PROJECT TITLE
HUMAN GENOME LIBRARIES

FINAL PROGRESS REPORT
for Period February 1, 1994 - August 31, 1997

Fa-Ten Kao, Ph.D.
Eleanor Roosevelt Institute for Cancer Research, Inc.
1899 Gaylord Street
Denver, CO 80206

January 1998

Prepared for
THE U.S. DEPARTMENT OF ENERGY
AGREEMENT NO. DE-FG03-94ER61819.A000-A003/M004

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

NOTICE
This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or 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.
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.
FINAL REPORT FOR PERIOD FEBRUARY 1, 1994 - AUGUST 31, 1997

This 3-year grant, entitled "Chromosome region-specific libraries for human genome analysis", was awarded for the period 3/1/94 to 1/31/97, with an unfunded extension to 8/31/97. This grant is a renewal of a preceding grant for the period 3/1/91 to 2/28/94.

The overall goal of this grant is to use a novel technology of chromosome microdissection and microcloning, previously developed in this laboratory funded by DOE, to construct chromosome region-specific libraries as resources for various human genome program studies, including physical mapping, STS preparation, clone isolation, microsatellite probes for genetic linkage mapping, positional cloning of disease genes, and reagents for sequencing.

During the grant period, we have achieved the stated goal of applying the microdissection and linker-adaptor microcloning technology to the construction of region-specific libraries for the entire human chromosomes 2 and 18. These high quality libraries are available not only for the DOE National Laboratories for their genome projects, but also for the scientific community as resources for their genome and genetic studies. The libraries have been deposited to the ATCC (American Type Culture Collection) for permanent maintenance and distribution. The following describe these accomplishments.


We have completed construction of all 11 region-specific microdissection libraries for the entire chromosome 2, including 4 libraries for the short arm, 6 libraries for the long arm, and one library for the centromeric region, as listed below:

<table>
<thead>
<tr>
<th>Library</th>
<th>Chromosome Region</th>
<th>Size (Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2P1</td>
<td>2p23-p25</td>
<td>25</td>
</tr>
<tr>
<td>2P2</td>
<td>2p21-p23</td>
<td>28</td>
</tr>
<tr>
<td>2P3</td>
<td>2p14-p16</td>
<td>22</td>
</tr>
<tr>
<td>2P4</td>
<td>2p11-p13</td>
<td>28</td>
</tr>
<tr>
<td>2CEN</td>
<td>2p11-q11</td>
<td>4</td>
</tr>
<tr>
<td>2Q6</td>
<td>2q11-q14</td>
<td>31</td>
</tr>
<tr>
<td>2Q5</td>
<td>2q21-q22</td>
<td>23</td>
</tr>
<tr>
<td>2Q4</td>
<td>2q23-q24</td>
<td>19</td>
</tr>
<tr>
<td>2Q3</td>
<td>2q31-q32</td>
<td>26</td>
</tr>
<tr>
<td>2Q2</td>
<td>2q33-q35</td>
<td>24</td>
</tr>
<tr>
<td>2Q1</td>
<td>2q35-q37</td>
<td>28</td>
</tr>
</tbody>
</table>
2. Characterization of 11 region-specific libraries of chromosome 2 for quality control

All microdissection libraries were carefully characterized to confirm their high quality. This quality control includes the following procedures:

1. Accurate microdissection of 20 chromosome fragments from the designated region of chromosome 2.

2. The chromosome fragments were treated with proteinase K, extracted with phenol, cleaved with MboI, and ligated to an MboI linker-adaptor. All were performed under the microscope.

3. The ligated sequences were amplified by PCR.

4. The amplified sequences were analyzed by gel electrophoresis to ensure that the library was in the right size range (100-800 bp).

5. Use of the PCR products in transformation to construct a pUC19 plasmid library.

6. Colony hybridization of microclones from the library using labeled total human DNA to ensure that the library contained the normal proportion (40-60%) of the microclones with highly repetitive sequences.

7. Isolation of a random sample of unique sequence microclones from the library for analysis by Southern blot hybridization using DNAs from human, human/CHO cell hybrid containing a single human chromosome 2, and CHO. A great majority of the microclones should be derived from chr. 2.

8. Library quality was confirmed to be excellent if the library met the above criteria.

3. Isolation and characterization of unique sequence microclones from each library for genomic uses.

For each library, between 26 and 66 unique sequence microclones were isolated and characterized to facilitate applications to genome analysis. These included 26 microclones for 2P1 Library, 60 microclones for 2P2, 66 microclones for 2P3, 35 microclones for 2P4, 37 microclones for 2CEN, 32 microclones for 2Q6, 26 microclones for 2Q5, 31 microclones for 2Q4, 33 microclones for 2Q3, 31 microclones for 2Q2, and 26 microclones for 2Q1.

These unique sequence microclones have been used as probes to isolate corresponding clones with large inserts, like YAC, P1, PAC, BAC, cosmid, and phage, to facilitate contigs assembly for the dissected region.

4. Detailed description of the production and characterization of the 11 libraries of chromosome 2 and the unique sequence microclones from the libraries.
The detailed characteristics of these libraries are summarized in Table 1. Details of the libraries have been described in the following publications:

(i) 2P1 Library (2p23-p25):

(ii) 2P2 Library (2p21-p23):

(iii) 2P3 Library (2p14-p16):

(iv) 2P4 Library (2p11-p13):

(v) 2CEN Library (2p11.1-q11.1):

(vi) 2Q6 Library (2q11-q14):

(vii) 2Q5 Library (2q21-q22):

(viii) 2Q4 Library (2q23-q24):

(ix) 2Q3 Library (2q31-q32):


(x) 2Q2 Library (2q33-q35):

(xi) 2Q1 Library (2q35-q37):

### TABLE 1. Summary of characteristics of 11 region-specific libraries of chromosome 2

<table>
<thead>
<tr>
<th>Library</th>
<th>Region</th>
<th>Size of Region (Mb)</th>
<th>Mean insert size (bp)</th>
<th>Potential library size*</th>
<th>Microclones derived from dissected region (%)</th>
<th>No. unique sequence microclones characterized</th>
<th>% clones conserved in rodents**</th>
<th>(No.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2P1</td>
<td>p23-p25</td>
<td>25</td>
<td>200</td>
<td>300,000</td>
<td>77</td>
<td>26</td>
<td>3.8 (1/26)</td>
<td></td>
</tr>
<tr>
<td>2P2</td>
<td>p21-p23</td>
<td>28</td>
<td>250</td>
<td>900,000</td>
<td>82</td>
<td>60</td>
<td>1.8 (5/28)</td>
<td></td>
</tr>
<tr>
<td>2P3</td>
<td>p14-p16</td>
<td>22</td>
<td>250</td>
<td>1,000,000</td>
<td>93</td>
<td>66</td>
<td>6.1 (4/66)</td>
<td></td>
</tr>
<tr>
<td>2P4</td>
<td>p11-p13</td>
<td>28</td>
<td>380</td>
<td>600,000</td>
<td>85</td>
<td>35</td>
<td>2.9 (3/135)</td>
<td></td>
</tr>
<tr>
<td>2CEN</td>
<td>p11.1-q11.1</td>
<td>4</td>
<td>180</td>
<td>200,000</td>
<td>90</td>
<td>37</td>
<td>2.7 (1/37)</td>
<td></td>
</tr>
<tr>
<td>2Q6</td>
<td>q11-q14</td>
<td>31</td>
<td>180</td>
<td>200,000</td>
<td>91</td>
<td>32</td>
<td>3.1 (1/32)</td>
<td></td>
</tr>
<tr>
<td>2Q5</td>
<td>q21-q22</td>
<td>23</td>
<td>180</td>
<td>200,000</td>
<td>85</td>
<td>26</td>
<td>0 (0/26)</td>
<td></td>
</tr>
<tr>
<td>2Q4</td>
<td>q23-q24</td>
<td>19</td>
<td>190</td>
<td>100,000</td>
<td>90</td>
<td>31</td>
<td>6.7 (2/30)</td>
<td></td>
</tr>
<tr>
<td>2Q3</td>
<td>q31-q32</td>
<td>26</td>
<td>190</td>
<td>200,000</td>
<td>85</td>
<td>33</td>
<td>3.3 (1/30)</td>
<td></td>
</tr>
<tr>
<td>2Q2</td>
<td>q33-q35</td>
<td>24</td>
<td>130</td>
<td>400,000</td>
<td>97</td>
<td>31</td>
<td>22.6 (7/31)</td>
<td></td>
</tr>
<tr>
<td>2Q1</td>
<td>q35-q37</td>
<td>28</td>
<td>270</td>
<td>20,000</td>
<td>77</td>
<td>26</td>
<td>7.7 (2/26)</td>
<td></td>
</tr>
</tbody>
</table>

*Estimated number of microclones that can be produced if the entire primary PCR products are used in transformation in DH5 bacterial host cells.

**Percentage of clones cross-hybridized to rodent species (CHO or mouse) by Southern blot hybridization.

5. Conversion of DNA (in PCR products) libraries to plasmid libraries for convenient use in genome research.

Initially, all of our region-specific libraries were prepared and stored as in the form of DNA after PCR amplification. In order to make the libraries more convenient to use by other investigators, we decided to convert the libraries into plasmid libraries. We took a subset of the DNA from the original PCR products in each library to clone into pUC19 to make a plasmid library. Each sub-library contained at least 20,000 independent, different species of microclones. Samples from each sub-library were analyzed to ensure excellent quality of the library, e.g. insert size, % of repetitive sequences, high percentage of microclones derived from dissected region,
etc. The titers of the libraries were generally maintained at approximately $1 \times 10^8$ microclones per ml.

6. Deposition of region-specific libraries to ATCC for permanent storage and general distribution.

We deposited all of the 11 region-specific libraries of chromosome 2 to ATCC. In order to facilitate the storage and distribution of these libraries by ATCC, we deposited the libraries to ATCC in plasmids, instead of DNA. This would eliminate the process of purification and transformation of DNA to make plasmid libraries by ATCC. Since the inserts of the microclones are short (mean length 200-300 bp), extreme care must be taken to reduce loss of the inserts during cloning, particularly in purification. Thus, we carried out this process in our lab because we have extensive experience dealing with these problems.

After we sent each plasmid sub-library to ATCC, their personnel expanded the library to large quantities and stored the aliquots in frozen vials for permanent maintenance and general distribution. They also sent us a vial to verify the authenticity of the expanded library before release. These chromosome region-specific libraries are listed in the ATCC's catalogue for general use. Thus, the libraries should have increased accessibility to a wider spectrum of the scientific community. The following ATCC Repository Numbers have been assigned to the 11 chromosome 2 libraries:

<table>
<thead>
<tr>
<th>Library</th>
<th>ATCC Repository No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2P1</td>
<td>#87188</td>
</tr>
<tr>
<td>2P2</td>
<td>#87189</td>
</tr>
<tr>
<td>2P3</td>
<td>#87103</td>
</tr>
<tr>
<td>2P4</td>
<td>#87104</td>
</tr>
<tr>
<td>2CEN</td>
<td>#87411</td>
</tr>
<tr>
<td>2Q6</td>
<td>#87410</td>
</tr>
<tr>
<td>2Q5</td>
<td>#87409</td>
</tr>
<tr>
<td>2Q4</td>
<td>#87310</td>
</tr>
<tr>
<td>2Q3</td>
<td>#87309</td>
</tr>
<tr>
<td>2Q2</td>
<td>#87308</td>
</tr>
<tr>
<td>2Q1</td>
<td>#77419</td>
</tr>
</tbody>
</table>

7. Construction of 3 region-specific libraries for the entire human chromosome 18

Three region-specific libraries for the entire human chromosome 18 have been constructed, including:
All these libraries are large, potentially comprising hundreds of thousands of recombinant microclones. The insert sizes ranged between 50-600 bp (mean 180-200 bp). The libraries contained approximately 40-60% microclones with unique sequence inserts. More than 30 unique sequence microclones from each library were analyzed by Southern blot hybridization to show that they are human specific and were derived from chromosome 18. These libraries and the unique sequence microclones are useful reagents for (1) isolating highly polymorphic microsatellite markers for refined linkage analysis, (2) identifying corresponding YAC or other clones with large inserts for contig assembly, (3) isolating cDNA clones from the dissected region as candidate genes for positional cloning, and (4) convenient sequencing of the short insert microclones. Since chromosome 18 is one of the less-exploited chromosomes in the human genome, these resources should facilitate high resolution mapping and sequencing of this chromosome. Details describing the 3 libraries of chromosome 18 have been published in:


8. Examples of applications of the chromosome region-specific microdissection libraries generated in this grant.

The grantee has made these resources available to many investigators for their genome-related and other genetic research. Some of the notable examples are summarized below.

(i) One of the most gratifying uses of our libraries was the contribution to the cloning of the gene responsible for hereditary nonpolyposis colorectal cancer (HNPCC). Right after the discovery of the location of this gene on the short arm of chromosome 2, region 2p15-p16, Dr. Bert Vogelstein of the Johns Hopkins Cancer Center contacted us about the use of our 2P3 Library (2p14-p16) to facilitate the gene search. We provided Dr. Vogelstein with the library and also supplied 66 unique sequence microclones from this region for use by his team. Through a world-wide collaboration, including our contribution of the library and the microclones, the first colon cancer gene hMSH2 was cloned which is a homolog of the bacterial mutS (and the yeast MSH2) mismatch repair gene. Our microclones from the region were also instrumental for isolating highly polymorphic markers for refined linkage analysis with the colon cancer gene. (Refer to publications in Cell 75, 1215-1225, 1993, and in Hum. Mol. Genet. 3, 2082, 1994.)

(ii) Another important application of the microdissection and microcloning technology developed in this grant was the collaboration between this lab and the biotech company BIOS in New Haven, CT. Since 1996, we have collaborated with BIOS to construct region-specific microdissection libraries from the most gene-rich regions (about 10 Mb per region) in the human genome. BIOS
would then convert these libraries into P1 libraries and make the P1 libraries commercially available (in the forms of individual P1 clones, DNAs from individual P1 clones, gridded P1 clones in plates, etc.) to investigators interested in these regions. These libraries can also be used as painting probes for FISH analysis involving the dissected regions. The collaboration has been funded by a Phase II SBIR grant (2 R44 GM50658) from NIH for two years (2/6/96-1/31/98), entitled "Band-specific libraries from microdissected chromosomes".

(iii) Dr. Max Muenke, Children's Hospital in Philadelphia and U. of Penn., used our 2P2 Library (2p21-p23) and unique sequence microclones in mapping and cloning of the gene for holoprocencephaly (HPE). HPE is an autosomal dominant disorder involving abnormal embryonic development in midbrain and midface. Some of the patients suffer an interstitial deletion in the 2p21 region. We sent Dr. Muenke unique sequence microclones from our 2P2 Library so he could use the microclones from this region to isolate corresponding YAC clones with large inserts for contig construction and physical mapping and cloning of the HPE gene. Details refer to the publication Hum. Mol. Genet. 5, 223-229, 1996.

(iv) Drs. Jeff Gingrich and Anthony Carrano of Lawrence Livermore National Laboratory used our 2Q1 and 2P1 libraries and the unique sequence microclones for their genome studies on chr. 2. In particular, this group sequenced microclones from 2Q1 library to prepare new STSs for the 2q35-q37 region. The unique sequence microclones were also used to screen genomic libraries with large inserts, such as cosmid, PAC, YAC and P1, for contig assembly and physical mapping.

(v) Dr. Eric Lai, University of North Carolina, used our 2P2 Library (2p21-p23) and 26 unique sequence microclones from the library to screen a chromosome 2-specific BAC library constructed in his lab. In one study, he used 20 microclones from 2P2 Library in the BAC screening and isolated 23 corresponding BAC clones. They were also developing methods to use the entire 2P2 library to screen the BAC library in order to make the screening highly efficient.

(vi) Professor Tom Strachan, University of Newcastle Upon Tyne, UK, used our 2P4 Library (2p11-p13) and 35 unique sequence microclones from the library to map and clone the gene underlying limb-girdle muscular dystrophy (LGMD). They established linkage of this gene to 2p13. They were sequencing our microclones to prepare primers for YAC library screening to isolate YACs from this region for refined physical mapping and the cloning of the disease gene.

(vii) Dr. Pudur Jagadeeswaran, University of Texas Health Sciences Center at San Antonio, used our 2Q1 Library to facilitate his development of a new technique for isolating coding sequences from genomic DNA by generating genomic-cDNA chimeras. He was interested in using our region-specific libraries in his search for cDNAs from defined genomic regions.

(viii) Dr. Nicholas Lemoine, Imperial Cancer Research Fund, Hammersmith Hospital, London, used our 2P2 Library to facilitate his search for a pancreatic cancer gene which was frequently rearranged near the 2p22 region.
(ix) Dr. Theodore Puck, Eleanor Roosevelt Institute for Cancer Research in Denver, used our 2Q1 Library in his studies of the genome structure and exposure in relation to gene expression and genome regulation.

(x) Because a number of genetic diseases have been mapped to specific regions of chromosome 2, the availability of our chr. 2 region-specific libraries should be valuable for the cloning of these disease genes. Fig. 1 presents some examples of the genetic diseases mapped to specific regions of chr. 2 but the genes responsible for these diseases have not yet been cloned.

**Figure 1.** Diagram showing examples of genetic diseases that mapped to specific regions in human chromosome 2 but the responsible genes have not yet been cloned.
9. Students and post-doctoral fellows trained on the project

(i) Jingwei Yu, M.D., Ph.D. - Dr. Yu received both graduate and post-doctoral training on this project. Dr. Yu has a medical doctor degree from Tongji University in China, with an M.S. degree in medical genetics from the same University. He recently completed his Ph.D. degree in Dr. Kao's lab in 1997, in the Department of Biochemistry and Molecular Genetics, University of Colorado Health Sciences Center in Denver. Dr. Yu had been a co-investigator of this grant since 1991. Dr. Yu is currently working at Yale University on a two-year Fellowship on the Clinical Genetics Training Program.

(ii) Suhong Tong, M.S. - Research Fellow trained on this project. Now working in Dr. Kao's lab.

(iii) Yiping Shen, M.S. - Research Fellow trained on this project. Now a Ph.D. student in the State University of New York at Syracuse.

(iv) Jianxin Qi, M.D. - Research Fellow trained on this project.

(v) Ben Tsai - Summer student trainee on this project. Now a pre-med student in the Northwestern University.

(vi) Amy Whittier, M.S. - Research Fellow trained on this project. Now working in a biotech company.

(vii) Patricia Way, B.S. - Research Fellow trained on this project. Now working in a biotech company.

(10) Publications resulted from this grant (a complete set of publications is enclosed in Appendix) —

(A) RESEARCH PAPERS


(B) ABSTRACTS


