459

DOE/ER/60692-2

MAPPING AND ORDERED CLONING OF THE HUMAN X CHROMOSOME

707489

Progress Report

November 1988 - October 1989

C. Thomas Caskey, M.D.
David L. Nelson, Ph.D.
David H. Ledbetter, Ph.D.

Institute for Molecular Genetics

Baylor College of Medicine

Houston, Texas 77030

REPOSITORY Oak Ridge Operations Off (ORO)
COLLECTION Procurement & Contracts Div. General Acquisition Br. AD-421
BOX No. Active Records Entered for Human Radiation Exps. Pjt.
FOLDER

October 1989

PREPARED FOR THE U.S. DEPARTMENT OF ENERGY
UNDER GRANT NUMBER DE-FG05-88ER60692

1021429

AD-421

A. Progress 12/1/88-11/30/89

Progress in this project has continued in several areas. These will each be discussed in detail below. Overall, the progress has been excellent, with some of the major goals of the application achieved. The areas currently underway are: 1) The establishment of a large number of cloned DNA segments from the region of the X chromosome retained in the X3000-11.1 cell line, 2) The development and use of a sub-chromosomal localization panel of somatic cell hybrids for mapping cloned probes from the X, 3) The development of *Alu* PCR as a technique for isolating human-specific DNA from a variety of sources, and 4) The establishment of methods for identification and isolation of expressed sequences.

Isolation and Mapping of Clones from X

In the previous progress report, the assignment of a large number of clones in phage lambda to regions of Xq24-qter was discussed. In the present period, similar results have been obtained with clones in the yeast artificial chromosome (YAC) vector pYAC4. This has been accomplished through the construction of two YAC libraries, and the development of the *Alu* PCR methodology for the preparation of DNA fragments as probe from each YAC clone.

Recombinant libraries in pYAC4 were constructed from DNA derived from both the X3000-11.1 and 4.12 cell lines, which retain Xq24-qter and X respectively, as their only human material on a hamster background. Screening a portion of each library with radiolabeled total human DNA by colony hybridization has provided a total of 400 human clones, approximately 200 from each library. Insert sizes average 125-150 kb in each library. This represents approximately three quarters of the human sequences in the X3000-11.1 cell line, and about one fifth of the X chromosome. Clones have been stored on plates and in glycerol stocks and stabs. DNA and agarose plugs have been prepared from each clone, and re-hybridized to confirm the presence of a human insert, and to estimate the size of each clone.

In order to characterize the YAC library, we sought to assign each clone to subregions of the X chromosome in a manner similar to that reported for the lambda clones, using the somatic cell hybrid mapping panels we have collected. Preparation of probes from individual YAC clones has been accomplished through the application of the *Alu* PCR method we have developed. The use of human-specific *Alu* primers in the polymerase chain reaction applied to total DNA from yeast cells harboring YAC clones has provided amplified sequences from approximately 80% of the clones attempted. The use of primers directed to the vector sequences adjacent to the human DNA insert has also allowed the isolation of end-specific probes. Many YAC clones demonstrate multiple PCR products in a single reaction. To regionally assign a clone, PCR product(s) are gel purified and labeled by random oligonucleotide labeling. These PCR probes are then hybridized to the mapping panel on a Southern filter using total human DNA to block repetitive sequence hybridization. About 80% of PCR products can be used as probes in this manner. The remainder represent sequences too repetitive to be used or contain contaminating hamster sequences (YAC clones are not routinely colony purified). Using this method, and two mapping panels (one for the Xq24-qter region, one for the entire X), we have assigned 46 clones in the Xq24-qter region, and 23 in the whole X library. This represents 7 Mbp in the Xq24-qter region (~20%) and 3.7 Mbp in the entire X (~2%) of assigned sequences. These results are shown in Figures 1 and 2, and the methods are described in the reprinted article from *PNAS* (Nelson *et al.*).

Alu PCR probes derived from YAC clones have also been used for the identification of overlapping YAC clones. Using the assigned clones as starting points, we hope to derive overlapping sequences to begin to develop the contigs required for mapping the entire X. DNA from each YAC clone has been immobilized on a "slot blot" as a convenient method of screening all clones at once. Using an *Alu* PCR product from the YAC clone RS88 (discussed in detail below) the overlapping clone RS41 was identified. As the clone inventory increases, this simple method for the identification of YAC contigs will be of greater utility as the likelihood of overlap identification will increase. The slot blots have also offered the opportunity to readily screen the library for the presence of

1021430

clones related to known cloned sequences (a Factor VIII YAC has been found), as well as those related to the Xq24-qter lambda clones and the expressed sequence clones (see below).

Thus far, two YAC clones have been identified that will contain or lie very close to disease loci. The first (RS88) identifies the breakpoint region of the X chromosome from a female Lowe's syndrome (Xq25) patient with a balanced X;3 translocation, presumed to have disrupted the gene. This patient's chromosome has been used to define a breakpoint for the somatic cell hybrid mapping panel (Lowe-3). *Alu* PCR product probes used to assign RS88 provided evidence for the presence of the breakpoint on this clone. In X chromosomes from the mapping panel, two *Eco*RI fragments hybridized with the probe when this region was present in a hybrid. Hybrids without the region (Xq25) showed no hybridization. However, in the Lowe-3 line, only one of the *Eco*RI fragments was detected. Further PCR products from RS88 identified restriction fragments of altered size in the Lowe translocation, and have allowed the breakpoint to be localized to a 700 base pair region. Portions of this region demonstrate cross-species hybridization, indicative of conserved, possibly coding sequences. Sequence analysis of this region is underway. A survey of a collection of 60 sporadic Lowe syndrome cases has been initiated in collaboration with Robert Nussbaum at U. Penn. to define the nature of these mutations. It is anticipated that if the gene defective in Lowe syndrome is contained within this YAC clone, it will be completely characterized by the next reporting period. The second YAC of interest was derived from the whole X library. Five YACs were found to be present in the Xp21.2-p22.3 interval (see Figure 2). These were tested (by Dr. Ed McCabe at Baylor) against a panel of five DNAs from patients with multiple diseases (contiguous gene syndromes) due to deletions of or distal to the Duchenne muscular dystrophy gene. *Alu* PCR product probes from one of these YACs (YHX39) are deleted in all patients with congenital Adrenal Hypoplasia (AHC), and can be excluded from the DMD and Glycerol Kinase genes on the proximal side and from distal regions defined by the probes C7, L1.4, B24 and 99-6 with the deletion panel. This shortest region of overlap, containing the AHC gene, is estimated to be no more than 1 Mbp. YHX39 contains over 200 kb, thus the gene defective in this disorder may be present in the YAC. At a minimum, identification of the YAC represents the first entry into this region.

Additional characterization of mapped YAC clones will proceed with sequence analysis of *Alu* PCR products derived from the clones, especially those clones located in interesting regions. Products from vector-adjacent PCR have been re-amplified using a single primer (the *Alu* primer) to create single stranded sequence. This has been used directly as template in sequencing reactions, using ^{32}P and the relevant vector primer. Over one hundred bases of sequence from both ends of the RS46 clone (located near the Fra(X) mutation in Xq27.3) has been generated. This represents the first direct sequence analysis of a YAC insert, and allows the development of sequence tagged sites (STS) from YAC clones. The use of fluorescent primers is under development along with more facile and rapid methods of isolating and sequencing portions of the YAC clone. The goal of this work is the development of a simple means for establishing sequence tagged sites (STS) from YAC clones.

Use of the *Alu* PCR Technique

One of the major goals of the original proposal was the establishment of the *Alu* PCR method. This goal has been met. Oligonucleotides derived from the consensus *Alu* repeat sequence from humans have been shown to amplify human DNA specifically from a variety of sources (see attached reprint, Nelson *et al.*). This method has proven extraordinarily useful in the analysis of YAC clones (see above). It has also been valuable for the characterization of somatic cell hybrids, including X-irradiation hybrids derived from the human X chromosome provided by Dr. Peter Goodfellow of London. Through the analysis of the pattern of PCR products amplified from various cell lines, it is possible to determine the amount and extent of overlap of the sequences retained in each hybrid. Furthermore, the isolation of individual fragments from among the *Alu* PCR products has allowed identification of region-specific probes. An extension of this technique which clearly demonstrates the concept is presented in the enclosed manuscript (Ledbetter *et al.*) describing amplification using a human-specific oligonucleotide derived from the L1 (*Kpn*) repeat. Multiple Xq28 probes were identified strictly through the analysis of patterns of amplification among somatic

1021431

cell hybrids retaining X regions, and confirmed by hybridization. The establishment of L1 repeat PCR extends and complements the *Alu* PCR method.

Refinement of Somatic Cell Mapping Panels

The expansion to characterization of clones from a whole X YAC library has required the addition of several somatic cell hybrids for mapping clones outside of the Xq24-qter region. These are shown in Figure 2. Several other breakpoints in distal Xp have been provided by Dr. Andrea Ballabio, an addition to the group, who has a long-standing interest in the contiguous gene syndromes in Xp22.3 and has collected a very extensive group of patients with cytogenetic abnormalities. Additional breakpoints for the Xq24-qter region have also been collected, especially in Xq28, and will be utilized for the sub-fractionation of groups of clones. Continued division of the chromosome into intervals defined by these chromosomes is the strategy we hope to apply to categorize clones as a strategy for the ultimate overlap cloning. The use of patient materials, as demonstrated by the Lowe syndrome and Adrenal Hypoplasia findings is important to the goals of defining disease genes of the X chromosome.

Development of Methods to Identify Expressed Sequences

This area represents a departure from the stated aims in the original proposal, but is of sufficient importance that we have devoted some resources to its development. Establishment of techniques for the identification and/or isolation of DNA sequences containing expressed genes will be important not only for genome projects in general, but also for the pursuit of disease loci in particular. In this regard, we have attempted to adapt the *Alu* PCR method to the isolation of expressed sequences from somatic cell hybrids, again utilizing X3000-11.1 as a test case. In this general protocol, oligonucleotides with human *Alu* specificity have been used to prime synthesis of cDNA from nuclear RNA isolated from either X3000-11.1 or its rodent parent, RJK88. Using the oligonucleotide TC-65, hncDNA is produced from X3000-11.1, but not from RJK88. This human specificity appears to derive from transcribed sequences from the human region contained in the hybrid. A library of over 100 clones has been established. In a sample of 25 clones, 21 contain human sequences from Xq24-qter, each specifically mapping to subregions. The remaining 4 clones represented hamster sequences (2) or clones with sequences hybridizing to repeats in both genomes (2). One clone was identified as containing an intron segment from the human HPRT gene (Xq26). These results are contained in an enclosed manuscript (Corbo, *et al.*). The ability to directly clone expressed sequences from a specific region has great potential for the isolation of disease genes, but also for the analysis of the genome. Expressed regions are of most interest to the definition of the genome, and the capacity to categorize genes by region is of obvious value.

B. Objectives 12/1/89-12/1/90

In the coming year, it is anticipated that efforts will continue on the projects addressed above. The isolation of additional YAC clones from the Xq24-qter and X libraries will be a primary goal. The previous year's progress with YAC clones has been quite remarkable, and with continued efforts along the same lines, it is conceivable that over 200 localized YAC clones from the X chromosome will be in hand by the end of the next reporting period, with the potential of .5-1x coverage of Xq24-qter in localized clones.

The current methods for localizing YAC clones proceed in a one-by-one fashion. Methods to allow group localization are under development. One method described above, is the use of currently localized clones to identify overlapping clones. In the case described, this has provided an overlapped region of ~300 kb containing two chromosome breakpoints (Lowe-3, and RJK734--see Figure 1). The use of multiple *Alu* PCR products from multiple YAC clones already assigned to probe slot blots containing the entire collection is being tested. It appears that the human repetitive sequences and yeast DNA contained in the probe will not inhibit this type of hybridization. In this way, overlapping clones and region-specific clones can be identified rapidly. Additional regionally

1021432

assigned probes will be employed in this manner as well. Over 50 lambda clones have been regionally assigned in previous experiments, and will be used to identify YACs containing sequences from the same regions. Again, this will be done with multiple probes to the YAC slot blots. A collaboration with Dr. Steve Warren of Emory University has been initiated to identify common sequences in our YAC libraries with a cosmid library he has developed from Xq28. Groups of cosmids will be used as probes to the slot blots, and Xq28 localized YACs will be used to identify related cosmids in his library.

Alu PCR products from somatic cell hybrids will also be employed for region-assignment of YACs. The products derived from somatic cell hybrids with specific X chromosome regions will be radiolabeled and used to probe YAC clones on slot blots. An example of this approach would be the amplification by *Alu* PCR of the region in the X3000-11.1 cell line, and the use of a probe derived from this product to identify Xq24-qter YAC clones in the whole X library. This would obviate the one by one approach to mapping these YACs. Such an approach could also be applied to screening an entire human genome YAC library. Dr. Maynard Olson has agreed to provide filters from the large insert human YAC library to test this approach.

Hybridization methods to identify overlapping clones will be useful. We are also exploring the use of *Alu* PCR to "fingerprint" individual YAC clones. Products with bands in common should represent overlapped clones. This method will be most useful for the analysis of clones known to map to specific regions, and for those where there is suspicion of overlap (by hybridization results, for example).

Further analysis of the Lowe and AHC YACs will be made. Of the two, the Lowe syndrome clone looks the most promising, but the AHC clone is also quite interesting. While it is not a primary goal of this project, the identification and analysis of disease genes in the region is an important overall goal, and cannot be ignored. These results further demonstrate the great utility of genome studies to human health.

The suggestion of the development of sequence tagged sites (STS) for cloned DNAs is an excellent one. We have demonstrated the ability to directly establish an STS from a YAC clone through *Alu* PCR and direct sequencing of PCR products. This required no subcloning of the human insert. Continued development of the procedures for accomplishing this from clones in pYAC4 will be made, and the eventual goal will be to obtain STS from all the YAC clones with regional assignments. This will require a protocol amenable to some degree of automation, and not subject to the random occurrence of *Alu* repeats in the insert.

Continued development of *Alu* PCR will be pursued. This work will proceed with three immediate goals. The first is increasing the number of products per unit DNA. This is currently a limitation in somatic cell hybrid and YAC PCR reactions. The capacity to provide a PCR product (hence a probe) every 10 kb or so would greatly enhance the ability to assign (and map, and overlap) YAC clones, and provide sufficient product from somatic cell hybrids (radiation hybrids, in particular) to allow the use of the PCR product to identify regionally localized cloned DNAs. A protocol whereby PCR products are derived from sequences surrounding an *Alu* through "inverse PCR" is being developed. The major advantage here is that there is no requirement for two closely spaced *Alu* repeats. The second goal will be the identification of additional regionally localized *Alu* PCR products. These can provide novel probes from regions, and with sequence analysis, STS from specific sites. Given the current frequency of amplification (approximately 1/100 kb), *Alu* PCR products may represent ideal STS. The third goal is the establishment of *Alu* PCR product probes for isolation of genomic clones from specific sites. This will be pursued using the X regions and current libraries as models, then extending to total human libraries.

Efforts to establish and exploit methods of identifying and isolating expressed sequences will continue. We will follow up the current hncDNA clones from the X3000-11.1 region by additional regional assignment, sequence analysis and identification of related genomic and cDNA clones. We anticipate that a wealth of newly identified genes from the Xq24-qter region will be present in this collection. Efforts to extend the methodology to the entire X and to smaller regions will be made. In addition, we will explore the utility of altering the pattern of expression in the somatic cell

1021433

hybrids (by drug treatments) to increase the number of transcripts represented. The capacity to identify expressed regions in the collection of YAC clones will also play an important role in the definition of the region. The use of hncDNA clones as probes to the YAC clone slot blots has already identified YAC clones containing an expressed region in Xq27.3, near the Fragile (X) mutation on the proximal side. We will also explore the possibility of transfection of YAC clones into mouse L cells followed by *Alu* primed hncDNA library construction as a means of scanning YAC inserts for the presence of transcription units. A clone containing the 5' end of the Factor VIII gene in Xq28 will serve as a model system for these experiments. We will extend our analysis of oligonucleotide probes designed to recognize common sequence motifs found in expressed regions for identification of transcribed regions of YAC clones.

1021434

Mapped YAC clones
X3000-11 Library

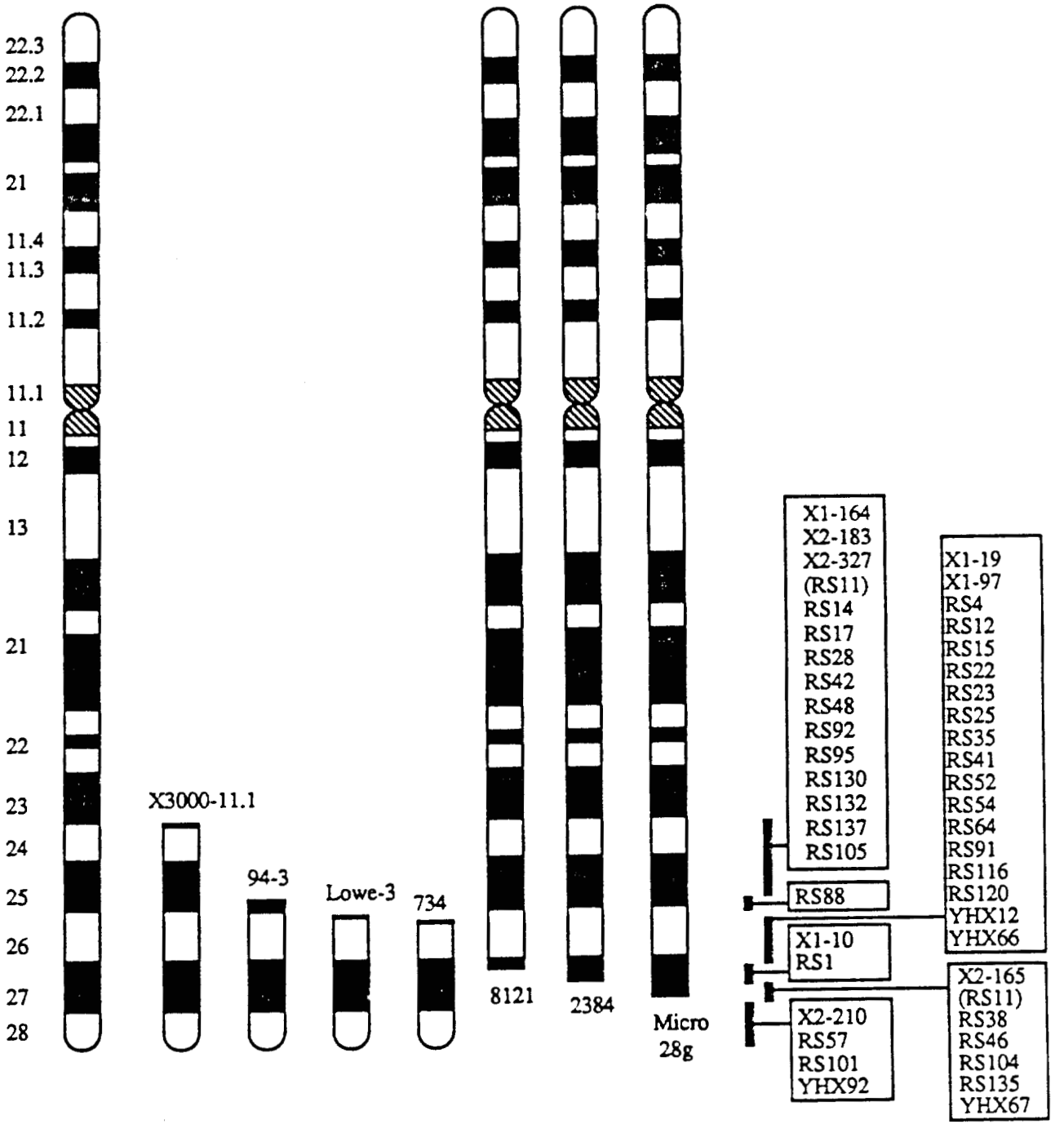

FIGURE 1

1021435

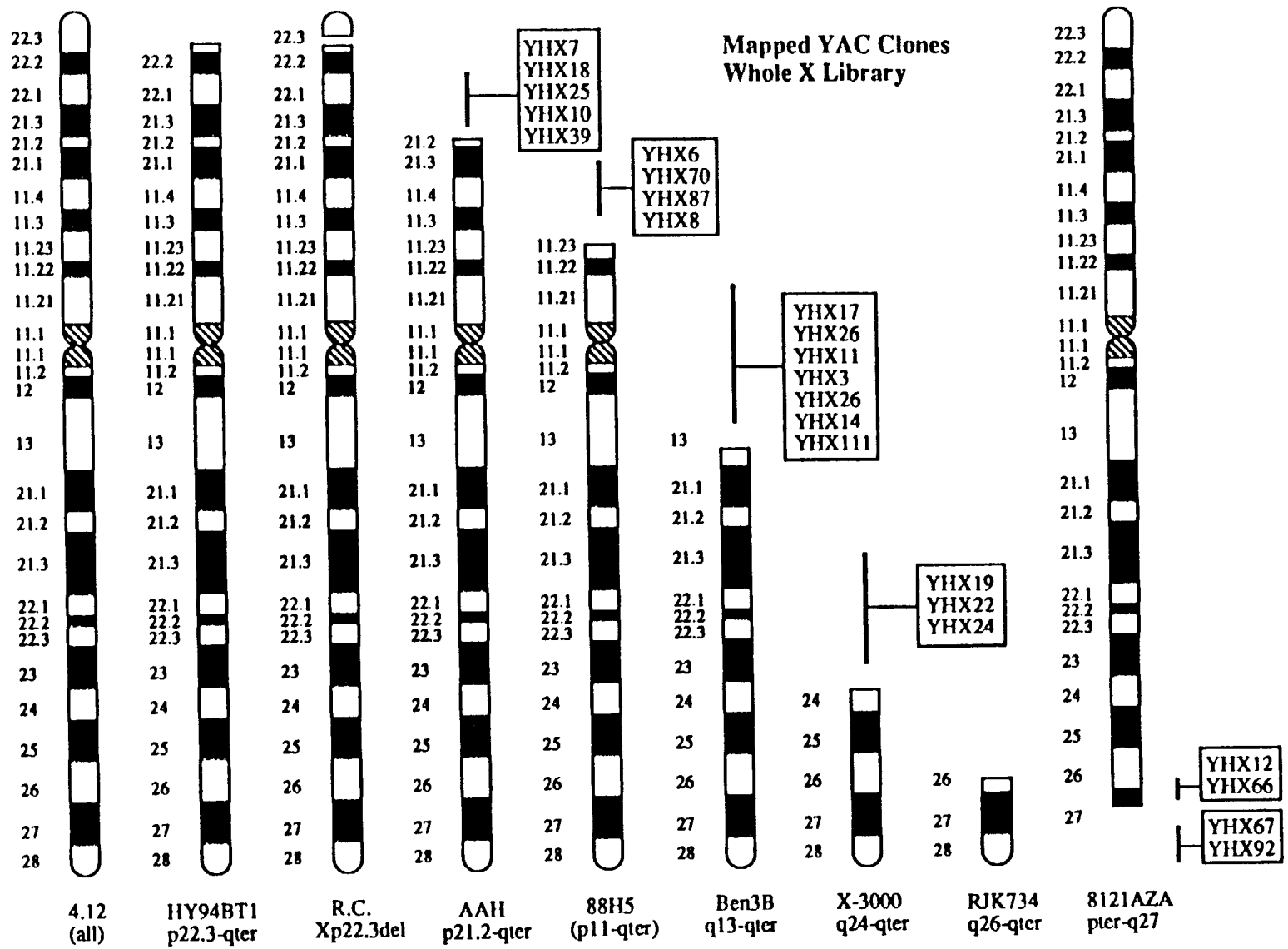

FIGURE 2

1021436