



Sequence-tagged site (STS) content mapping of human chromosomes: theoretical considerations and early experiences.

E D Green and P Green

Genome Res. 1991 1: 77-90

Access the most recent version at doi:[10.1101/gr.1.2.77](https://doi.org/10.1101/gr.1.2.77)

References This article cites 82 articles, 33 of which can be accessed free at:
<http://genome.cshlp.org/content/1/2/77.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

A horizontal banner advertisement with a teal background. On the left, the text "CRISPR and RNAi Genetic Screening. Your new superpower." is written in white. In the center, there is a white-bordered box containing the words "LEARN MORE" in blue. On the right, there is a photograph of a woman wearing a red superhero mask and a red cape over a white shirt. To the right of the photo is the Cellecta logo, which consists of a cluster of green dots of varying sizes, with the word "CELLECTA" in white capital letters below it.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Copyright © Cold Spring Harbor Laboratory Press

Sequence-tagged Site (STS) Content Mapping of Human Chromosomes: Theoretical Considerations and Early Experiences

Eric D. Green^{1,2} and Philip Green¹

Departments of ¹Genetics, ²Pathology, and ²Internal Medicine, Washington University School of Medicine, St. Louis, Missouri 63110

Major goals of the Human Genome Project include the isolation of the entire human genome in overlapping clones and the development of physical maps of the cloned DNA.⁽¹⁾ Construction of these physical maps involves specifying the coordinates of DNA-based landmarks, whose nature, to a great extent, has reflected the experimental tools that are available for their detection within complex sources of DNA. The advent of polymerase chain reaction (PCR) technology has facilitated the evolution of a new type of landmark—the *sequence-tagged site* (STS)—for constructing physical maps of chromosomes. Here we review the basic concepts that guide the use of STSs as DNA landmarks, discuss a strategy for developing clone-based *STS-content maps* of chromosomes, and describe the role that such approaches are playing in the early phases of the Human Genome Project.

OVERVIEW OF STS-CONTENT MAPPING

The process of developing long-range physical maps of complex genomes involves the identification and localization of landmarks within recombinant DNA clones. In the typical case, this process can be viewed as occurring in four steps: (1) fragmenting the genome into smaller pieces (e.g., by restriction digestion); (2) isolating and purifying each fragment (i.e., cloning); (3) examining the individual DNA fragments for the presence of common landmarks; and (4) constructing physical maps of the cloned DNA based on information about the relative positions

of the landmarks. Recent technological advances provide powerful new tools for carrying out these interrelated activities. In particular, the development of yeast artificial chromosomes (YACs) as a cloning system in conjunction with the utilization of PCR for the detection of physical landmarks has yielded powerful approaches for the physical mapping of entire genomes.

YAC cloning provides the means to isolate segments of DNA that are large (100–1000 kb) yet experimentally manageable.^(2–5) The implementation of YAC cloning has been greatly facilitated in recent years by the development of efficient methods to construct^(6–13) and screen^(14–20) YAC libraries as well as the establishment of effective approaches to manipulate^(21–28) and characterize^(23,26–33) YAC clones, including a number of different strategies for isolating their insert ends.^(2,30,31,34–43) In their short history, YACs have been used to isolate and map contiguous stretches of DNA that are significantly larger than those isolated using bacterial cloning systems^(21,35,44–51) and to play important roles in the identification of genes causing several human genetic diseases, such as neurofibromatosis,⁽⁵²⁾ fragile X syndrome,^(53–57) and familial adenomatous polyposis coli.^(48,49) Based on the experience to date, YACs are generally considered the best available cloning system for constructing long-range physical maps of DNA.

The most serious problem thus far associated with YAC cloning is the presence of clones containing two or more unrelated pieces of DNA (i.e., seg-

ments arising from different regions of the genome of origin). Such *chimeric* YACs have been detected either by obvious differences between the physical map of the YAC and that of the corresponding region of the source genome or by discrepant chromosomal origins of the two insert ends^(21,34,35,39,46) The frequency of chimeric clones seems to vary widely among different libraries and may vary from locus to locus in the same library. For example, in one library derived from total human DNA,^(6,10,17) 40–60% of the YAC clones are estimated to be chimeric,⁽³⁴⁾ while in another library made from a human–hamster hybrid cell line, the fraction of chimeric YACs is estimated at 10–20%.⁽⁵⁸⁾ The mechanism(s) by which such anomalous YACs form remains an area of investigation,⁽³⁴⁾ as is the development of methods to eliminate their formation. One proposed mechanism of chimeric YAC formation involves the recombination in the transformed yeast cell of repetitive sequences (e.g., *Alu*) located within unrelated DNA molecules. This model is based on the detailed characterization of one chimeric YAC⁽³⁴⁾ and the collective experiences with libraries derived from human–rodent hybrid cell lines, which seem to contain fewer chimeric clones. In the latter cases, the presence of a background of irrelevant rodent DNA likely minimizes the frequencies with which recombination events can occur between unrelated human–YAC clones. A much less frequent problem with YAC cloning is that created by DNA segments that are unclonable or unstable. In some cases

in which instability has been encountered, stable overlapping YACs containing the problematic DNA have been isolated.⁽⁵⁸⁾

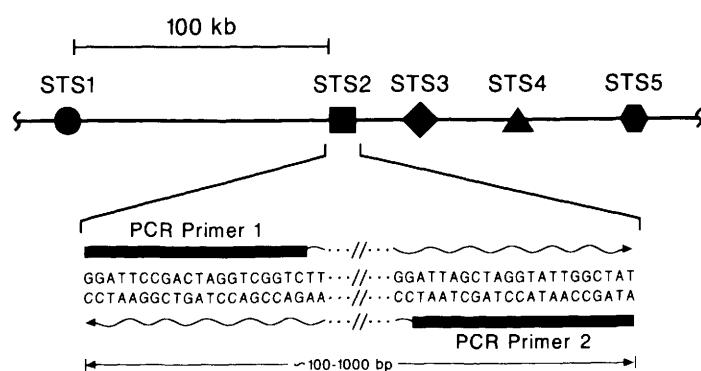
The development of YAC cloning has largely occurred in parallel with that of PCR technology. For many applications, PCR has been used as the front-line tool for both isolating^(14,15) and characterizing^(21,35,39,46,50) YAC clones in efforts to construct physical maps of DNA. These early studies helped to motivate the proposal to use STSs as the landmarks in physical maps of DNA,⁽⁵⁹⁾ although the case in favor of STSs transcends any particular mapping method. By definition, an STS is a short stretch of DNA that can be spe-

cifically detected by a PCR assay (see Fig. 1A). Thus, the physical landmark is the DNA sequence itself, with PCR being the method used to detect the landmark. Implicit in this definition is that the STS be operationally unique (i.e., that the PCR assay amplifies the STS from a single site in the genome), although absolute demonstration of this feature is not always a simple task (see ref. 60).

A major advantage of using STSs as landmarks is their interlaboratory "transportability." The defining features of an STS can be reduced to information which can be stored in a computer database (e.g., oligodeoxynucleotide primer sequences, PCR as-

say conditions, STS size). Once the features are entered into such a database, any investigator can easily recover the DNA segment represented by that STS by synthesizing the relevant oligodeoxynucleotide primers and performing PCR under the described conditions.⁽⁶⁰⁾ No physical exchange of materials is necessary. This situation is in contrast to clone-based landmarks, which require the storage and distribution of biological samples. The problems associated with the latter include the expense of storing and exchanging clones, the inconvenience and inefficiency of such exchanges, the high risk of errors relating to human mistakes (e.g., misidentified clones, im-

A. STS-Content Map of a Human Chromosome



B. YAC Isolation and Contig Assembly

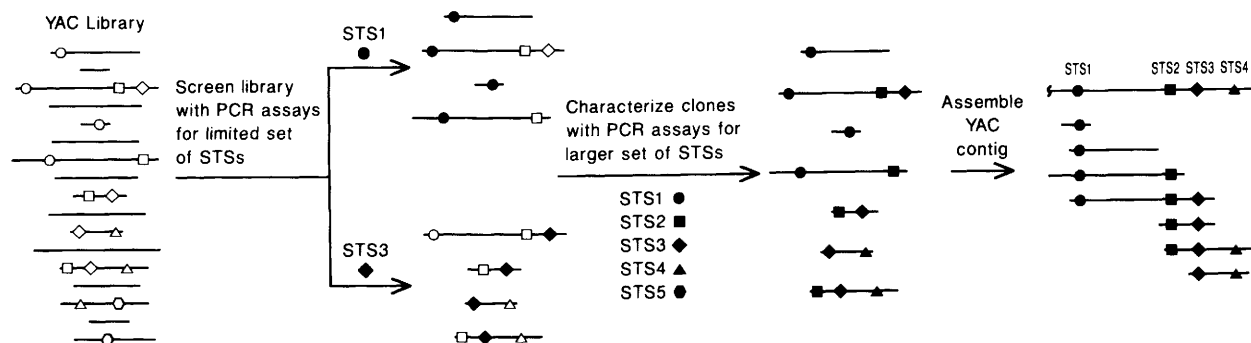


FIGURE 1 General strategy for constructing STS-content maps. (A) A physical map corresponding to a region of a human chromosome is represented with STSs as landmarks (depicted as unique symbols). Each STS consists of a short DNA sequence that can be specifically detected by a PCR assay employing two oligodeoxynucleotide primers. (B) A YAC-based STS-content mapping scheme is depicted schematically. A YAC library, consisting of YAC clones of various sizes and unknown compositions of STSs (indicated as open symbols), is screened for a subset of available STSs using a PCR-based screening method.⁽¹⁴⁾ For each group of isolated clones, only the STS whose corresponding PCR assay was used to perform the library screen is known to be present in the YACs (indicated as filled-in symbols). However, the isolated YACs can then be tested for the presence of additional STSs as well as sized electrophoretically. The resulting information can be used to establish the overlap relationships among the clones and to assemble a YAC contig, from which an STS-content map can be abstracted. In the example illustrated, the relative order of the STSs as well as some information about the spacings between them could be deduced. Of note, STS5 could not be included in the STS-content map, since YACs bearing that STS were not isolated from the library. Subsequent isolation of the depicted STS5-containing YACs using the corresponding PCR assay would be one route toward expanding the contig and placing this STS on the map.

proper storage), clone instability, and, most importantly, the rapid obsolescence of particular vector–host systems. The ability to access the DNA landmarks on which the physical maps are based without dependence on stored biological materials is a key notion behind the proposal to use STSs as the "common language" in physical maps of DNA.⁽⁵⁹⁾

The fundamental features of a YAC-based STS-content mapping strategy are highlighted in Figure 1. Dispersed across a large DNA interval (e.g., a human chromosome) reside numerous unique DNA sequences (i.e., STSs), each of which can be detected by a specific PCR assay (Fig. 1A). One route toward defining the relative order and spacings of such STSs is schematically depicted in Figure 1B. A YAC library contains clones bearing one or more of the indicated STSs. Using appropriate PCR assays, YACs containing specific STSs can be readily isolated.⁽¹⁴⁾ The subsequent characterization of the resulting YACs includes both testing for a larger number of STSs and electrophoretic assessment of YAC size. Knowledge of the *STS content* and size of each YAC can be used to construct a *clone map*, which is a collection of clones aligned in an overlapping manner (a set of contiguous, overlapping clones is also known as a *contig*). This information can in turn be used to establish an *STS-content map* indicating the relative locations of the various STSs within the contig. Incorporation of contig maps on STS-content maps of the whole chromosome will require, in addition, the use of long-range methods (such as linkage mapping and radiation hybrid mapping) to order and orient the contigs along the chromosome.

In the following, we propose a strategy for developing STS-content maps of whole chromosomes and discuss some of its theoretical properties. We then review relevant practical experience, with particular emphasis on activities in the Washington University Human Genome Center, where this strategy is being applied to construct physical maps of human chromosomes 7 and X. Several alternative physical mapping strategies have been successfully used by other investigators, as reviewed recently by Evans.⁽⁶¹⁾

A STRATEGY FOR CONSTRUCTING STS-CONTENT MAPS OF WHOLE CHROMOSOMES

Conceptual Basis of Strategy

The development of high-resolution physical maps typically involves the assembly of overlapping clones into a contig, from which a map can be derived. Two classical strategies have been used to produce such overlapping clone maps: (1) *fingerprinting*, in which a characteristic sequence-dependent pattern is generated for each randomly selected clone (e.g., by sizing restriction fragments) and clone overlaps are inferred based on shared parts of the pattern; and (2) *chromosome walking*, in which unique sequence probes are used to obtain clones, screening the clone library with an initial probe, and then, iteratively, with probes derived from the ends of clones obtained in the previous screen, until the region of interest is fully isolated. Chromosome walking has the advantages of being targeted to the region of interest and of being able to detect small overlaps between clones, thereby allowing the construction of larger contigs. However, this approach has generally been considered unsuitable for mapping large genomic regions, because each walking step depends upon the successful completion of the preceding one and requires both probe generation and library screening.

The STS-content mapping strategy (Fig. 1B) represents a hybrid approach that can be used for constructing overlapping YAC maps of large, targeted genomic regions, such as whole human chromosomes (the details of this strategy were originally presented by P. Green at the 1990 "Genome Mapping and Sequencing" meeting at Cold Spring Harbor Laboratory). In this approach, STSs randomly located throughout the region to be mapped fill three roles: (1) a set of random STSs serve as anchor points for multiple, short chromosome walks; (2) STSs near the ends of the growing YAC contigs serve as *surrogate ends* for expanding the contigs by walking; and (3) the STS contents of the YACs are used to establish overlaps among the YACs (i.e., the STSs in a YAC constitute its "fingerprint" for use in constructing contig maps, as illustrated in Fig. 1B). Two

disadvantages of chromosome walking, the complete dependence on clone end isolation and the need for multiple sequential steps, are largely avoided. As with other targeted mapping approaches, numerous library screens are required; however, their number is minimized by prescreening the previously isolated YACs for the presence of newly generated STSs and performing library screens either with STSs absent from the previously isolated YACs or with surrogate-end STSs. Thus, library screens are only performed with STSs likely to add a significant amount of new DNA to the clone map. Moreover, a single library can, in principle, be used for mapping any genomic region.

Because contig assembly is based on STSs already known to be located on the chromosome being mapped, this strategy, unlike pure fingerprint-based approaches, is relatively insensitive to the presence of chimeric YACs. A portion of a YAC that lies on another chromosome is not detected with any of the STS-specific PCR assays, and therefore will not cause spurious mergers of non-overlapping contigs. (Chimeric YACs can affect the distance estimates between STSs, however.) Of course, chimeric YACs containing two discontinuous segments from the same chromosome can cause problems, but these should be manageable if they are not too frequent (see Prospects and Limitations, below).

Implementation of Strategy

The proposed strategy comprises three phases (Fig. 2). During phase 1, a collection of randomly located, chromosome-specific STSs are developed and used to screen the YAC library. The YACs obtained are placed into a chromosome-specific *YAC panel*. As phase 1 progresses, new library screens have an increasing probability of detecting previously isolated YACs. When this probability becomes significant (the precise criterion is discussed below), phase 2 begins. In phase 2, each new STS is used to screen the YAC panel. If this panel screen is negative, then a full library screen is performed with the STS. If the panel screen is positive, then the information about which YACs contain the STS is used to assem-

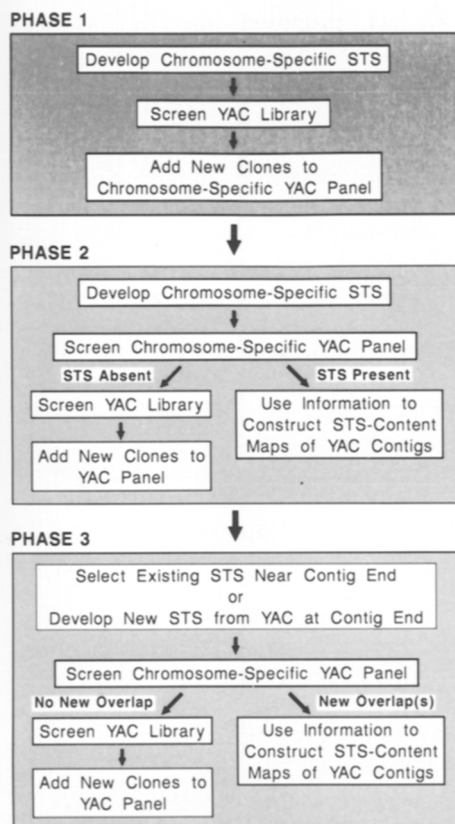


FIGURE 2 Experimental strategy for constructing STS-content maps of whole chromosomes. The general aspects of the three phases of an STS-content mapping project are outlined, with details provided in the text.

ble the YACs into contigs and to construct STS-content maps.

In essence, phases 1 and 2 initiate multiple chromosome walks from random locations on the chromosome. Phase 3, which begins when the YAC panel contains most of the DNA in the chromosome, aims to extend the existing contigs, both by identifying any undetected overlaps among the isolated YACs and by obtaining additional overlapping YACs from the library. The contig maps developed in phases 1 and 2 are utilized in a directed manner for selecting the STSs to be used for library screens. First, *surrogate ends*, STSs which have previously been used to screen the YAC panel but not the full library and which are known from the STS-content maps to lie near the ends of contigs, are retested against the YAC panel. (This retesting is necessary because the YAC panel is continuously

expanding, so it will usually contain clones added subsequent to the initial panel testing with that STS.) Sometimes a new panel screen with a surrogate end will detect a previously unknown overlap between two contigs, thereby allowing the contigs to be merged. If no new overlap is detected, then the STS is used for a library screen.

An obvious modification of this scheme is to use *true ends* (developed from the YACs at the ends of contigs) instead of surrogate ends as the STSs in phase 3. True ends are somewhat more likely to reveal overlaps and detect new YACs. However, the generation of STSs from true ends is likely to involve more effort, since it generally requires isolating both insert ends of several terminal YACs and testing them against other YACs in the contig to determine which YAC end represents the true contig end. Moreover, the presence of chimeric YACs implies that in some instances the true contig end will, in fact, not be at the end of a YAC insert. The decision about the type of "end" to use in phase 3 will ultimately depend upon the relative costs and the desired contig sizes, as discussed below.

In describing this model for developing STS-content maps, we have considered all the mappable landmarks to be STSs, implying that all mapping data would be obtained using PCR as the analytical technique. In actual practice, this will likely not be the case. For example, any single-copy hybridization probe (e.g., λ , cosmid, or cDNA clones) can be viewed as a "potential" STS and used in the identical fashion for mapping as has been illustrated in Figures 1 and 2 for true STSs, the major difference being the mode of detection (hybridization for probe sequences vs. PCR for STSs). Of note, the conversion of hybridization probes (potential STSs) into functional PCR assays is relatively straightforward. Thus, for the screening of YAC libraries and panels described above, either hybridization detection of probe sequences or PCR detection of STSs can be utilized and the resulting information used in an identical manner for constructing physical maps.

The choice between PCR- and hybridization-based protocols is really an experimental issue rather than a map-

ping one. In practice, we envision a mixture of experimental activities, with the overall efficiency, accuracy, and cost of each process dictating the extent of its use. We also expect that such a mixture will change throughout the course of a project. For example, in our opinion high-throughput screening of YAC libraries can more efficiently be performed by PCR than by hybridization.⁽¹⁴⁾ PCR assay development would therefore likely predominate during phase 1, when large numbers of library screens are being performed. In contrast, in phases 2 and 3, most of the analyses are being performed on relatively small numbers of YACs clones (the YAC panel), which may be more efficiently analyzed by hybridization. Overall, there is an incentive to tip the balance toward PCR assay development because of the desirability of using true STSs as landmarks.^(59,60) An additional factor which may favor the emphasis of PCR is the existence of duplicated segments of DNA within a chromosome. Evolutionarily older duplications are likely to have diverged over time, and PCR offers better prospects than does hybridization for discriminating between different copies. Such duplications may turn out to be relatively common in the human genome, with examples including pseudogenes, gene families, and several large regions of the X chromosome.^(58,62) These considerations are also relevant to the important issue of how best to integrate chromosomal STS-content maps with existing long-range maps developed using other methods, such as linkage mapping and radiation hybrid mapping. We envision performing this integration by incorporating the markers from the long-range maps into the STS-content maps. This can be done either early in the project, by generating STSs from the markers and using them to screen the YAC library, or later, by using the markers to screen the YAC panel by hybridization. The former approach is likely to be less efficient, since it requires the generation of DNA sequences from templates cloned into a diversity of vectors; moreover, the number of such markers available for individual chromosomes is small relative to the needs of the overall project (see below).

Map Quality

What can be expected of the physical maps produced by the above strategy? That is, what will be the sizes of the contigs, how much of the chromosome's DNA will they contain, and what level of map resolution will be achieved, as a function of the number of STSs and YAC library characteristics?

As noted previously, there are two variations on phase 3 of the proposed mapping strategy: One relies on surrogate ends and the other on true contig ends for the walking steps that occur in phase 3. It can be demonstrated that the contigs obtained using the surrogate-end version should contain the same cloned DNA as the contigs obtained using a "brute force" approach of screening the YAC library for all random STSs. Theoretical formulas for estimating the average contig sizes and chromosomal coverage produced by a brute force approach have been derived by several investigators (P. Green, unpubl.; and refs. 63–65) using the assumption that the YAC size distribution in a library is of a simple form (constant, uniform, normal, or exponential). Unfortunately, none of these forms is applicable to the YAC libraries currently in use, which have a highly skewed but nonexponential size distribution. Thus, these theoretical predictions are of limited applicability. Therefore, we have performed simulation studies using an actual YAC size distribution, as estimated from a random sample of 1000 clones in the Xq24–q28 library of Schlessinger and colleagues.^(58,66,67) We based the simulations on this library because it is well characterized and known to have a relatively low proportion of chimeric YACs.⁽⁵⁸⁾ In libraries with a high proportion of chimeric clones, it is the size of the chromosomally contiguous portion of each YAC that represents the relevant information; however, little data of this type is as yet available. In this regard, the quantitative and qualitative aspects of the size distribution are likely to be specific for each YAC library. Thus, extrapolation of the results provided below to projects using other libraries should be done cautiously.

The results of our simulations for mapping a 150-megabase pair (Mb) chromosome using the surrogate-end

version of the strategy are shown in Figure 3. The predicted average contig size as a function of the number of STSs used is depicted in Figure 3A. The fraction of the chromosome contained within contigs as a function of the number of STSs employed is shown in Figure 3B.

These simulations reveal that 99% chromosome coverage should be obtained with about 1400 STSs. The desirable average contig size is largely dictated by the resolution of the long-range methods (linkage mapping, radiation hybrid mapping, in situ hybridization) that will be utilized to order the contigs along the chromosome and to orient each contig with respect to the centromere and telomere. Specifically, ordering contigs involves determining the relative chromosomal order of an STS from each contig, while orienting a contig requires determining the chromosomal order of at least two STSs within the contig (preferably from the ends). Thus, contigs should be at least as large as the minimum distance currently resolvable with the methods listed above, which is roughly 1–2 Mb.^(68,69) For mapping a 150-Mb chromosome, the corresponding number of STSs necessary to obtain contigs in this size range can be inferred from Figure 3A to be about 700–1850.

Another consideration is the map resolution (i.e., the average spacing between landmarks). It is the resolution of a physical map which dictates its value for most applications, such as searching for genes. Figure 3C indicates the average spacing between uniquely ordered STSs, as a function of the total number of random STSs. It should be noted that once adequate contiguity is obtained, resolution can be improved in either of two ways: by developing STSs from YAC ends or by screening the YAC panel with additional random STSs. Of these, developing YAC end-specific STSs is likely to be more cost-effective, because the resulting STS can always be uniquely placed simply by testing other YACs in its contig. In contrast, for a random STS, the entire YAC panel must be analyzed, and a unique position for the STS on the map will not always be obtained. These issues will be important in striving toward the development of physical maps with STSs every 50–100

kb, as proposed for the Human Genome Project.⁽⁵⁹⁾

Laboratory Effort Required

The predicted number of panel screens and library screens required using the

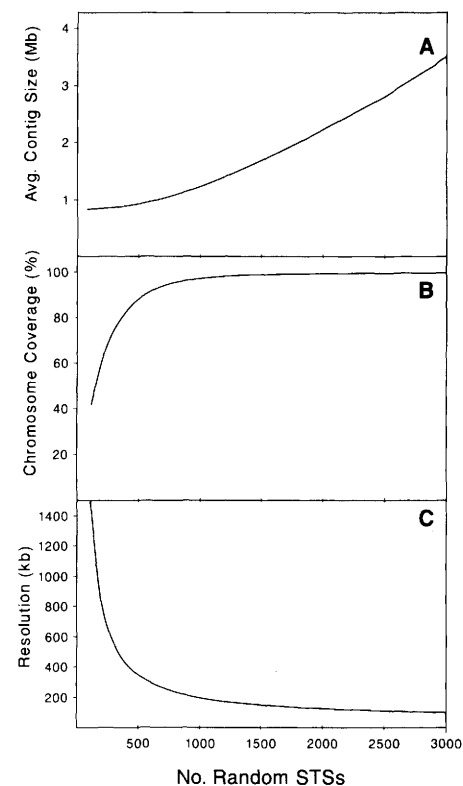


FIGURE 3 Expected map quality produced by STS-content mapping projects using surrogate ends. The predicted average contig size (A), chromosome coverage (B), and map resolution (C) that are obtained using the indicated numbers of random STSs to map a 150-Mb chromosome with the STS-content mapping strategy described in the text and in Fig 2 are indicated. Simulations assumed a YAC library containing six genome equivalents of cloned DNA and no chimeric YACs and having the same size distribution as a sample of 1000 random YACs selected from the Xq24–q28 library of Schlessinger et al.^(58,66,67) (The average YAC size in this sample is 241 kb; the minimum and maximum YAC sizes are 30 and 970 kb, respectively). For each multiple of 50 random STSs between 100 and 3000, 100 simulated mapping projects were performed and the results averaged. In these simulated projects, phase 1 was terminated when at least 2 out of 20 successive screens detected a previously isolated YAC (this occurred, on average, at the 21st library screen), and phase 2 was terminated when the indicated number of random STSs had been processed. Additional details of the computer simulations are available upon request.

surrogate-end version of the strategy, as a function of the number of STSs, is shown in Figure 4, A and B, respectively. Note that the total number of library screens needed is considerably less than the total number of STSs. For example, for 2000 random STSs, the proposed strategy would require about 390 library screens. In contrast, a brute force approach of screening the library with all STSs would require 2000 library screens. This constitutes an important advantage of the proposed strategy relative to a brute force approach.

What dictates the proper time to move from phase 1 to phase 2 and from phase 2 to phase 3 (Fig. 2)? The phase 2 to phase 3 transition point is relatively straightforward: Phase 2 ends when the complete set of random STSs (a number decided in advance based on simulation studies, such as those in Figs. 3 and 4) has been generated and utilized in the scheme shown in Figure 2. In contrast, the optimal

phase 1 to phase 2 transition point depends upon the relative "costs" of panel screens versus library screens. Roughly speaking, phase 2 should begin when the fraction of library screens that detect already isolated YACs (as estimated using the most recent library screen data) begins to exceed the ratio of a panel screen cost to a library screen cost. For example, if a panel screen cost is 1/10 that of a library screen, then phase 2 should begin when more than one out of 10 library screens detects a previously isolated YAC, because at that point 10 panel screens would save performing at least one library screen. The cost estimates should include labor and material costs for maintaining the panel and library, performing the screens (whether by hybridization- or PCR-based protocols), and the extent of STS development required for each type of screen. The cost of YAC library screening will critically depend upon the library size. While chromosome-specific YAC libraries (such as those derived from human-rodent hybrid cell lines) are smaller and thus less costly to screen than whole genomic libraries, they do require greater initial costs to construct, because the resulting human-YAC clones must be isolated from the more abundant rodent-YAC clones. Also to be considered is the relative reliability of each screening method and how this affects the total cost (e.g., the frequency with which experiments need to be repeated because of poor data quality, the "cost" of false-negative screens).

To compare the surrogate- and true-end versions of the proposed strategy, we considered the relative laboratory effort required to obtain a chromosome map with 99% coverage and a 2-Mb average contig size. In the surrogate-end version (Figs. 3 and 4), 1850 random STSs, 1970 panel screens, and 395 library screens are necessary and produce a map with a resolution of 125 kb. With the true-end version, an essentially arbitrary number of random STSs can be used in phases 1 and 2, with as many walking steps as necessary being performed in phase 3 to obtain the desired map; the more random STSs are used initially, the fewer walking steps are required. Figure 5 indicates the relative laboratory effort and

map resolution, as a function of the number of random STSs using true ends during phase 3. For example, if 1000 random STSs are used, then approximately 1360 panel screens (Fig. 5A) and 310 library screens (Fig. 5B) will need to be performed and 230 true contig ends (Fig. 5C) isolated to obtain 99% coverage and a 2-Mb average contig size. This will result in a map with an average distance between uniquely ordered STSs of 150 kb (Fig. 5D).

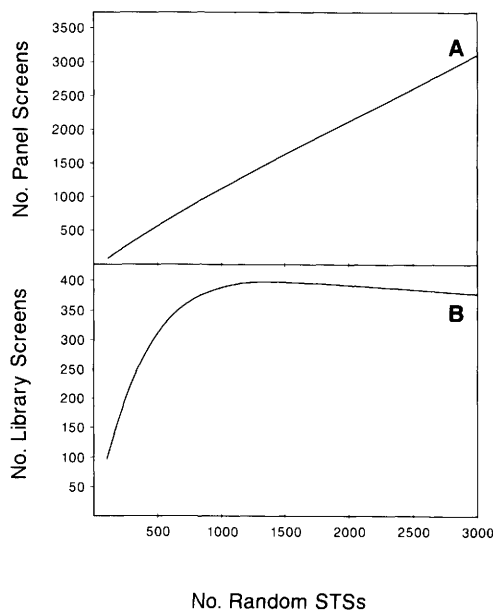


FIGURE 4 Expected laboratory effort of STS-content mapping projects using surrogate ends. The predicted number of YAC panel screens (A) and YAC library screens (B) that are required for chromosome mapping projects employing surrogate ends in phase 3 are indicated. Simulations were performed as described in the legend to Fig. 3, processing the indicated number of random STSs through phases 1 and 2 and terminating phase 3 when all surrogate ends had been screened against the YAC library and/or YAC panel.

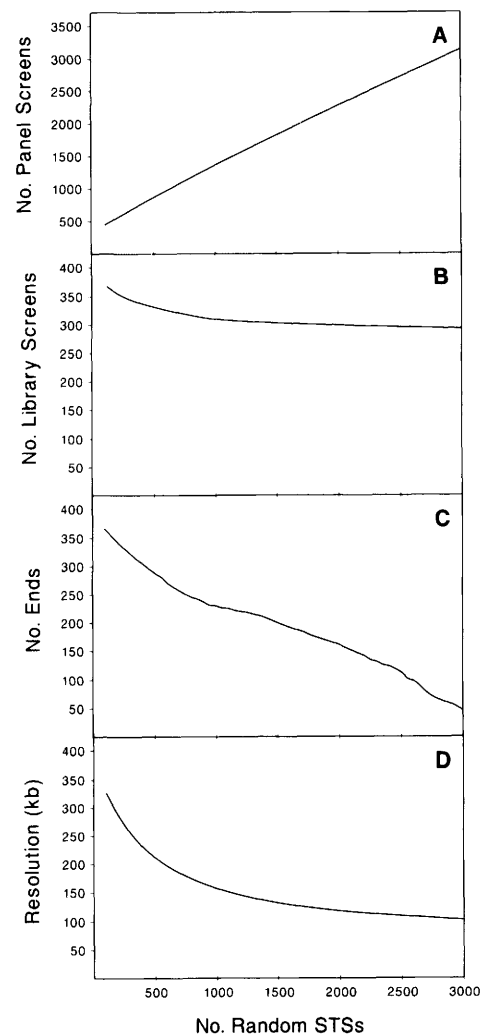


FIGURE 5 Expected results of STS-content mapping projects using true ends. The predicted number of YAC panel screens (A), YAC library screens (B), true ends developed into STSs (C), and map resolution (D) that are required for chromosome mapping projects employing true ends in phase 3 are indicated. Simulations were performed as described in the legend to Fig. 3, processing the indicated number of random STSs through phases 1 and 2 and terminating phase 3 when 99% chromosome coverage and 2 Mb average contig size were attained.

If one knows the relative costs of library screens, panel screens, and isolation of contig ends, then Figures 4 and 5 can be used to estimate the total cost of a mapping project for any given number of random STSs. This information can be used both to determine the "optimal" number of random STSs and to compare the surrogate-end and true-end versions of the strategy. At present, library screens are by far the most expensive of these laboratory operations, while YAC end isolation is becoming relatively efficient. This tends to favor the true-end version of the strategy, which at the cost of isolating a few hundred YAC ends, requires somewhat fewer library screens and substantially fewer panel screens than the surrogate-end version.

Data Analysis

The primary data required for constructing STS-content maps include information about the YACs known to contain each STS, the YACs known not to contain each STS, and the YAC sizes. This information is used to deduce the relative order and spacings of the STSs and the YAC ends. Since there may be several different STS orders and spacings consistent with the data, "the map" must represent these inherent ambiguities as accurately and succinctly as possible.

Under the assumption that the presence of all STSs within all YACs is known, then the STSs in a contig can be shown to fall into groups, with the precise order of the groups unambiguously established. Within a group, any STS order would be consistent with the data. Of note, a group may (and often does) consist of a single STS, in which case there is no ambiguity. The YAC ends also fall into groups, and these groups also have a unique order. The information about STS and YAC end order may then be used for constructing linear programming problems to find bounds for the distance between any pair of points (STSs or YAC ends), and these problems are solvable by standard methods (e.g., the simplex method).⁽⁷⁰⁾ This analysis procedure has been implemented in the computer program SEGMAP (P. Green and Y. Xu, unpubl.). An example of an STS-content map con-

structed by SEGMAP is shown in Figure 6.

One of the principal challenges in analyzing STS-content mapping data is detecting data errors. Possible sources of error include false-positive or false-negative results during the panel and library screens (PCR- or hybridization-based), chimeric YACs, and nonunique STSs (reflecting multicopy chromosomal DNA segments). At present, SEGMAP identifies many chimeric YACs based on length inconsistencies, since large cocloned fragments often prevent the YAC from "fitting" into the contig map (see Fig. 6, where 8 of 13 known chimeric YACs in this contig were detected by SEGMAP). Chimeric YACs with small cocloned segments will typically go undetected by the program. The other types of data errors are to some extent detectable by SEGMAP in that a consistent group order cannot be established. Such errors can, in principle, be pinpointed to particular STSs or YACs by omitting them one by one and reanalyzing the remaining data until a consistent group order is found.

METHODS FOR GENERATING STSs

Intrinsic to any large STS-content mapping project, such as that described above for the mapping of whole human chromosomes, is the need to develop significant numbers of PCR assays, each amplifying a specific STS. This process will most typically involve the generation of STSs from a defined region of the genome (a specific chromosome or portion thereof). Given the paucity of mapped DNA markers in the human genome, virtually all of the STSs will need to be derived from random (anonymous) pieces of DNA obtained from the region of interest. The process of generating STSs from "targeted" genomic regions can broadly be viewed as occurring in three major steps: (1) the purification and sequencing of random DNA fragments from the relevant region; (2) the selection and synthesis of oligodeoxynucleotide primers for use in the PCR assays; and (3) the optimization and characterization of the STS-specific PCR assays.

Generation of Random DNA Sequences from Targeted Regions

A number of approaches have been

utilized to isolate target-specific DNA segments suitable for DNA sequencing. Many of the underlying strategies are similar to those employed by numerous investigators for the development of region-specific hybridization probes. However, for STS generation, a small amount (100–300 bp) of DNA sequence is obtained from the isolated DNA and used for designing a PCR assay, rather than testing the DNA segment for its performance as a hybridization probe.

One approach to enrich for DNA from a single chromosome employs the technique of flow-sorting.⁽⁷¹⁾ Flow-sorted DNA can be subcloned into a standard cloning vector (e.g., M13) and DNA sequence obtained from random clones.⁽⁶⁰⁾ Although this method has proven effective at enriching for DNA from a particular chromosome, its major drawback is the almost universal presence of contaminating, irrelevant DNA segments. The "purity" of a flow-sorted DNA preparation depends upon a number of factors, including the specific chromosome of interest and the cell type utilized (e.g., human cell lines vs. human-rodent hybrid cell lines). M13 libraries derived from flow-sorted DNA have been used effectively in the early efforts to generate human chromosome 7-specific STSs.⁽⁶⁰⁾

A second strategy for generating STSs from a defined genomic region employs microdissection techniques.⁽⁷²⁾ In this approach, cytogenetically distinct subchromosomal regions are physically excised and isolated from the remaining genomic DNA. Although microdissection was initially refined for use with the large and more easily manipulated polytene chromosomes of *Drosophila*,^(72–74) the advent of PCR has markedly enhanced the power of this approach. For example, microdissected DNA can be either cloned into a plasmid and amplified by PCR using universal primers,^(75–79) ligated to double-stranded linkers and amplified by PCR using linker-specific primers,^(73,80–81) or directly amplified using degenerate primers⁽⁷⁴⁾ or primers homologous to repetitive elements (e.g., *Alu*).⁽⁷⁶⁾ While most of the previous studies have used the isolated material as hybridization probes, the same DNA templates can readily be se-

quenced and used for STS development. Microdissection-based strategies are increasingly being used in efforts to isolate targeted segments of the human genome.^(75-79,81) Microdissection is of

more limited value in mapping efforts directed at the whole chromosome, since information on contig localization is most efficiently obtained later in such projects through long-range

mapping of selected STSs by appropriate methods (linkage mapping, radiation hybrid mapping, in situ hybridization). However, microdissection could serve an important role in obtaining

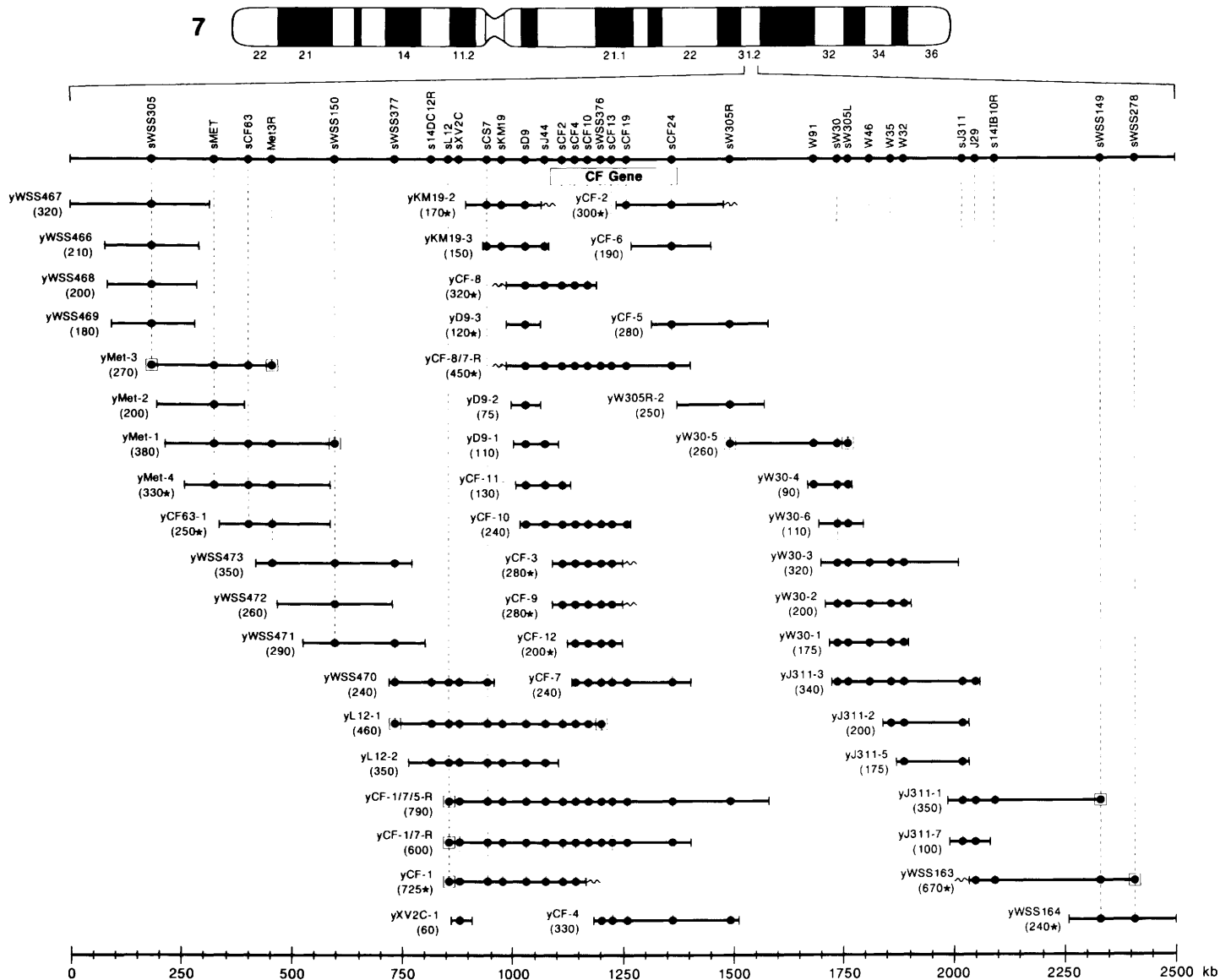


FIGURE 6 A YAC-based STS-content map of the CF region of human chromosome 7. The indicated physical map is based on the STS content and sizes of 50 YAC clones (depicted as horizontal bars). All of the YACs were isolated from the Washington University YAC library,^(6,10,17) except for yL12-1 and yL12-2, which were obtained from the YAC library of Albertson et al.,⁽⁹⁾ and yCF-8/7-R, yCF-1/7-R, and yCF-1/7/5-R, which were constructed by the in vivo recombination of overlapping YACs.⁽²¹⁾ Along the top are indicated the STSs and probe sequences (Met3R, W91, W46, W35, W32, J29); the latter were detected in the YACs by DNA-DNA hybridization rather than PCR. Also indicated is the position of the CF gene. The positioning of the STSs and probe sequences on the map as well as the alignment of the YACs within the contig were determined using the computer program SEGMAP (P. Green and Y. Xu, unpubl.), employing an algorithm designed to maximize the distances between STSs. The presence of an STS or probe sequence in a YAC is indicated by a closed circle. YAC-end-specific STSs or probe sequences are depicted with open squares at the appropriate positions on the YAC(s) of origin. The measured sizes of the YACs (in kb) are provided in parentheses under the clone's name. The presence of an asterisk after the YAC size indicates that the clone is likely to be chimeric. For some YACs, the position of the unrelated DNA segment was established (indicated by a wavy line). Some of the putative chimeric YACs were detected by the SEGMAP program, which found a length inconsistency (see text for details); in these cases, the depicted horizontal length of the clone corresponds to the size the program calculated the YAC could be without causing an inconsistency with other data. In some cases (yMet-4, yCF63-1, yKM19-2, yWSS164), the presence of a chimeric clone was detected only during the detailed characterization of the YACs (e.g., YAC end recovery, restriction mapping), without SEGMAP flagging a length inconsistency. SEGMAP found the maximum and minimum possible sizes of the contig (as measured by the distance between the terminal STSs) to be 2600 and 2400 kb, respectively. The total cloned DNA contained within the YAC contig thus represents roughly 1/60 of human chromosome 7.

STSs in poorly represented regions, should the distribution of otherwise unselected STSs turn out to be non-random.

Somatic hybrid cell lines, such as those made with human and rodent cells, provide a valuable source of DNA from a limited genomic region (e.g., a single human chromosome or chromosomal region). A number of approaches have been employed to "extract" the human DNA from the rodent DNA background. For example, λ or cosmid clones derived from somatic hybrid cell lines can be screened for the presence of human-specific repetitive sequences. Since most segments of human DNA present in large-insert λ (15–25 kb) or cosmid (45 kb) clones should contain at least one repetitive motif (e.g., *Alu*), the clones that contain human DNA can be identified by hybridization with radioactive probes generated with total human DNA. For example, this approach has been utilized to generate STSs from human chromosome 7.⁽⁶⁰⁾ Alternatively, segments of human DNA contained within somatic hybrid cell lines can be amplified by *Alu* element-mediated PCR. In this strategy, total DNA from the hybrid cell line is used as template in PCR reactions employing oligodeoxynucleotide primers specific to consensus *Alu* sequences. The PCR products generated are enriched for DNA segments residing adjacent to *Alu* motifs.^(30–31) The feasibility of this approach was recently demonstrated by Cole et al.,⁽⁸²⁾ with the generation of STSs from two specific regions of the human X chromosome. These general approaches can be tailored to the development of STSs from even more limited genomic regions by use of hybrid lines produced by irradiation-fusion techniques.⁽⁶⁸⁾ For example, irradiation of a rodent cell line containing a single human chromosome yields a set of new cell lines, each of which contains one or more smaller segments of the initial human chromosome and can serve as a more limited target for STS development.

A variant type of targeted DNA segments is represented by the sequences contained in expressed genes. Typically, clones from a tissue-specific cDNA library are randomly selected and sequenced. Specific portions of the

resulting sequence (e.g., the 3' untranslated regions) are used to develop gene-specific STSs^(83,84) (or ESTs, for *expressed sequence tags*, as proposed by Adams et al.⁽⁸³⁾), which in turn are mapped to specific chromosomes. The application of such strategies to whole chromosome mapping projects is hampered by the difficulty of selecting cDNA clones originating from a specific region of the genome.

For one version of the STS-content mapping strategy outlined in Figure 2, an important source of new STSs during phase 3 is the true ends of YACs at the ends of contigs. Various approaches are available for recovering the insert ends of YACs, including conventional subcloning methods,^(34–37) PCR-based strategies,^(30,31,38–42) and schemes based on the recovery of bacterial plasmids containing the YAC vector-insert junctions.^(2,43) Of note, the relatively high frequencies of chimeric YACs require the use of caution during the development of STSs from YAC ends, with the determination of the chromosomal origin required in virtually every instance.

Selection of PCR Primers

The approaches described above can be used effectively to generate tracts of DNA sequence from targeted genomic regions. The best strategy to progress from such a DNA sequence to the development of a corresponding PCR assay remains an area of ongoing investigation. Ideally, such a scheme would both yield robust and specific PCR assays in a large fraction of the cases and be conducive to the type of high-throughput environment inherent to large mapping projects. Furthermore, the required effort and cost of both synthesizing oligodeoxynucleotide primers and testing new PCR assays underscore the need for high success rates in the design of PCR assays. Toward that end, a number of computer programs have been developed to guide the selection of PCR primers.^(85–87) These programs attempt to account for the physical properties of DNA that are believed to affect the efficiency of PCR. Using one such program, OSP,⁽⁸⁵⁾ success rates of >90% have been achieved using random DNA sequences derived from human

chromosome 7⁽⁶⁰⁾ (E.D. Green, unpubl.).

Optimization and Characterization of STS-specific PCR Assays

The need to generate significant numbers of STSs within the context of a large mapping project has necessitated the development of routine procedures for optimizing and characterizing PCR assays. Thus, the testing of various PCR parameters (e.g., thermal cycling conditions, incubation buffers) must be kept to a minimum. This has been found to be relatively straightforward if during the design of PCR assays attempts are made to ensure somewhat uniform behavior among different assays (e.g., by the selection of PCR primers with calculated melting temperatures confined to a narrow range).⁽⁶⁰⁾ Furthermore, for YAC-based mapping projects, it is essential to evaluate the performance of the PCR assay on templates containing significant amounts of yeast DNA.⁽⁶⁰⁾ Demonstration of acceptable behavior of a PCR assay in the presence of such templates is critical for ensuring its ability to screen YAC libraries and to analyze YAC clones. Once established, each new PCR assay is typically characterized to confirm the genomic origin of its corresponding STS. In the case of human DNA-specific STSs, this test is typically performed by applying the PCR assay to a panel of human-rodent hybrid cell lines that differ in their compositions of human chromosomes.⁽⁶⁰⁾

EXPERIENCES WITH YAC-BASED STS-CONTENT MAPPING

The Cystic Fibrosis Region of Human Chromosome 7

One of the earliest tests of a YAC-based STS-content mapping approach was performed by Green and Olson⁽²¹⁾ using the cystic fibrosis (CF) region of human chromosome 7 as a model system. This DNA interval was chosen due to the significant amount of preexisting DNA sequence information that was available for developing STSs. Thus, the study's emphasis was to demonstrate the feasibility of the general strategy outlined in Figure 1B, rather than addressing issues related to STS generation.

For the initial phase of the study, seven PCR assays (corresponding to

STSs sXV2C, sKM19, sCF2, sCF10, sCF24, sW30, and sj3.11 in Fig. 6) were used to screen a human-YAC library, resulting in the isolation of 30 YAC clones. Subsequent analysis of these clones using additional PCR assays (corresponding to known STSs and YAC-end STSs) as well as a number of hybridization probes allowed the assembly of a YAC contig stretching over 1.5 Mb of contiguous DNA.⁽²¹⁾ Subsequently, additional YAC clones have been isolated and incorporated into the contig. For example, using newly developed YAC-end STSs (sWSS305, sWSS150, sL12, sWSS149), overlapping clones were isolated and used to expand the contig from both ends. The present configuration of the YAC contig from the human CF region is shown in Figure 6. Included within the expanded contig, which now spans roughly 2.5 Mb in size and has an average resolution of one uniquely ordered STS every 80 kb, is the known genetic marker MET (sMET), which along with J3.11 (sj3.11) served as a key flanking landmark during the search for the CF gene.⁽⁸⁸⁾ It is worth noting that most of the STS data shown in Figure 6 were generated by performing PCR directly with yeast cells⁽⁸⁹⁾ (E.D. Green, unpubl.) rather than with purified DNA, thereby improving the efficiency of the analyses. The YAC contig of the CF region shown in Figure 6 as well as a similar one assembled by Anand et al.⁽⁴⁷⁾ contain significantly larger amounts of contiguous cloned DNA than was achieved with bacterial-based cloning systems during the successful search for the CF gene.⁽⁸⁸⁾

Other Regions of the Human Genome

A number of other large regions of the human genome have been isolated and mapped using YAC clones. These include several segments of the X chromosome (most of 50 Mb in Xq24-Xq28, including one contig over 8 Mb in size within Xq26⁽⁵⁸⁾; the dystrophin gene^(50,51); the fragile X site^(53,56,90,91); and a 2.5-Mb segment of Xq26 (C.G. Cole and D.R. Bentley, pers. comm.), the major histocompatibility complex of chromosome 6,^(35,92) the MYCN proto-oncogene region of

chromosome 2,⁽⁹³⁾ the BCL2 proto-oncogene region of chromosome 18,^(39,46) and the Wilms' tumor locus of chromosome 11.⁽⁹⁴⁾ Of note, only a small number of these studies^(35,39,46,50,58) have employed strategies similar to the STS-content mapping approach described above, with the remainder employing more conventional hybridization-based schemes. If these experiences prove to be typical of the rest of the human genome, then the continuity achieved with a YAC-based STS-content mapping approach should be adequate for the mapping of whole human chromosomes by the strategies described above.

PROSPECTS AND LIMITATIONS

What possible obstacles to attaining chromosome maps can be foreseen? First, failure of our theoretical assumptions could affect the estimates of the total effort required. Of these, the assumption that STSs and YACs are derived from random points along the chromosome is perhaps most in danger of being incorrect. The most extensive data available on YAC distribution come from the analysis of screens performed on the YACs from human Xq24-q28 with several hundred single-copy probes.^(58,66,67) The number of YACs containing these various probes fit Poisson expectations fairly well, suggesting that the YAC distribution is close to random in this chromosomal region. In particular, there is little evidence for the existence of unclonable or poorly clonable DNA segments, although there is a suggestion that a small region of Xq28 is underrepresented in several YAC libraries. Of greater concern is the potential for a non-random distribution of STSs. For example, some of the methods being used for generating STSs⁽⁶⁰⁾ may be prone to problems relating to the uneven distribution of repetitive elements in the genome. If there are biases in the distribution of STSs, then the above strategy would produce larger than expected contigs from regions of the chromosome that are overrepresented for STSs, since with a higher density of STSs, there is a greater likelihood that YAC overlaps will be identified (Fig. 3A). Conversely, underrepresented regions may require more

extensive walking efforts than expected to ensure adequate coverage.

The presence of chimeric YACs potentially represents another obstacle. Of particular concern are clones having discontinuous segments derived from the same chromosome. In a library with only 20% chimeric YACs, this problem would be manageable; for example, of 3000 chromosome-specific YACs, roughly 30 would be expected to contain two unrelated segments from the same chromosome. Most of the latter clones could not be assembled into a consistent contig map, and, therefore, would be easily detected. A few chimeric YACs might spuriously merge contigs from two different regions of the chromosome; however, nearly all such cases would be detected during the long-range mapping of STSs from each contig end that will be necessary to orient the contigs. Unfortunately, the frequency of chimeric YACs in many whole genome YAC libraries appears to be substantially higher (e.g., 40-60% in the current Washington University YAC library^(6,10,17) made from total human DNA⁽³⁴⁾). In addition to increasing the likelihood of causing spurious contig mergers, such a high rate implies that the average size of contiguous inserts may be considerably smaller than expected. This would substantially decrease the contig sizes obtainable using a given number of STSs (Fig. 3A). Moreover, a high frequency of chimeric YACs hampers any strategy that relies heavily on the isolation of true contig ends. Thus, it is much more preferable to use YAC libraries with lower frequencies of chimeric clones.

A major technical challenge will be to obtain the high laboratory throughput necessary to process a thousand or more STSs per chromosome through the scheme shown in Figure 2. At present, we find that the rate-limiting step is performing library screens, rather than generating random STSs or isolating YAC ends; the goal of reducing the required number of library screens was, in fact, a major consideration in designing the STS-content mapping strategy. Nonetheless, performing even the few hundred library screens required under this strategy will be impractical without automation of much of the screening process, particularly

those steps involved in preparing and analyzing PCR reactions.

Finally, we have to expect that there will be additional obstacles due to the unpredictable nature of the genome itself. For example, a major unknown is the frequency of duplicated chromosomal regions, some of which may result in nonunique STSs. Like chimeric YACs, nonunique STSs should be detected if they do not occur too frequently, because they will usually prevent the assembly of consistent contig maps.

SUMMARY

The magnitude of the effort required to complete the human genome project will require constant refinements of the tools available for the large-scale study of DNA. Such improvements must include both the development of more powerful technologies and the reformulation of the theoretical strategies that account for the changing experimental capabilities. The two technological advances described here, PCR and YAC cloning, have rapidly become incorporated into the standard armamentarium of genome analysis and represent key examples of how technological developments continue to drive experimental strategies in molecular biology. Because of its high sensitivity, specificity, and potential for automation, PCR is transforming many aspects of DNA mapping.⁽⁹⁵⁾ Similarly, by providing the means to isolate and study larger pieces of DNA, YAC cloning has made practical the achievement of megabase-level continuity in physical maps. Taken together, these two technologies can be envisioned as providing a powerful strategy for constructing physical maps of whole chromosomes. Undoubtedly, future technological developments will promote even more effective mapping strategies. Nonetheless, the theoretical projections and practical experience described here suggest that constructing YAC-based STS-content maps of whole human chromosomes is now possible. Random STSs can be efficiently generated and used to screen collections of YAC clones, and contiguous YAC coverage of regions exceeding 2 Mb can be readily obtained. While the predicted laboratory effort required for mapping whole human chromosomes

remains daunting, it is clearly feasible.

ACKNOWLEDGMENTS

We thank Drs. Maynard Olson, David Schlessinger, and David Chaplin for critical review of this manuscript, David Schlessinger for providing YAC length information used in the simulations, and Maynard Olson and David Schlessinger for many illuminating discussions. We also thank Rose Mohr Tidwell, Jacquelyn Idol, Valerie Braden, and Y. Xu for technical assistance. This work was supported in part by National Institutes of Health grant P50-HG00201. E.D.G. is a Lucille P. Markey Scholar and this work was supported in part by a grant from the Lucille P. Markey Charitable Trust. P.G. was supported in part by a New Faculty Award from the Lucille P. Markey Charitable Trust.

REFERENCES

- Green, E.D. and R.H. Waterston. 1991. The human genome project: Prospects and implications for clinical medicine. *J. Am. Med. Assoc.* (in press).
- Burke, D.T., G.F. Carle, and M.V. Olson. 1987. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**: 806-812.
- Hieter, P., C. Connelly, J. Shero, M.K. McCormick, S. Antonarakis, W. Pavan, and R. Reeves. 1990. Yeast artificial chromosomes: Promises kept and pending. In *Genome analysis* (ed. K.E. Davies and S.M. Tilghman), pp. 83-120, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Schlessinger, D. 1990. Yeast artificial chromosomes: Tools for mapping and analysis of complex genomes. *Trends Genet.* **6**: 248-258.
- Burke, D.T. 1990. YAC cloning: Options and problems. *Genet. Anal. Tech. Applic.* **7**: 94-99.
- Burke, D.T. and M.V. Olson. 1991. Preparation of clone libraries in yeast artificial-chromosome vectors. *Methods Enzymol.* **194**: 251-270.
- Anand, R., A. Villasante, and C. Tyler-Smith. 1989. Construction of yeast artificial chromosome libraries with large inserts using fractionation by pulsed-field gel electrophoresis. *Nucleic Acids Res.* **17**: 3425-3433.
- Anand, R., J.H. Riley, R. Butler, J.C. Smith, and A.F. Markham. 1990. A 3.5 genome equivalent multi access YAC library: Construction, characterization, screening and storage. *Nucleic Acids Res.* **18**: 1951-1956.
- Albertsen, H.M., H. Abderrahim, H.M. Cann, J. Dausset, D. Le Paslier, and D. Cohen. 1990. Construction and characterization of a yeast artificial chromosome library containing seven haploid genome equivalents. *Proc. Natl. Acad. Sci.* **87**: 4256-4260.
- Imai, T. and M.V. Olson. 1990. Second-generation approach to the construction of yeast artificial-chromosome libraries. *Genomics* **8**: 297-303.
- McCormick, M.K., J.H. Shero, M.C. Cheung, Y.W. Kan, P.A. Hieter, and S.E. Antonarakis. 1989. Construction of human chromosome 21-specific yeast artificial chromosomes. *Proc. Natl. Acad. Sci.* **86**: 9991-9995.
- McCormick, M.K., J.H. Shero, C.J. Connelly, S.E. Antonarakis, and P. Hieter. 1990. Methods for cloning large DNA segments as artificial chromosomes in *S. cerevisiae*. *Technique* **2**: 65-71.
- Larin, Z., A.P. Monaco, and H. Lehrach. 1991. Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci.* **88**: 4123-4127.
- Green, E.D. and M.V. Olson. 1990. Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. *Proc. Natl. Acad. Sci.* **87**: 1213-1217.
- Heard, E., B. Davies, S. Feo, and M. Fried. 1989. An improved method for the screening of YAC libraries. *Nucleic Acids Res.* **17**: 5861.
- Kwiatkowski, T.J., Jr., H.Y. Zoghbi, S.A. Ledbetter, K.A. Ellison, and A.C. Chinault. 1990. Rapid identification of yeast artificial chromosome clones by matrix pooling and crude lysate PCR. *Nucleic Acids Res.* **18**: 7191-7192.
- Brownstein, B.H., G.A. Silverman, R.D. Little, D.T. Burke, S.J. Korsmeyer, D. Schlessinger, and M.V. Olson. 1989. Isolation of single-copy human genes from a library of yeast artificial chromosome clones. *Science* **244**: 1348-1351.
- Traver, C.N., S. Klapholz, R.W. Hyman, and R.W. Davis. 1989. Rapid

- screening of a human genomic library in yeast artificial chromosomes for single-copy sequences. *Proc. Natl. Acad. Sci.* **86**: 5898-5902.
19. Lai, E. and C. Cantrell. 1989. Rapid colony screening of YAC libraries by using alginate as matrix support. *Nucleic Acids Res.* **17**: 8008.
 20. Mendez, M.J., S. Klapholz, B.H. Brownstein, and R.M. Gemmill. 1991. Rapid screening of a YAC library by pulsed-field gel Southern blot analysis of pooled YAC clones. *Genomics* **10**: 661-665.
 21. Green, E.D. and M.V. Olson. 1990. Chromosomal region of the cystic fibrosis gene in yeast artificial chromosomes: A model for human genome mapping. *Science* **250**: 94-98.
 22. Silverman, G.A., E.D. Green, R.L. Young, J.I. Jockel, P.H. Domer, and S.J. Korsmeyer. 1990. Meiotic recombination between yeast artificial chromosomes yields a single clone containing the entire BCL2 protooncogene. *Proc. Natl. Acad. Sci.* **87**: 9913-9917.
 23. Pavan, W.J., P. Hieter, and R.H. Reeves. 1990. Generation of deletion derivatives by targeted transformation of human-derived yeast artificial chromosomes. *Proc. Natl. Acad. Sci.* **87**: 1300-1304.
 24. Srivastava, A.K. and D. Schlessinger. 1991. Vectors for inserting selectable markers in vector arms and human DNA inserts of yeast artificial chromosomes (YACs). *Gene* (in press).
 25. Cellini, A., R.M. Lacatena, and G.P. Tocchini-Valentini. 1991. Detection of homologous recombination between yeast artificial chromosomes with overlapping inserts. *Nucleic Acids Res.* **19**: 997-1000.
 26. Pavan, W.J., P. Hieter, and R.H. Reeves. 1990. Modification and transfer into an embryonal carcinoma cell line of a 360-kilobase human-derived yeast artificial chromosome. *Mol. Cell. Biol.* **10**: 4163-4169.
 27. Reeves, R.H., W.J. Pavan, and P. Hieter. 1990. Modification and manipulation of mammalian DNA cloned as YACs. *Genet. Anal. Techn. Appl.* **7**: 107-113.
 28. Campbell, C., R. Gulati, A.K. Nandi, K. Floy, P. Hieter, and R.S. Kucheralapati. 1991. Generation of a nested series of interstitial deletions in yeast artificial chromosomes carrying human DNA. *Proc. Natl. Acad. Sci.* **88**: 5744-5748.
 29. Nelson, D.L. 1990. Current methods for YAC clone characterization. *Genet. Anal. Techn. Appl.* **7**: 100-106.
 30. Nelson, D.L., A. Ballabio, M.F. Victoria, M. Pieretti, R.D. Bies, R.A. Gibbs, J.A. Maley, A.C. Chinault, T.D. Webster, and C.T. Caskey. 1991. Alu-primed polymerase chain reaction for regional assignment of 110 yeast artificial chromosome clones from the human X chromosome: Identification of clones associated with a disease locus. *Proc. Natl. Acad. Sci.* **88**: 6157-6161.
 31. Nelson, D.L. 1991. Interspersed repetitive sequence polymerase chain reaction (IRS PCR) for generation of human DNA fragments from complex sources. *Methods: Companion Methods Enzymol.* **2**: 60-74.
 32. Monaco, A.P., V.M.S. Lam, G. Zehetner, G.G. Lennon, C. Douglas, D. Nizetic, P.N. Goodfellow, and H. Lehrach. 1991. Mapping irradiation hybrids to cosmid and yeast artificial chromosome libraries by direct hybridization of Alu-PCR products. *Nucleic Acids Res.* **19**: 3315-3318.
 33. Pieretti, M., R. Tonlorenzi, and A. Ballabio. 1991. Rapid assembly of lambda phage contigs within YAC clones. *Nucleic Acids Res.* **19**: 2795-2796.
 34. Green, E.D., H.C. Riethman, J.E. Dutchik, and M.V. Olson. 1991. Detection and characterization of chimeric yeast artificial-chromosome clones. *Genomics* (in press).
 35. Bronson, S.K., J. Pei, P. Taillon-Miller, M.J. Chorney, D.E. Geraghty, and D.D. Chaplin. 1991. Isolation and characterization of yeast artificial chromosome clones linking the HLA-B and HLA-C loci. *Proc. Natl. Acad. Sci.* **88**: 1676-1680.
 36. Huxley, C., Y. Hagino, D. Schlessinger, and M.V. Olson. 1991. The human HPRT gene on a yeast artificial chromosome is functional when transferred to mouse cells by cell fusion. *Genomics* **9**: 742-750.
 37. Gnirke, A., T.S. Barnes, D. Patterson, D. Schild, T. Featherstone, and M.V. Olson. 1991. Cloning and in vivo expression of the human GART gene using yeast artificial chromosomes. *EMBO J.* **10**: 1629-1634.
 38. Ochman, H., M.M. Medhora, D. Garza, and D.L. Hartl. 1990. Amplification of flanking sequences by inverse PCR. In *PCR protocols: A guide to methods and applications* (ed. M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White), pp. 219-227, Academic Press, San Diego.
 39. Silverman, G.A., R.D. Ye, K.M. Pollock, J.E. Sadler, and S.J. Korsmeyer. 1989. Use of yeast artificial chromosome clones for mapping and walking within human chromosome segment 18q21.3. *Proc. Natl. Acad. Sci.* **86**: 7485-7489.
 40. Riley, J., R. Butler, D. Ogilvie, R. Finnear, D. Jenner, S. Powell, R. Anand, J.C. Smith, and A.F. Markham. 1990. A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.* **18**: 2887-2890.
 41. Breukel, C., J. Wijnen, C. Tops, H. v/d Klift, H. Dauwerse, and P. Meera Khan. 1990. Vector-Alu PCR: A rapid step in mapping cosmids and YACs. *Nucleic Acids Res.* **18**: 3097.
 42. Lagerstrom, M., J. Parik, H. Malmgren, J. Stewart, U. Pettersson, and U. Landegren. 1991. Capture PCR: Efficient amplification of DNA fragments adjacent to a known sequence in human and YAC DNA. *PCR Methods Applic.* **1**: 111-119.
 43. Hermanson, G.G., M.F. Hoekstra, D.L. McElligott, and G.A. Evans. 1991. Rescue of end fragments of yeast artificial chromosomes by homologous recombination in yeast. *Nucleic Acids Res.* (in press).
 44. Coulson, A., R. Waterston, J. Kiff, J. Sulston, and Y. Kohara. 1988. Genome linking with yeast artificial chromosomes. *Nature* **335**: 184-186.
 45. Garza, D., J.W. Ajioka, D.T. Burke, and D.L. Hartl. 1989. Mapping the Drosophila genome with yeast artificial chromosomes. *Science* **246**: 641-646.
 46. Silverman, G.A., J.I. Jockel, P.H. Domer, R.M. Mohr, P. Taillon-Miller, and S.J. Korsmeyer. 1991. Yeast artificial chromosome cloning of a two-megabase-size contig within chromosomal band 18q21 establishes physical linkage between BCL2 and plasminogen activator inhibitor type-2. *Genomics* **9**: 219-228.
 47. Anand, R., D.J. Ogilvie, R. Butler, J.H. Riley, R.S. Finnear, S.J. Powell, J.C.

- Smith, and A.F. Markham. 1991. A yeast artificial chromosome contig encompassing the cystic fibrosis locus. *Genomics* **9**: 124–130.
48. Joslyn, G., M. Carlson, A. Thliveris, H. Albertsen, L. Gelbert, W. Samowitz, J. Groden, J. Stevens, L. Spirio, M. Robertson, L. Sargeant, K. Krapcho, E. Wolff, R. Burt, J.P. Hughes, J. Warrington, J. McPherson, J. Wasmuth, D. Le Paslier, H. Abderrahim, D. Cohen, M. Leppert, and R. White. 1991. Identification of deletion mutations and three new genes at the familial polyposis locus. *Cell* **66**: 601–613.
49. Kinzler, K.W., M.C. Nilbert, L.-K. Su, B. Vogelstein, T.M. Bryan, D.B. Levy, K.J. Smith, A.C. Preisinger, P. Hedge, D. McKechnie, R. Finnear, A. Markham, J. Groffen, M.S. Boguski, S.F. Altschul, A. Horii, H. Ando, Y. Miyoshi, Y. Miki, I. Nishisho, and Y. Nakamura. 1991. Identification of FAP locus genes from chromosome 5q21. *Science* **253**: 661–665.
50. Coffey, A.J., R.G. Roberts, E.D. Green, C.G. Cole, R. Anand, F. Giannelli, and D.R. Bentley, in preparation.
51. Monaco, A.P., A.P. Walker, I. Millwood, Z. Larin, and H. Lehrach, in preparation.
52. Wallace, M.R., D.A. Marchuk, L.B. Andersen, R. Letcher, H.M. Odeh, A.M. Saulino, J.W. Fountain, A. Brereton, J. Nicholson, A.L. Mitchell, B.H. Brownstein, and F.S. Collins. 1990. Type 1 neurofibromatosis gene: Identification of a large transcript disrupted in three NF1 patients. *Science* **249**: 181–186.
53. Heitz, D., F. Rousseau, D. Devys, S. Saccone, H. Abderrahim, D. Le Paslier, D. Cohen, A. Vincent, D. Toniolo, G. Della Valle, S. Johnson, D. Schlessinger, I. Oberle, and J.L. Mandel. 1991. Isolation of sequences that span the fragile X and identification of a fragile X-related CpG island. *Science* **251**: 1236–1239.
54. Kremer, E.J., M. Pritchard, M. Lynch, S. Yu, K. Holman, E. Baker, S.T. Warren, D. Schlessinger, G.R. Sutherland, and R.I. Richards. 1991. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)_n. *Science* **252**: 1711–1714.
55. Yu S., M. Pritchard, E. Kremer, M. Lynch, J. Nancarrow, E. Baker, K. Holman, J.C. Mulley, S.T. Warren, D. Schlessinger, G.R. Sutherland, and R.I. Richards. 1991. Fragile X genotype characterized by an unstable region of DNA. *Science* **252**: 1179–1181.
56. Verkerk A.J.M.H., M. Pieretti, J.S. Sutcliffe, Y.-H. Fu, D.P.A. Kuhl, A. Pizzuti, O. Reiner, S. Richards, M.F. Victoria, F. Zhang, B.E. Eussen, G.-J.B. van Ommen, L.A.J. Blonden, G.J. Riggins, J.L. Chastain, C.B. Kunst, H. Galjaard, C.T. Caskey, D.L. Nelson, B.A. Oostra, and S.T. Warren. 1991. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**: 905–914.
57. Oberle, I., F. Rousseau, D. Heitz, C. Kretz, D. Devys, A. Hanauer, J. Boue, M.F. Bertheas, and J.L. Mandel. 1991. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* **252**: 1097–1102.
58. Schlessinger, D., R.D. Little, D. Freije, F. Abidi, I. Zucchi, G. Porta, G. Pilia, R. Nagaraja, S.K. Johnson, J.-Y. Yoon, A. Srivastava, J. Kere, G. Palmieri, A. Ciccodicola, V. Montanaro, G. Romano, A. Casamassimi, and M. D'Urso. 1991. Yeast artificial chromosome-based genome mapping: Some lessons from Xq24-q28. *Genomics* (in press).
59. Olson, M., L. Hood, C. Cantor, and D. Botstein. 1989. A common language for physical mapping of the human genome. *Science* **245**: 1434–1435.
60. Green, E.D., R.M. Mohr, J.R. Idol, M. Jones, J.M. Buckingham, L.L. Deaven, R.K. Moyzis, and M.V. Olson. 1991. Systematic generation of sequence-tagged sites for physical mapping of human chromosomes: Application to the mapping of human chromosome 7 using yeast artificial chromosomes. *Genomics* (in press).
61. Evans, G.A. 1991. Combinatoric strategies for genome mapping. *BioEssays* **13**: 39–44.
62. Freije, D. and D. Schlessinger, in preparation.
63. Torney, D.C. 1991. Mapping using unique sequences. *J. Mol. Biol.* **217**: 259–264.
64. Barillot, E., J. Dausset, and D. Cohen. 1991. Theoretical analysis of a physical mapping strategy using random single-copy landmarks. *Proc. Natl. Acad. Sci.* **88**: 3917–3921.
65. Arratia, R., E.S. Lander, S. Tavare, and M.S. Waterman. 1991. Genomic mapping by anchoring random clones: A mathematical analysis. *Genomics* (in press).
66. Abidi, F.E., M. Wada, R.D. Little, and D. Schlessinger. 1990. Yeast artificial chromosomes containing human Xq24–Xq28 DNA: Library construction and representation of probe sequences. *Genomics* **7**: 363–376.
67. Wada, M., R.D. Little, F. Abidi, G. Porta, T. Labella, T. Cooper, G. Della Valle, M. D'Urso, and D. Schlessinger. 1990. Human Xq24–Xq28: Approaches to mapping with yeast artificial chromosomes. *Am. J. Hum. Genet.* **46**: 95–106.
68. Cox, D.R., M. Burmeister, E. Roydon Price, S. Kim, and R.M. Myers. 1990. Radiation hybrid mapping: A somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* **250**: 245–250.
69. Lichter, P., C.-J.C. Tan, K. Call, G. Hermanson, G.A. Evans, D. Housman, and D.C. Ward. 1990. High resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science* **247**: 64–69.
70. Danzig, G.B. 1963. *Linear programming and extensions*. Princeton University Press, Princeton.
71. Deaven, L.L., M.A. Van Dilla, M.F. Bartholdi, A.V. Carrano, L.S. Cram, J.C. Fuscoe, J.W. Gray, C.E. Hildebrand, R.K. Moyzis, and J. Perlman. 1986. Construction of human chromosome-specific DNA libraries from flow-sorted chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* **51**: 159–167.
72. Saunders, R.D.C. 1990. Short cuts for genomic walking: Chromosome microdissection and the polymerase chain reaction. *BioEssays* **12**: 245–248.
73. Johnson, D.H. 1990. Molecular cloning of DNA from specific chromosomal regions by microdissection and sequence-independent amplification of DNA. *Genomics* **6**: 243–251.
74. Wesley, C.S., M. Ben, M. Kreitman, N. Hagag, and W.F. Eanes. 1990. Cloning regions of the Drosophila genome by microdissection of

- polytene chromosome DNA and PCR with nonspecific primer. *Nucleic Acids Res.* **18**: 599-603.
75. Ludecke, H.-J., G. Senger, U. Claussen, and B. Horsthemke. 1989. Cloning defined regions of the human genome by microdissection of banded chromosomes and enzymatic amplification. *Nature* **338**: 348-350.
 76. Djabali, M., C. Nguyen, I. Biunno, B.A. Oostra, M.-G. Mattei, J.E. Ikeda, and B.R. Jordan. 1991. Laser microdissection of the fragile X region: Identification of cosmid clones and of conserved sequences in this region. *Genomics* **10**: 1053-1060.
 77. MacKinnon, R.N., M.C. Hirst, M.V. Bell, J.E.V. Watson, U. Claussen, H.-J. Ludecke, G. Senger, B. Horsthemke, and K.E. Davies. 1990. Microdissection of the fragile X region. *Am. J. Hum. Genet.* **47**: 181-187.
 78. Buiting, K., M. Neumann, H.-J. Ludecke, G. Senger, U. Claussen, J. Antich, E. Passarge, and B. Horsthemke. 1990. Microdissection of the Prader-Willi syndrome chromosome region and identification of potential gene sequences. *Genomics* **6**: 521-527.
 79. Davis, L.M., G. Senger, H.-J. Ludecke, U. Claussen, B. Horsthemke, S.S. Zhang, B. Metzroth, K. Hohenfellner, B. Zabel, and T.B. Shows. 1990. Somatic cell hybrid and long-range physical mapping of 11p13 microdissected genomic clones. *Proc. Natl. Acad. Sci.* **87**: 7005-7009.
 80. Saunders, R.D.C., D.M. Glover, M. Ashburner, I. Siden-Kiamos, C. Louis, M. Monastirioti, C. Savakis, and F. Kafatos. 1989. PCR amplification of DNA microdissected from a single polytene chromosome band: A comparison with conventional microcloning. *Nucleic Acids Res.* **17**: 9027-9037.
 81. Kao, F.-t. and J.-w. Yu. 1991. Chromosome microdissection and cloning in human genome and genetic disease analysis. *Proc. Natl. Acad. Sci.* **88**: 1844-1848.
 82. Cole, C.G., P.N. Goodfellow, M. Bobrow, and D.R. Bentley. 1991. Generation of novel sequence tagged sites (STSs) from discrete chromosomal regions using Alu-PCR. *Genomics* **10**: 816-826.
 83. Adams, M.D., J.M. Kelley, J.D. Gocayne, M. Dubnick, M.H. Polymeropoulos, H. Xiao, C.R. Merril, A. Wu, B. Olde, R.F. Moreno, A.R. Kerlavage, W.R. McCombie, and J.C. Venter. 1991. Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science* **252**: 1651-1656.
 84. Wilcox, A.S., A.S. Khan, J.A. Hopkins, and J.M. Sikela. 1991. Use of 3' untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: Implications for an expression map of the genome. *Nucleic Acids Res.* **19**: 1837-1843.
 85. Hillier, L. and P. Green. 1991. OSP: A computer program for choosing PCR and DNA sequencing primers. *PCR Methods Applic.* **1**: 124-128.
 86. Lowe, T., J. Sharefkin, S.Q. Yang, and C.W. Dieffenbach. 1990. A computer program for selection of oligonucleotide primers for polymerase chain reactions. *Nucleic Acids Res.* **18**: 1757-1761.
 87. Rychlik, W. and R.E. Rhoads. 1989. A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing, and in vitro amplification of DNA. *Nucleic Acids Res.* **17**: 8543-8551.
 88. Rommens, J.M., M.C. Iannuzzi, B.-S. Kerem, M.L. Drumm, G. Melmer, M. Dean, R. Rozmahel, J.L. Cole, D. Kennedy, N. Hidaka, M. Zsiga, M. Buchwald, J.R. Riordan, L.-C. Tsui, and F.S. Collins. 1989. Identification of the cystic fibrosis gene: Chromosome walking and jumping. *Science* **245**: 1059-1065.
 89. Huxley, C., E.D. Green, and I. Dunham. 1990. Rapid assessment of *S. cerevisiae* mating type by PCR. *Trends Genet.* **6**: 236.
 90. Dietrich, A., P. Kioschis, A.P. Monaco, B. Gross, B. Korn, S.V. Williams, D. Sheer, D. Heitz, I. Oberle, D. Toniolo, S.T. Warren, H. Lehrach, and A. Poustka. 1991. Molecular cloning and analysis of the fragile X region in man. *Nucleic Acids Res.* **19**: 2567-2572.
 91. Hirst, M.C., K. Rack, Y. Nakahori, A. Roche, M.V. Bell, G. Flynn, Z. Christadoulou, R.N. MacKinnon, M. Francis, A.J. Littler, R. Anand, A.-M. Poustka, H. Lehrach, D. Schlessinger, M. D'Urso, V.J. Buckle, and K.E. Davies. 1991. A YAC contig across the fragile X site defines the region of fragility. *Nucleic Acids Res.* **19**: 3283-3288.
 92. Ragoussis, J., A. Monaco, I. Mockridge, E. Kendall, R.D. Campbell, and J. Trowsdale. 1991. Cloning of the HLA class II region in yeast artificial chromosomes. *Proc. Natl. Acad. Sci.* **88**: 3753-3757.
 93. Schneider, S.S., B.A. Zehnbaauer, B. Vogelstein, and G.M. Brodeur. 1991. Yeast artificial chromosome (YAC) vector cloning of the MYCN amplified domain in neuroblastomas. *Adv. Neuroblastoma Res.* **3**: 71-76.
 94. Bonetta, L., S.E. Kuehn, A. Huang, D.J. Law, L.M. Kalikin, M. Koi, A.E. Reeve, B.H. Brownstein, H. Yeger, B.R.G. Williams, and A.P. Feinberg. 1990. Wilms tumor locus on 11p13 defined by multiple CpG island-associated transcripts. *Science* **250**: 994-997.
 95. Rose, E.A. 1991. Applications of the polymerase chain reaction to genome analysis. *FASEB J.* **5**: 46-54.