). In

Polymerase Chain Reaction I: A Practical Approach, M. J. McPherson, P. Quirke, and G. R. Taylor, Eds. IRL Press, Oxford.

Oliva, R., Lawrence, S. K., Wu, Y., and Smith, C. L. (1991) In *Encyclopedia of Human Biology,* R. Bulbecco et al., Eds., pp. 475–488. Academic Press, San Diego, CA.

Ricke, D. O., Liu, Q., Gostout, B., and Sommer, S. S. (1995) Nonrandom patterns of simple and cryptic triplet repeats in coding and noncoding sequences. *Genomics,* 26:510–520.

Schalkwyk, L. C., Francis, F., and Lehrach, H. (1995) Techniques in mammalian genome mapping. *Curr. Opin. Biotechnol.* 6:37–43.

Smith, C. L., and Condemine, G. (1990) New approaches for physical mapping in small genomes. *J. Bacteriol.* 172:1167–1172.

———, Warburton, P., Gaal, A., and Cantor, C. R. (1986) Analysis of genome organization and rearrangements by pulsed field gradient gel electrophoresis. In *Genetic Engineering,* Vol. 8, J. K. Setlow and A. Hollaender, Eds., pp. 45–70. Plenum Press, New York.

———, Lawrance, S. K., Gillespie, G. A., Cantor, C. R., Weissman, S. M., and Collins, F. S. (1987) Strategies for mapping and cloning macroregions of mammalian genomes. *Methods Enzymol.* 151:461–489.

———, Klco, S., Zhang, T. Y., Fang, H., Oliva, R., Fan, J. B. Bremer, M., and Lawrance, S. (1993) Analysis of megabase DNA using pulsed field gel electrophoresis. In *Methods in Molecular Genetics,* K. W. Adolph, Ed., pp. 155–196. Academic Press, San Diego, CA.

———, Wang, D., Broude, N., Bukanov, N., Monastryskaya, G., and Sverdlov, E. (1995) Parallel processing in genome mapping and sequencing. In *GenoMethods: A Companion to Methods in Enzymology.* M. Schwab, Ed. Academic Press, San Diego, CA.

Van der Ploeg, L. H. T., Gottesdiener, K. M., Korman, S. H., Weiden, M., and Le Blancq, S. (1992) Protozoan genomes. In *Methods in Molecular Biology, Pulsed-Field Gel Electrophoresis. Protocols, Methods, and Theories,* M. Burmeister and L. Ulanovsky, Eds. Humana Press, Totowa, NJ.

Wang, D., and Smith, C. L. (1994). Large-scale structure conservation along the entire long arm of human chromosome 21. *Genomics,* 20:441–51.

Welsh, J., and McClelland, M. (1994) Fingerprinting using arbitrarily primed PCR: Application to genetic mapping, population biology, epidemiology, and detection of differentially expressed RNAs. In *The Polymerase Chain Reaction,* K. B. Mullis, F. Ferre, and R. A. Gibbs, Eds., pp. 295–303. Birkhäuser, Boston.

Williams, N. (1995) Closing in on the complete yeast genome sequence. *Science,* 268:1560–1561.