Introduction - Overview of the Project

Our efforts on the human genome project were initiated in September, 1990, to contribute towards completion of the human genome project physical mapping effort. In our original application, we proposed a novel strategy for constructing a physical map of human chromosome 11, based upon techniques derived in this group and by others. The original goals were to 1) produce a set of cosmid reference clones mapped to specific sites by high resolution fluorescence in situ hybridization, 2) produce a set of associated STS sequences and PCR primers for each site, 3) isolate YAC clones corresponding to each STS and, 4) construct YAC contigs such that >90% of the chromosome would be covered by contigs of 2 mb or greater. Since that time, and with the advent of new technology and reagents, the strategy has been modified slightly but still retains the same goals as originally proposed. We have added a project to produce chromosome 11-specific cDNAs and determine the map location and DNA sequence of a selected portion of them.

Progress to date:

At this point, we have achieved a number of our original goals. We have produced a set of over 340 landmark clones, consisting of known genes and cosmid markers from our libraries mapped to a regional band location. Each marker is assigned a chromosome position (such as distal 11q23.3 with estimated resolution of 2-5 mb) and a fractional chromosome length (FLpter - fractional length from 11pter). Each marker also has an associated DNA sequence. The DNA sequences were determined by manual or automated sequencing directly from cosmid templates. We have prepared and tested PCR primers from about 80% of these and the remainder are expected to be completed before the end of the summer, 1992. At this point, we will submit the entire collection of data to GDB, and submit a publication on the chromosome 11 reference map based on these markers. All of this information will be made publically available at the Chromosome 11 workshop in La Jolla, in September, 1992. Finally, we have isolated and characterized over 175 YAC clones from the St. Louis and CEPH YAC libraries corresponding to these markers. Most have been characterized in detail and about 60 have been characterized by in situ hybridization. A publication has recently been submitted on this work, detailing the high level of chimeric and otherwise artifactual clones in these libraries.

While the majority of effort has been in the preparation of landmark clones, some contig construction has been carried out. We have utilized a useful approach of preparing 32p-YAC DNA and hybridizing this probe to high density arrays of chromosome-11-specific cosmids. The resulting pattern of spots has three important consequences: 1) it allows construction of YAC contigs by a hybridization-based fingerprinting approach, 2) it allows for only the portion of a chimeric YAC containing chromosome 11 to be included in the resulting map, and 3) it generates a set of cosmids which can be used to construct a cosmид contig in the region of the YAC. This is of great significance since the isolation of genes is vastly expedited by cloning in bacterial rather than yeast vectors. Contigs of megabase size have been constructed in the region of the CD3-Thy-1 genes,
DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.
DISCLAIMER

Portions of this document may be illegible electronic image products. Images are produced from the best available original document.
the NCAM-DRD2 region, the telomeric region flanked by random markers 23.2 and 5.8, and the INT2-PYGM region.

Detailed mapping was carried out in the CD3-Thy-1 region because this region is the site of a number of chromosome translocation breakpoints in human leukemias. We constructed a 3 mb contig from CD3 delta through Thy-1 and spanning 9 markers linked by in situ hybridization. We and others found that the YAC containing the CD3 genes contained the t(4;11) breakpoint of acute lymphocytic leukemia (ALL) and the t(9;11) breakpoint of acute myelocytic leukemia. We have subsequently subcloned the region into cosmids, defined the breakpoint in a 5 kb region, and are currently sequencing the region. We identified four transcripts in this region, a 1.5 kb transcript expressed in pancreas, heart and brain, a 4.5 kb transcript found in most tissues, a 1.2 kb transcript expressed in placenta and embryonic cell lines, and a 11.5 kb transcript expressed in a number of tissues including lymphocytes. The 1.5 and 4.5 kb transcripts have been completely sequenced and demonstrate no known homologies. The 11.5 kb transcript is interrupted at the 5' end by the breakpoint and limited sequencing shows that it is homologous to the homeo/pou family of transcription factors. A manuscript is now in preparation on these results.

Also, in collaboration with Bryan Young (ICRF, London) we found that the t(4;11), t(9;11), t(6;11), t(11;14) and t(X;11) breakpoints are all with the same 50 kb cosmid. Thus, one approach we intend to pursue is complete sequencing of this region in normal and translocation chromosomes.

In addition, STS sequencing from the end of cosmids has detected a number of possible similarities which may identify new genes. The most notable, is a sequence of >85% identity with hamster UDP-adenosyl phosphotransferase. The human enzyme has not been cloned or mapped. The cosmid containing this sequence maps between CD3 and Thy-1 and thus it is highly likely that the human UDP-adenosyl phosphotransferase gene is located 1.2 mb telomeric of the CD3 gene cluster. A manuscript on this result was recently submitted for publication.

Future of the genome center

Major developments in the past year have caused a great deal of consideration of the future of this project. During the past few months, impressive progress has been reported by groups at CEPH/Genathon in Paris, as well as others, using new cloning reagents. Discussions with Dr. Daniel Cohen and Dr. Michael James of CEPH have led to the possibility that copies of the 1 mb YAC genomic library and the chromosome 11-specific YAC library will be available to us before the end of this year. Should this be the case, we will be anxious to incorporate these into the current strategy to aid in completion of the map.

Most notable, the development of a 1.2 mb average size YAC library in CEPH/Genathan in Paris, and the potential distribution of this library to many centers including our laboratory, would mean that this project and the chromosome 11 physical map may be completed far faster than anticipated. Also, the number of clones necessary for completion of the map as defined in the original grant application, due to the increase size of the YACs, would be far fewer. We anticipate making use of these reagents when they are available to us and in completing the goals of this project as soon as possible.

In addition, in the future, we feel that the best effort of this group will be in the following areas:
1. Completion of the first level physical map of chromosome 11, consisting of 1 mb YAC contigs covering more than 90% of the chromosome, as soon as possible.

2. Large scale DNA sequencing in regions of biological interest. We have already begun to sequence a 40 kb region containing a cluster of translocation breakpoints found in acute leukemias using entirely automated protocols. Additional sequencing could cover regions containing important disease genes. However, we do not anticipate sequencing mb regions but would ideally leave that to future large scale sequencing groups or companies.

3. We feel that a major area for future development is in the area of automation, and our immediate goal is the development of high throughput PCR robots. The goal of this project, which we now propose as the future direction of project 2, will be the development of automated YAC library screening by robot with a throughput of 10,000 samples per day. A prototype instrument is already in construction by Dr. Skip Garner and his associates at the Institute for Advanced Technology of General Atomics Corporation.

4. While the immediate goal will be completion of the human chromosome 11 physical map, the development of powerful automation for genome analysis and high throughput PCR would allow the eventual automated physical map construction of entire genomes. Should this development become a reality within the next two years, we would anticipate using this technology to rapidly complete the regions of the human genome not the subject of major efforts. However given recent progress on the human genome, it may be that first level physical maps will be complete before this technology is completed. Should this be the case, we would envision a "Complete Biomass" genome effort to make physical and genetic maps of most organisms. While this may seem unfeasible, the development of genomic YAC libraries with 1 mb average insert size coupled with automated high throughput PCR instruments would conceivably allow completion of a whole genome map of a mammal within 6 months. With these developments, completion of genome maps of hamster, mouse, rat, cow, pig, dog mutjac, grey whale, California condor and other species is perhaps not far off.