DOE/ER/60692-4

458

MAPPING AND ORDERED CLONING OF THE HUMAN X CHROMOSOME

707488

Progress Report

September 1990 — August 1991

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

Institute for Molecular Genetics

Baylor College of Medicine

Houston, Texas 77030

10214 23

September 1991

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

REPOSITORY Oak Ridge Operations-ORO
COLLECTION Procurement + Contracts Div. General Acquisitions Br. AD-421
BOX No. Active Records
FOLDER

AD-421

A. Progress 9/1/90–8/30/91

Progress in the past year has been excellent. Significant advances have been made toward several goals. In particular, the accumulation of additional large insert clones of known location on the X, identification of disease loci and associated genes, and generation of new cDNAs, polymorphic markers and sequence-tagged sites.

YAC libraries

We have constructed two libraries in the pYAC4 vector. The first (X3000 library) was made with DNA from the hybrid cell line X3000-11.1, which retains the human X chromosome region from Xq24 to Xqter as its only human DNA in a Chinese hamster fibroblast cell. The second (4.12 library) was constructed from DNA derived from the 4.12 cell line, which retains an intact X chromosome as its only human material in the same background cell line. Human positive clones were identified at 0.4% in the X3000 library, and ~2% in the 4.12 library. Approximately 500 human YACs have been isolated from these two libraries (Nelson *et al.*, *PNAS* 1991).

We also have available the total human YAC library from Olson and Schlessinger at Washington University (St. Louis). Approximately 56,000 clones has been transferred to Baylor and stored in triplicate. In addition, the total human YAC library from the CEPH in Paris is almost completely assembled at Baylor. This 7x library (~50,000 clones) with 400 kb average insert size will be available for screening in one-two months. These efforts have already resulted in the establishment of the facilities and methods needed for routine storage and screening of large collections of YAC clones and derived information.

Efforts to continue to produce YAC libraries in house continue, with the aim of isolated larger clones from hybrid cell lines. One attractive hybrid has been obtained from the Willard group. This cell retains multiple X chromosomes (an average of 6) in a mouse cell background. This line has also been made available to Larry Deavan of Los Alamos National Labs to attempt to produce YACs from sorted X chromosomes. This library will also become available to us upon completion. Collection of various clones and of additonal libraries will supplement the current materials.

YAC assignment

Figure 1 shows the current YACs of known location. Over 200 clones have either been localized by random assignment or isolated with specific PCR assays. The somatic cell hybrid libraries have yielded 113 localized clones, 40 from the 4.12 library and 73 from the X3000 library. This represents about 1/4 of the total clones isolated. These YACs average ~130 kb in length. These assignments have been made by Southern hybridization of *Alu* PCR products derived from the YACs to a hybrid mapping panel for localization (Nelson *et al. PNAS* 1991). Individual PCR assays of the St Louis and CEPH libraries have yielded an additional 85 clones for numerous specific loci. A few other clones have been obtained from other investigators' (A. Monaco and R. Nussbaum) collections. Five YACs from the St. Louis library are represented in figure 1 by underlining. These are random clones identified by X-specific bulk *Alu* PCR probes and subsequently sub-localized.

YAC isolation

To identify specific YACs from the St. Louis library, we have adopted a three step strategy similar to that of Green and Olson. This first requires the isolation of DNA from 150 pools composed of about 400 YAC clones each. Screening for the presence of specific PCR-generated products on 30 "master pools" containing 5 such individual pools is followed by secondary screening on the individual pools making up the positive masters. A "row and column" PCR approach allows narrowing of the positive to four colonies out of 384 in the pool, and this is followed by amplification of crude lysates from each colony to identify the clone. This has proven superior to the hybridization method in reproducibility and ease. Several regions are being studied intensively (see below) and YACs are being isolated in successive "chromosome walking" experiments to generate contigs. Ease of PCR screening is an important determinant of the efficiency of these efforts.

1021424

A major goal of our technique development area is to determine the feasibility of subdividing total libraries into specific sub-libraries derived from individual chromosomes using *Alu* PCR product probes derived from single chromosome (or more reduced) somatic cell hybrids. Preliminary results using Alu PCR product probes derived from single-chromosome hybrids to screen colony hybridization filters have identified YACs from the St. Louis library that have subsequently been confirmed to map to the chromosome used to generate the probe. Figure 1 shows 5 such clones from the X, which are designated by underlining. These results are preliminary, but very encouraging for the dissection of total human libraries into chromosome specific sub-libraries, and for further sub-fractionation of chromosome specific clones. Currently, we are attempting to use such a protocol to assign the 500 X clones to sub-regions. This promises to be far more efficient than the individual assignment of clones has been (see below). Numerous YAC clones have already been assigned to chromosomes or sub-chromosomal regions through the "dot-blot" procedure described for the characterization of the chromosome content of somatic cell hybrids. This employs *Alu* PCR product probes derived from the YAC to hybridize dot blots composed of *Alu* PCR products from various hybrid cell lines. The major advantage of the dot blot method is the increased number of clones that can be analyzed and the rapidity with which the analysis can be carried out. This will allow the rapid determination of the chromosomal locations of candidates for chromosome or region specific clones.

Specific loci

Fragile X Syndrome. One of the major successes of the past year has been the identification of the fragile X site and an associated gene, FMR-1. This was accomplished in close collaboration with Drs. Steve Warren of Emory University in Atlanta, and Ben Oostra of Erasmus University in Rotterdam (Verkerk *et al. Cell* 1991). Fragile X syndrome is the most common form of inherited mental retardation, with a frequency of ~1/1200 males. One of the randomly assigned clones from the X3000 library, RS46, was found to be close to the fragile X mutation site as defined by a series of fragile X associated X chromosome breakpoints generated by Dr. Warren. This led us to further investigate RS46 and the physical map of the region. An overlapping 450 kb YAC from the CEPH library (none of the libraries present at Baylor contained overlapping clones) allowed extension of the cloned region across the fragile site, and eventual isolation of the fragile X mutation region and associated gene.

The fragile site has turned out to be one of the most interesting loci in the human genome from the point of view of mutational mechanisms. The fragile site consists of a repetitive sequence present in an exon of the FMR-1 gene. The repeat is $(CGG)_n$ and encodes arginine in the cDNA reading frame. We have recently determined by PCR that the number of CGG repeats is polymorphic in the human population, varying from a low of 6 to a high of 46 repeats, with an average length of 29 repeats (Fu *et al.*, submitted). Females are heterozygous for allele length ~65% of the time, and we have observed 28 different alleles thus far.

Fragile X families have two types of mutations. The first, a "premutation" involves an expansion of the number of CGG repeats above the normal range. We have observed premutation alleles from a low of 52 to a high of ~250 repeats. These have no phenotypic effect in males or females, and explain the "Sherman paradox" of so-called normal transmitting males and their daughters. Meiotic transmission of the premutation results in alterations in size in nearly every case studied thus far. These can be subtle (60 repeats to 75 repeats) or dramatic (60 repeats to a fragment suggesting several hundred to thousands of repeats--in which case the repeats cannot be assayed by PCR). The dramatic changes are always observed in female meiosis, never in male meiosis, further explaining the observed risks of offspring in fragile X pedigrees (daughters of transmitting males are never retarded). The "full" mutation is associated with the Fragile X phenotype in males and in the minority of affected females. Alleles of the "full" mutation are highly mosaic; this region appears to be quite unstable mitotically in the expanded form. Continued study of the fragile site region should allow determination of the nature of the mutations and insight into the propensity of the CGG repeat to high mutation frequencies.

The FMR-1 gene found at the fragile site appears to be involved in generation of the syndrome (Pieretti, *et al. Cell* 1991). The majority of fragile X males show no mRNA for this gene in their peripheral blood leukocytes. This correlates with hypermethylation of the CpG island adjacent to the CGG repeats. A minority of fragile X patients show near normal levels of FMR-1 mRNA in blood cells, and this correlates with incomplete methylation at the CpG island. However, this latter group is highly mosaic for both the methylation and the size of the allele, often

1021425

with a "premutation" sized allele in addition to the full mutation alleles. The fact that these individuals are affected would suggest that the blood cells analyzed are not the appropriate tissue or developmental stage at which to assay the presence of *FMR-1*.

Lowe's syndrome. Of the 97 YAC clones assigned to sub-regions of Xq24-qter, one clone was identified which crosses the breakpoint in the X chromosome retained in Lowe-3. This chromosome was isolated from a female patient with the oculocerebrorenal syndrome of Lowe, and is a translocation of X to chromosome 3 with a breakpoint in Xq25. The *Alu* PCR products used to assign RS88 were found to map to either side of the translocation breakpoint. This finding led to further analysis of the RS88 clone, with the identification of a 4.3 kb *Alu-Alu* PCR product that contains the region of the translocation in Lowe-3, and we have narrowed the breakpoint to a 100 bp region in this fragment. In collaboration with Robert Nussbaum of U. Penn, we are attempting to define the gene involved in this disease. Several candidate cDNAs have been found using the RS88 YAC as a probe in cDNA libraries. One of these appears to fulfill the criteria for involvement in Lowe's syndrome. It spans the translocation breakpoint in Lowe-3, is absent from cell lines from the translocation patient, is absent from other Lowe's patients' cell lines, and is altered in size in another patient. The mRNA is 5.5 kb and is found in numerous tissues. Sequence analysis is commencing to allow guesses at the function of this gene.

Other X-linked loci of interest at Baylor. Contig assembly is proceeding rapidly in several other regions surrounding disease loci of interest to Baylor researchers. Among these are the XIST, KAL, XLA, XLP, GK/AHC, IDS, and Rett syndrome-translocation loci. Additionally, efforts to span Xq28 will provide materials for the numerous disease loci in this popular region of the genome. Candidate genes for the X-inactivation center and for the KAL (Kallmann) syndrome defect have been identified, and significant YAC contigs have been defined surrounding these areas by Andrea Ballabio's group at Baylor. Collaboration with Bakary Sylla and Gilbert Lenoir of IARC in Lyon has allowed construction of a 1 Mbp contig of YACs near the X-linked lymphoproliferative syndrome locus (XLP) in Xq25. Significant progress in bridging the scarce markers in the Xp21.2-21.3 regions distal to DMD has been made by the efforts of Huda Zoghbi and Ed McCabe's groups at Baylor. These focussed efforts provide highly detailed description of X chromosome regions of great interest.

Alu PCR and variations

One of the major goals of the original application was the development of a novel technique that has become known as "Alu PCR". This method has become a standard one for analysis of the human genome. Recent work at Baylor has defined additional repeats that can be used (from human as well as mouse and even bacteria) for similar purposes, and has provided numerous variations on the general theme for ever more useful applications of the technique. Many of these are reviewed in *Methods* (Nelson, 1991).

HncDNA development

We recently reported the development of a method for isolation of human-specific transcripts from somatic cell hybrid RNAs (Corbo *et al. Science* 1990). Additional efforts in this area include sequence analysis of the hncDNA clones isolated for development of transcript-tagged, sequence-tagged sites. Of 72 clones available, 35 have been sequenced thus far (using the Pharmacia A.L.F. automated sequencer). Of these, 15 have over 350 bp of sequence data. Primers for PCR are being designed for STS development. The PCR tests will allow simple sublocalization and identification of corresponding YAC clones.

Novel polymorphic markers

Sequence analysis of the human HPRT locus has identified a short tandem repeat (STR) that is polymorphic in repeat frequency. This particular sequence is CTAT repeated in 8 to 15 times in tandem. Single copy primers flanking the repeated sequence have been utilized to rapidly type DNA samples for alleles. The advantage of the trimer and tetramer STRs is ease of interpretation due to the spacing of alleles (3 or 4 base pairs instead of 2). We have developed oligonucleotide probes that can identify clones containing such repeat sequences. Methods for the rapid isolation of STRs and surrounding sequences by amplification are being developed. Since such polymorphisms

1021426

are plentiful (estimated one per 100 kb), reasonably simple to identify and very simple to assay, they are bound to become standard genetic markers. We are investigating methods for isolating STRs directly from YAC clones and for identifying potential STRs within the YAC clones by hybridization. Although some such repeats will be found in the yeast genome, it should be possible to identify ones without yeast equivalents, and to use these for rapid identification of YAC clones with polymorphic sites, allowing definition of genetic markers linked to the physical locations defined by the YAC clones.

From a flow-sorted library of the X chromosome in bacteriophage lambda, 28 different STR loci have been identified and subcloned into M13 for sequencing. These will be used to develop primers capable of amplifying the STR and determining the polymorphism of each locus. These loci will be placed on the physical map by amplification of the somatic cell hybrid mapping panel, and by analysis of the YAC libraries. Genetic mapping will be performed in the CEPH families.

B. Objectives 9/1/91–8/30/92

No significant alterations from the proposed studies are contemplated. Major efforts will concentrate on improving the genetic and physical maps of the X chromosome, and support of the regional mapping efforts ongoing at Baylor and collaboratively in various centers worldwide will continue. The newly established NIH-funded Human Genome Center at Baylor provides further support of X chromosome efforts, with core facilities for YAC isolation, sequencing and DNA synthesis, computation and mapping support. Smooth interactions between the Genome Center cores and this project already exist, and these are expected to continue. It is expected that the next year of funding will generate significant new insights into the organization of the human X chromosome and associated disease loci, while continuing to produce cloned materials useful for these analyses.

C. Publications 9/1/90–8/31/92

Nelson DL 1990. Current methods for YAC clone characterization. *Genetic Analysis: Techniques and Applications* 7, 100-106.

Corbo L, Maley JA, Nelson DL and Caskey CT 1990. Direct cloning of human transcripts using hnRNA from hybrid cell lines. *Science* 249, 652-655.

Ledbetter SA and Nelson DL 1991. Genome amplification using primers directed to interspersed repetitive sequences (IRS-PCR). In *Polymerase Chain Reaction: A Practical Approach* (M McLaren, Ed) London: IRL Press, pp 107-119.

Verkerk AJMH, Pieretti M, Sutcliffe JS, Fu Y-H, Kuhl DPA, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang F, Eussen BE, van Ommen G-JB, Blonden LAJ, Riggins GJ, Chastain JL, Kunst CB, Galjaard H, Caskey CT, Nelson DL, Oostra BA and Warren ST 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.

Nelson DL 1991. Applications of polymerase chain reaction methods in genome mapping. *Curr Opin in Genetics and Development* 1, 62-68.

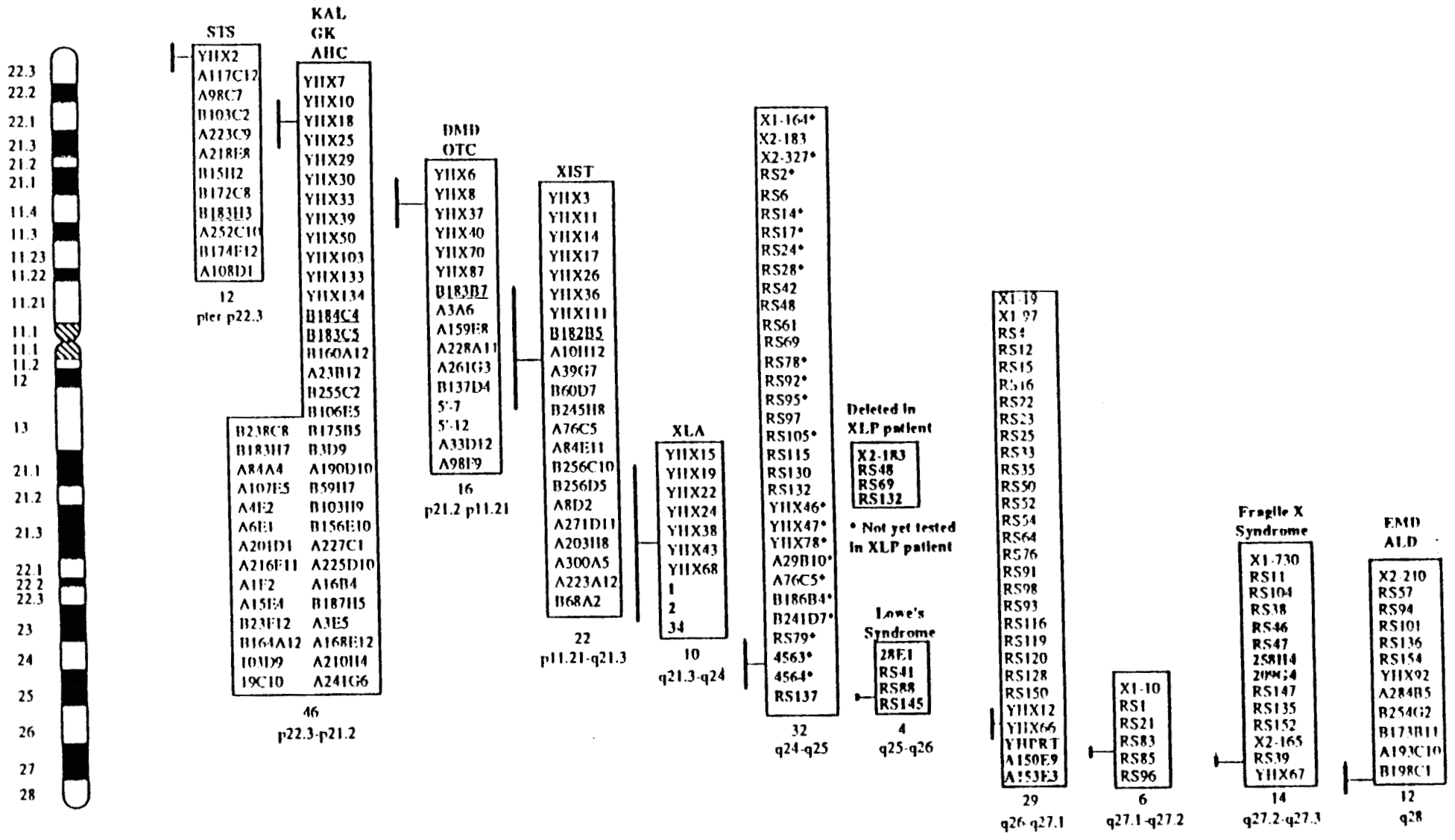
Nelson DL, Ballabio A, Victoria MF, Pieretti M, Bies RD, Gibbs RA, Maley JA, Chinault AC, Webster TD and Caskey CT 1991. *Alu*-primed polymerase chain reaction for regional assignment of 110 yeast artificial chromosome clones from the human X chromosome: identification of clones associated with a disease locus. *Proc Natl Acad Sci USA* 88, 6157-6161.

Nelson DL 1991. Interspersed repetitive sequence polymerase chain reaction (IRS PCR) for generation of human DNA fragments from complex sources. *Methods* 2, 60-74.

Purtilo D, Sumegi J, Victoria MF, Davis J, Grierson H, Nelson D and Caskey CT (1991). Toward cloning of the X-linked lymphoproliferative disease (XLP) susceptibility to Epstein-Barr virus (EBV) gene. *Miami Short Reports Volume 1: "Advances in Gene Technology: The Molecular Biology of Human Disease"*, p. 2.

Pieretti M, Zhang F, Fu YH, Warren ST, Oostra BA, Caskey CT and Nelson DL (1991). Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* 66, 817-822.

>200 YAC Clones from the Human X Regionally Assigned



1021428