DOE, DE-FC03-96ER62242

"Toward an Integrated BAC Library Resource for Genome Sequencing and Analysis"

Progress Report (8/15/96 – 9/30/99)

We have devoted significant effort toward two major goals: 1. To develop a large insert BAC cloning system and stably propagate genomic DNA in bacterial cloning vectors, construct high quality genomic libraries for human and mouse, and annotate the library resources by end sequencing, restriction fingerprinting, and mapping the clones to EST and other markers; 2. To develop approaches to high throughput library screening and efficient clone characterization for rapid construction of sequence-ready physical BAC contig maps using the annotated libraries.

Currently, large insert BAC libraries play a central role in genome projects. BACs provide reliable access to specific pieces of genomic DNA that can easily be used for standard laboratory procedures including DNA sequencing and FISH mapping. BAC library resources will become much more valuable and efficient tools once the majority of the clones in the libraries are characterized and annotated. Several laboratories including TIGR, the genome center at University of Washington, and ourselves have been involved in determining the clone end sequences and restriction fingerprint patterns from each BAC in human library D and BAC libraries from other libraries. In the past three years, we have been focusing on a human BAC-EST project that aims to correlate unique ESTs or Unigenes to human BAC library D clones under the sponsorship of the DOE.

1. <u>BAC library construction and characterization</u>

YACs carrying large pieces of genomic DNA greatly facilitated long-range physical mapping, leading to the completion of a genome-wide YAC-STS map. However, there was a basic, widespread need in the community for stable, non-chimeric clones with relatively large insert size that can be easily manipulated and sequenced. In 1992, we successfully developed the first BAC cloning system (Shizuya et al., 1992; Kim et al., 1992) based on a single copy *E. coli* F-factor replicon and subsequently constructed a series of large insert human BAC libraries (Kim et al., 1995; summarized in Table 1). BAC clones are stable and non-chimeric, and typically maintain genomic fragments of 100-350 kb in size. The procedures for BAC DNA preparation are relatively simple, robust, and can easily be scaled for higher throughput. BAC DNA preps are excellent material for FISH mapping, direct end sequencing, and shotgun sublibrary construction. BAC libraries of 130-250 kb average insert sizes are well suited to bridge the gap between higher order genome maps such as cytogenetic or YAC maps and the need to perform detailed analysis of specific clones. BAC libraries serve to integrate genetic, STS, and cytogenetic map information while providing direct access to stable materials that could be utilized for the purpose of molecular analyses, medical diagnostics, and ultimately for genome sequencing.

Our first BAC library A (Kim et al., 1995) was constructed using normal fibroblast cells from a male as a source of material. We chose untransformed fibroblast cells to avoid possible chromosome alterations that are usually associated with transformed cell lines. Subsequently, we have switched to sperm DNA to further reduce the chance of somatic mutations and any

possible DNA rearrangement associated with differentiation. Programmed gene rearrangements in the immunoglobulin gene clusters are well known. Therefore, we avoid using blood samples although they are easier to obtain in large quantities. For the same reason, we choose to use mouse germline cells (Embryonic Stem cells) rather than somatic tissues as a DNA source for mouse BAC library construction. Current status of Caltech BAC libraries is summarized in accordance with 96-NHGRI-DOE Guidance. Library D represents one of the highest quality genomic libraries that have been generated. Updated information on the status and usage of Caltech BAC libraries are available from our web site (http://www.tree.caltech.edu). Almost all of the clones in library D have been end sequenced (the results are posted in DOE-web site http://www.ornl.gov/meetings/bacpac/95bac/html) and restriction fingerprinted by HindIII by Dr. Lee Hood's laboratory at University of Washington.

Table 1. Current status of Caltech human and mouse BAC libraries

Human BAC Libraries	Source material	Plate numbers	Average Insert Size	Availablity	Cloning Site	NHGRI-DOE approval
A	987SK cells	1-1,000 (96-well)	115 kb	Caltech	HindIII	
B	987SK cells	1,001-1,194 (384-well)	120 kb	Research Genetics	HindIII	
C	human sperm (HSP)	1,195-1,879 (384-well)	125 kb	Research Genetics	HindIII	
D1	human sperm (HSP)	2,001-2,423 (384-well)	130 kb	Research Genetics	HindIII	Approved
D2	human sperm (HSP)	2,501-2,565 (384-well)	202 kb	Research Genetics	EcoRI	Approved
	human sperm (HSP)	2,566-2,671 (384-well)	182 kb	Research Genetics	EcoRI	Approved
	human sperm (HSP)	3,000-3,253 (384-well)	143 kb	Research Genetics	EcoRI	Approved

Mouse BAC library	Source Material	Plate Numbers	Acerage Insert Size	Availability	Cloning Site
CitCJ7	ES cell line/129Sv	1 - 612 (384-well)	130 kb	Research Genetics	HindIII

2. Physical mapping of human chromosome 22q

Our first target region for high resolution mapping using BACs was the chromosome 22q arm that spans approximately 45 Mb. In collaboration with us, the Sanger Center team led by Dr. Ian Dunham completed a high resolution YAC-STS map (Collins et al. 1995). The map was based on nearly 1,000 markers for this chromosome with a density greater than 1 marker per 50 kb. Approximately 500 markers were selected from the map and screened against Caltech BAC library A by both high density filter hybridization and pooled library PCR. We also used more than 100 cosmids and Fosmids that were localized to specific subregions on chromosome 22, by FISH mapping as additional probes. The work was rather straightforward, and two technicians

in our laboratory completed all the work necessary to generate enough data for the assembly of the contig map that consisted of 613 BACs in 113 contigs within four months, a relatively short time period for chromosome-scale mapping. In the map, BAC contigs were on the average 370 kb in size and were separated by gaps that were estimated to be approximately 90 kb. At least 80% of the 45 Mb chromosome 22q arm was covered with BACs (Kim et al., 1996). The map and related mapping data are also posted on our web site. The project involved extensive library screening with probes and PCR primer pairs after preparation of high density colony filters and library pools using a Biomek 2000 robotic workstation (Beckman Instrument, Fullerton, CA). All of the BACs mapped to chromosome 22 were restriction fingerprinted and the extent of contiguity between clones determined. All the chromosome 22 BACs mapped in this project have been shipped to the Sanger Center, Dr. Bruce Roe's laboratory at University of Oklahoma, and Research Genetics, which is now distributing BAC libraries as well as mapped BAC resources to the public at nominal fees. Currently the laboratories at the Sanger Center and University of Oklahoma are undertaking further map development and complete sequencing of chromosome 22.

3. Physical mapping of human chromosome 16

In the past two years, Caltech has been building sequence ready physical contig maps on the centromeric half of the chromosome 16p arm (16p13.1-11.2) and the telomeric half of the 16q arm (16q22.1-24.3) corresponding to 20 Mb and 30 Mb, respectively, jointly with The Institute for Genomic Research (TIGR) and Los Alamos National Laboratory (LANL) to provide TIGR with minimally overlapping sets of BAC clones selected from the contig maps. First the 3.5 X pooled library A was screened by STS-PCR with 112 ordered STS primer pairs in the 20 Mb 16p13.1-11.2 region that have been taken from the integrated chromosome 16 YAC-STS map constructed by LANL (Doggett et al., 1995). In this initial screening, 80 STS markers identified 206 positive BACs. These BACs provided a framework for library walking and further map building. Inserts from these positive BACs were used for screening library B and C (approximately 10X genomic coverage) by colony hybridization. The libraries A, B, and C were also screened with approximately 90 IMAGE cDNA probes that have been localized to the 20 Mb region, and with approximately 250 IMAGE cDNA probes mapped to the rest of chromosome 16. Shotgun subclones derived from the ends of completely sequenced BACs were used as probes to identify BACs that overlap minimally with the sequence BACs. More than 120 OVERGO probes have been designed from BAC end sequences and used for library walking. A complete set of chromosome 16 mapping data is available through our web pages. We have thus far identified over 2,500 putative BACs belonging to the 20 Mb region, and approximately 4,000 BACs over the entire chromosome 16. Of these, 126 BACs have been sequenced at TIGR yielding a total of 18 Mb sequences, and almost all putative chromosome 16 BACs have been end sequenced at Caltech and TIGR. We recently completed a 12.5 Mbp BAC contig in the target region by matching the BAC end sequences to completed BAC sequences as well as by the analysis of restriction fingerprint patterns, STS contents, and FISH mapping data from individual BACs. We have posted the map along with all supporting data on our web site. The characteristics of each BAC on this map and other chromosome 16 BACs are also listed on our web pages. As a collaboration with JGI, we are currently developing contigs in the 6-8 Mbp pericentric region in the chromosome 16p arm. JGI is currently sequencing the BACs selected from our contig map to achieve a large contiguous sequence in the target region.

4. Genome-wide BAC-EST mapping in human

One of our major goals in recent years has been to generate genome-wide BAC contig resources using EST markers. Mapped ESTs are excellent resources for both biomedical research and physical mapping. They represent expressed sequences with a known map location that have been derived from almost the entire euchromatic region and serve as excellent probes that permit rapid access to genomic clones in the target regions by colony hybridization. Our approach was to exhaustively screen a 4X coverage Caltech BAC library D that have been approved by NHGRI/DOE Guidance using human Unigene EST probes that have been made available by public repositories. Research Genetics has selected more than 40,000 unique EST clones representing Unigene clusters from the IMAGE cDNA collection in collaboration with Dr. Greg Schuler at NCBI and shipped them to us. To date, cDNA inserts from approximately 24,000 Unigene cDNA clones have been isolated, and 15,000 of these preps have been used for library screening in 100x100 or 20x20 matrix pools. Experimental details for high throughput, pooled probe hybridization and deconvolution of the positive addresses for individual probes are described below and in our web site (http://www.tree.caltech.edu/protocols/hi_densityscreen.txt). We have also designed and synthesized approximately 500 OVERGO primer pairs corresponding to the unique 36 bp from 3' untranslated regions of human genes associated with cancer and signal transduction. Figure 2 is an example of the autoradiograms from high density filter hybridized with pooled probes consisting of 100 cDNA inserts and 20 OVERGO probes, respectively. High quality data are routinely obtained from almost all of such hybridizations, consistently yielding expected statistical characteristics in the numbers of positive BACs. Initially, approximately 5,000 Unigenes were screened against Caltech library C before the approved library D was available. These have been re-screened against library D. We keep updated status posted through our web server (http://www.tree.caltech.edu). The web-based database is built on Oracle for Windows NT, and is capable of cataloging and searching annotation as well as mapping data associated with ESTs and positive BACs. It is also capable of automatically deconvoluting positive BAC addresses for individual probes from the incoming pooled probe positive BAC lists.

The initial assessment of the human Unigene clones suggests that a great majority of them are from unique genomic loci as indicated by the results from the 90 primary screenings of library D with 100 pools (Table 1). Please visit our web site.

Preparation of cDNA insert probes

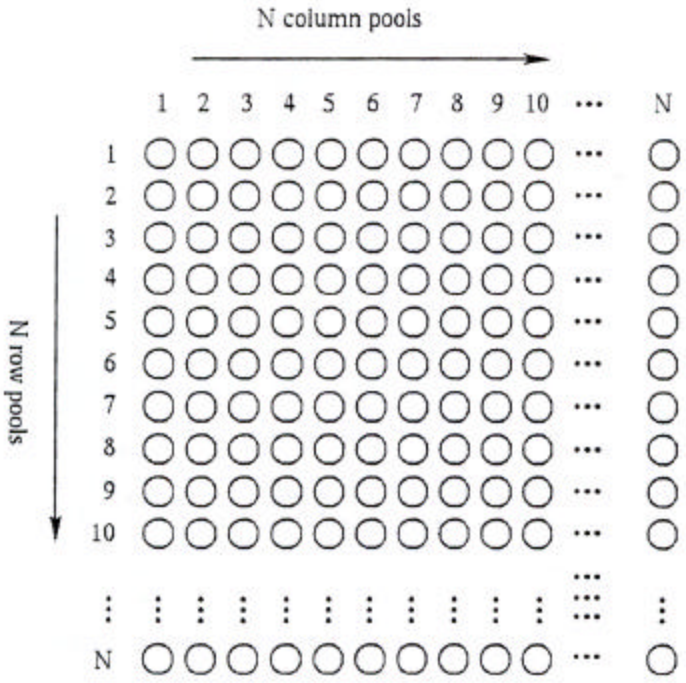
More than 42, 000 Unigene cDNA clones were provided by Research Genetics in five shipments over the period of two years. Earlier shipments consisted of lower quality Unigenes according to Troy Moore at Research Genetics who collaborated with Dr. Greg Schuler at NCBI on selecting Unigene cDNA clones from EST clusters based on sequence analysis. Approximately 20,000 "high quality" Unigene clones were sent to us through February 1998. The cDNA inserts were isolated from approximately 5,000 Unigene clones shipped earlier, and 19,000 Unigene clones shipped more recently by PCR amplification and gel purification according to the procedure described in our web site. Over 70% of the clones yielded strong major PCR bands on agarose gels. Approximately a 10% aliquot from each of the cDNA insert preps was used for probe pooling. The rest of the preps are being stored in microfuge tubes at –80°C for further use. Since preparing high quality cDNA inserts from such a large number of

clones is an extraordinarily laborious and time consuming task, our cDNA insert prep is a highly valuable resource for microchip development and gene expression studies as well as genome mapping.

Library screening with pooled cDNA insert probes

cDNA inserts represent relatively unique sequences as compared with average genomic DNA, and the majority of them are free of long stretches of repetitive sequences or low repeats that are difficult to suppress during pre-annealing of the probes. By virtue of this, we cold develop and implement a robust, highly efficient scheme to screen BAC libraries with a large number of Unigene probes. We are able to screen BAC libraries with probe pools that represent rows or columns of the N X N probe matrix (Figure 1) such that 2N screenings (N row pools + N column pools) are required for the library screening rather than N^2. This significantly reduces the number of hybridizations as N becomes large. The key technology in this scheme is hybridizing high density colony filters with up to 100 cDNA clone inserts that have been labeled in pools (see web site). In our ongoing human BAC-Unigene project, four 22 cm X 22 cm high density colony filters corresponding to 4X genome coverage are being used for each hybridization. It is possible to increase the number of cDNA probes to be used for single hybridization, although it will increase the total number of the probes involved in each cycle, and lengthen the cycle time required for the completion of library screening, i.e. until all of the probes are deconvoluted. Figure 2 is an example of the autoradiogram from such hybridization with 100 cDNA inserts. High quality data are routinely obtained in more than 95% of such hybridizations, consistently yielding expected numbers of positive BACs. Approximately 5,000 Unigene probes were used for screening Caltech human BAC library C before the approved library D was available. We performed 200 hybridizations in this scheme (20,000 total probes) on human library D, read the positive addresses from the autoradiograms, and deconvoluted the positive BACs to individual probes. Weekly throughput of ten such hybridizations was routinely performed by three technicians. One technician isolated 500 probes from the agarose gels every week after the PCR amplification of the cDNA inserts. We post updated status reports through our web servers (http://www.tree.caltech.edu/). A 20x20 matrix was also employed to generate reduced size pools for 400 cDNA insert probes. In this scheme, 40 hybridizations with 20 cDNA pools are required to complete the cycle. To date, we have completed five 20x20 hybridization cycles and obtained deconvolution data for additional 2,000 probes.

Figure 1. Generation of N probe pools from N^2 probes in an N X N matrix. A total of 2N pools are generated from such arrays (N row pools + N column pools).



An important issue to be addressed regarding the usage Unigene cDNA as probes for BAC library screening as genome-wide markers is how unique the cDNA sequences are, and at what specificity the probes would hybridize with legitimate BACs. Unlike the STSs that are designed to be unique in the genome and are used for PCR screening of pooled libraries, a certain fraction of Unigene cDNAs would face the problem of redundancy, i.e. more than one Unigene cDNA could have been derived from the same gene, and the problem of gene multiplicity, in which a cDNA probe may represent a gene that exists in multiple copy or as a member of a gene family. A Unigene may even contain repetitive sequences, and share homologies with genomic regions other than the locus from which it has been derived, albeit only the minor fraction of these genomic copies match the Unigene probe perfectly. They can potentially generate complex hybridization patterns, and make straightforward deconvolution difficult. The majority of problems arising from cross-hybridization due to Alu and other suppressible repeats can be circumvented by employing experimental conditions for stringent hybridization and efficient suppression.

Deconvolution of pooled probe hybridization data

Autoradiograms were read and the positive addresses for each pooled hybridization submitted to our informatics team led by Robert Xu. The data was automatically deconvoluted using and ORACLE database and the results are posted to the web through CGI programs. Approximately 96% of probes in a 100x100 scheme and 92% of probes in a 20x20 scheme found at least one hit. Statistical breakdown of the data obtained from the deconvolution of these hybridizations shows that the overall hybridization profiles from 10,000 and 2,000 cDNA probes deconvoluted in a 100x100 matrix and a 20x20 scheme, respectively, simulate the single unique

probes (Table 2A and 2B). The data demonstrate the distribution of 10,000 markers on BACs with average insert size of 130 kb (Table 2A). Of the 26,797 unique BACs, approximately 2,500-3,000 or 10% show aberrant behavior, i.e. detection by higher number of probes than the predicted Poisson values. BACs detected more than 5-7 times might contain ribosomal repeats or other microsatellite sequences. Data from a 100x100 scheme closely resemble random Poisson distribution of 10,000 single probes over the BAC clones. However, some complications are seen in both data. Relatively stronger tendency of multiple probe-BAC correlation is seen in 20x20 deconvolution as demonstrated by multiple peaks in the number of BACs along the number of detections.

Over 70% and 80% of the deconvolution data from 100x100 and 20x20 hybridization, respectively, agreed with the control experiment data. Approximately 200 of these Unigenes contained sequences, oncogenes, and other genes that were separately screened against Caltech BAC libraries. Comparison of these data supports the results from the control experiments. Restriction fingerprint analysis also indicates that the majority of the BAC clusters correlated to the individual probes formed single contigs. However, despite the overall uniqueness of the probes and the deconvolution accuracy, two serious problems emerged in the usage of cDNA probes. First, many of the cDNA probes contain DNA sequences that non-specifically hybridizes to multiple genomic loci, and coding sequences with homologies to multiple genome loci. Second, many of the cDNAs suffer from sample tracking errors, i.e. the clones are not what they are supposed to be. We have found a solution to circumvent both of these problems: using OVERGOs for the problematic cDNA clones instead.

Development of arrayed BAC resources for the analysis of chromosome aberration related to cancer.

Under the sponsorship of NCI and in collaboration with Drs. Gray and Pinkel at UCSF, Dr. Pieter de Jong at Rosewell Park Cancer Institute, Dr. David Bentley at the Sanger Center, and Dr. Barbara Trask at University of Washington, we have been developing a BAC resource covering the human genome at an approximate interval of 1 Mbp. The resource will serve as a framework for fine resolution mapping of over 27,000 cancer-related chromosome aberrations that have been documented to date. Our role is to provide the consortium with high quality BACs corresponding to all known oncogenes, cancer-related genes, signal transduction genes including receptors, kinases and heterotrimeric G-protein gene subtypes. We have thus far screened libraries with over 200 onco-probes provided by Dr. Gray and 200 signal transduction gene probes. More than 500 OVERGO primer pairs for these genes have been designed from the 3' UTR region of the consensus EST sequences and synthesized. Since designing a large number of OVERGOs is a time consuming task, and we are currently developing an OVERGO designing software tool to automate the queries and selection, and to implement a convenient GUI base on local sequence search routines to speed up this process.

Figure 2. An autoradiogram of a high density filter generated by the Genetix robot (also called Q-Bot). There are 55,296 spots on the 22 cm x 22 cm filter shown here representing 27,648 double spotted BAC colonies. Filters are hybridized with pooled probes containing 100 labeled inserts from Unigene cDNA clones. Our web site posts additional autoradiograms with pooled cDNA and OVERGO probes hybridization.

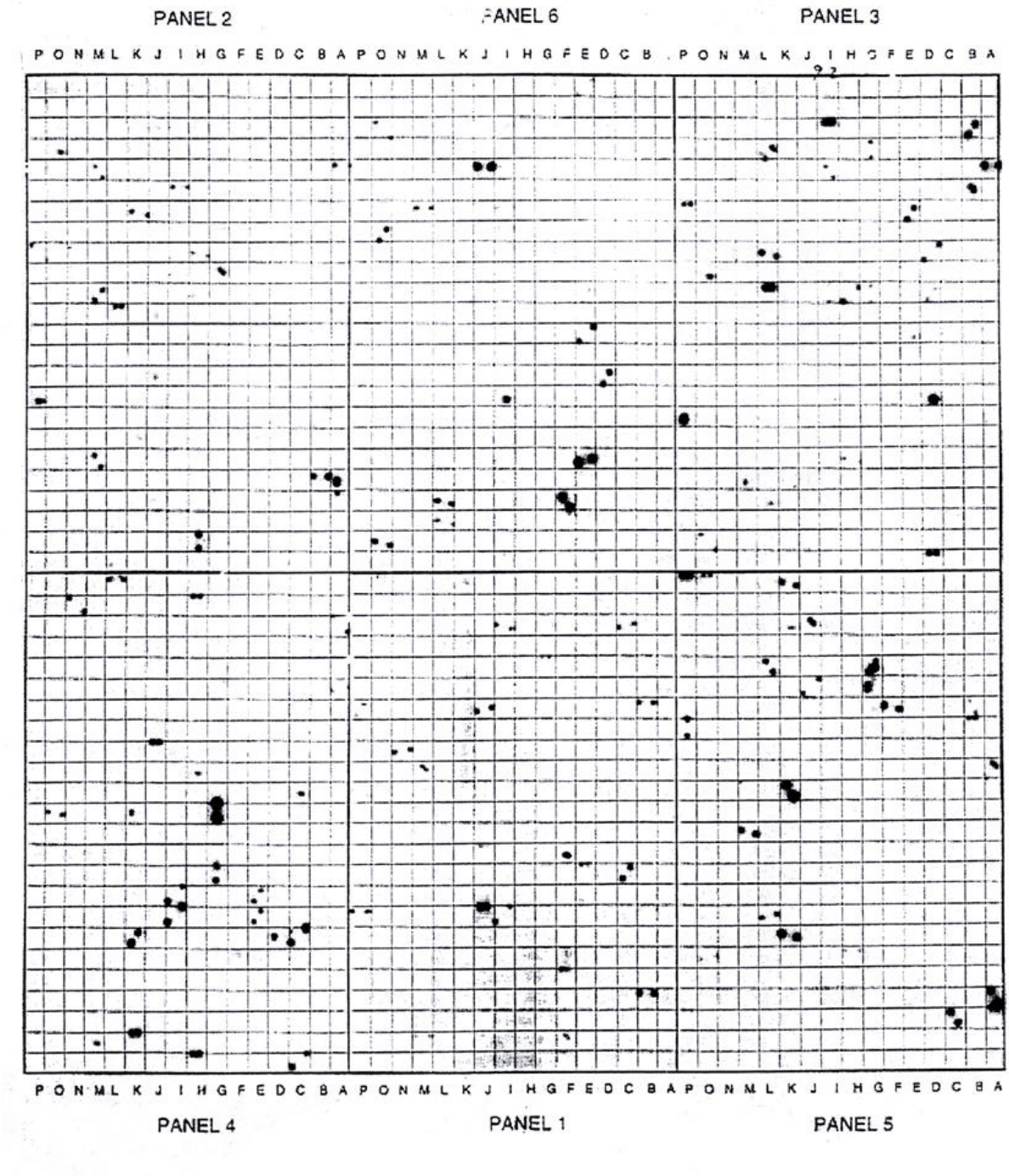


Table 2A. Redundancy of positive BACs in screens of 4X human library D with 10,000 Unigene probes in 100x100 pools. Total number of BACs identified in this pooled hybridization is 41,620 (26,797 unique addresses). *Poisson values are calculated based on $(e^{-m}m^x/x!)/\Sigma(e^{-m}m^x/x!)$ X 26,797 where x is the number of hits excluding 0 hits (i.e. 1,2,3…,14) and m = 130 (average BAC insert site in kb)/260 (average size of intervals of Unigene markers in euchromatin).

Number of detections	Number of BACs	Predicted Poisson values*
0	N.A.	N.A.
Once	18,115	20,654
Twice	5,443	5,162
3 times	1,857	858
4 times	775	109
5 times	291	14
6 times	143	1
7 times	65	0
8 times	38	0
9 times	30	0
10 times	13	0
11 times	12	0
12 times	5	0
13 times	2	0
14 times	1	0

Table 2B. Redundancy of positive BACs in the screens of 4X human library D with 2,000 Unigene probes in 20x20 pools. Total number of BACs identified in this pooled hybridization is 10,038 (6,261 unique addresses). *Poisson values are calculated base on $(e^{-m}m^x/x!)/\Sigma(e^{-m}m^x/x!)$ X 6,261, where x is number of hits excluding 0 hits (i.e. 1,2,3,…14) and m = 130 (average BAC insert size in kb)/1300 (average size of intervals of Unigene markers in euchromatin).

Number of detections	Number of BACs	Predicted Poisson values*
0	N.A.	N.A.
Once	5,518	5,951
Twice	1,044	296
3 times	65	13
4 times	172	1
5 times	2	0
6 times	85	0
7 times	0	0
8 times	29	0
9 times	12	0
10 times	5	0
11 times	0	0
12 times	7	0
13 times	0	0
14 times	2	0

In Summary

We developed a great deal of expertise in building large BAC libraries from a variety of DNA sources including humans, mice, corn, microorganisms, worms, and Arabidopsis. We greatly improved the technology for screening these libraries rapidly and for selecting appropriate BACs and mapping BACs to develop large overlapping contigs. We became involved in supplying BACs and BAC contigs to a variety of sequencing and mapping projects and we began to collaborate with Drs. Adams and Venter at TIGR and with Dr. Leroy Hood and his group at University of Washington to provide BACs for end sequencing and for mapping and sequencing of large fragments of chromosome 16. Together with Dr. Ian Dunham and his co-workers at the Sanger Center we completed the mapping and they completed the sequencing of the first human chromosome, chromosome 22. This was published in *Nature* in 1999 and our BAC contigs made a major contribution to this sequencing effort. Drs. Shizuya and Ding invented an automated highly accurate BAC mapping technique. We also developed long-term collaborations with Dr. Uli Weier at UCSF in the design of BAC probes for characterization of human tumors and specific chromosome deletions and breakpoints. Finally the contribution of our work to the human genome project has been recognized in the publication both by the international consortium and the NIH of a draft sequence of the human genome in *Nature* last year. Dr. Shizuya was acknowledged in the authorship of that landmark paper. Dr. Simon was also an author on the Venter/Adams Celera project sequencing the human genome that was published in *Science* last year.

References

Chi, A. A., H. C. Marrone, et al., An integrated physical map of human chromosome16. *Nature (Supplement)* **377**, 335-365.

Collines, J. E., C. G. Cole, L. J. Smink, C. L. Garrett, M. A. Leversha, C. A. Soderlund, G. L. Maslen, L. A. Everett, K. M. Rice, A. J. Coffey, S. G. Gregory, R. Gwilliam, A. Dunham, A. F. Davies, S. Hassock, C. M. Todd, H. Lehrach, T. J. M. Hulsebos, J. Welssenbach, B. Morrow, R. S. Kucherlapati, R. Wadey, P. J. Scambler, U.-J. Kim, M. I. Simon, M. Peyrard, Y.-G. Xie, N. P. Carter, R. Durbin, J. P. Dumanski, D. R. Bentley, and I. Dunham. A high density YAC contig map of human chromosome 22. *Nature* (*Supplement)* **377**, 367-379, 1995.

Kim, U.-J., H. Shizuya, X.-N. Chen, L. Deavan, S. Speicher, J. Solomon, J. Korenberg and M. I. Simon. Characterization of a human chromosome 22 enriched bacterial artificial chromosome sublibrary. *Genetic Analysis* **12**, 73-79, 1995.

Kim, U.-J., H. Shizuya, P. J. de Jong, B. Birren and M. I. Simon. Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucl. Acids Res.* **20**, 1083-1085, 1992.

Kim, U.-J., H. Shizuya, H.-L. Kang, S.-S. Choi, C. L. Garrett, L. J. Smink, B. W. Birren, J. R. Korenberg, I. Dunham and M. I. Simon. A bacterial artificial chromosome-based framework contig map of human chromosome 22q. *Proc. Natl. Acad. Sci. USA* **93**, 6297-6301, 1996.

Shizuya, H., B. Birren, U.-J. Kim, V. Mancino, T. Slepak, Y. Tachiiri and M. I. Simon. Cloning and stable maintenance of 300-kilobase pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**, 8794-8797, 1992.

Papers resulting from this project.

Cao, Y., H. L. Kang, X. Xu, M. Wang. S. H. Dho, J. R. Huh, B.-J. Lee, F. Kalush, D. Bocskai, Y. Ding, J. G. Tesmer, J. Lee, E. Moon, V. Jurecic, A. Baldini, H.-U. Weier, N. A. Doggett, M. I. Simon, M. D. Adams and U.-J. Kim. A 12-Mb complete coverage BAC contig map in human chromosome 16p13.1-p11.2. *Genome Res.* **9**, 763-774, 1999.

Choi, S. S., Y. S. Kang, Ung-Ji. Kim, K.-H. Lee and Hee-Sup Shin. Chromosomal localization of ESTs obtained from human fetal liver via BAC-mediated FISH mapping. *Molecules and Cells* **9**, 403-409, 1999.

Dunham, I., N. Shimizu, B. A. Roe, S. Chissoe *et al.* The DNA sequence of human chromosome 22. *Nature* **402**, 489-495, 1999.

Johnson A., S. Minoshima, S. Asakawa, N. Shimizu, H. Shizuya, B. A. Roe and H.E. McDermid. A 1.5-mb contig within the cat eye syndrome critical region at human chromosome 22q11.2. *Genomics* **57**, 306-309, 1999.

Loftus, B. L., U.-J. Kim, V. P. Sneddon, F. Kalush, R. Brandon, J. Furhmann, T. Mason, M. L. Crosby, M. Barnstead, L. Conin, P. Harris, E. Eichler, J. C. Venter and M. D. Adams. Genome Duplication and other Features in 12 Mpb of DNA Sequence from Human Chromosome 16p. *Genomics* **60**, 295-308, 1999.

Udar, N., S. Xu, J. Bay1, S. Dandekar, N. Patel, X. Chen, T. Liang, N. Uhrhammer, I. Klisak, H. Shizuya, H. Yang, G. Samara, J. Nelissen, M. Sawicki, P. Concannon and R. Gatti. Physical map of the region surrounding the ataxia telangiectasia gene on human chromosome 11q22-23. *Neuropediatrics* **30**, 176-180, 1999.

Freimuth, R. R., R. B. Raftogianis, T. C. Wood, E. Moon, U.-J. Kim, J. Xu, M. J. Siciliano and R. M. Weinshilboum. Human sulfotransferase SULT1C1 and SULT1C2: cDNA characterization, gene cloning, and chromosomal localization. *Genomics* **65**, 157-165, 1999.

Han, C S., P. B. Jewett, M. L. Campbell, L. J. Meincke, J. G. Tesmer, M. O. Mundt, R. D. Sutherland, U.-J. Kim, L. L. Deaven and N. A. Doggett. Construction of a BAC contig map of chromosome 16q by two-dimensional overgo hybridization. *Genome Research* **10**, 714-721, 2000.

Cheung, V. G., N. Nowak, W. Jang, I. R. Kirsch, S. Zhao et al. The BAC Resource Consortium: Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature* **409**, 953-958, 2001.

Choi, S and U.-J. Kim (2001) Construction of large insert Bacterial Artificial Chromosome (BAC) libraries. In: Methods in Molecular Biology: Genomics Protocols, M. P. Starkey and R. Elaswarapu, eds., Humana Press, Vol. 175, pp 57-68.