U.S. Department of Energy

Human Genome Program

Contractor-Grantee Workshop III

Santa Fe, New Mexico
February 7-10, 1993

ORNL/M-2588

Human Genome Program

U.S. Department of Energy

Contractor-Grantee Workshop III
February 7-10, 1993
Santa Fe, New Mexico

Date Published: January 1993

Prepared for the
U.S. Department of Energy
Office of Energy Research
Office of Health and Environmental Research
Washington, D.C. 20585
under budget and reporting code KP 0404000

Prepared by
Human Genome Management Information System
Oak Ridge National Laboratory
Oak Ridge, TN 37831-6050

Managed by
MARTIN MARIETTA ENERGY SYSTEMS, INC.
for the
U.S. DEPARTMENT OF ENERGY
UNDER CONTRACT DE-AC05-84OR21400

Contents

Workshop Agenda v-vii

Poster Presentation Times viii

Introduction To The Santa Fe Workshop ix

Abstracts*

U.S. Department of Energy Laboratories
  Human Genome Centers
    LANL—Los Alamos National Laboratory (1-20)
    LBL—Lawrence Berkeley Laboratory (21-48)
    LLNL—Lawrence Livermore National Laboratory (49-75)

  Ames Research Center (76)
  ANL—Argonne National Laboratory (77-80)
  BNL—Brookhaven National Laboratory (81)
  LANL—Los Alamos National Laboratory (82-90)
  LBL—Lawrence Berkeley Laboratory (91-96)
  ORNL—Oak Ridge National Laboratory (97-109)
  PNL—Pacific Northwest Laboratory (110-111)

Other Institutions (112-198)

Appendices
  A. Subject Index 199
  B. Author Index 200
  C. Anticipated Workshop Attendees 207

Each section alphabetized by first author.

Workshop Agenda
DOE Human Genome Program
Contractor-Grantee Workshop III
Santa Fe, NM
February 7-10, 1993

Plenary sessions are in the Eldorado, poster sessions are in the Hilton.
Each speaker and demonstration in the plenary sessions will have an abstract number, and thus a poster, associated with the talk. Schedule correct as of January 15, 1993. Agenda subject to change.

Sunday, February 7, 1993

2:00-8:00 p.m.	**On-site Registration**	**Eldorado Hotel**
6:00-8:00 p.m.	**Reception**	**Eldorado Hotel**

Monday, February 8, 1993

7:30 a.m.	**On-site Registration continues**	
8:00	**Welcome: David Galas**	
8:15	**Opening Comments: Senator Pete Domenici**	
8:25	**Report from HUGO: C. Thomas Caskey**	
8:40 a.m.	**ELSI**	**Chair–Michael Yesley**
8:40	Michael Yesley (20)	
8:55	Joe McInerney (197)	
9:25	Troy Duster (117)	
9:55	Phil Reilly (164)	
10:25	Betsy Fader (131)	
10:55 a.m.	**Break**	
11:25 a.m.	**Sequencing I**	**Chair–Deborah Nickerson**
11:25	Lloyd Smith	
11:40	**Current Efforts–prospects for scale-up and time frame for 300MB/yr**	
11:40	Chris Martin (37)	
12:05	Robert Weiss (138)	
12:30 p.m.	**Lunch**	
2:00 p.m.	**Sequencing I (cont.)**	
2:00	Leroy Hood (149)	
2:25	Ron Davis (128)	

2:50 p.m.	**Sequencing Goals: Panel** David Galas (chair), Lloyd Smith, Raymond Gesteland, Jasper Rine, Leroy Hood, Tom Caskey, Tony Carrano, Robert Moyzis	
4:00-7:00 p.m.	**Poster Session I**	**Hilton Ballroom**

Tuesday, February 9, 1993

8:00 a.m.	**Sequencing II**	**Chair-Deborah Nickerson**
8:00 a.m.	**Technology Developments for Large Scale Sequencing**	
8:00	Lloyd Smith (185)	
8:25	Richard Mathies (150)	
8:50	Klaus Schneider (179)	
9:15	William Studier (81)	
9:40	Steve Fodor (135)	
10:05 a.m.	**Break**	
10:30 a.m.	**Informatics**	**Chair-Diane C. Smith**
10:30	Diane Smith: Informatics review and comment	
10:45	Software Demonstration: Michael Cinkosky (4)	
11:00	Ed Uberbacher (106, 108)	
11:20	Jim Fickett (85-87)	
11:40	Software Demonstration: Tom Marr (127)	
12:00 noon	**Lunch**	
1:30 p.m.	**Informatics (cont.)**	
1:30	Ken Fasman (132)	
1:50	Software Demonstration: Tom Slezak (72)	
2:05	Software Demonstration: Richard Douthart (110)	
2:20	Manfred Zorn (48)	
2:40	Charles Lawrence (148)	
3:00	Software Demonstration: Suzanna Lewis (32)	
3:15	End of Session	
4:00-7:00 p.m.	**Poster Session II**	**Hilton Ballroom**

Wednesday, February 10, 1993

8:00 a.m.	**Mapping**	**Chair-Robert Moyzis**
8:00	Bruce Birren (119, 154)	
8:25	Pieter de Jong (63)	
8:50	Natasha Kouprina (12)	
9:20	Mary Kay McCormick (5, 13, 17)	
9:45 a.m.	**Break**	

10:15 a.m.	**Mapping (cont.)**	
10:15	Jasper Rine (42)	
10:40	Glen Evans (186)	
11:05	Daniel Pinkel (95)	
11:30	Richard Woychik (109)	
11:55 a.m.	**Lunch**	
1:30-4:30 p.m.	**Poster Session III**	**Hilton Ballroom**
7:00 p.m.	**cDNA**	
7:00 p.m.	**1. cDNA Update**	**Chair–Mihael Polymeropoulos**
7:00	M. Bento Soares (120)	
7:15	Mark Adams (112)	
7:30	James Sikela (184)	
7:45	Joseph Gatewood (8, 18)	
8:00	Chris Fields (133)	
8:15 p.m.	**2. cDNA Technologies**	**Chair–Greg Lennon**
8:15	Jan-Fang Cheng (22)	
8:27	Radoje Drmanac (78, 79)	
8:39	David L. Nelson (172)	
8:51	Greg Lennon (65, 66)	
9:03	Lisa Stubbs (107)	
9:15 p.m.	**Break**	
9:30 p.m.	**Issues in cDNA**	**Chair–Thomas Caskey**

Each speaker will have a short introduction (5 minutes), followed by an open dialog with panel and audience. The purpose is to stimulate discussion and seek input from the broader community. The session can accommodate ad hoc presentations (1-2 min.) from the audience. A series of questions will be provided to each speaker as a basis for the presentation.

David Galas–Statement of DOE cDNA Goals and Opportunities
M. Bento Soares–Library Needs
Glen Evans–Full Length cDNA Sequencing
Julie Korenberg–Faster cDNA Mapping Methods

11:00 p.m.	**Close**

Additional funding has been provided by the Perkin Elmer Corporation.

Poster Presentation Times

The number assigned to the investigator's abstract determines when his or her poster will be presented:

- Numbers with a remainder of 1 (Poster Session I)
- Numbers with a remainder of 2 (Poster Session II)
- Numbers evenly divisible by 3 (Poster Session III)

Poster Sessions

Hilton Ballroom

Session I	Monday (4–7 p.m.)
Session II	Tuesday (4–7 p.m.)
Session III	Wednesday (1:30–4:30 p.m.)

All posters should be mounted and ready for display before Session I begins.

Introduction to the Santa Fe Workshop

Welcome to the Third Contractor-Grantee Workshop sponsored by the Department of Energy (DOE) Human Genome Program. This meeting, designed to foster interaction among investigators and facilitate project coordination, offers DOE-supported genome researchers, program managers, and invited guests the opportunity to become familiar with current research, assess progress, and initiate collaborations.

The DOE Human Genome Program has grown tremendously, as shown by the marked increase in the number of genome-funded projects since the last such workshop, held in 1991. The 115 abstracts and 208 attendees of 1991 have grown to almost 200 abstracts and 400 attendees in 1993. Numerous collaborations resulting from the previous workshop have already borne fruit, and we expect that this meeting will prove equally successful.

The abstracts in this book describe the genome research of DOE-funded grantees and contractors and invited guests, and all projects are represented at the workshop by posters. The 3-day meeting includes plenary sessions in the Eldorado Hotel's Anasazi Ballroom on ethical, legal, and social issues pertaining to the availability of genetic data; sequencing techniques; informatics support; and chromosome and cDNA mapping and sequencing. All poster exhibits in the Hilton Mesa Ballroom will be open in the evenings throughout the meeting to maximize their availability to all attendees. New material resources and software are also on exhibit in a separate area of the Hilton.

With its multidisciplinary capacities, DOE is uniquely positioned to exploit the exciting opportunities presented by the Human Genome Project and ultimately to provide some measure of understanding of the genomic effects of radiation and chemicals. Serving as research sites for a multitude of interdisciplinary mapping and sequencing efforts are the three DOE human genome centers at Lawrence Berkeley, Lawrence Livermore, and Los Alamos national laboratories, as well as other DOE-supported laboratories and more than 40 different universities and research organizations. The Office of Health and Environmental Research appreciates the hard work and commitment of all contributors who, by their efforts, are advancing genome research toward the goals established over 2 years ago.

David A. Smith, Director
Health Effects and Life Sciences Research Division

Los Alamos National Laboratory
Center for Human Genome Studies

Strategies for the Development of a Gene Map of Human Chromosome 4

Michael R. Altherr [1&2], James F. Gusella [3] and Alan J. Buckler [3]. [1] Genomics and Structural Biology Group, LANL, [2] Department of Biological Chemistry, University of California Irvine. [3] Neurogenetics Laboratory, Massachusetts General Hospital.

Individual chromosomes provide the skeletal framework on which genomic data is organized. The genetic map provides the sinew for each chromosome and a contiguous array of molecular clones represents the dermis. Recently, this level of resolution was achieved for two human chromosomes and is near completion for several others. With these tools in hand investigators are positioned to begin more detailed analysis of chromosome structure and gene organization. One interesting piece of data in which to dress these developing bodies would be the localization of chromosome specific expressed sequences. To this end we have initiated two strategies to identify expressed sequences on human chromosome 4. Our first strategy employs 96-well plate pools of DNA from the flow sorted and arrayed chromosome 4 cosmid library as substrate for exon trapping. This technique has proved extremely efficient at identifying expressed sequences in the Huntington disease region and in the vicinity of several disease loci. The second strategy employs the hybridization of degenerate oligonucleotides capable of recognizing specific functional domains (eg. homeobox domain) to grided arrays of the chromosom be sub-localized rapidly using a somatic cell hybrid mapping panel with an average bin size representing 8% of chromosome 4. In addition, we have constructed a collection of chromosome 4 radiation hybrids and we are currently in the process of constructing a radiation hybrid map using the chromosome 4 genetic marker reference map. Once completed, we will be able to provide a relatively high resolution map of all the expressed sequences detected by our screening procedures. This will provide investigators who have previously identified linkage to disease genes using chromosome 4 reference markers additional markers and candidate genes to analyze in their families.

POOLING YAC LIBRARIES FOR SCREENING WITH UNIQUE SEQUENCES

D. Balding, D. Bruce, W. Bruno, N. Doggett, A. Ford, C. Macken, M. McCormick, D. Torney, C. Whittaker

University of London and Los Alamos National Laboratory

Efficiently accessing libraries of cloned DNA and clones in physical maps is a technical challenge with far-reaching ramifications for mapping and sequencing [E. Green and M.V. Olson; E. Barillot, et al.; P. deJong]

t-designs [Beth et al.] have been found to be optimal in that they enable one to locate all clones in a library containing a unique sequence via a minimal number of PCR reactions and pools. t-design pools include subsets of the clones not constrained to lie in rows, columns, or plates, and they have the advantage of allowing redundancy for correcting false-negative or other erroneous PCR results.

Our poster gives the details of the application of t-designs to the pooling of several YAC libraries created at LANL. One of our pooling schemes is similar to those proposed by Barillot et al. and deJong; these have the property of being nearly as efficient as t-designs and they can be easier to construct.

Novel pooling schemes, not necessarily involving t-designs, are optimal when the number of library screenings becomes sufficiently large and when the number of positive clones expected per screening is *e* or greater. We also describe pooling schemes for this case, where it is appropriate to pool both the library and the unique sequences that one wishes to screen the library with.

We have been programming robots to make the aforementioned pools for each of several YAC libraries consisting of 500 to 40,000 clones, for the purpose of screening for unique sequences. The programming of the Beckman and Packard robots will be described.

Barillot, E., B. Lacroix, and D. Cohen. Theoretical analysis of library screening using a N-dimensional pooling strategy. Nucleic Acids Res.,19,6241 (1991)

Beth, T., D. Jungnickel, and H. Lenz. Design Theory, Cambridge U. Press (1986)

Green, E. D. and M. V. Olson. Systematic screening of yeast artificial chromosome libraries by use of the polymerase chain reaction. PNAS USA,81,1213 (1990)

Chromosome 5 specific yeast artificial chromosome library from flow sorted chromosomes: construction and characterization. M. Campbell, M. K. McCormick, P. Schor, J. Fawcett, E. Martinez, L. L. Deaven and R. K. Moyzis. Center for Human Genome Studies and Life Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Yeast artificial chromosomes (YACs) have been constructed using DNA isolated from flow sorted human chromosome 5. Chromosomes for flow sorting were isolated from the somatic cell hybrid Q826-20 which contains chromosome 5 as the only human chromosome. DNA isolated from sorted chromosomes was restricted with Cla1, ligated to YAC vectors pJS97 and pJS98 and transformed into *S. cerevisiae* strain YPH 250. Approximately 925 human YACs, with an average size of 213 kb, have been generated, which is estimated to represent ~1X coverage of human chromosome 5. Additional YACs are being constructed and the chimera frequency will be estimated by fluorescent *in situ* hybridization of YACs to human metaphase chromosomes. (Supported by U.S. DOE grant # DOE RPIS BO4408)

SIGMA:
System for Integrated Genome Map Assembly

**Michael J. Cinkosky, Michael A. Bridgers, William M. Barber,
Charles D. Troup, Mohamad Ijadi and James W. Fickett**

The Human Genome Information Resource
Theoretical Biology and Biophysics Group and
The Center for Human Genome Studies
Los Alamos National Laboratory, Los Alamos, NM 87545

SIGMA (System for Integrated Genome Map Assembly) is an X Windows-based software tool for creating, editing and viewing integrated genome maps. One of the most significant aspects of the tool is that SIGMA maps contain both a "draft map" (essentially the picture) and the data on which it is based. This helps map builders keep their maps consistent with all of their data and it helps map users by giving them direct access to the underlying data.

With SIGMA it is possible to create maps that combine data from physical, cytogenetic and linkage maps in a single, unified map. Views can be created that emphasize particular aspects of these maps while retaining the full, underlying map intact. In addition, SIGMA is able to handle new types of map elements, as well as new map scales, without reprogramming. SIGMA can automatically generate and evaluate maps based on objective data supplied by the experimentalists. Built on an object-oriented database, SIGMA enables the easy sharing of information both within and between working groups.

SIGMA is available freely from Los Alamos and is currently in use at a number of sites around the genome community.

Construction of DNA libraries from flow sorted human chromosomes. L. L. Deaven, M. K. McCormick, D. L. Grady, D. L. Robinson, J. M. Buckingham, N. C. Brown, E. W. Campbell, M. L. Campbell, J. J. Fawcett, J. L. Longmire, A. Martinez, L. J. Meincke, P. L. Schor, and R. K. Moyzis. Center for Human Genome Studies and Life Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545.

The National Laboratory Gene Library Project is a cooperative project between the Los Alamos and Lawrence Livermore National Laboratories. At Los Alamos, a set of complete digest libraries has been cloned into the EcoRI insertion site of Charon 21A. These libraries are available from the American Type Culture Collection, Rockville, MD. We are currently constructing sets of partial digest libraries in the cosmid vector, sCos1, and in the phage vector, Charon 40, for human chromosomes 4, 5, 6, 8, 10, 11, 13, 14, 15, 16, 17, 20, and X. Individual human chromosomes are sorted from rodent-human hybrid cell lines until approximately 1 µg of DNA has been accumulated. The sorted chromosomes are examined for purity by *in situ* hybridization, DNA is extracted, partially digested with Sau3A1, dephosphorylated, and cloned into sCos1 or Charon 40. Partial digest libraries have been constructed for chromosomes 4, 5, 6, 8, 10, 11, 13, 14, 16, 17, and X. Purity estimates from sorted chromosomes, flow karyotype analysis and plaque or colony hybridization indicate that most of these libraries are 90-95% pure. Additional cosmid library constructions and 5-10X arrays of libraries into microtiter plates are in progress. Libraries have been constructed in M13 or bluescript vectors (chromosomes 5, 7, 17) to generate STS markers for selection of chromosome specific inserts from a genomic YAC library. We have also been able to clone sorted DNA into YAC vectors and have constructed YAC libraries for chromosomes 9, 16 and 21. Supported by the U.S. Department of Energy under contract W-7405-ENG-36.

Construction of a cosmid, YAC, and STS physical map of human chromosome 16.

N.A. Doggett[1,2], D.F. Callen[4], R.L. Stallings[5], M.K. McCormick[1,2], C.E. Hildebrand[1,2], D.C. Torney[2,3], J.W. Fickett[2,3], M. Cinkosky[2,3], L.L. Deaven[2], G.R. Sutherland[4] and R.K. Moyzis[2]. [1]Life Sciences Division, [2]Center for Human Genome Studies, [3]Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545. [4]Dept. of Cytogenetics, Adelaide Children's Hospital, Adelaide, South Australia. [5]Deptartment of Human Genetics, University of Pittsburgh, Pittsburgh, PA 15261.

We have constructed a detailed physical map of human chromosome 16. This map was generated by the fingerprinting of 4032 chromosome 16 cosmid clones and the hybridization of 411 chromosome 16 YAC clones (both constructed from flow sorted chromosomes) to high density membrane grids of these cosmids. The map is comprised of 462 islands having an average size of 218 kb. The coverage of the chromosome in these contigs is 94%. Islands have been regionally localized to a somatic cell hybrid breakpoint map using STSs derived from the end clones of cosmid contigs. A total of 140 islands (cosmids and YACs) and singleton cosmids covering ~24 Mb or 25 % of chromosome 16 has been regionally localized on the chromosome. The somatic cell hybrid panel consists of 54 different hybrids that produce a cytogenetic-based physical map with an average resolution of 1.6 Mb. A 1.6 Mb average resolution sequence-tagged site (STSs) map has been developed from contigs mapping to each region of this cytogenetic map. More than 50 known genes and markers have been integrated into the map by fingerprinting of cosmid clones and hybridization of probes to the high density cosmid membranes. This map is facilitating the cloning of various disease loci and fragile sites on chromosome 16. Supported by US DOE contract W-7405-ENG-36.

CONSTRUCTION OF A 1.6 MEGABASE AVERAGE RESOLUTION STS MAP OF HUMAN CHROMOSOME 16.

N.A. Doggett, L.A. Duesing, J.G. Tesmer, D.F. Callen*, R.I. Richards*, R.L. Stallings, C. E.. Hildebrand, and R.K. Moyzis, Life Sciences Division and Center for Human Genome Studies, Los Alamos National Laboratory, Los Alamos NM 87545, and *Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, North Adelaide, SA 5006, Australia.

We are constructing a detailed cosmid, YAC, and STS physical map of human chromosome 16. This map currently consists of 4032 fingerprinted cosmid clones, and 334 YAC clones (hybridized to high density arrays of cosmid clones) which make up 462 islands of 218 kb average size and cover 94% of this chromosome. Regional localization of these islands is accomplished by multiplex PCR deletion analysis of sequence tagged sites in somatic cell hybrids. The hybrid panel contains 50 breakpoints with an average resolution of 1.6 megabases. Sequence tagged sites are developed from the end clones of cosmid contigs. 408 cosmid end clones have been sequenced and more than 100 STSs have been developed and localized to the hybrid breakpoint panel, resulting in a 1.6 Mb average resolution STS map of chromosome 16. GRAIL analysis indicates that 14% of the 408 cosmid sequences may contain open reading frames. One sequence has high similarity to inosine monophosphate dehydrogenase, suggesting that it may be a third locus for this gene family. Supported by US DOE grant 005063 (F137) under contract W-7405-ENG-36.

Sequencing and Mapping of cDNAs from a Paternally Encoded Human Pregnancy

J.M. Gatewood, K.S. Denison, C.L. Lemanski, and R. Lobb.
Los Alamos National Laboratory, Los Alamos, New Mexico, 87545.

We are examining gene expression in an unusual human pregnancy (hydatidiform mole) by cDNA library construction and cDNA sequencing. This pregnancy is exclusively of paternal genetic origin and results from paternal genetic imprinting. A goal of this project is to determine if the genes expressed uniquely in hydatidiform mole are encoded for by a subset of sperm chromatin which hypothetically may be involved in paternal imprinting.

Complete hydatidiform moles are abnormal diploid pregnancies which result from the fertilization of an enucleated ovum. The pregnancies are characterized by hypertrophy of the trophoblast, hydropic degeneration of all placental villi, absence of a fetus, and a propensity to become malignant. The moles have maternal mitochondrial DNA but paternal chromosomes. Diploidy is restored by either fertilization by two sperm, or in the majority of cases fertilization by one sperm which undergoes duplication without cytokinesis.

An analysis of over 1200 cDNA clones from a directionally cloned and arrayed nonamplified cDNA library indicates that approximately 10% are homologous to known DNA sequences, 10% contain repetitive elements, less than 1% lack inserts, and the remainder represent newly identified genes. Of the newly identified genes, approximately 40% contain regions of limited homology which may indicate function. The homologous sequences are also providing information on tissue specific RNA splicing.

In collaboration with Dr. Julie Korenberg, chromosomal localization of cDNA clones using fluorescent *in situ* hybridization (FISH) has been initiated. This approach has successfully been used to map single copy genes with high fidelity. STSs are generated for the cDNAs which are mapped using FISH. The STS primers are then used to determine the organization of the gene in human sperm chromatin by a PCR based assay. Using this assay, cDNAs have been localized to distinct sperm chromatin components.

Systematic generation of chromosome 5 specific sequence-tagged sites (STSs). D. L. Grady[1], J.D. McPherson[2], D. L. Robinson[1], L.L. Deaven[1], J.J. Wasmuth[2], R.P. Wagner[1], and R. K. Moyzis[1].

[1]Los Alamos National Laboratory, Los Alamos, NM 87545, and [2]University of California, Irvine, CA.

The systematic generation of chromosome 5 specific sequence-tagged sites (STSs) is in progress, utilizing flow-sorted human chromosomes. Chromosomes were purified from either normal cells or a monochromosomal cell line. Purity estimates of approximately 95% were obtained, yielding a 50-fold enrichment. DNA from approximately 200,000 chromosomes was digested with two restriction enzymes (usually BamH1 and Hind III) and cloned directly into the bacteriophage M13 mp18. One pass sequencing was conducted with a Dupont Genesis 2000 automated sequencer. Approximately 15% of the greater than 500 random DNA sequences analyzed, contain putative coding regions, as determined by DNA homology searches, or Grail analysis (Oak Ridge National Laboratory). Following further analysis to identify common human repetitive sequences, appropriate PCR oligomers were synthesized in unique regions. An acceptable STS-PCR assay yielded the appropriate size amplification product from hybrid cell line DNA containing only human chromosome 5. Eleven cell hybrids which retain chromosome 5 from individuals with deletions or translocations of the long arm have been isolated. These eleven hybrids, along with 12 hybrids with deletions of 5p, represent a natural deletion mapping panel which defines 30 discrete segments of chromosome 5. To date over 100 of the chromosome 5 STSs have been regionally ordered using this mapping panel. The STS markers appear to be randomly distributed along the length of chromosome 5. The exact order of a subset of STSs was determined by the use of radiation hybrid panels of chromosome 5. These results demonstrate that a 0.5-1Mb STS map of chromosome 5 is achievable using these approaches. Supported by the U.S. Department of Energy under contract W-7405-ENG-36.

DNA FRAGMENT SIZING BY FLOW CYTOMETRY

M.E. Johnson[1], P.M. Goodwin[1], W.P. Ambrose[1], J.C. Martin[2], B.L. Marrone[2], J.H. Jett[2],
and
R.A. Keller[1]

Chemistry and Laser Sciences Division[1] and Life Sciences Division[2]

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

We have demonstrated flow cytometric detection and sizing of single pieces of fluorescently stained lambda DNA (48.5 kb) and individual *Kpn I* restriction fragments of lambda DNA (17.05 kb and 29.95 kb). DNA was stained stoichiometrically with an intercalating dye such that the fluorescence from each fragment was directly proportional to DNA fragment length. Laser powers up to 100 mW were used. Transit times through the focused laser beam were on the order of milliseconds. Measurements were made using time-resolved single photon counting of the detected fluorescence emission from individual stained DNA fragments. Samples were analyzed at rates of about 50 fragments per second. Measured fluorescence intensities of ethidium homodimer stained DNA were linearly correlated with DNA fragment length over the range measured. Initial sizing accuracy was ~1%. Preliminary size resolution for full sized lambda DNA was ~ 4 kb. Results using other DNA stains will also be presented. Detection sensitivity and resolution needed for analysis of small and large pieces of DNA will be discussed. The projected detection range of this technique is from ~1 kilobase to multiple megabases of DNA. Several applications of this methodology including DNA sizing and DNA fingerprinting will find immediate application in molecular biology where rapid sizing of DNA fragments is needed. For example, it will be possible to determine the location of a transposon insertion site in a cosmid by analysis of a rare cutter digest because the restriction digest will contain only 2 or 3 fragment sizes. As resolution improves, it will be possible to determine the restriction fingerprint of a cosmid clone in which an average of 10 fragments are present. When the measurement resolution is limited by photoelectron statistics, the fractional resolution will be better for larger sized fragments, in contrast to gel analysis where the resolution degrades as DNA size increases. The time required to analyze a sample to obtain adequate statistical accuracy of the peak locations is on the order of 10 minutes. This work is an outgrowth of our single molecule detection and rapid DNA sequencing programs.

(Supported by the US Department of Energy .)

DNA Sequencing by Single Molecule Detection of Labeled Nucleotides Sequentially Cleaved from a Single Strand of DNA

Richard A. Keller
Center for Human Genome Studies,
Los Alamos National Laboratory
Los Alamos, NM 87545

John D. Harding
Life Technologies, Inc.
8717 Grovemont Circle
Gaithersburg, MD 20898

We are developing a laser-based technique for the rapid sequencing of 40 kb or larger fragments of DNA at a rate of 100 to 1000 bases per second. Our approach relies on fluorescent labeling of the bases in a single fragment of DNA, attachment of this labeled DNA fragment to a support, movement of the supported DNA into a flowing sample stream, and detection of the individual fluorescently labeled bases by laser-induced-fluorescence as they are cleaved from the DNA fragment by an exonuclease. The ability to sequence large fragments of DNA will reduce significantly the amount of subcloning and the number of overlapping sequences required to assemble megabase segments of sequence information. Progress in single molecule detection, measurement of the lifetimes of single molecules, fluorescent labeling of long strands of DNA, and suspension of single strands of DNA will be discussed.

Genetic Analysis of YAC Integrity Following Yeast Transformation and Mitotic Growth. N. Kouprina, M. Eldarov, V. Larionov, R. Moyzis, and M. Resnick. National Institute of Environmental Health Sciences, Research Triangle Park, NC; and Center for Human Genome Studies, Los Alamos National Laboratory, Los Alamos, NM.

It is essential for the accurate characterization of the genome that the DNA within YACs be an exact replica of that in human cells, that there not be cloning artifacts and that the YACs be stable within yeast. However, present YAC cloning systems produce several types of cloning "errors" including co-cloned DNA sequences (chimeras from different regions of the genome), rearrangements during transformation and instabilities during growth. We have developed genetic systems that can signal physical changes in YACs. In these studies a cassette Alu-HIS3-Alu has been targeted to YAC12 which contains 360 kb from chromosome 21 (McCormick et al., 1990). These model YACs (telomere TRP1--HIS3--CEN URA3 telomere) were examined for loss of the HIS3 marker during transformation and subsequent mitotic growth. The frequency of internal deletions in mitotically growing cells was approximately 10^{-4} to 10^{-3}. Physical analysis of 30 $his3^{-}$ $TRP1^{+}$ $URA3^{+}$ colonies revealed the replacement of the 360 kb parental YAC by a smaller YAC, 50 to 310 kb in size. The new YACs contained the original telomeres. A likely source of deletions is recombination between repeats such as Alu's. These results are in sharp contrast to those obtained with YACs undergoing transformation. We found that nearly 30% of the transformed YACs lacked the HIS3 marker. The transformation-associated loss was also due to deletion, with the size varying between 50 and 180 kb. As expected a high frequency of deletions was observed even in YACs retaining the HIS3 marker; 10% of the HIS^{+} YACs exhibited deletions in the range of 50-100 kb. Similar results were obtained with another YAC containing a 360 kb insert of mouse DNA. Possible sources of the unexpected high level of transformation-associated rearrangements include methods of YAC isolation/preparation and transformation. Regardless of the reasons, the high level was subject to RAD52 control. The RAD52 mutant exhibited a greater than 10- to 20-fold lower level of human YAC instability during transformation as well as during mitotic growth as compared to an isogenic RAD^{+} strain. These results demonstrate that YAC instability can be genetically characterized. We are pursuing the genetic control of YAC instability to improve system for YAC cloning.

HIGH-PRECISION MAPPING OF YAC CLONES ONTO CHROMOSOME 21 BY FLUORESCENCE *IN SITU* HYBRIDIZATION AND IMAGE ANALYSIS.

Babetta L. Marrone, Evelyn W. Campbell, Thomas M. Yoshida, Sarah L. Anzick, Mary K. McCormick, and Larry L. Deaven.

Life Sciences Division and Center for Human Genome Studies, Los Alamos National Laboratory, Los Alamos, NM 87545.

YAC clones containing human chromosome 21 DNA inserts (100-400 kb, 200 kb average) from flow sorted chromosomes have been mapped onto human diploid fibroblast metaphase chromosomes by fluorescence *in situ* hybridization (FISH) and digital imaging microscopy. The goals are: 1) to provide a complete long-range physical map of chromosome 21 with 1 Mb resolution between YACs; and 2) to provide sub-regional location and ordering of known and unknown markers on the long arm of chromosome 21, particularly in the Down syndrome region (q22). Mapping of the fluorescent DNA probe to its location on the chromosome in a metaphase spread is done using a procedure modified from Lichter *et al.* (Science 247:64, 1990). YAC clones are indirectly labeled with fluorescein and the total DNA of the chromosome is counterstained with propidium iodide. Using a 520 nm long pass emission filter both the FISH signal and the whole chromosome can be viewed together. One image is acquired for each chromosome of interest containing the fluorescent probe signal in a metaphase spread. From the digitized image the fluorescence intensity profile through the long axis of the chromosome gives the total chromosome length and the probe position. The map position of the probe is expressed as the fractional length (FL) of the total chromosome relative to a fixed reference point which is designated FLpter. From each hybridization, 20-40 chromosomes/images are analyzed. Using this rapid and simplified procedure 47 YACs have been mapped onto chromosome 21 thus far. The mapped probes include 12 known markers (superoxide dismutase, collagen) and structural landmarks (telomeres and centromere). To confirm the order of a dense population of YACs within the Down syndrome region, a two color mapping strategy is used which involves locating an anonymous YAC relative to one or two known markers on metaphase chromosomes. Our chromosome 21 metaphase map has a 1-2 Mb resolution and the FL measurement of each probe has a typical standard error of 0.5 Mb.

Automated Methods for Large-Scale Physical Mapping

Patricia A. Medvick and Tony J. Beugelsdijk
Los Alamos National Laboratory
MS J580, MEE-3
Los Alamos, NM 87545

The Center for Human Genome Studies (CHGS) at Los Alamos National Laboratory (LANL) needs to produce high-resolution gridded membranes for screening their cosmid and YAC libraries. The demand for these membranes has been steadily increasing and is projected to drastically increase with the construction and distribution of additional libraries. Since March of 1992, a prototype automated system developed at LANL has been producing gridded membranes from our Chromosome 16 cosmid library. The system produces eight duplicate membranes in 2.5 hours, with each membrane representing up to 16 plates of the cosmid or YAC library. A technician is required only to provide the initial setup. Hardware and software improvements to the prototype, to improve user friendliness, permit random-order presentation of microtiter plates and list correctly completed plates at the end of a gridding session. The addition of blunt-tipped gridding tools now allow the gridding of YAC libraries. The ability to replicate libraries, another frequently repeated task at the CHGS, has been added. This addition required that another pair of microtiter plate dispenser/restackers, a second barcode reader, and software routines also be added.

Ongoing modifications to three system components on the prototype will increase system throughput. Increasing the size of the membrane holders will permit the gridding of twelve membrane copies from each plate presentation, and, thus, will reduce setup and cycling time. An additional tool drying station, along with a third gridding tool, will permit increasing robot speed, while allowing sufficient time for the gridding tools to dry between microtiter well sampling. We are also exploring the possibility ofconverting to high-density microtiter plates (384 well) with matching gridding tools. This will result in an immediate factor of 4 increase in throughput.

Steps are being taken to establish a Cooperative Research and Development Agreement (CRADA) with an industrial partner to provide commercial availability of our gridding system.

Our second generation system will be flexible and modular and will allow for the facile implementation of different tasks. The components for this system include a selective compliance assembly robotic arm (Adept Model 604S); a plate stacking system, which is easily modified to accept different types of microtiter plates; and an automated subsystem for tool cleaning. An expert system, as the user interface, will encourage the exploration of rules for automatic control of the biological decisions. Data flow between computer databases will be incorporated.

STRAND-SPECIFIC ORIENTATION OF SATELLITE DNA SEQUENCES DETERMINED BY FLUORESENT IN SITU HYBRIDIZATION.

Julianne Meyne and Edwin H. Goodwin
Los Alamos National Laboratory, Los Alamos, NM, USA.

A modification of the FISH procedure was used to demonstrate that satellite DNA sequences have distinct patterns of orientation on specific chromosomes. Some chromosomes, usually those with large blocks of tandem repeats, have sequences aligned in both the 5' to 3' and 3' to 5' directions. Other chromosomes have sequences aligned in the same orientation in a head-to-tail fashion. As the resolution of the FISH method is not absolute, the latter pattern must be considered to be the predominant orientation of the repeat. The method employed here would not detect very low numbers of repeats oriented in the opposite direction. It is known that many of the ubiquitous repeats such as Alu are aligned in both the 5' to 3' and 3' to 5' directions. As expected, the in situ hybridization pattern of these repeats were essentially the same in appearance with the strand-specific method as with the standard protocol for FISH. The alpha satellite, beta satellite, and simple sequence repeats of the classical human satellite DNA sequences demonstrated the chromosome dependent pattern described above. That is, some of the chromosomes showed hybridization to only one chromatid, while a few of the major heterochromatic sites of other chromosomes, most notably 1, 9, and/or 16, had fluorescent signals on both chromatids. The major satellite DNA of mouse and the non-telomeric locations of the telomeric sequence in some mammalian species also have a strand-specific orientation.

This work was supported by grants from the U.S. Department of Energy.

INTERVAL GRAPH ALGORITHMS FOR PHYSICAL MAP ASSEMBLY

Mark O. Mundt, Vance Faber, Carol A. Soderlund, Reid Rivenburgh, Mark Goldberg, Robert M. Pecherer, Amanda A. Ford, David C. Torney

Los Alamos National Laboratory and Rensselaer Polytechnic Institute

Interval graphs can be used to integrate pairwise overlapping clones into a physical map by enforcing the constraint that clones are fragments from a linear DNA molecule (Benzer). A graph represents the connectivity of a physical map when we take clones for vertices and overlaps for edges.

We developed algorithms to detect 'forbidden features' of a map, including either cyclical structures (involving more than three clones) or branching. We implemented the Modified PQ-tree algorithm of Korte and Mohring to provide an interval representation for a physical map.

Furthermore, since some overlaps participate in 'forbidden features', we have performed Monte Carlo experiments in which a map is constructed by including edges according to their probabilities and the resulting graph is kept if it is an interval graph. Thus, the probabilities of edges can be refined to reflect the local properties of many overlapping clones, and this has been achieved in the cosmid map of Human chromosome 16.

So that a user could easily use the interval graph algorithms, along with accessing data and alternative algorithms, the MAP software system was developed, based on a previous system for manipulating graphs. MAP depicts the aforementioned graphs and allows the user to edit these depictions. An initial clone layout is generated by an algorithm designed to work with noisy data; the clones are arranged to give the maximum overlap score (Churchill et al.). Using the algorithms mentioned above, MAP highlights 'forbidden structures' in a graph and gives suggestions for repairs. We have been connecting MAP with a database, allowing the user to interactively build, store, and reach consensus on multilevel genome maps.

We have used this software to identify potential problems in the cosmid map of chromosome 16, largely due to repeated DNA sequences. The same algorithms can aid in the integration of YAC clones into the map, using STS sequences or Alu-PCR for anchoring. Chimeric YACs can be detected if they lead to 'forbidden structures'. Our algorithms can be used to explore maps produced by other techniques as the genetic algorithm of Fickett and Cinkosky. Other generalizations of a graph representing a map may prove useful.

Benzer, S. On the topology of genetic fine structure. PNAS, USA, 45,1607 (1959)

Chruchill, G., C. Burks, M. Eggert, M. Engle, and M. Waterman Assembling DNA sequence fragments by shuffling and simulated annealing. In preparation

Fickett, J. W. and M. J. Cinkosky. A genetic algorithm for assembling chromosome physical maps. In preparation

Korte, N. and R. H. Mohring. An incremental linear-time algorithm for recognizing interval graphs. SIAM J. of Computing,18,68 (1989)

Construction and characterization of a human genomic YAC library with a low frequency of chimeric clones. C. Munk, E. Saunders, E. Campbell, K. Shera, P. Schor, M.K. McCormick, L. Deaven, R. Moyzis. Center for Human Genome Studies and Life Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545

Yeast artificial chromosomes (YACs) have been constructed using DNA isolated from an apparently normal, spontaneously transformed human lymphoblast cell line, GM 130 (46XY). High molecular weight DNA was prepared in low melt agarose plugs, partially digested by competition of EcoRI with EcoRI Methylase, ligated to YAC vectors pJS97 and pJS98, and transformed into *S. cerevisiae* strain YPH 250. Approximately 40,000 YACs with an average size of 180 kb have been identified, which is estimated to be 2X coverage of the genome. The frequency of chimeric YACs was estimated by fluorescent *in situ* hybridization (FISH) to human metaphase chromosomes. Fifteen of sixteen YACs resulted in a single, discrete FISH signal on homologous chromosomes. The remaining YAC resulted in two FISH signals in ~50% of the metaphases. This could be due to a chimeric clone or to the presence of repetitive sequences on the YAC, since the FISH signals were located at the telomeres of the chromosomes. If the YAC is a chimera, then the frequency of chimeric clones may be up to 6% in this library. The availability of non-chimeric YAC clones will facilitate construction of accurate physical maps of human chromosomes.

A Loosely-Coupled, Distributed Laboratory Notebook for the cDNA Community

Robert M. Pecherer and Joe M. Gatewood.
Los Alamos National Laboratory, Los Alamos, New Mexico, 87545.

There are many precedents for the use of commercial Database Management Systems (DBMS) to provide laboratory support for mapping and sequencing, and several, central repositories for map and sequence data. We are now developing a distributed, loosely-coupled database to link collaborating cDNA researchers into a network for localized project support, data sharing across projects, and global resource management. A database schema and a collection of interactive tools have been designed to provide management of cDNA clones, sequencing experiments, sequences, mapping experiments, and features. The database schema consists of a core of objects, relationships and attributes common to all projects, and a non-core portion customizable by individual labs. The interactive tools are being built around the schema core, and are customizable in terms of behavior. We have taken this approach in the interests of using the database and interface as a linking node at multiple sites engaged in various activities.

When sequence data are entered, vector sequences, cloning artifacts and repeats are tagged. Sequence similarity and homology detection are time-consuming operations and the subject of much research. Our approach is to perform comparisons using packaged software (such as BLAST or FASTA) and record results as time-stamped sequence features with annotation for begin, length, statistical similarity, etc. All values are obtained by parsing BLAST/FASTA output and are subject to user editing and confirmation. Since we can generate databases for any interesting sequences for BLAST/FASTA, we can use a single, uniform mechanism for detecting vector, repeats, cloning artifacts, previously encountered sequences and GenBank homologies. Further, we use the results of prior repeat homology comparisons to filter GenBank homologies to eliminate homologies due only to repeats. Time-stamping allows us to rerun homology comparisons when BLAST or FASTA databases are regenerated.

With respect to the community of collaborating cDNA researchers, commonality in the core database schema and the interface tools promote data sharing (as well as shared software) between and among projects. Sequence similarity to remote collections is detected by the same mechanism used to insure sequences are locally unique, and this should avoid duplication of sequencing effort. Centralized data management suggests one model for managing distributed, megabase sequencing projects. We view our current effort as prototype investigation of a loosely-coupled "federation" of locally controlled databases that promotes cooperation and collaboration.

Construction of restriction maps using simulated and real data

C. Soderlund[1], C. Hildebrand[2], N. Doggett[2] and C. Burks[1,3]
[1]Theoretical Biology and Biophysics Group,
[2]Center for Human Genome Studies,
Los Alamos National Laboratory, Los Alamos, New Mexico 87545

GRAM (Genomic Restriction Map Assembly) is a software tool that takes as input the fragment data for a set of clones and outputs one or more possible partial restriction maps. The algorithm uses two steps: (1) determine the set of unique fragments, and (2) permute the fragments such that the maximum number of fragments within each clone are contiguous. Due to the uncertainty in the data, step 1 cannot always be determined perfectly; the user can query the data through an interactive interface and detect these anomalies. Step 1 uses an algorithm based on clustering techniques[1], step 2 uses a modification of the shuffle routine used for sequence assembly [2], and the graphics are a modification of the graphics written for a previous restriction map assembly system[3].

We will show results from simulated data and from experimentally derived chromosome 16 data [4], as follows: (1) We will show, through simulated input data, that if there are no uncertainties in the data, GRAM gets the correct solution. (2) By adding uncertainty to the data, we will show that GRAM still arrives at a solution that is close to optimal, and that it is generally easy to detect the source of uncertainty in the data using the interactive graphics. (3) We will show examples of where GRAM was able to detect anomalies in real data, such as repeats and chimeric clones. (4) We will show some restriction maps assembled from real data.

[1] C. Soderlund, D. Torney, and C. Burks, "Calculating Shared Fragments for the Single Digest Problem", to be published in the *Proceedings of the Twenty-sixth Hawaii International Conference on System Science.*

[2] G. Churchill, C. Burks, M. Eggert, M. Engle, and M. S. Waterman, "Assembling DNA sequence fragments by shuffling and simulated annealing," (in preparation), 1992.

[3] C. Soderlund, P. Shanmugam, and C. Fields, "User documentation for the contig assembly (CA) programs," Technical Report MCCS-90-190, Computing Research Laboratory, New Mexico State University, 1990.

[4] R. L. Stallings, D. C. Torney, C. E. Hildebrand, J. L. Longmire, L. L. Deaven, J. H. Jett, N. A. Doggett, and R. K. Moyzis, "Physical mapping of human chromosomes by repetitive sequence fingerprinting," *Proc. Natl. Acad. Sci. USA*, vol. 87, 1990, pp. 6218-6222.

Assistance for Ethical, Legal, and Social Issues Projects

Michael S. Yesley
Los Alamos National Laboratory, Los Alamos, NM 87545
505/665-2523 Fax 505/665-4424, Internet: yesley_michael_s@ofvax.lanl.gov

Michael S. Yesley, working closely with Daniel W. Drell at the Human Genome Program Office, coordinates the DOE program on the ethical, legal, and social issues (ELSI) raised by the Human Genome Project. In this role, Yesley is involved in establishing the direction of the DOE ELSI program, serving as liaison with grant recipients and potential grant applicants, reviewing grant preapplications, arranging peer-review and technical evaluations of grant proposals, developing additional activities in support of the program including assembling the group of contractors into an informal consortium focusing on privacy and confidentiality issues, and representing DOE at meetings of the DOE-NIH Joint ELSI Working Group and other ELSI conferences and workshops.

LBL

Lawrence Berkeley Laboratory
Human Genome Center

Development of Mass Spectrometry Detectors

W. H. Benner, K. Hom, G. Rose and J. Jaklevic
Human Genome Center Instrumentation Group and Engineering Division
Lawrence Berkeley Laboratory, University of California, Berkeley CA 94720

Two major problems must be addressed before mass spectrometry can be routinely applied to the sizing of large DNA fragments. The size of fragments that can be ionized and vaporized must be increased beyond the current limits. The second problem, which has been our focus, is the development of impact ion detectors that can respond to impacts of large ions. Future mapping and sequencing efforts employing mass spectrometry would benefit from the development of ion detectors that would be sensitive to ions larger than several hundred thousand daltons, the approximate upper limit reported for the mass spectrometric analysis of proteins.

We have approached these problems by constructing a mass spectrometer test stand that can be optionally adapted for DNA fragment ionization studies or detector testing. We have constructed a matrix-assisted laser desorption time-of-flight (MALD-TOF) mass spectrometer and an electrospray time-of-flight (ES-TOF) mass spectrometer. Each is equipped with a detector chamber in which custom-made or commercial detectors can be tested.

The following types of detectors are being evaluated. Aluminum oxide and lead oxide based ion multipliers are being incorporated into a detector module so that their efficiency for large ions can be compared. Thin metal-oxide-silicon (MOS) structures are being evaluated as devices for direct capacitive detection of ion impact. Bolometric detectors measure the temperature jump that occurs when the kinetic energy of an ion is converted to heat during impact. New bolometers display a timing resolution of about 100 ns, a marked improvement over older versions. Several triboluminescent materials have been identified for use as substrates for molecular impacts. In this sensor, light is emitted from the triboluminescent crystal when it is fractured by an impacting ion. Electron-trapping (ET) materials capture low energy electrons in centers that can be stimulated to release the electron when the trap is vibrated by a phonon. As the electron falls to its ground state, a photon is emitted and detected by a photomultiplier tube. The phonon is produced when the kinetic energy of an ion is converted to heat following impact with the ET material. We will present testing methods, modelling approaches, and sensitivity analyses for these detectors.

This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Genome Program, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

THE ISOLATION AND MAPPING OF CHROMOSOME 21 cDNA CLONES

J.-F. Cheng, Y. Zhu, T. Torok, and D. Scott,
Human Genome Center, Lawrence Berkeley Laboratory, Berkeley, California.

Two methods based on hybridization of homologous sequences have been used to isolate chromosome 21 specific cDNA clones. The first method is derived from the cDNA selection method previously described by Lovett et al. (1991) and Parimoo et al. (1991), in which genomic DNA is immobilized on a solid support and used to capture cDNA sequences. The second method employs dot blot or plaque hybridizations of individual cDNA clones with genomic DNA, in which cDNA is immobilized on nylon filters.

Three cDNA libraries (fetal whole brain, HeLa cell and normalized thymus) are being screened for chromosome 21 specific clones. A flow-sorted chromosome 21 cosmid library is the source of genomic DNA probes. This cosmid library, which contains 6-fold coverage of chromosome 21, is free of mouse and rDNA clones. Twelve cosmids were pooled, and DNA was prepared from these pools.

In our cDNA selections, pooled cosmid DNA was biotinylated and hybridized to the cDNA libraries. The cDNA and cosmid DNA hybrids were then immobilized on strepavidin coated magnetic beads, and the hybridized cDNAs were recovered by PCR amplification. This method is straightforward and works on both non-normalized and normalized libraries. This method, however, requires secondary screening to distinguish the selected cDNA from clones which bound non-specifically.

In the filter hybridizations, pooled cosmid DNA was labeled and hybridized to the cDNA libraries on filters. Multiple copies of the same cDNA filter were prepared, and each was hybridized to a different set of cosmid pools. The cDNA clones hybridized to only one set of cosmid probe were examined further by probing back to the cosmid clones. This method is even simpler than the selection method, and the positive cDNA hybridizations are easily distinguishable from those due to the repetitive sequences.

All cDNA isolates are being hybridized to a set of chromosome 21 YAC clones whose chromosomal locations were determined by a STS contig map (Chumakov et al. 1992).

Lovett et al., (1991) Proc. Natl. Acad. Sci. USA 88, 9628-9632.
Parimoo et al., (1991) Proc. Natl. Acad. Sci. USA 88, 9623-9627.
Chumakov et al., (1992) Nature 359, 380-387.

Physical Mapping of Human Chromosome 21 in YACs

Jeffrey C. Gingrich, Farideh Shadravan and John Thomson
Human Genome Center, Lawrence Berkeley Laboratory, Berkeley, CA 94720

YAC contigs on chromosome 21q22.3 compiled using a combination of the screening approaches described below will be presented. The results confirm mapping of many of the clones on the chromosome 21 contig map recently published by Chumikof *et al.* (Nature Genetics, 359, 380-387). However, we have been unable to confirm some clones and have added additional clones to the published contig map. Furthermore we find a different map order for some of the STSs.

YAC contigs on chromosome 21q22.3 are being generated from clones derived from four YAC libraries by both hybridization and STS-based screening strategies. In the hybridization-based strategy, *Alu*-PCR products from cosmids and YACs that have been regionally assigned on chromosome 21 were used as probes for Southern dot blot hybridization to identify overlapping clones from the LANL flow-sorted chromosome 21 YAC library. DNA from pools of YACs from the ~2,000 member library, along with control positive YACs and YACs from a chromosome 21 library made at LBL (Gingrich *et al.* Genomics, in press) were PCR-amplified and probed. Overlapping clones have been confirmed by electrophoretic separation of the *Alu*-PCR amplified products from the clones in parallel to PCR products used as the probe. The high throughput of this screening method allowed screening 50 probes per week by one person.

The STS-based screening strategy recently published by Amemiya *et al.* (NAR, 20, 2559-2563) for the Washington University YAC library has been extended to include STS screening of the total genomic CEPH YAC library. In essence, the ~55,000 member YAC library arrayed on microtiter plates are pooled by plate (Dimension 1) and plate position (Dimension 2) into 12 plate pools. The 12 plate pools are further pooled by row to obtain 96 "superpools". STS screening is in two steps. First, DNA from the 96 superpools is PCR-amplified. Positives from the initial screening indicate which row (12) of a pool plate is to be PCR-amplified. Positives from the second round of PCR identify the plate(s) and plate position(s) of the clone(s) containing the STS. Since all of the DNAs are arrayed in microtiter plates, all steps of the screening are easily automated. We have used the pools from the Washington University and CEPH YAC libraries to identify clones for known STSs on chromosome 21. Collaborations with other groups have identified a number of YACs on numerous other chromosomes.

A Fluorescence In-situ Hybridization (FISH) Map of Human Chromosome 21 Consisting of 30 Genetic and Physical Markers on the Chromosome: Localization by FISH of 137 Additional YAC and Cosmid Clones With Respect to These Markers

Jeffrey C. Gingrich, Farideh Shadravan, and Stephen R. Lowry.
Human Genome Center, Lawrence Berkeley Laboratory, Berkeley, CA 94720

In order to rapidly map new anonymous clones on human chromosome 21, a FISH map of human chromosome 21 was created from DNA markers that have been used in physical and genetic maps of the chromosome. The map was compiled using yeast artificial chromosome (YAC) probes that encode markers distributed along the q arm of the chromosome. Fourty-three YACs that encode one of 28 different genetic and physical loci were used to compile this map. Two additional probes that recognize the centromere and the rDNA repeat sequences in the p-arm were also placed as reference markers on the map. For each probe, the location of the fluorescent hybridization signal was measured on metaphase chromosomes with respect to fractional chromosome length from p-ter (FL). The standard error (Sx) from 9 to 63 measurements of FL for the marker DNA probes indicate that the location of the chromosome markers is established to ± 1.9 Mb, assuming the chromosome is 50 Mb in size. As determined by FISH, the relative order and separation of the markers corresponds to other physical maps of the chromosome. Fifty-one additional YAC and 86 cosmid clones were also localized by FISH with respect to the markers on the FISH map. The cosmids, chosen at random from a flow-sorted chromosome 21 cosmid library, show some biases in chromosome distribution.

Physical Mapping of Human Chromosome 21 in Bacteriophage P1

Jeffrey C. Gingrich, Farideh Shadravan and Stewart Scherer
Human Genome Center, Lawrence Berkeley Laboratory, Berkeley, CA 94720

In preparation for a large scale DNA sequencing project of a 3-4 Mb region of human chromosome 21q22.3, we are generating a STS-based physical map of the region rooted first in YAC clones (see abstract by Gingrich, Shadravan, and Thomson). The STSs are being used to isolate the homologous bacteriophage P1 clones. The P1 clones will then be used as the sequencing template by the directed strategy developed at LBL for the Drosophila genome project.

Bacteriophage P1 clones are being identified by STS screening pools of clones from a 3X coverage human genomic P1 library obtained from DuPont. The DuPont P1 library is arrayed in ~125 microtiter plates. Each plate well contains twelve clones. The arrayed library has been pooled by plate and by plate position. Clones in the P1 library from chromosome 21q22.3 are being identified by STS screening first using available STS sequences. Additional STSs are being generated with the goal of one STS every 75 kb, about three times the number available now. New STSs are being generated from the sequence of M13 subclones from cosmids known to be contained within the YAC contigs and by sequencing ends of YACs within the contigs. Another source of STSs are a subset of the chromosome 21 polymorphic repeats being characterized here (see abstract by Lowry *et al.*). STSs are also being generated from ends of P1 clones for closure of the P1 contigs. Together, these approaches will identify the clones in the bacteriophage P1 library which come from chromosome 21q22.3.

A Fast Thermal Cycler Utilizing Standard Microtiter Plates

A. D. A. Hansen, M. Hugentobler, W. L. Searles and J. M. Jaklevic
Human Genome Center Instrumentation Group and Engineering Division
Lawrence Berkeley Laboratory, University of California, Berkeley CA 94720

We have developed a fast thermal cycler designed to perform the protocols for PCR (R) amplification in standard, thin-walled polycarbonate 96-well microtiter plates in a manner amenable to robotic handling. After loading with reagents, the plate is clamped with a heated lid without the need for individual well caps or other seals requiring manual intervention. Heat is transferred to the plate by rapidly-flowing circulating water in direct contact with the underside of the plate. This water flow may be switched among four reservoirs, normally held at 94° C, 55° C, 72° C and 4° C. The three hot flows provide the appropriate temperatures for DNA denaturization, primer annealing and extension, while the cold flow is used only at the end of the cycling process to terminate any further reactions and to allow the plate to be held in the apparatus until it is removed. Direct temperature measurements show that a 50 µL volume of liquid in a well reaches thermal equilibrium (within ~1° C of end-point) in 10 to 12 seconds under these switching conditions. All water flow components have low thermal mass, and the water flow rate is equal to several under-plate volumes per second. This allows a three-temperature cycle to be completed in less than one minute: an amplification to be completed in less than half an hour. In the present apparatus, one set of reservoirs can supply controlled-temperature flows to six plate stations. Tests using a 600 bp fragment showed uniform amplification and no detectable undesired sequences in wells at all locations in the plate.

The combination of (i) use of standard plates, (ii) no need for special plate caps, (iii) amenability for robotic loading and unloading, and (iv) fast cycling mean that this apparatus has the potential to greatly increase the throughput of DNA amplifications as required to meet the goals of the Human Genome Project.

This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Genome Program, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Automation of a Large-Scale Directed Sequencing Strategy

J. M. Jaklevic
Human Genome Center Instrumentation Group and Engineering Division
Lawrence Berkeley Laboratory, University of California, Berkeley CA 94720

A significant portion of the instrumentation development efforts at the LBL Human Genome Center is directed toward large-scale automation and system integration of procedures for genetic mapping and sequencing. This emphasis reflects our appreciation of the magnitude of the goals of the genome project, but, more importantly, the realization that there exist a number of bottlenecks in various existing and proposed protocols which must be addressed in order for more efficient mapping and sequencing strategies to be developed.

In order for us to develop an efficient design for automated systems, a thorough understanding of various alternative approaches in terms of throughput, cost, and amenability to automation must be carried out. Our approach is to develop a close, collaborative interaction between the biologists and instrumentation specialists in the iterative development of the various protocols used for specific tasks. A recent example which illustrates this approach is the design and implementation of an automated approach to transposon-assisted, directed sequencing of 100 kbase genomic fragments.

This specific activity illustrates many of the more general features of the instrumentation development program. Protocols are designed around a series of discrete modules which are then integrated into a complete system capable of carrying out specific protocols required in the experiments. These modules are either custom designed and fabricated within the laboratory or adapted from existing commercial instruments where appropriate. The instruments are then integrated into a functioning system with computer control of individual modules coupled with automated sample tracking, data acquisition, and record-keeping.

Functions that are currently carried out with existing modules include automated colony picking, library replication and pooling, preparation of polymerase chain reactions, and loading of agarose gels for assays. A significant improvement in system throughput and reliability has been achieved with the development of a dedicated image acquisition and data analysis system which facilitated the accurate determination of transposon locations in individual, mapped P1 clones. Sequence analysis is performed using a commercial sequencer following preparation of sequencing reactions using a protocol developed from a commercial robot using recently developed software. An overall approach to system integration of sample handling and data manipulation has been designed and partially implemented.

The modules developed for this protocol are adaptable for use in other applications. In addition, there are other modules under development that will eventually be integrated into the system to improve throughput. These include a multiple, 96-well microtiter plate thermal cycler capable of simultaneous, high-speed cycling of multiple microtiter plates, and a 96-well custom oligosynthesizer.

The application of this overall systems approach will be discussed in the context of large-scale directed sequencing together with reference to other potential applications. Individual posters presented at the conference will describe many of the individual modules in detail. In addition, longer range projects designed to develop advanced instruments to supplant existing modules will also be described.

This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Genome Program, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

An Imaging Station Designed to Operate in an Integrated System Performing Large Scale Directed Sequencing and Physical Mapping

J. E. Katz, J. M. Jaklevic, W. F. Kolbe and J. D. Meng
Human Genome Center Instrumentation Group and Engineering Division
Lawrence Berkeley Laboratory, University of California, Berkeley CA 94720

An imaging station capable of automated data acquisition, analysis, and interpretation has been designed to facilitate biological procedures including transposon insert analysis and STS pool analysis. The station can stand alone, but is intended to operate as part of a larger integrated system performing large scale directed sequencing and physical mapping tasks.

The imaging system is based upon a cooled CCD camera for low light level detection with an associated computer controlling sample manipulation, image acquisition, instrument operation, and data processing. The analysis of images is facilitated through the use of standard sample holders which register objects (e.g., gels, microtiter plates) at specific locations within the CCD image field. The sample holders are designed to be compatible with other automated components of the system.

Image processing algorithms have been integrated into several automated analysis programs. These programs can automatically find gel lanes, locate peaks, and provide line scans of data. Multiple calibration lanes are fitted to an exponential function which is then used as a basis for computation of fragment sizes. For transposon insert analysis, 14 x 14 cm agarose gels with 32 lanes are used. In the case of STS pool analysis, 96 data lanes are automatically characterized on a similar size gel previously loaded under robotics control. Data from the image analysis are then processed to determine particular biological parameters of interest. An important component of the analysis is the precise assignment of fragment length using multiplex calibration lanes. Experience with both transposon and pool analysis has demonstrated that the results obtained from the automated system yield superior accuracy and reproducibility in significantly less time than existing manual methods and provide output data directly compatible with our database system.

Details of the hardware and software design will be presented together with examples of data acquired and analyzed in a large scale directed sequencing and physical mapping project. We will also discuss other elements of the integrated automation approach that relate to the imaging system, including gel loading and electrophoresis optimization.

This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Genome Program, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Construction of a Bacteriophage P1-based Physical Map of the *Drosophila melanogaster* Genome. William J. Kimmerly*, Cheryl Ericsson*, Oroon Hubbard*, Victor Stevko*, Karen Stultz*, Gerald M. Rubin#*, Christopher H. Martin*, and Michael J. Palazzolo*#. Human Genome Center*, Lawrence Berkeley Laboratory, Berkeley CA 94720 and Department of Molecular, Cell and Developmental Biology#, University of California, Berkeley, CA 94720.

We are constructing a physical map of the 125 megabase euchromatic genome of the fruit fly *Drosophila melanogaster* based on a five-hit P1 genomic library consisting of approximately 10,000 clones (of 90 kb average insert size) by a double-ended clone limited approach. We use a PCR screening assay to map sequence-tagged site (STS) markers to various memebers of the genomic library. The STS markers come from three sources: (a) known genes from databases, (b) genomic DNA flanking P element insertions recovered from lethal enhancer trap lines, (c) and the terminal sequences from the genomic inserts of individual P1 clones (hence the term, double-end clone-limited strategy). For this approach, we have devised a P1 DNA template preparation protocol which allows direct sequencing using fluorescently-labelled dideoxynucleotide chain terminators and the ABI 373A DNA Sequencer. The STS markers obtained from the above three sources are used to devise primers for PCR, which are in turn used to screen pools of P1 clones to determine the members of the library that share a particluar STS. This information allows the assembly of P1 contigs, overlapping P1 clones which cover large regions of the fly genome. To establish the validity of this approach to genome mapping, and to provide a measure of the completeness of the P1 genomic library, we are currently building P1 contigs to three regions of the fly genome. Two of these regions, the *Bithorax* complex (BX-C) and the *Antennapedia* complex (ANT-C) encode several key developmental regulatory genes. The third region we are studying is the 1.8 megabase 34D-36A region of chromosome 2L, which contains the well-studied alcohol dehydrogenase (*Adh*) gene. The P1 clones assigned to each of these contigs will later be used to generate sequencing templates for our long-term goal, the elucidation of the sequence of the entire euchromatic genome of the fly.

Automation of Gel-Based PCR Assays

W. F. Kolbe, J. M. Jaklevic, J. E. Katz, M. J. Palazzolo and M. J. Pollard
Human Genome Center Instrumentation Group and Engineering Division
Lawrence Berkeley Laboratory, University of California, Berkeley CA 94720

Assays based on the detection of PCR amplification play a key role in almost all of the mapping and sequencing projects currently underway in the LBL Human Genome Center. In the next few years, biologists in the Center plan to map thousands of sequence tagged sites (STSs) requiring hundreds of thousands of PCR assays. A similar large number of PCR reactions will be required to support large-scale sequencing efforts. The setup and analysis of PCR reactions are in fact the major rate-limiting bottleneck in the Center's biological experiments.

For this reason, the instrumentation group has set out to develop automated procedures for performing gel-based PCR assays. Specific steps in the protocol which have been automated or are in the process of automation include preparation of PCR reactions, loading of agarose gels with PCR products, running gels in multiplex systems designed for high throughput and direct digitization of ethidium bromide stained gels. The resulting images are then analyzed for the presence or absence of PCR products together with precise measurements of fragment sizes for transposon sequencing and mapping.

A key to the approach is the use of standardized formats for gel casting and lane loading. This facilitates robotic gel handling and improves the efficiency with which the software can locate individual lanes within the image. The PCR reactions and gel loading are performed using a Beckman Biomek robot which has been modified to accommodate the specific tools developed at our laboratory. Custom software modules, used in the development of control programs, provide access tothese specific tools.

A modular, cooled multi-gel system capable of running five, 32-lane gels at a time has been developed. Special consideration has been given to the stability and uniformity of the agarose gel separations in order to reduce "smiles" in individual bands and curvature in the lanes. The system can also be adapted for running 96 lanes per gel in applications where precise band size is less important.

This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Genome Program, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098. Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

Automated Determination of PCR-Amplified DNA Concentrations by Fluorescence Detection

W. F. Kolbe, M. J. Palazzolo, J. E. Katz and J. M. Jaklevic
Human Genome Center and Engineering Division
Lawrence Berkeley Laboratory, University of California, Berkeley CA 94720

PCR-based STS content assays are a powerful and proven tool for the construction of physical maps. The mapping of even moderately sized genomes can require the performance of tens, and even hundreds of thousands of individual PCR reactions. One of the rate-limiting factors in this type of analysis is the characterization of the PCR products by gel electrophoresis, even though it is only the presence or absence of a PCR product and not its precise size that is of interest.

Recently others have developed a qualitative scoring system for the detection of amplified PCR products based on the enhancement of ethidium bromide fluorescence in the presence of dsDNA. We have incorporated this technique into a machine for automatically analyzing PCR samples loaded into 96-well microtiter plates. Ethidium is added to each well and a UV fluorescence image is obtained with a cooled CCD camera. The digitized image is then analyzed by processing software which locates the central area of each well, measures its integrated intensity and predicts the DNA content based upon comparison with concentration standards loaded into a few of the wells.

We will discuss the results obtained with this instrument and compare them with the more traditional gel electrophoresis approach. The range of DNA concentrations which can be determined as well as the precision of the measurements will be described. Measurements using fluorescent dyes other than ethidium bromide as a means of determining DNA content will also be discussed.

This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Genome Program, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

HGCdb: A Database for Human DNA Sequence Information*

S. Lewis, J. McCarthy, E. Theil, A. Aggarwal, D. Davy, S. Pitluck, M. Palazzolo

Human Genome Computing Group
Lawrence Berkeley Laboratory, Berkeley, CA. 94720

HGCdb is a variant of ACeDB, a suite of database and display software originally developed by Richard Durbin and Jean Thierry-Mieg to meet the needs of the C. elegans project. It includes all the functionality found within ACeDB and extends those capabilities to meet the new requirements of the LBL chromosome 21 sequencing project. It is being used to maintain and provide information for both laboratory personnel and the chromosome 21 research community.

There are three aspects — schema design, data presentation, collaboration — of ACeDB that make it an attractive approach. First, one unique feature of ACeDB is the ability to continuously refine the database schema to match ongoing research needs. This is a critical feature enabling timely responses to laboratory requirements. Second, numerous graphical displays of genomic data are already available, saving us considerable development effort. Furthermore, ACeDB provides an independent simple graphics library that is completely portable across many platforms. This enables us to quickly develop our own customized data displays as required (as we have done for flyDB developed for the Drosophila physical mapping project). Finally, ACeDB is also being used for human sequencing data at the Sanger Sequencing Center in Cambridge, U.K. We are collaborating with developers there.

LBL is exploiting a directed sequencing technique on this project. The implications for laboratory data management are significant. A directed strategy requires, by definition, that biologists know the complete heritage of each DNA sequence and its position in relationship to other DNA sequences. It is this knowledge that simplifies the sequence assembly process and makes it a more tractable problem. The database must record: all the subclones derived from each P1 clone, the P1 subclone map, the transposon insertion map of each subclone, descriptions of all the transposon inserted priming sites derived from each subclone, and the sequencing status and results for every priming site. The recorded data are available both graphically and computationally.

This physical map and sequence data on chromosome 21 generated by the project are available to the community through this database. The data includes not only LBL's P1 physical map but also the corresponding linkages to the Genethon YAC map provided by STS's derived at both laboratories. The database incorporates high resolution maps of individual P1's made before sequencing and the sequence data itself. Collaborative work on a mechanism for public access service is under way. Emphasis is on a graphical presentation that looks and feels natural for biologists.

---

* This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Genome Program, of the US Department of Energy under Contract No. DE-AC03-76SF00098.

A method for heterozygous carrier screening using an E.coli mismatch binding protein, mutS.

A.I.Lishanskaya and J.Rine, Human Genome Center, Lawrence Berkeley Laboratory, Berkeley, CA 94720

The proposed method makes use of the ability of MutS, one of the E. coli mismatch repair proteins, to recognize and bind to DNA molecules containing mismatches. The experimental strategy is as follows. The genomic DNA sequence of interest is PCR amplified using 5'-end labeled primers, and the resulting PCR products are denatured and reannealed. In the case of heterozygote carriers of any mutation within the amplified sequence, 4 different hybrid molecules are formed: 2 homoduplexes and 2 heteroduplexes. MutS protein recognizes mismatches within heteroduplex molecules. The binding is detected by a gel mobility-shift assay. This strategy was first used for detection of the ΔF508 3-bp deletion in exon 10 of the cystic fibrosis gene (A.Lishanskaya, E.Ostrander and J.Rine, in preparation). The size of the PCR product with the site of ΔF508 in the middle is 100 bp. Heteroduplexes with this deletion in one of the strands have a bulge formed by 3 looped out bases in the complementary strand. MutS binding to such heteroduplexes is consistently much stronger than to homoduplexes. In a blind experiment using this strategy, mutS was able to correctly detect all five heterozygous carriers of ΔF508 mutation among 15 individuals. This method was also used to detect heterozygosity for the point mutation G542 in exon 11 of CFTR gene resulting in G/A and T/C single-base mismatches in a reannealed PCR product 141 bp long. Several non-gel versions of the mutS-based technique are being tested. In one of them using a membrane impregnated with protein, mutS was able to distinguish between mismatched and perfectly matched short synthetic oligonucleotides, offering the potential of a gel-free assay for heterozygosity.

Isolation and Characterization of Dinucleotide Repeat Polymorphisms from Human Chromosome 21. S.R. Lowry, R. Blajez, K. M. Wilson, J.R. Rine, and E.A. Ostrander. Lawrence Berkeley Labotatory. Berkeley, CA 94720

Dinucleotide repeat polymorphisms specific for human chromosome 21 have been isolated and characterized. A chromosome 21 cosmid library was initially provided by Lawrence Livermore National Lab. This library was prepared from flow-sorted chromosome 21 material which was obtained from a hybrid mouse cell line whose sole human component is chromosome 21. In an initial screening, clones containing mouse sequences and ribosomal repeats were identified and removed. From this reduced library four pools containing 400 cosmids each were prepared. Sublibraries, containing 300-800 bp inserts, were then prepared from each pool. A portion of these libraries were arrayed and probed with a ^{32}P-labeled $(CA)_{15}$ oligonucleotide. 120 positives were selected from the initial screening for further study. A subset of approximately forty of these 120 have been sequenced and PCR primers have been made which identify polymorphic STSs. These markers appear to represent new STSs which do not appear in current database listings. A compilation of these markers and corresponding primer sequences is shown. The utility of these markers for genetic mapping has been investigated using a nonradioactive mapping technique. In this assay, one primer from each marker set is tagged with fluorescein and PCR reactions are performed for 35 cycles. PCR products are then rapidly sized using an automated DNA sequencer (Pharmacia A.L.F.). Exemplary data from CEPH family 1341 with marker 21.H1, which has six alleles in this family, is shown.

These genetic markers are being placed on two physical maps by PCR screening. Initially, markers are being placed on a YAC contig map which spans most of chromosome 21. Location of the mapped markers is shown. Markers will also be localized using chromosome 21 radiation hybrids in collaboration with David Cox. The correlation of polymorphic genetic markers with substantially compete physical maps should facilitate the identification of loci of interest on chromosome 21.

Data Management Tools For Genomic Applications: A Status Report*

Victor M. Markowitz and Arie Shoshani

Data Management Research and Development Group
Lawrence Berkeley Laboratory, Berkeley, CA 94720

We briefly review the SDT and QST tools and discuss their use in developing genomic data management systems. SDT and QST aid scientists in defining and in querying databases using object-oriented concepts. These tools increase the productivity, clarity and correctness of the database development process, and allow scientists to deal with concise application-specific (e.g. genomic) structures and operations rather than large, hard to comprehend, database definitions and queries. Consequently, scientists are insulated from the underlying database management system, and thus can avoid learning database management specific concepts and query languages such as SQL. SDT is a tool for developing database definitions for relational database management systems (DBMS). SDT provides a powerful and easy to use design interface for non-technical users, based on a version of the Extended Entity-Relationship (EER) model. EER schemas can be specified using either a regular text editor, or using a graphical schema editor called ERDRAW, and are subsequently mapped by SDT into relational DBMS schema definitions, including procedures for maintaining referential integrity constraints. Additionally, SDT generates a metadatabase that contains information (metadata) on the EER schema, the relational schema generated by SDT, and the mapping of these two schemas. SDT and ERDRAW have been implemented on Sun SPARCstations, and currently target SYBASE 4.0, INGRES 6.3, INFORMIX 4.0, and ORACLE 6.0 DBMSs. QST is a tool for specifying database queries in terms of objects. Users need to be aware only of the existence of objects and attributes, and are guided in the process of specifying queries. QST allows both textual and graphical specifications of queries. QST is based on a query language developed by us, called the Concise Object Query Language (COQL). COQL is unique in its conciseness, in its support of inheritance, and in the capabilities it provides for defining application-specific structures. Queries specified using QST are subsequently translated into SQL queries or procedures. QST uses the metadatabase generated by SDT for inferring the connections (paths) between the objects specified in object-level queries, and for mapping information to translate these queries into SQL queries. QST has been implemented on Sun SPARCstations for SYBASE 4.0, and will be implemented for other relational DBMSs as well. The research related to database schema and query translations is described in several papers published in scientific journals and conference proceedings. ERDRAW and SDT are currently used at over 35 locations worldwide. QST was released recently and is currently used at 5 locations. ERDRAW, SDT and QST have been used for the development of several genomic data management systems, such as the Chromosome 22 database at the University of Pennsylvania, the Integrated Genomic Database at the German Cancer Research Centre, Heidelberg, and databases at Genethon, Paris.

Acknowledgments. SDT was implemented by Weiping Fang. Ernest Szeto contributed to the design of ERDRAW and QST and implemented both. The outstanding quality of their work is greatly appreciated.

---

* This work was supported by the Applied Mathematical Sciences Research Program and the Office of Health and Environmental Research Program, of the Office of Energy Research, U.S. Department of Energy, under Contract DE-AC03-76SF00098.

Data Management Tools for Laboratory Information Management Systems*

Victor M. Markowitz and I-Min Chen

Data Management Research and Development Group
Lawrence Berkeley Laboratory, Berkeley, CA 94720

We are developing data management tools for molecular biology Laboratory Information Management Systems (LIMS). The goal of these tools is to ensure the rapid development of LIMS data management applications that (1) are flexible and easily adaptable to a variety of protocols, (2) can be easily modified and maintained, and (3) allow users to interact with LIMS in terms of their own frame of reference, that is, molecular biology objects and protocols. We have developed a data model called the Object-Protocol Model (OPM), that provides constructs necessary for defining LIMS-specific data structures in terms of objects and protocols. In OPM objects and protocols are classified into classes and are qualified by attributes that are associated with value classes. Connections of object and protocol classes are expressed in OPM via attributes. Protocols often involve a series of steps, which are also protocols. OPM supports the specification (expansion) of protocols in terms of alternative and sequences of component (sub) protocols, thus allowing protocol designers to progressively specify the protocol in increasing levels of detail. We plan to develop graphical OPM schema and query editors that will allow specifying LIMS data structures and queries in terms of objects and protocols. The OPM schema and query editors will be coupled with tools for translating OPM schemas and queries into definitions and queries for commercially available database management systems (DBMS). For relational DBMSs, these tools will first map OPM schemas and queries into intermediary Extended Entity-Relationship (EER) schemas and queries, and then use existing tools for generating relational database schemas from EER schemas (EER schemas are translated into relational database definitions by a tool developed by us, called SDT, also presented at this workshop), and for generating relational queries from EER queries (EER queries are expressed in the Concise Object Query Language and are translated into SQL queries by a tool developed by us, called QST, also presented at this workshop). The OPM schema editor has been implemented on Sun SPARCstations. The mapping underlying the OPM schema translator is fully specified and is currently implemented in C++ on Sun SPARCstations. The research related to OPM and the OPM to EER translation is described in two LBL reports. We have tested the capabilities of OPM by modeling the large scale sequencing LIMS application in Leroy Hood's laboratory.

Acknowledgments. We want to thank Tim Hunkapiller for assisting us in modeling the large scale sequencing application, and Frank Olken for his contribution to the early work on modeling protocols. Ofer Ben-Shachar contributed to the design of the OPM editor. The OPM editor was implemented by Ofer Ben-Shachar, Francis Pang, and Carol Jean Smith. We greatly appreciate their excellent work.

---

* This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Genome Program, of the US Department of Energy under Contract No. DE-AC03-76SF00098.

Biological Aspects of Directed Large Scale Genomic Sequencing at LBL

Chris Martin, Carol Mayeda, Cheryl Davis, Mike Strathmann, Kaoru Yoshida, Bill Kimmerly, Mike Palazzolo; Lawrence Berkeley Laboratory, Human Genome Center, Berkeley, CA 94720

Our major long term goals are to elucidate the sequence of the entire *Drosophila* genome as well as large sections of human chromosome 21 (see Rine et al.). To accomplish these goals it is essential to: (1) develop resilient protocols for the conversion STS content maps into sets of mapping clones that are capable of faithfully maintaining large cloned eucaryotic inserts, (2) design and implement a directed sequencing strategy, (3) develop the automation of the molecular biological elements of this directed strategy, and (4) develop computer programs that will facilitate assembly, editing, and rapid analysis of the sequencing data. In the past year we have solved many of the biological problems that have limited large-scale sequencing efforts. As a proof of principle we sequenced an 80 kb segment of the *Drosophila* BX-C with a four month, 4 FTE effort that also included significant amounts of new technology development. This is among the first examples of a large sequence determined exclusively in a directed fashion. The automation of these procedures has started to come on line, with more expected in the coming year (see Jaklevic et al.). The design of specific computer programs that are tailored for these directed strategies have recently been initiated (see Theil et al.).

A variety of data suggest that the bacteriophage P1 cloning system can play a key role in providing a source of DNA for both high resolution physical mapping and large-scale sequencing (see Kimmerly et al.). Protocols have been developed whereby STS markers taken from both the literature and our own physical mapping experiments can be used to convert a YAC-based map into sets of overlapping P1 clones. Gaps can be filled in a straightforward fashion by direct sequence analysis from the ends of the cloned P1 inserts.

A key problem in sequencing is the gap between the size of the cloned inserts (40 kb - 1 megabase) and the size of the sequence that can be obtained from a single priming site (350-400 bp). We had previously developed a successful approach of transposon-facilitated DNA sequencing that uses mobile priming sites to sequence 3 kb fragments in a directed fashion. We realized that the final problem in developing an effective nonrandom sequence procedure was to develop an efficient mechanism for organizing a set of 3 kb subclones made from fragmented P1 clone into a set of 3 kb minimally overlapping transposon targets.

To accomplish this goal we developed a novel strategy of high resolution physical mapping. The experimental methodology produces a set of minimally overlapping clones whose average size is 3 kb. Every 3 kb there is a sequenced region of about 400 base pairs. This strategy shares some formal similarities with STS content mapping. However, it differs in several important ways. The sequence tags can be mapped in relation to each other with precise distance information, orientation, and with gene size resolution. To differentiate the features of these high resolution sequence tags from STSs we call them DOGtags (Distance, Orientation, Gene size resolution). We refer to the mapping strategy as DOGtagging.

By juxtaposing STS content mapping, DOGtagging, and transposon-facilitated DNA sequencing we have developed a four-stage assembly line approach to directed large-scale sequencing. The mapping information is recorded at each stage with increasing resolution. This approach was developed to minimize the problems in shotgun sequencing, including: (1) the minimization of redundant sequencing, (2) the elimination of the assembly problems,.(3) the minimization of the need for custom synthesis of one-time sequencing primers,.(4) the elimination of useless sequencing of the vector, the overlaps of the mapping clones, and the regions sequenced by the community,.(5) the introduction of well characterized sequencing primers into the target sequences , and (6) finally, and perhaps most importantly, this set of procedures is well-suited to automation and a truly effective and cost-efficient scale up.

A reliable DNA template preparation protocol for plasmids with transposon insertions.

Carol A. Mayeda, Christopher H. Martin, Cheryl A. Davis, and Michael J. Palazzolo.

Human Genome Center, Lawrence Berkeley Laboratory
Department of Molecular and Cell Biology
University of California
Berkeley, Ca. 94720

A major objective in our *DROSOPHILIA* Genome Center Lab at Lawrence Berkeley Labortory is the development and implementation of a directed strategy for large-scale genomic sequencing. A key element of this approach is the utilization of the transposon gamma delta, to mobilize sequencing priming sites throughout a target DNA fragment.

Initially, plamids containing this transposon were prepared using various standard protocols, but were found to be poor templates for automated fluorescent sequencing. Subsequently, we have modified a basic boiling plasmid preparation procedure, which now produces consistently reliable templates. We will present comparative results of sequencing reactions using several different DNA preparation protocols, as well as the modified boiling prep method we are currently employing in our project.

System Integration of Automated Modules for Large-Scale Mapping and Sequencing

J. D. Meng, J. M. Jaklevic, S. E. Lewis, E. H. Theil, D. C. Uber and M. D. Zorn
Human Genome Center and Engineering Division
Lawrence Berkeley Laboratory, University of California, Berkeley CA 94720

A key element in the development of an automated laboratory for biological research is the design of a control system with which to integrate the various steps involved in a typical protocol. We have adopted an approach in which a protocol is broken into a sequential series of modules, each of which performs an individual hardware or software task in a configuration determined by the protocol. System integration then consists of defining the specifics of tasks performed by the individual modules and tracking the flow of data and materials through the linked series of modules. We describe a computerized system designed to perform these functions in a variety of applications within the mapping and sequencing projects at the LBL Human Genome Center. The system consists of a central control computer with a dispatcher for maintaining and recording the identity of materials and data as they pass through the system and for supplying modules with protocol-specific setup parameters when material or data arrives. Support for the dispatcher includes the ability to edit protocols, using a library of existing protocols as a template.

In the current stage of development, the system design is being carried out with individual modules replaced with software simulators that mimic the eventual hardware components. As the system approaches completion, drivers for individual instruments will replace the simulators in the appropriate steps in the protocols. In addition to facilitating the design of the system integrator, this approach has the advantage that realistic simulations of candidate protocols can be performed using a complete mock-up of the proposed hardware and software modules. Results of a simulation of a large-scale, clone-limited STS mapping project will be presented.

This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Genome Program, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Atomic Force Microscopy of Biochemically Tagged DNA

Matthew N. Murray[1], Helen Hansma[2], D. Frank Ogletree[3], William F. Kolbe[1]
Sylvia Spengler[1], Cassandra Smith[4], Charles Cantor[4], Miquel Salmeron[3]

[1] Human Genome Center
Lawerence Berkeley Laboratory
Berkeley, CA 94720

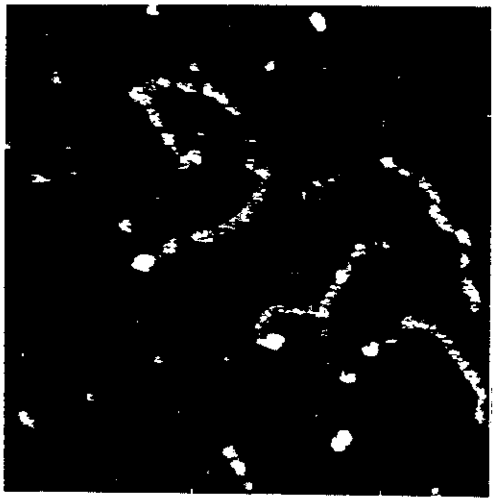
[2] Department of Physics
University of California
Santa Barbara, CA 93106

[2] Material Science Division
Lawerence Berkeley Laboratory
Berkeley, CA 94720

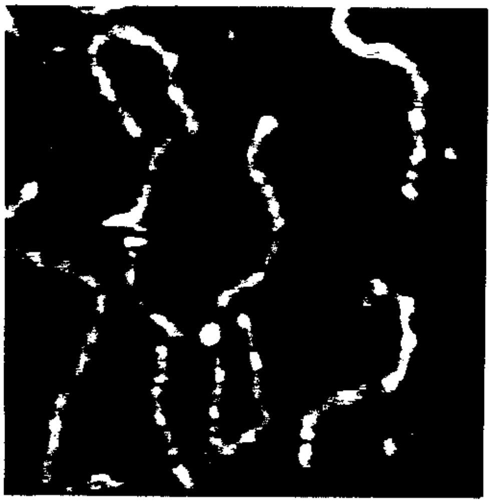
[4] Chemical Biodynamics Division
Lawerence Berkeley Laboratory
Berkeley, CA 94720

Small fragments of DNA of known length were made using the polymerase chain reaction. These fragments had biotin molecules (vitamin H) covalently attached to each end and were then labeled with streptavidin. This tetrameric complex was expected to bind up to four DNA molecules via their attached biotin molecules. The DNA was then imaged with the atomic force microscope. Images revealed the protein at the end of the DNA strands as well as the presence of dimers, trimers, and tetramers of DNA bound to a single protein, as expected theoretically. Imaging time was approximately one minute.

With these results we believe that we have shown that the AFM does have sufficient resolution to map DNA. In its simplest form mapping involves the measurement of the physical distance between two points of the DNA. In this experiment we have demonstrated the ability of the AFM to perform this task by attaching a large protein marker to genetically engineered pieces of human DNA and then using the AFM to locate the marker and measure the known length from the protein to the other end of the DNA.



DNA fragments with protein tag



Trimer of DNA fragments

This work has been supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, Materials Science Division of the US Department of Energy under contract No. DE-AC03-76SF00098.

Beckman Biomek Workstation Program Development

M. J. Pollard and G. Gonzalez
Human Genome Center Instrumentation Group and Engineering Division
Lawrence Berkeley Laboratory, University of California, Berkeley CA 94720

An important component of our overall automation effort is the development of unique applications using a Biomek 1000 robotic workstation as the instrumental platform. A general feature of these applications is the use of the robotic pipetting capabilities, in the low volume range of 1-20 µl, of the Biomek workstation. These applications have required extensive hardware development that is incompatible with the standard Genesis software package available with the instrument. In order to circumvent this limitation, we initiated the development by the Stanley Reifel Company of a versatile language (BiomekQB) capable of accessing the specialized hardware we have developed. At the core of the language is a library of low-level control functions that can be linked to the Microsoft QuickBASIC programming language. BiomekQB has virtually all of the capabilities of Beckman's Biotest3 control software but now permits programming in a high level programming language environment.

Four programs developed in this environment and currently in use are:

1. *PCR setup.* This program dilutes the overnight growths of bacterial cells/DNA templates and then combines them with the primers prior to PCR thermal cycling. Samples are prepared in the Perkin Elmer 9600 reaction tube trays.

2. *Agarose gel loading.* After samples have been amplified by PCR thermal cycling, they are loaded into precast agarose gels. For transposon insert analysis, two 14 x 14 cm agarose gels, each with 32 lanes, are automatically loaded with samples from a Perkin Elmer 9600 reaction tube tray. To implement STS pool analysis, 96 lanes are loaded in the same size gel. The gels are loaded dry and in a format that is compatible with subsequent imaging.

3. *DNA sequencing setup.* After double stranded DNA preparation, each sample is combined with each of the four dye-labeled nucleotides. Samples are prepared in a Perkin Elmer 9600 reaction tube tray for subsequent PCR processing to attach the dye-labeled nucleotides to the DNA templates.

4. *DNA sequencing sample pooling.* After PCR processing, there are four samples of each template, each of which has been labeled with one of the four nucleotides. The four samples of each template are then pooled together prior to introduction into the ABI sequencer.

This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Genome Program, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098. Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

A PHYSICAL AND GENETIC MAP OF HUMAN CHROMOSOME 21: A PRELUDE TO SEQUENCE

J. Rine, R. Blajez, J.F. Cheng, J. Gingrich, S. R. Lowry, E. Ostrander, S. Scherer, D. Scott, F. Shadravan, T. Torok, K.M. Wilson, and Y. Zhu, Human Genome Center, Lawrence Berkeley Laboratory, Berkeley, California 94720

The advances in physical and genetic maps have set the stage for the next biological goals of genome research: obtaining genomic sequence from substantial regions of the genome and preparing the genetic infrastructure for genotyping populations. Toward these ends we are preparing to sequence a 3-4 megabase region from a medically significant portion of chromosome 21 and are working toward saturating the chromosome with genetic markers.

In evaluating potential sources of DNA for sequencing templates the YACs used in the chromosome 21 STS map present serious problems. Only 28% of YACs spanning three or more STSs fail to exhibit obvious deletions. Thus, YACs appear to be poor choices for sequencing studies. The rapid success of the directed genome sequencing strategy at LBL has focused our attention on the human P1 library of Shepherd and Sternberg. We have constructed pools 1000 clones deep from this library and have screened this library with STSs from the 22.2 to 22.3 region from chromosome 21. These P1s are now in the pipeline for large scale directed DNA sequencing. This process involves the construction of physical maps in which the distance (d) and orientation (o) of each gene-sized sequencing template (g) is known and is tagged by sequence from each end. This procedure, known as dog tagging, offers the best known method for large scale genome sequencing and avoids the sequence assembly headaches of random strategies.

We have used marker selection techniques to isolate a large number of simple sequence repeats from chromosome 21 as a source of genetic markers. STSs produced from approximately 40 of these repeats have been assigned to the map with resolution of a few hundred kilobases. We have approximately doubled the density of genetic markers on this chromosome, making it the most densely marked human chromosome.

Our physical mapping efforts have focused on the distal third of the q arm. In this region we have mapped approximately 280 YACs plus cosmids by FISH, and have constructed contig maps. This mapping has allowed us to detect and correct mapping errors in the recently published map (Chumakov et al, 1992) including misplacement of genes by as much as 2 megabases. Corrected maps for these regions will be presented.

We have developed methods for physical selection of cDNAs corresponding to mapped YACs and cosmids. We have mapped 21 new cDNAs to their respective location on chromosome 21. By sequence analysis each of these define new genes and pioneer proteins. The cDNA effort is now focused on saturation of the multimegabase target of the genomic sequencing effort.

The capacity to produce genome information has outstripped the capacity of formal publication procedures to disemminate it to the community. To help close the information gap between producers and consumers, all of our unpublished cDNA sequences have been deposited in the cDNA Inform database (LANL), all genetic markers in GDB, and the sequences from which they are derived in GenBank. In addition, we are establishing a public database at LBL that will serve as an open notebook for the mapping data on chromsome 21 and for the sequence information from the P1 clones. Our mapping data enter this database directly and the sequence enters as each 3 kilobase dog unit is complete.

Informatics for LBL's Human Genome Center*

E. Theil, D. Davy, S. Lewis, V. Markowitz, J. McCarthy, S. Pitluck, E. Veklerov, M. Zorn

Human Genome Computing Group
Lawrence Berkeley Laboratory, Berkeley CA 94720

Lawrence Berkeley Laboratory is focusing on large scale sequencing, with the immediate aim of sequencing scientifically and medically important regions of chromosome 21. This requires a significant computational component. Our goal is to provide that component in close collaboration with biologists and instrumentation scientists. Among the informatics projects currently underway at LBL are:

1. **Sequence assembly software:** The transposon-based approach to sequencing used at LBL is facilitated by assembly software whose model of the data strongly reflects the particular strategy. For this reason, we are collaborating on a full suite of customized software, with an emphasis on a unified model and consistency in the user interface at all levels of the assembly process.
2. **Sequence Analysis:** Research on empirical statistical analysis of sequence data is under way. In addition, a novel approach to homology searches and gene determination will be described.
3. A **community access data base** for sequence and associated data for C21.The goal is to provide early, informal access to results as they are generated. The emphasis is on a graphical interface with a look and feel natural for biologists. Included in the database will be a LBL's P1 physical map with linkages to the Genethon YAC map provided by STS's derived at both laboratories.
4. **Laboratory automation:** Together with our Instrumentation colleagues, we are in the process of integrating the flow of laboratory materials and sequence data. Methods for tracking the state of the experiment and the use of robots for automatic operation will be discussed.
5. The development and use of **data management tools and techniques** for more effective design and implementation of database systems. Several examples will be described, including a high level query language and a way to automatically generate user interfaces, based on a metadata description of the underlying database.

This large complement of projects would challenge the resources of even a much larger group than ours. Consequently, where possible, our strategy is to seek leverage by collaborating with other groups who have developed software that we can adapt and to which we can supply additional value in return. These collaborations will be described as potentially useful prototypes for other projects. Finally, we will provide schedules for the completion of these tasks.

---

* This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Genome Program, of the US Department of Energy under Contract No. DE-AC03-76SF00098.

Enhancements to the LBL Colony Picker

D. C. Uber, M. J. Pollard, J. M. Jaklevic, G. L. Granados and V. C. Kirk
Human Genome Center Instrumentation Group and Engineering Division
Lawrence Berkeley Laboratory, University of California, Berkeley CA 94720

Our early experience with the colony picking machine designed and built at the Lawrence Berkeley Laboratory Human Genome Center indicated that, although the system worked very well, there were areas in which further automation would improve throughput and reduce labor. Consequently, we have enhanced the system in two major areas.

First, we have added a robotic arm which loads and unloads petri dishes and microtiter plates on demand from the colony picker. With this side loader, we can pick 30 dishes without attention from an operator.

Second, we have written new image processing software which is also capable of fully automated operation. The software automatically performs background correction and thresholding on each petri dish image, and selects non-overlapping colonies suitable for picking based on area, aspect ratio, circularity, neighbor clearance, edge clearance, and the shape of the density profile along the major axis. The operator may still edit the results simply by clicking the mouse on objects in the image. The program is written in the high-level Optimas (Bioscan, Edmonds, WA) Windows-based language that facilitates rapid development of complex image analysis tasks.

This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Genome Program, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098. Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

Use of a General-Purpose Robot at the LBL Human Genome Center

D. C. Uber, W. L. Searles and J. M. Jaklevic
Human Genome Center Instrumentation Group and Engineering Division
Lawrence Berkeley Laboratory, University of California, Berkeley CA 94720

We describe our progress in customizing the ORCA (Hewlett-Packard, Avondale PA) general-purpose laboratory robot for use in the LBL Human Genome Center for large-sale mapping and sequencing studies. The ORCA robot has the advantage of a large, accessible work area and the capability to act as a side-loader for several stand-alone instruments. The tasks initially automated have been replication of DNA libraries in 96- and 384-well microtiter plate formats, conversion between these two formats, and production of high-density colony blot filters. Adaptation of the commercial robot for these purposes required the design and fabrication of custom tools and workstations in order to accommodate the microtiter plate format. We demonstrate the use of the gripper tool, plate stackers, workstations, plate filling station, replication tools, and sterilizer that have been developed and are currently in use. Future development plans including the use of the ORCA as a platform for system integration will be discussed.

This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Genome Program, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098. Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

Software for Sequence Assembly Based on the Directed Approach*

E. Veklerov, S. Lewis, C. Martin, S. Pitluck, E. Theil

Human Genome Computing Group
Lawrence Berkeley Laboratory, Berkeley, CA 94720

Existent software packages do not fully support the directed DNA sequencing strategy developed at LBL. Specifically, they are inadequate in the following areas:

Algorithms: Their assembly algorithms were originally motivated by the shotgun strategy. They were not designed to take advantage of all the information available to the biologist when a directed strategy is used. Algorithms that properly use this information can overcome performance difficulties when the sequences become very long or when repeated sequences cause ambiguities.

Data Model: The sequencing strategy developed at LBL relies on a hierarchy of maps of increasingly higher resolution. The various pieces of the sequencing software must be able to make use of all these maps by incorporating them into a comprehensive data model.

User Interface: The large volume of data generated by large–scale sequencing requires that all the data be available in a simple graphical form. The most time–consuming operations should be fully automated, while still allowing the biologist to override the automatic procedures.

We have written several programs that alleviate some difficulties of applying the Staden `xdap` package to our strategy. They perform several disjoint functions including:

- graphical display of the output of the `xdap` alignment algorithm;
- assembly of 3-4 kb fragments into a P1 clone;
- locating the positions in the consensus line in which gel readings are inconsistent.

These programs will be incorporated into a new, much more flexible software package meeting the requirements above. Because of its advantages in producing fast prototypes and superior data modeling capabilities, we are implementing the system using Smalltalk. We will discuss developments to date and future work.

---

* This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Research Program, of the U. S. Department of Energy under Contract DE–AC03–76SF00098.

Automatic Generation of User Interfaces for Genomic Databases*

M. Zorn, O. Ben-Shachar, V. Markowitz, F. Olken, and J. McCarthy

Human Genome Computing Group
Lawrence Berkeley Laboratory, Berkeley CA 94720

Databases for genomic data are subject to continuing evolution to cope with scientific advances. Modifications in the database definition invariably trigger changes in the user interface. Thus a significant effort is spent in constantly adapting the database to new requirements and the user interface to the modified database definition. To break this vicious cycle we have developed a user interface that is automatically generated from the database definition, i.e., the metadata.

A generic user interface guides the user through a standard flow of actions, from object selection and query formation to viewing the details of an object and following connections to linked objects. A plain text configuration file read upon program start-up provides information for a specific database, e.g., names of objects, attributes, labels, thus customizing the generic interface for a particular application. This creates an object-oriented view of a relational database implemented using an Extended Entity Relationship model.

The configuration file has been automatically created from the metadata. It defines object appearance in the user interface, defines mappers that translate between the database representation of objects and the interface representation, database specific help, and provides for extensive user customizing. In addition, the configuration file defines stored procedures that retrieve data from the database. These procedures are generated from metadata using a query language, Complex Object Query Language (COQL), based on the same Extended Entity Relationship model that has been used in the database design. COQL queries are subsequently translated into SQL procedures.

Thus a working user interface, i.e., static layout of buttons and fields, data retrieval from the database, data conversion between the database output and the user interface data structures, is automatically generated from the database definitions. Modifications are propagated by simply regenerating the interface. This approach, originally developed for the Chromosome Information System, has been applied to an image database, and several other databases.

---

* This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Genome Program, of the US Department of Energy under Contract No. DE-AC03-76SF00098.

Large Scale Sequence Analysis*

M. Zorn[+], *J. McFarlane*[†], *S. Scherer*[+], *and R. Armstrong*[‡]

[+]Human Genome Center
[†]Information and Computing Sciences Division
Lawrence Berkeley Laboratory, Berkeley CA 94720,
[‡]Sandia National Laboratories, Livermore, CA 94550

The rate at which new sequences are being generated has increased dramatically in the past year. This presents two challenges for sequence analysis.

First, the data flow of sequence data has to be organized in such a way as to allow for automatic sequence analysis using a standard set of programs. The results have to be stored in a database to avoid recomputing them. Results from sequence fragments need to be tied to the assembled sequence.

The second challenge stems from ever larger sequences that will have to be analyzed at one time, i.e., finished sequences are becoming larger than 100kb in size. The size of the databases used for sequence similarity searches doubles almost every year now. The available sophisticated computing technology to tackle these problems, e.g., faster machines, parallel processing, distributed computing, exists already. However, the use of these resources requires detailed knowledge of the particular resources to optimally access them.

The Parallel Object-oriented Environment and Toolkit, POET, is modeled after the X11 toolkit and enables both high and low level control of the computational methods. The object-oriented programming paradigm allows data encapsulation and methods to hide implementation details so as to present a unified object view to the user. Existing software can be adapted to exploit the power of parallel processing. Thus sequence analysis can be performed transparently to the user in reasonable time where POET divides either the query sequence or the database in multiple pieces to run on parallel computers or a number of workstations in a distributed environment.

We will present a prototype system that integrates sequence analysis into the sequencing protocol and performs comparisons of sequences between 50 – 100Kb. A user interface will allow parameter specification for several analysis options and launch the analysis program. A graphical display will present the results to the user.

---

* This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Genome Program, of the US Department of Energy under Contract No. DE-AC03-76SF00098.

LLNL

Lawrence Livermore National Laboratory Human Genome Center

CLOSURE OF CHROMOSOME 19 USING BACTERIAL ARTIFICIAL CHROMOSOME (BAC) CLONES.

Michelle Alegria-Hartman[1], Mark A. Batzer[1], Chris T. Amemiya[1], Jeffrey A. Garnes[1], Chira Chen[1], Benjamin S. Wong[1], Hiroaki Shizuya[2], Bruce Birren[2], Ung-Jin Kim[2], Melvin I. Simon[2], Jennifer S. McNinch[1], Anthony V. Carrano[1] and Pieter J. de Jong[1]. [1]Human Genome Center, L-452, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA 94551; [2]California Institute of Technology, Division of Biology, 147-75, Pasadena, CA 91125.

A number of different cloning systems are currently being used to generate contiguous physical maps of individual chromosomes (*e.g.* Yeast Artificial Chromosomes, cosmids and bacteriophage P1). Recently a bacterial F-factor based system has been developed (Shizuya *et al.*, 1992, *Proc. Natl. Acad. Sci. USA*, **89**: 8794-8797). We are currently constructing a four-fold redundant total human genomic BAC library in collaboration with Dr. Melvin I. Simon at the California Institute of Technology. In an effort to isolate chromosome 19-specific BAC clones for closing the chromosome we have constructed a set of high-density colony filters from a portion of the BAC library using a Biomek workstation. The filters have been screened using a variety of chromosome19-specific probes including inter-*Alu* PCR products, "Degenerate Oligo Primed" PCR (DOP-PCR) products and a 37 bp repetitive element (PE670) [Das *et al.*, 1987, *J. Biological Chem.*, **262**: 4787-4793] . We have isolated a number of chromosome 19-specific clones. A total of fifteen putative PE670 positive BAC clones have been identified. Based on the assumption of a random distribution of PE670 repeats on chromosome 19, screening a 0.5 fold redundant filter set should result in the localization of 27 PE670 positive BACs. The isolation of 15 PE670 positive BACs probably results from the non-random distribution of the PE670 repeat along chromosome 19. The BACs which contained PE670 were subsequently used as probes to screen chromosome 19-specific high-density cosmid filters. Individual BAC clones hybridized to a number of chromosome 19 cosmid contigs. The hybridization of the BACs to chromosome 19-specific cosmids is presently being confirmed. These data demonstrate the utility of BAC clones for the generation of a contiguous chromosome 19 physical map.

Work at Lawrence Livermore National Laboratory was performed under the auspices of the U.S. Department of Energy contract No.W-7405-ENG-48. Work at the California Institute of Technology was supported by DE-FG03-89ER60891 from the Department of Energy to MIS.

DEVELOPMENT OF AN INTEGRATED *IN SITU*, YAC AND COSMID MAP OF THE p13.2 BAND OF CHROMOSOME 19.

Susan Allen, Emilio Garcia, Lorie Devlin, Anne Fertitta, Barbara Trask, Brigitte Brandriff, Laurie Gordon, Anthony Carrano and Anne Olsen. Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore CA.

The p13.2 band of chromosome 19 spans about 6 Mb and includes a number of interesting genes such as the receptors for LDL, insulin and erythropoetin, the oncogenes JUNB and VAV, a tyrosine kinase, and a subset of the olfactory receptor gene family (OLFR). We describe a focused effort to develop an integrated map of this region at multiple levels of resolution.

An initial framework map was developed by fluorescence *in situ* hybridization. Hybridization to metaphase chromosomes was used to identify cosmids located in p13.2. These were then ordered by hybridization to interphase nuclei, resulting in a set of 24 marker cosmids distributed along the band. The set included cosmids for all the genes currently mapped to p13.2. In certain regions, such as the OLFR, *in situ* hybridization to sperm pronuclei is being used to produce a higher resolution *in situ* map.

YACs are being identified by screening a library (CEPH) either by PCR, using STS markers for sequences mapped to p13.2, or by colony hybridization with cDNAs or probes generated from cosmids mapped to this band. Twenty-one YACs, ranging in size from 175 to 550 kb have been isolated for nine p13.2 markers. To integrate these YACs into the cosmid map, Alu-PCR products from the YACs are hybridized to cosmid colony filters. In many cases, positive cosmids are already members of contigs previously established by the fingerprinting procedure which constitutes the foundation of the chromosome 19 cosmid map. The pattern of positive contigs serves to order the YACs for a given marker and also permits some ordering of the set of contigs identified by these YACs.

To achieve closure at the level of cosmids where possible, the contigs identified by a YAC, or otherwise regionally localized, are being extended and merged by a multiplex walking procedure. Pools of probes made from cosmids at the ends of contigs are hybridized to high density cosmid colony filters. The contig assignment of positive clones is then accomplished by fingerprinting. Since the fingerprints analyzed in this case are restricted to the subset of positive clones and starting contigs, contig assembly can be done at a lower stringency, resulting in more sensitive overlap detection. Finally, Eco RI digestion of selected cosmids yields a restriction map which serves to verify the established contigs and to determine the length of DNA spanned by the overlapping clones.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. W-7405-ENG-48.

MAPPING AND CHARACTERIZATION OF A LARGE NUMBER OF PUTATIVE ZINC FINGER-ENCODING GENES ON CHROMOSOME 19.

Chris T. Amemiya[1], Mark A. Gonzalgo[1], Eric Bellefroid[2], Anthony V. Carrano[1] and Pieter J. de Jong[1], [1]Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA 94551, 510-423-3634 (ph), 510-423-3608 (fax); [2]Laboratoire de Biologie Moléculaire et de Génie Génetique, Université de Liège, B-4000 Sart-Tilman, Belgium.

The human genome contains large numbers of zinc finger (ZF) protein-encoding genes (Bellefroid *et al.*, 1989, DNA 8: 377-387; Crossley and Little, 1991, PNAS 88: 7923-7927). The Cys_2-His_2 family of ZF genes similar to the Krüppel gene of *Drosophila* has been studied in the most detail, and there are good indications that there may be more than 300 such genes in the human genome (Bellefroid *et al.*, 1991, PNAS 88: 3608-3612). We have used degenerate oligonucleotide probes to the conserved "H/C-link" region of adjacent Krüppel-type zinc fingers to isolate putative chromosome 19-specific cosmid clones from a library of flow-sorted material. Over 1100 clones were identified as giving moderate to strong hybridization signals from a library of 10-12X coverage of the chromosome. These data suggest that the total number of zinc finger genes on chromosome 19 could be in excess of 100 and support the view that chromosome 19 contains a disproportionate number of zinc finger genes relative to the rest of the genome (Hoovers *et al*, 1992, Genomics 12: 254-263; Lichter *et al.*, 1992, Genomics 13: 999-1007). All clones have been rearrayed into 96-well microtiter dishes and high-density colony filters have been generated of these clones. We are currently correlating all known chromosome 19 ZF genes to the corresponding cosmids using this zinc finger cosmid "sublibrary." We have localized several cDNAs to cosmid contigs that have been established using fluorescent fingerprinting (Carrano *et al.*, 1989, Genomics 4: 129-136). Zinc finger-containing regions appear to be distributed largely in clusters in 19p12, p13.12–p13.2, q13.1 and qter.

Work performed by Lawrence Livermore National Laboratory under the auspices of the U.S. Department of Energy under contract no. W-7405-ENG-48.

AN OVERVIEW OF LLNL'S CHROMOSOME 19 PHYSICAL MAP

L. Ashworth, C. Amemiya, M. Alegria–Hartman, S. Allen, R. Barlett, M. Batzer, A. Bergmann, B. Brandriff, E. Branscomb, M. Burgin, C. Chen, M. Christensen, A. Copeland, P. de Jong, L. Devlin, J. Elliott, R. Esposito, J. Eveleth, A. Fertitta, E. Garcia, J. Garnes, A. Georgescu, D. Georgi, L. Gordon, S. Hoffman, C. Kwan, J. Lamerdin, G. Lennon, K.Lieuallen, A. Martin–Gallardo, J. McNinch, H. Mohrenweiser, M. Montgomery, D. Nelson, A. Olsen, D. Ow, S. Roquier, L. Scheidecker, M. Seagraves, T. Slezak, S. Tsujimoto, M. Wagner, B. Wong, M. Yeh, A. Carrano. Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA.

Our chromosome 19 physical map is comprised of a set of overlapping cosmid and YAC clones which we estimate span about 90% of the chromosome. The foundation cosmid map was assembled using an automated fluorescence–based fingerprinting strategy. 12,828 chromosome 19–specific cosmids have been assembled into 930 contigs. Cosmids on the map have been associated with 140 loci, 109 genes and 31 anonymous markers.

Our current effort is focussed on closing gaps between contigs by a combination of cosmid and YAC walking using both STSs and Alu–PCR probes. We have closed regions ranging from 650 Kb –1 Mbp around the carcinoembryonic antigen gene family in q13.2, the ryanodine receptor in q13.1, the D19S11 locus in p13.2, and the myotonic dystrophy region in q13.3.

Fluorescence *in situ* hybridization has been used to localize contigs on the cytological map and to order cosmids and measure distances between them. Over 550 cosmids representing 219 contigs have been localized. This subset of contigs spans ~20 Mb or ~40% of the non–centromeric regions of chromosome 19. 110 of these contigs have one or more genes or markers assigned to them. In addition, FISH mapping to sperm pronuclear DNA has allowed us to determine distances between cosmids as close as 90 Kb.

We are identifying chromosome 19 cDNAs by screening cDNA libraries using cosmids from our map as probes. Additional cDNAs are being selected, sequenced and mapped after identifying them through random selection protocols.

STSs/ESTs are being generated from cDNAs and from pre–mapped cosmids. In addition, we are sequencing selected cosmids containing genes of biological interest from chromosome 19. These include XRCC1 (and the mouse counterpart), ERCC2, ERCC1, and telomeric regions.

Our physical map information is stored in a relational database and is accessible to end–users via SQL or graphical query. A graphical interface is routinely used for display and analysis of physical map data. Physical map data are also exported via Internet to GDB for public access.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. W–7405–ENG–48.

ALU REPEATS AS MARKERS FOR HUMAN POPULATION GENETICS.

Mark A. Batzer[1], Jennifer Alleman[1], Pieter J. de Jong[1] and Prescott L. Deininger[2]. [1]Human Genome Center, L-452, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA 94551; [2]Department of Biochemistry & Molecular Biology, Louisiana State University Medical Center, 1901 Perdido St., New Orleans, LA 70112.

The Human-Specific (HS) subfamily of Alu sequences is comprised of a group of 500 nearly identical members which are almost exclusively restricted to the human genome. Individual subfamily members share 97.9% nucleotide identity with each other and 98.9% nucleotide identity with the HS subfamily consensus sequence. HS Alu family members are thought to be derived from a single source "master" gene, and have an average age of 2.8 million years old. Using an oligonucleotide probe complementary to the HS subfamily we have identified approximately 15 chromosome 19-specific HS Alu family members. Twelve of the 15 HS Alu family members are located on the q arm of chromosome 19 with a slight bias toward the centromere. We have developed a Polymerase Chain Reaction (PCR) based assay using primers complementary to the 5' and 3' unique flanking DNA sequences from each HS Alu that allows each locus to be assayed for the presence or absence of an Alu repeat. Using this assay individual HS Alu family members were found to be dimorphic or monomorphic for the presence or absence of individual HS Alu family members. The dimorphic HS Alu sequences represent a unique source of information for human population genetics and forensic DNA analyses. HS Alu family member insertion dimorphism differs from other types of polymorphism (*e.g.* Variable Number of Tandem Repeat [VNTR] or Restriction Fragment Length Polymorphism [RFLP]) because individuals share HS Alu family member insertions based upon identity by descent from a common ancestor as a result of a single unique event which occurred one time within the human population. The VNTR and RFLP polymorphisms may arise multiple times within a population and are identical by state only. The distribution of a number of dimorphic HS Alu insertions will be presented.

Work at Lawrence Livermore National Laboratory was performed under the auspices of the U.S. Department of Energy contract No.W-7405-ENG-48.

Long-Range Chromosome Mapping By 3-Color Fluorescence in situ Hybridization to Pronuclear Interphase Chromatin Targets.

B. F. Brandriff, L. A. Gordon, A. Bergman, E. Branscomb, and A. V. Carrano. Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA

The fluorescence in situ hybridization system employing interphase human sperm cell nuclei, "pronuclei", as chromatin targets allows us to order cosmids and to make estimates, in kbp, of unknown genomic distances separating cosmids. Distance estimates of <50 kbp to over 1 Mbp are reliably determined by reference to a calibration curve relating physical and genomic distances. We are developing a scheme exploiting this system for long-range mapping and ordering of contigs on chromosome 19. Our hypothesis is that cosmids separated by 0.5 to 1 or more Mbp can be placed on pronuclear chromatin targets to form a ladder into which unknown cosmids can be placed to obtain, first, their position within the ladder, and second, an estimate of genomic distance separating the unknown cosmids from their neighbors. Cosmids estimated to be separated by 2 to 5 Mbp, identified by their proximity in metaphase, will be used as anchor points in this scheme. These cosmids will be hybridized to pronuclear targets, with additional cosmids placed to make ladders of cosmids approximately 0.5 to 1 Mbp apart. Ladders will be labeled in two alternating colors, with unknowns labeled in a third color. Thus, 6 cosmids separated by 1 Mbp intervals will form a ladder spanning a 5 Mbp chromosome band. The success of this scheme depends on being able to identify, with reasonable certainty, the sequence of cosmids constituting the initial ladder.

To test our hypothesis, we assembled proof-of-principle data in three regions, 19p13.1, 19p13.2 and 19q13.2. A sequence of signals from pools of 6 cosmids estimated to be separated by 0.5 Mbp intervals and spanning about 2.5 Mbp, and pools of 4 cosmids estimated to be separated by 1 Mbp intervals could be distinguished. Preliminary data showed that in about 40 to 60% of pronuclei, the order and separation of cosmids was sufficient to place unknown cosmids with reasonable certainty. Alternatively, a stretch of DNA was "painted" by simultaneous hybridization with 14 cosmids previously determined to be located along a 2.5 Mbp span on 19q. In approximately one-half to two-thirds of pronuclei the FISH signals created traceable "backbone" strands extending to over 100μm. These approaches will allow us to obtain order and accurate genomic distance information for unknown cosmids. Both approaches illustrate the reason why pronuclear interphase targets will be used for this work. Their decondensed chromatin configuration yields a percentage of nuclei in which the DNA is extended in a relatively linear fashion over megabase distances, maintaining the integrity of cosmid order along the linear DNA molecule, whereas somatic interphases lose "linearity" above 1 Mbp.

Our goal is to create a set of ladders consisting of a total of approximately 60 to 120 cosmids spanning the 60 Mbp of chromosome 19 into which any contig, gene or marker can be inserted for order and genomic distance information.

Work performed by Lawrence Livermore National Laboratory under the auspices of the U.S. Department of Energy under Contract W-7405-ENG-48.

High Resolution Fluorescence in situ Hybridization Maps of Three Regions On Chromosome 19. B. F. Brandriff, L. A. Gordon, A. S. Olsen, S. M. G. Hoffman, H. W. Mohrenweiser, and A. V. Carrano. Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA

We have constructed high-resolution FISH maps for three regions of chromosome 19: (a) the D19S11 locus, a 560 kbp region on 19p13.1; (b) one cluster of the olfactory receptor gene family (OLFR) on 19p13.1 spanning 840 kbp and overlapping the D19S11 locus, and another cluster spanning 520 kbp on p13.2; and (c) the carcinoembryonic antigen and pregnancy-specific protein gene families (CEA and PSG), a 1.2 Mbp region on 19q13.2. In each case, we used individual cosmids from contigs as starting points. Order of, and genomic distances between contigs positive for these gene families or loci were initially unknown. Fluorescence in situ hybridization of selected cosmids to metaphase and somatic interphase chromatin targets provided initial localization and preliminary order information of contigs along chromosome 19.

Definitive order and estimates of genomic distances separating contigs were subsequently established by in situ hybridizations to pronuclear chromatin targets. Distance estimates were derived by reference to a calibration curve previously constructed by relating physical distances separating cosmid pairs as measured in pronuclei to known genomic distances obtained by PFGE or restriction enzyme mapping. Pronuclear estimates formed the basis for chromosome walking and restriction mapping strategies for closing gaps between contigs in the three regions.

Pronuclear estimates of genomic distance were compared to restriction map estimates subsequently obtained in the D19S11 and CEA regions for cosmid pairs separated by 20 to 150 kbp. In fifteen out of sixteen instances, the distance estimates derived in pronuclei fell within the restriction map distance between cosmid midpoints and the restriction map distance between cosmid ends. For instance, the pronuclear estimate of 34 kbp separating two D19S11 cosmids was consistent with restriction estimates of 21 kbp from midpoint to midpoint and 63 kbp from end to end. In the only discrepant case, the pronuclear estimate was 75 kbp, slightly longer than the end-to-end restriction estimate of 69 kbp.

Work performed by Lawrence Livermore National Laboratory under the auspices of the U.S. Department of Energy under Contract W-7405-ENG-48.

INFORMATION THEORETIC MEASURES OF VALUE AND PROGRESS IN GENOMIC MAPS

Elbert Branscomb (branscomb1@llnl.gov), Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA 94550.

Is there a simple but fair way of defining numerical measures of "value" and "progress" for genomic maps? I propose that potentially useful measures for this purpose can be based on standard statistical concepts of information content. This approach provides a definition of the "informativeness" of an experimental observation for answering a given specific question (see, e.g. S. Kullback, "Information Theory and Statistics", 1968, Dover Press). On these terms, two maps could be validly compared, for example, as to their cost per unit of "mapping" information provided. One obvious qualification is that physical maps are used for many different purposes so no single type of question captures all varieties of value. However, most uses of such maps rely on their ability to answer questions concerning localization ("in this interval", "on this clone") and on this basis possibly useful approximate formulae measuring value may be derived. For example, we might consider that an "interval map" is used primarily to help answer questions of the form "is a probe within a specified interval in the map or not?". With this question, and taking the map to be composed of n intervals l_i covering a region of length L, the informativeness of the map in Kullback's sense may be approximated by the expression:

$$I = \sum_{i}^{n} \frac{l_i}{L} \log_b \left[\frac{L}{l_i / c_i} \right]$$

where "b" is the base of the logarithms used (which merely sets the units of "information"), and c_i is the (estimated) confidence probability for localization in the *i'th* interval (i.e., the probability an inference based on the map that a locus is in the interval l_i would turn out to be correct). Given a specification of the ideal "complete" map, the level of progress achieved by any real map can then be defined naturally as the ratio of its informativeness to that of the complete map. The mathematical justification for expressions of this type, and examples of their application to real, and realistic sets of map data, will be given.

(This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.)

IDENTIFICATION OF REGION–SPECIFIC COSMID, BAC AND PAC CLONES BY HYBRIDIZATION WITH MICRODISSECTION–DERIVED PROBES.

Chira Chen[1], Peter Kroisel[1,2], Chris Amemiya[1], Panos Ioannou[1], Michelle Alegria–Hartman[1], Jennifer McNinch[1], Fa–Ten Kao[3], Paul Meltzer[4], Mark Batzer[1] and Pieter de Jong[1], [1] Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA, [2] Institute of Medical Biology & Human Genetics, University of Graz, Austria, [3] Eleanor Roosevelt Institute for Cancer Research, Denver, CO and [4] Department of Radiation Oncology, University of Michigan Medical Center, Ann Arbor, MI.

Recently, various approaches for micro–dissection cloning have been employed to prepare complex region–specific PCR product libraries. We are interested in the use of these libraries for the isolation of region–specific large–insert clones from genomic and chromosome-specific libraries. Such region–specific clone collections would contain from 1–10% of the clones of the originating BAC (Bacterial Artificial Chromosome), PAC (P1 Artificial Chromosome) or cosmid libraries. The smaller sub–libraries would facilitate the mapping of defined regions throughout the human genome as well as streamline the dissemination of these valuable resources. In our inititial experiments, we are using microdissection libraries from human chromosomes 19, 2 and 12 to isolate region–specific clones. Three libraries have been produced for specific regions on human chromosome 2 by ligation of PCR–adapters (Kao et al, unpublished results), an additional library was produced for 2q11–q12 by Degenerate Oligonucleotide Primer PCR (DOP-PCR). We are currently using these chromosome 2 libraries and derivative micro-clones for the probing of cosmid, BAC and PAC high–density colony filters. In addition, we are generating similar micro–dissection libraries from chromosome 19 to assist in the isolation of region specific mapping reagents consisting of BAC and PAC clones. In an initial test we used micro–dissection PCR products from a homogeneous staining region (HSR) of chromosome 12 involved in tumor development as probes. These probes hybridized to identical clones in duplicate demonstrating the specificity of this approach.

Work performed by the Lawrence Livermore National Laboratory under the auspices of the U.S. Department of Energy under contract No.W–7405–ENG–48.

TOOLS FOR MORE EFFICIENT LIBRARY REPLICATION AND COLONY FILTER PRODUCTION.

A. Copeland, B. Wong, M. Alegria-Hartman, M. Batzer, C. Kwan, A. Georgescu, P. de Jong, A. Olsen, G. Lennon. Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA.

The effort to produce a physical map of chromosome 19 involves a modest effort to automate and streamline genomic and chromosome specific library manipulation. These efforts have focused to date on the automation of library replication and production of high density filter arrays for colony hybridization.

We have previously reported on the application of the Beckman high density replicating system to plate 1536 to 3456 clones on an 8 X 11 centimeter filter. We have recently made several modifications to this system to increase the throughput and decrease the space required for clone storage. We have designed a new microtiter plate for cosmid storage. The new plate has the same outer dimensions as a standard 96-well mictotiter plate but it has 384 square wells. The wells are 3.5 mm across at the top and 12.0 mm deep and taper slightly which gives them a maximum volume of 130 microliters, and a working volume of 75 microliters. The relatively large cross-sectional area of the square wells makes the new plates easy to dispense into compared to other 384-well plates. We are using a 384-pin replicating tool from Helix Corporation with the Beckman Biomek to produce filters containing 6144 clones from libraries rearrayed into 384-well plates.

We have also developed an inexpensive device to facilitate the transfer of arrayed libraries from 96-well microtiter plates to 384-well plates. A plastic tablet with a rectangular cutout in its center (large enough to contain four microtiter plates) has been fitted with adjustable vertical braces mounted at the corners of the cutout. The braces are adjusted so that at each corner of the cutout, a brace will guide the replicating tool into a different offset position of the 384-well plate when the replicating tool is held flush against the brace. By positioning the target 384-well plate in each corner of the cutout, then inoculating the plate and sterilizing the pin tool, four 96-well plates will be transferred to the correct offsets of a 384-well plate. The device may also be used for manually replicating 384-well plates.

We are also adapting a Hewlett-Packard ORCA robot for high density colony filter production and library replication. There are currently no commercially available tools for making colony filters with the ORCA, so we have modified Beckman Biomek tools. Flexible and extensible programs for controlling the robot are easily written using the Methods Development Software provided with the ORCA. On the basis of preliminary results generating cosmid colony filters, we anticipate achieving inoculation rates of over 100,000 positions per hour.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. W-7405-ENG-48.

Chromosome 19 Closure: Long range YAC and cosmid mapping of regions q12 and p13.2.

Jeffrey M. Elliott, Lorie G. Devlin, Jane Lamerdin, Anthony Carrano, Briggitte Brandriff, Laurie Gordon and Emilio Garcia. Human Genome Center. Biology and Biotechnology Research Program. Lawrence Livermore National Laboratory, Livermore, CA 94550.

The present chromosome 19 physical map consists of approximately 900 cosmid contigs with spanning paths that range between 2 to 30 cosmids, and with and estimated coverage of greater than 80% of the total chromosome. A systematic effort at closing the gaps in the cosmid contig map has been started by focusing primarily on the q12 and p13.2 regions of the chromosome. A two-tiered approach for the identification of YACs spanning these gaps has been used. For the q12 region, a series of cosmids previously assigned to this band by fluorescent in situ hybridization (FISH) were used in Alu PCR reactions to generate probes that were directly hybridized to a high density YAC array of the CEPH YAC library (Albertson, et al. PNAS 87: 4526, 1990). For the p13.2 region, a region of chromosome 19 that contains a diverse number of important genes, such as; low density lipoprotein, erythropoietin and insulin receptor; oncogenes, such as VAV, JUNB, MEL and LYL; complement cascade system members, such as C3 etc., the approach has been to use preexisting or specifically designed sequenced tagged sites (STS) for identification of corresponding YACs from DNA pools of the same library. Twenty one YACs mapping to this region and comprising nine distinct loci have been isolated by the STS approach and 32 YACs corresponding to the 12q region have been isolated by the direct Alu-PCR probe hybridization. The YACs obtained by this combined approach have been used as template to generate Alu-PCR products, which in turn, have been used as probes against chromosome 19- specific cosmid arrays. Examination of the pattern of shared contigs present in the YACs allowed preliminary ordering of the previously defined cosmid contig into metacontigs (contigs defined by a combination of YAC-cosmid overlaps). For a selected area on the q12 region, we have confirmed the order, distance and orientation of the contigs by using a combination of two color interphase and pronuclei FISH analysis on key clones of the neighboring contigs. Contig assembly by this method consolidates an average of 4 cosmid contigs per YAC isolated, and has allowed a detailed continuous megabase mapping of this area of chromosome 19. Further physical map assembly using this approach is being carried out by "directing" the closure effort using probes derived from end clones of the expanding contig, and/or using probes derived from clones pre-mapped and ordered to a given band by FISH analysis.

This work was performed under the auspices of the U.S. DOE by LLNL under contract no. W-7405-Eng-48.

RAPID ACQUISITION AND ANALYSIS OF DISTANCE MEASUREMENTS FROM PHOTOGRAPHIC IMAGES USING A DIGITIZING BOARD.

R.J. Esposito, G. van den Engh, B. J. Trask, B. Brandriff, R.G. Langlois. Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, Ca.

The large-scale application of fluorescence *in situ* hybridization techniques (FISH) for physical mapping has created a need for a simple and rapid method of extracting distance information from probe hybridization spots. An analysis system that utilizes a 35mm slide projector to project images on a large (1.2 m X 0.9 m) digitizing board has been successfully developed and applied to FISH images. The digitizer board and associated computer allow one to perform point-to-point and contour line distance measurements. All functions of the measurement process, including advancement of the slide projector, are implemented using a multi-button mouse cursor to select items on the board's surface. This allows extremely rapid data input of up to 1000 point-to-point measurements per hour. The spatial resolution of the cursor on the board's surface is 0.2 mm. For typical photomicrographs, a distance of 50 μm on a microscopic slide corresponds to a projected distance of approximately 0.6 m on the board, so one can easily measure object distances down to less than 0.1 μm. Measurement results for each experiment are automatically stored on a computer as a text file for easy editing and statistical analysis. To date the system has been applied to FISH measurements on metaphase chromosomes, interphase nuclei, and hamster-human pronuclei, and to analysis of image bands from gel electrophoresis. In summary, this system provides a rapid tool for the analysis of FISH images that can be easily adapted to other simple image analysis tasks. (Work performed under the auspices of the U.S. Deptartment of Energy by LLNL under contract W-7405-ENG-48)

CONSTRUCTION OF CHROMOSOME SPECIFIC COSMID AND LAMBDA LIBRARIES FROM FLOW SORTED CHROMOSOMES USING NANOGRAMS OF DNA.

Jeffrey A. Garnes[1], Jennifer S. McNinch[1], Jennifer Alleman[2], Benjamin Wong[1], Hillary Massa[3], Jerry Eveleth[1], Barbara J. Trask[3], Ger van den Engh[3], Richard Langlois[1], and Pieter J. de Jong[1]. [1]Human Genome Center, L-452, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA 94550; [2]Genetech, Inc. 460 Point San Bruno Blvd., South San Francisco, CA 94080; [3]University of Washington School of Medicine, Department of Molecular Biotechnology, GJ-10, Seattle, WA 98195.

As part of the National Gene Library Project and to assist in human genome mapping and gene analysis, we have constructed cosmid and lambda libraries from flow-sorted chromosomes. Human chromosomes have been isolated from human/rodent hybrid cell lines using the Livermore Modular Instrument for High Resolution Flow Karyotype Analysis and High Speed Chromosome Sorting. For cosmid cloning, a lambda-replicon vector (Lawrist 5 or Lawrist 16) with double cos sites has been used to clone 100-200 nanograms of partially-digested chromosomal DNA. Large cosmid libraries (20 - 100 fold redundant) have been prepared for chromosomes 2, 7, 9, 12, 19, 21, 22, Y and X. Similar quantities of partially-digested chromosomal DNA from chromosomes 9, 22 and X have been cloned into an *E.coli* F factor replicon vector (pFOS1) containing double cos sites (Kim et al., 1991, Nucleic Acids Research, Vol. 20, No. 5: 1083-1085). For most of the chromosomes, five to tenfold redundant sets of cosmids from each of the libraries have been arrayed into microtiter dishes for subsequent characterization and distribution throughout the scientific community. The purity of these libraries has been assessed by colony hybridization with human or rodent DNA probes and is in the range of 70 - 95% for all of the libraries. In addition to the cosmids, we have prepared lambda libraries in the Charon40 vector following partial digestion with MboI. Currently, we have prepared high purity lambda libraries for chromosomes 2, 7, 9, 12, 18, 19, 21, 22, Y and X. Most of the lambda libraries have been deposited in the ATCC repository for distribution to the scientific community.

Work at Lawrence Livermore National Laboratory was preformed under the auspices of the U.S. Department of Energy contract No.W-7405-ENG-48.

Organization of the three cytochrome P450 subfamilies on 19q13.2

Hoffman, S. M. G., and Mohrenweiser, H. W. Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore CA 94550

The cytochrome P450 superfamily of mono-oxygenases is of central importance in the metabolism of xenobiotic compounds, including drugs, pollutants and plant metabolites. More than 110 mammalian P450 proteins are grouped into subfamilies by sequence identity of >60%. In all known cases, the genes within each subfamily are clustered, but the clusters are scattered throughout the human genome. Three of these subfamilies, CYP2A, CYP2B and CYP2F, were previously localized to chromosome 19q. We have developed a cosmid contig map of the region as part of the effort to characterize the number and arrangement of the CYP genes in each of these subfamilies.

One cDNA from each of the A, B and F subfamilies was used to probe chromosome 19-enriched cosmid libraries (~10X). Cosmids were analyzed using a fingerprinting strategy and assembled into contigs. At present, all of the CYP genes on chromosome 19 are in two contigs. These contigs span 137 and 210kb, separated by a gap; previous PFGE analysis suggests that the gap should be less than 10 kb.

Cosmids positive for any of the CYP probes were further analyzed by restriction enzyme digestion and Southern blotting. The region contained nine full-sized genes, including five CYP2A, two CYP2B and two CYP2F loci; at least one additional B locus and one additional F locus were incomplete pseudogenes. The genes are arranged in a complex fashion, with the A and F genes intermingled. The orientations of most of the full-size genes have been established. Two of the A genes appear to be the product of a recent tandem duplication, since they are within contiguous blocks of ~20kb that have nearly identical distributions of sites for several restriction enzymes.

We are currently screening additional libraries to find larger clones that can span the gap between the contigs. The sequence of the CYP gene region on 19q is being determined in a collaborative effort to identify regualtory elements of the extensive CYP2 gene family.

Work performed under auspices of the US DOE by the Lawrence Livermore National Laboratory, contract number W-7405-ENG-48.

BACTERIAL ARTIFICIAL CHROMOSOMES (BACs) and P1 ARTIFICIAL CHROMOSOMES (PACs): BAC-PAC ING THROUGH THE HUMAN GENOME.

Panayiotis A. Ioannou [1,2], Mark A. Batzer[1], Jeffrey A. Garnes[1], Chris T. Amemiya[1], Peter Kroisel[1] and Pieter J. de Jong[1]. [1]Human Genome Center, L-452, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA 94551; [2]The Cyprus Institute of Neurology & Genetics, P. O. Box 3462, Nicosia, Cyprus.

Recently, new procedures have been developed for the efficient cloning of large DNA fragments (100-300kb) as P1 Artificial Chromosomes [PACs] (Ioannou *et al.*, submitted) and Bacterial Artificial Chromosomes [BACs] (Shizuya *et al.*, 1992, *Proc. Natl. Acad. Sci. USA*, **89:** 8794-8797). Total human genomic DNA libraries are currently being constructed using BACs (in collaboration with Dr. Melvin I. Simon at the California Institute of Technology) and PACs. Both of these vectors allow for the efficient cloning of large DNA fragments (100-300kb) for intermediate physical mapping between cosmids (40kb in size) and Yeast Artificial Chromosomes [YACs] (200kb-1Mb in size). Each of these cloning systems allows the propagation of individual clones within bacterial hosts facilitating their analysis and distribution throughout the scientific community. The PAC vector offers the advantages of a positive selection system for recombinant clones, as well as an inducible high copy number origin of replication for the isolation of individual clone DNA as compared to the BAC system. The construction of the PAC vector (pCYPAC1) as well as a comparison to the BAC system will be presented. In addition, an update on the construction and analysis of total human genomic libraries generated using BAC and PAC vectors will be presented.

Work at Lawrence Livermore National Laboratory was performed under the auspices of the U.S. Department of Energy contract No.W-7405-ENG-48.

SEQUENCE DETERMINATION AND ANALYSIS OF THE HUMAN AND MOUSE XRCCI DNA REPAIR GENE REGIONS

J.E. Lamerdin, M. Montgomery, S. Stilwagen, L. Scheidecker, R. Tebbs, L.H. Thompson, and A.V. Carrano. Human Genome Center, Biology and Biotechnology Research Program , Lawrence Livermore National Laboratory, Livermore, CA 94550.

The human XRCC1 (X-ray repair cross complementing) gene is involved in the efficient repair of DNA single strand breaks formed by exposure to ionizing radiation and alkylating agents. Both the human and complementary mouse cDNAs have been isolated and the products are being studied. The genomic regions are not well characterized and such information is important for developing targeting vectors for gene knockout experiments. The human gene maps to 19q13.2 and is a member of a contig spanning approximately 150 kbp in our foundation cosmid map for human chromosome 19. We selected one of the cosmids (F5050) which was positive by hybridization for the human XRCC1 cDNA probe (pXR1-30) to sequence. We selected a mouse cosmid containing a 38 kbp XRCC1-positive genomic fragment cloned into the sCos-1 vector. The two cosmids were sonicated, end-repaired, size-selected (1-4 kbp), and cloned into the SmaI site of pBluescript KS+. Double-stranded templates were prepared by the Qiagen alkaline lysis procedure and sequenced using the Taq cycle sequencing protocol on a Catalyst 800 workstation (Applied Biosystems Inc.). Sequence data were collected on an ABI 373A DNA Sequencer. Editing, analysis, and sequence assembly were performed using the ABI SeqEd program and Intelligenetics suite. Approximately 75% of each cosmid has been assembled by random selection of clones sequenced using fluorescent primers. Some small gaps have been closed by primer walking using fluorescent DyeDeoxy^{TM} Terminators. Preliminary data analysis of 28.7 kbp of the mouse sequence and 18.4 kbp of the human sequences was performed. For the human sequence 13 Alu sequences were found for a density of 1 Alu per 1.4 kbp. For the mouse repetitive B1 element , 8 copies were found for a density of 1 per 3.6 kbp. No L1 sequences were found in either cosmid. The chromosome 19-specific 37 bp repetitive element, pe670, was identified in multiple copies in two different sites in the human cosmid but no sites were identified in the mouse. The dinucleotide CA repeat was found in both mouse and human. In general, the repeat length was longer in the mouse (up to 16 repeats) than in the human (up to 7 repeats). Further analysis for coding regions and exon structure will be performed using GRAIL as soon as sufficient closure is obtained. Comparisons to the cDNAs are also being done to identify splice-junctions. (This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. W-7405-ENG-48.)

ISOLATION, SEQUENCING, AND MAPPING OF HUMAN CHROMOSOME 19 CODING REGIONS

Gregory G. Lennon, Neil S. Ghiso, and Kimberly Lieuallen
Human Genome Center, Biology and Biotechnology Research Program,
Lawrence Livermore National Laboratory, Livermore, CA

The goal of this effort is to isolate, sequence, and map coding regions in the form of cDNAs or exons, and focuses on human chromosome 19. Four aspects of this work will be presented. First, ESTs or other cDNAs appearing to map to chromosome 19 are verified with respect to their true localization, and if on 19, are fine-mapped using the cosmid spanning set filters generated here. Both 5' and 3' ends of cDNAs are used to help verify contig assembly and determine transcriptional orientation. Second, methods for the identification and generation of full-length cDNAs are being developed. These include a method for isotopically labelling mRNA caps, and for enriching for full-length first-strand cDNA. Third, as high-throughput mapping to chromosomes and sub-chromosomal regions is increasingly needed, we are producing stamp-sized filters containing flow-sorted chromosome dots, and are working on optimizing hybridization conditions for both cDNA and genomic (YAC, BAC, etc.) clones. The distribution of these filters should allow a rapid means of localizing probes to chromosomes. Fourth, a relational database for the catalog of chromosome 19 genes is in place to keep track of information related to the over 150 genes placed to date on the chromosome 19 map.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. W-7405-ENG-48.

IDENTIFICATION OF CHROMOSOME 19 cDNAs BY ARRAY HYBRIDIZATION

K. Lieuallen, J. Lamerdin, A. Carrano, and G. Lennon
Human Genome Center, Biology and Biotechnology Research Program,
Lawrence Livermore National Laboratory, Livermore, CA

As part of the Human Genome Project, an effort has been launched to locate and identify novel cDNAs on chromosome 19. One method used has consisted of hybridization to high density colony filters with genomic probes known to be from chromosome 19. The 20x20cm filters consist of 9,216 cDNA clones arranged in a 96X96 matrix (Trends In Genetics 7, 314 (1991)). The library used is a human fetal brain cDNA library, constructed at the Imperial Cancer Research Fund. Probes used on the fetal brain cDNA library were selected from a chromosome 19 cosmid library, constructed at Lawrence Livermore National Laboratory. Close to 1,000 randomly selected cosmid library clones were screened for GC rich sequences by performing Sac II restriction digests. Those cosmids with 3 or more Sac II sites were used as probes on the human fetal brain cDNA library. Over 3,000 cosmids from the spanning path subset of chromosome 19 have been similarly analyzed for other GC-rich sites based on both restriction digestion (SacII, ApaI) and hybridization to probes composed of GC-rich sequences isolated from flow-sorted chromosome 19 chromosomes. To date, forty hybridizations have been performed. Hybridizations were analyzed on a PhosphorImager and the positives identified and sequenced. Additionally, the assignment of GC-rich cosmids to cytogenetic bands has allowed us to generate a preliminary map of the distribution of the most gene-rich areas of this chromosome.

Over four novel cDNAs from chromosome 19 have been identified, in addition to cDNAs corresponding to previously identified genes. New cDNAs are being analyzed for tissue expression via Northern blot hybridization. Full-length clones of these cDNAs are also being isolated employing traditional library screening and plaque purification as well as 5' RACE and solution - phase library screening. The effort involved underscores the need for high-quality cDNA libraries, and for more efficient techniques for full-length cDNA isolation. We are also arraying more libraries in order to increase the percentage of cDNA-positive hybridizations.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. W-7405-ENG-48.

Physical mapping and sequencing of telomeric regions from human chromosome 19.

Antonia Martin-Gallardo, Jeffrey M. Elliott, Jane Lamerdin, Anthony Carrano, and Emilio Garcia. Human Genome Center. Biology and Biotechnology Research Program. Lawrence Livermore National Laboratory, Livermore, CA 94550.

The telomeres of mammalian chromosomes represent interesting and challenging regions of the genome to characterize. In addition to their role in chromosomal replication, the telomeric regions contain polymorphic sequences, the characterization of which may contribute to understand genetic diversity and evolution. In order to study the structural organization of these regions, we are undertaking both a mapping and sequencing effort of the telomeric regions of human chromosome 19. By using first a telomere-associated probe to identify chromosome 19 cosmids, and then inter Alu-PCR products derived from these cosmids a number of YACs were isolated. These YACs map to the telomeric regions on the p and q arm of chromosome 19, as determined by fluorescence in situ hybridization (FISH). End-probes from these YACs are being generated by inverse PCR and used to identify and order further clones in these areas. In addition to the mapping effort, we are developing a sequencing project aimed at determining the DNA sequence of specific subtelomeric regions. In this context, we have determined the DNA sequence of a 2 kb telomere-associated fragment that was derived from a half YAC clone and shown by FISH to hybridize at the telomere of 19p among other chromosomes. Sequence data from a subtelomeric, chromosome 19-specific cosmid have also been obtained. FISH analysis of this cosmid revealed hybridization sites at the telomeres of chromosomes 19p, 3q, 11p and 15q in all individuals analyzed. Furthermore, this analysis detected polymorphisms at the telomeres and interstitial sites of other human chromosomes. Analysis of the DNA sequences should provide information on these polymorphisms and also on subtelomeric structural features which are either chromosome 19-specific or common to other chromosomes. This work was performed under the auspices of the U.S. DOE by LLNL under contract no. W-7405-Eng-48.

COMPARING THEORY AND PRACTICE OF PROBABILISTIC FINGERPRINTING --- THE LLNL EXPERIENCE

David O. Nelson (daven@stille.llnl.gov) and Elbert W. Branscomb.
Human Genome Center, Biology and Biotechnology Research Program,
Lawrence Livermore National Laboratory, Livermore, CA 94550

Our approach to constructing a physical map of Chromosome 19 from a cosmid clone library begins by producing a restriction fingerprint for each clone and then comparing the fingerprints of each pair of clones to estimate the probability of pairwise overlap. The effectiveness of this approach rides on a number of unknown parameters and simplifying assumptions about our data and the fingerprinting process.

In the past, we have examined several issues pertaining to the data generation process, including methods for detecting signal peaks corresponding to restriction fragments, and potential sources of error in data gathering and data reduction. We now have enough experience to begin to examine critically the behavior of our approach in practice.

As one example we now have available a complete restriction map of 111 clones covering approximately 600 kilobases of the P arm of Chromosome 19. This region seems to contain a representative sample of the problems and idiosyncracies which have arisen while analyzing Chromosome 19 DNA, including the existence of repeat sequences and gene families.

Using this large region where the overlap information is known in advance, we address such questions as:

* What is the actual distribution of the likelihood ratio statistic, as a function of overlap percentage?

* How much does the assumption of equal length clones affect the results?

* How much do errors in fingerprint data reduction degrade the sensitivity of the likelihood ratio as an overlap detector?

(This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory Contract No. W-7405-Eng-48.)

STRATEGIES FOR MAPPING GENE FAMILIES: THE PREGNANCY-SPECIFIC GLYCOPROTEINS ON CHROMOSOME 19.

Anne Olsen, Alex Copeland, Elbert Branscomb, David Nelson, Brigitte Brandriff, Laurie Gordon and Anthony Carrano. Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA

Detailed physical maps of gene families are an essential resource for investigations into the regulation of expression of these genes and their evolution. However, the mapping of such families often presents a challenge to many standard mapping techniques, because the homology among various members of a family can produce misleading indications of clone overlap. One large gene family encountered in the chromosome 19 mapping project is the pregnancy-specific glycoproteins (PSGs). The PSG family consists of 12 or more highly homologous genes located in a 500 kb region of 19p13.2.

To establish cosmid contigs in this region, it has been necessary to modify the standard procedure used to assemble contigs from fingerprinting data. This involved analyzing all PSG cosmid fingerprints as a separate subset, and then modifying the algorithm for overlap determination so that the contribution of each fragment in the fingerprint was weighted inversely to the prevalence of that size fragment in the fingerprints of all family members in the subset. This has enabled us to assemble all the PSG cosmids into two large contigs.

Attempts to verify contig assembly and determine contig length by Eco RI restriction digests were also complicated by the presence of homologous regions. In order to establish an accurate restriction map of the entire PSG region, we have used a novel partial digest mapping strategy in which specific linkers are ligated to the Sfi I sites on either side of the insert site in the Lawrist vector. Using this technique, in combination with complete digests, we have established a complete Eco RI map for most of the region. Individual PSG gene coding regions were identified by hybridization of Southern blots with a probe for the constant domain of these genes. Hybridization with gene-specific oligos (in collaboration with S. Hammarstrom, University of Umea) led to localization of all previously identified PSG genes. Our results suggest the existence of at least three additional members of the family.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. W-7405-ENG-48.

Displaying Restriction Maps in Postscript

David J. Ow (ow@tornak.llnl.gov), T. Mimi Yeh, Thomas R. Slezak. Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA 94550

We have developed a program that generates restriction map displays in postscript. These restriction maps may be displayed on any workstation by using a postscript displayer or printed on any postscript printer. The restriction maps are formatted in Encapsulated Postscript (EPS) and follow Adobe's Document Structuring Conventions (DSC). They may be imported by any applications which accept EPS files. This tool enables the biologist to see the restriction maps stored in our database as well as maps that have been created in spreadsheets (i.e., direct database or ASCII input).

We will demonstrate how this code is used via live Internet access to our database and via local files. We plan to incorporate this tool as a display option within our existing database browser. Future extensions may include generating postscript output for portions of integrated maps which include a user-specified collection of objects in an area of interest. The current restriction map display code is available from the authors.

(This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory Contract No. W-7405-Eng-48.)

Closure of the chromosome 19 physical map: A Cosmid and Yeast Artificial Chromosome Contig containing the complete Ryanodine Receptor Gene.

S. P. Rouquier[1], D. G. Giorgi[1], D. H. MacLennan[2], M. S. Phillips[2] and P. J. DeJong[1].
[1]Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA; [2]Banting and Best Department of Medical research, Charles H. Best Institute, University of Toronto, Toronto, Canada.

Presently the chromosome 19 physical map established at the LLNL is comprised of a set of overlapping cosmid and YAC clones (organised in contigs) which is estimated to represent about 85% of the chromosome. Our effort focuses on closing the gaps between all these contigs to get a complete map of the chromosome. Malignant Hyperthermia (MH) is an inherited disease of man and various animals. In patients, anesthesia can induce skeletal muscle rigidity, hypermetabolism and high fever and can lead to death. Biochemical and genetic studies have shown that the disease is related to abnormalities in the Ca^{2+} release channel of skeletal muscle sarcoplasmic reticulum (the ryanodine receptor (RYR1) gene containing approximately 100 exons and located on human chromosome 19q13.1). The full length RYR1 cDNA is composed of 15.4 kilobasepairs (kbp) corresponding to a mRNA encoding a protein of 5032 amino acids. We have chosen RYR1 gene as a pilot region in order to: a) establish a large physical map of the RYR1 gene and surrounding regions, b) ascertain the validity of already established contigs, and c) link these different contigs by filling the gaps between them. We screened individual libraries (cosmids, Bacterial and Yeast Artificial Chromosomes (BACs and YACs)) spotted on high-density filters. We have isolated genomic DNA from the RYR1 region as a set of overlapping cosmid and yeast artificial chromosome (YAC) clones. Ninety-eight cosmid clones were identified by screening three chromosome 19-specific cosmid libraries using RYR1 cDNA probes and subsequently cosmid inserts for chromosome walking. Among these clones, 23 were assembled in a minimal overlap overlapping set according to restriction analysis and hybridization data. We also isolated three YAC clones by screening a human YAC library with selected cosmid inserts. Two of the three YACs contained chimeric inserts but nevertheless filled a remaining gap in the cosmid contig, and further enlarged the contig. The overlaps between the YACs and the cosmid contig were determined by hybridizing YAC Alu-PCR products to cosmid DNAs. The RYR1 gene spans about 205 kb of the more than 800 kb of the RYR1 region cloned in cosmids. We have also isolated a 150 kbp BAC which overlaps with the 3' end of the RYR1 gene. These clones provide reagents for the complete characterization of the RYR1 gene fine structure and other markers possibly related to Central Core Disease (CCD), a genetic myopathy whose candidate gene was described in the RYR1 region by linkage analysis.

This work was performed under the auspices of the U.S.Department of Energy by Lawrence Livermore National Laboratory under Contract No. W-7405-ENG-48 and the MRC (Canada) and the MDAC.

Genomic Cartography on a Flat-Earth Budget

Tom Slezak (slezak@llnl.gov), Mark C. Wagner, Elbert Branscomb.
Human Genome Center, Biology and Biotechnology Research Program,
Lawrence Livermore National Laboratory, Livermore, CA 94550

This seems like a good time for those of us involved in the informatics aspects of "mapping DNA" to take stock of the current situation:

- There is no generally accepted non-political definition of what a *useful* genome map would be. There is nearly universal demand for better access to map information and no general agreement on how to represent maps (in a computer database or visually).

- Every biologist seems to have their own opinion of exactly how they want to see their data on the "final map". All of these viewpoints are valid and nobody wants to wait long for results to appear on their workstation.

- Every major mapping project seems to be in the "90+% completed" phase. Some recently even advertise completion. There are almost as many different approaches to mapping as there are labs doing it. Some involve bits of paper stuck on large spans of walls. The target resolution for different mapping efforts ranges over nearly two logs, from less than 10Kb to hundreds of kilobases.

Problems due to size, complexity, and fluidity exist in almost every aspect of the DNA mapping problem. Easy examples include integrating data of varying confidence, tracking how map placements were decided, detecting and dealing with conflicts, and physically drawing and navigating the final results. Major choices have to be made at each step of the way, dictated by local data, local needs, and available resources.

We will discuss and illustrate some of the choices that we have made in order to explore some of the more important tradeoffs involved in highly-automated approaches to genome mapping.

(This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory Contract No. W-7405-Eng-48.)

Organization of the repeat elements and polymorphic sites at the D19S11 locus

S. Tsujimoto, S. M. G. Hoffman, B. Brandriff, L. Gordon, A. V. Carrano and H. W. Mohrenweiser. Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, California 94550.

The D19S11 locus, an anonymous marker on chromosome 19, was originally shown to be a complex set of six nonallelic polymorphic sites located at 19p13.1. Three probes, 13-1-82, 13-2-21, and 13-1-25, derived from a single cosmid, 1-13, were used to screen a human chromosome 19-enriched cosmid library. The resulting probe-positive cosmids, which included several cosmids simultaneously positive for the MEL gene, along with additional non-probe positive cosmids from the library, were initially assembled into six contigs. The order and orientation of these contigs were determined by fluorescence *insitu* hybridization analysis, using G1 interphase and sperm pronuclei as targets. Adjacent contigs joined by end probe walking resulted in a single contig of 650 kb.

In order to obtain finer resolution and greater detail in this area, an EcoRI restriction map of cosmids spanning the contig was created. The EcoRI fragments with homology to the D19S11 probes were then identified, and it was found that some cosmids have homology to all three probes. Cosmids positive for the MEL gene were found at the centromeric end of the contig. At least four OLFR genes map to this contig, but in no case are the OLFR genes and D19S11-positive EcoRI fragments coincident.

Analysis of the nonallelic variant sites revealed that several different types of polymorphisms including RFLPs, insertions/deletions, and a VNTR are tightly clustered into a 40 kb region and are interspersed among the repetitive elements which are detected with the three probes. Within this region, three RFLPS and an insertion/deletion polymorphism are found within 15 kb of each other.

Preliminary data suggest that the genetic diversity seen at the D19S11 locus may be evolutionarily significant, since the probes also hybridize to Southern blots of chimpanzee and baboon DNA.

Work preformed under auspices of the US DOE by the Lawrence Livermore National Laboratory under contract No. W-7405-ENG-48

Recent Enhancements to the LLNL Genome Browser

Mark C. Wagner (mwagner@kooler.llnl.gov), Tom Slezak, Roy Cantu III, Elbert Branscomb. Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA 94550

We have developed a highly-automated system for creating physical maps of DNA from human chromosome 19 and a graphical contig browser to assist in comprehending and manipulating this data. The LLNL Genome Browser has undergone extensive modification since the last DOE Workshop in February 1991. Our current version is in daily use helping us to analyze the over 13,000 human chromosome cosmid clones fingerprinted to date. The integrated mapper portion of the browser has enabled us to relate our cosmid clones to a total of over 100,000 other biologically significant entities (YACs, Loci, Restriction Fragments, Sequences, cDNAs, etc.). The system currently handles 14 entity classes and the user has complete control over which classes and which interconnections are to be viewed.

Biologists working on the project can also display global ordering information obtained by Fluorescence Insitu Hybridization (FISH) mapping and other methods. We have developed software to take all order and orientation data and generate a global partial ordering. Ordering information between entities of all types is readily displayed by the mapper.

Our design seeks to provide access and visual display of all data generated by our Human Genome mapping project free of any prejudgment as to what questions are of interest or what objects, properties, and relationships should be displayed at any time. The integrated mapper portion of the browser now enables us to display all data objects and relationships stored in our database (currently over 150,000 properties and relationships are known) using a flexible object-based mechanism. We are currently in the process of integrating all the mapping data into single-window views that can be readily customized to show only the desired types of objects and levels of detail.

(This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory Contract No. W-7405-Eng-48.)

Second Generation Data Entry Scripts for a Genome Database.

Mimi Yeh (mimi@krab.llnl.gov), Linda K. Ashworth, Tom Slezak, Elbert Branscomb.
Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA.

Large amounts of Chromosome 19 mapping data generated at the LLNL Human Genome Center are stored on a Sybase relational database. To date, 140 tables contain well over 100 Mb of data and indicies. These tables include descriptive information about libraries, clones, genetic loci, and experimental design which are linked to tables containing results from fingerprinting, hybridizations, PCR, restriction mapping, sequencing, and various re-assemblies of this data. Timely, correct input of this data is a major concern.

Input of biological data is the responsibility of individual generators of that data. To make this process as painless as possible, several scripts have been written which prompt users for data. In each script program, underlying keys, linkages, and triggers are built and validated automatically without the need for the biologist to comprehend these concepts. The scripts are designed to require the minimum amount of human input and to provide the earliest possible detection of entry errors, as well as to allow easy maintenance.

More recently we have begun development of second generation data entry programs. The newer design allows one 'generic' script for each experiment type (such as hybridization, FISH, sequencing). Our first such program was written to accommodate all current 'varieties' of hybridization experiments. It branches to different modules based on data input regarding which variety of probe and target types were used in the experiment. Depending on which modules are used, the program may write data in as many as a dozen different tables. This new program replaces eight previous scripts that accreted over time as experimental approaches evolved for different users.

Our new design has required modularization of sections of old scripts and has eliminated duplicating efforts when additions are made. It also offers the biologist the advantages that users do not need a new program written each time a slightly different experiment is designed and fewer programs must be learned and used.

This approach to data entry has been used successfully for hybridization data and is being expanded to other type of experiments within the LLNL Genome Center. We note that our data input approach is decidedly and deliberately "low-tech" and that we have tried and rejected fancy graphics-based data input methods for a variety of economic and human-factor reasons.

(This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory Contract No. W-7405-Eng-48.)

Human Genome Research at Other U.S. Department of Energy Laboratories

Ames Research Center
ANL—Argonne National Laboratory
BNL—Brookhaven National Laboratory
LANL—Los Alamos National Laboratory
LBL—Lawrence Berkeley Laboratory
ORNL—Oak Ridge National Laboratory
PNL—Pacific Northwest Laboratory

Ames Research Center

MULTIPLEXED FLUORESCENCE DETECTOR FOR CAPILLARY ELECTROPHORESIS USING AXIAL OPTICAL FIBER ILLUMINATION

John A. Taylor and Edward S. Yeung, Ames Laboratory-USDOE and Department of Chemistry, Iowa State University, Ames, IA 50011

It is obvious that irrespective of whichever basic technology is eventually selected to sequence the entire human genome, there are substantial gains to be made if a high degree of multiplexing of parallel runs can be implemented. Such multiplexing should not involve expensive instrumentation and should not require additional personnel, or else the main objective of cost reduction will not be satisfied even though the total time for sequencing is reduced. A corollary is that if a certain basic technology can be readily and efficiently highly multiplexed, then it may ultimately become the method of choice to sequence the entire human genome.

In the last two years, several research groups have shown that capillary gel electrophoresis (CGE) is an attractive alternative to slab gel electrophoresis (SGE) for DNA sequencing. A 25-fold increase in the sequencing rate per capillary (per lane) has already been demonstrated. Part of the improvement in sequencing speed in CGE is counteracted by the inherent ability of slab gels for accommodating multiple lanes in a single run. So, unless capillary gels can be highly multiplexed and run in parallel, the above-mentioned advantages cannot lead to real improvements in sequencing the human genome. Parallel sequencing runs in a set of up to 24 capillaries have been demonstrated recently by others. To provide sensitive laser-excited fluorometric detection, a confocal illumination geometry couples a single laser beam to a single photomultiplier tube. Observation is one capillary at a time, and the capillary bundle is translated across the excitation/detection region at 20 mm/s by a mechanical stage. Since data acquisition is sequential and not truly parallel, the ultimate sequencing speed will be determined by the observation time needed per DNA band for an adequate signal-to-noise ratio. Having more capillaries in the array or being able to translate the array across the detection region faster will not increase the overall sequencing speed.

Recently, we have developed an axial beam excitation scheme for capillary electrophoresis. The excitation laser is coupled via an optical fiber which in turn is inserted into the capillary tube. Observation is outside the capillary walls perpendicular to the axis. Detection at the pM level was demonstrated. We report here the use of this excitation geometry to simultaneously monitor 10 capillary tubes undergoing electrophoresis. This represents a truly parallel multiplexing scheme for monitoring large arrays of capillaries.

Several important observations can be made. First, truly simultaneous multiplexing of capillary electrophoresis was achieved because the CCD camera looks at all capillaries at all times, with data rates fast enough for sequencing at > 1 base per s per lane, or 1000 bases per s per CCD system. Second, there are no moving parts and the injection end of the capillary bundle can be freely manipulated without affecting alignment. Third, the excitation laser simply irradiates the entrance of the optical fiber bundle without critical alignment of the optics to achieve distribution of energy into each capillary. Fourth, there are variations in the excitation energies reaching each capillary, but that can be calibrated for in the same way the individual CCD pixels are normalized. Fifth, there are variations in absolute and relative migration times for the test compounds. This is expected due to the uncontrolled nature of the capillary walls (electroosmotic flow), different temperatures, and different geometries in each capillary. However, recently we developed a migration index and an adjusted migration index to specifically correct for these variations. Sixth, there are variations in the relative peak heights and areas among the capillaries for the same injected sample concentration. This is again expected due to bias at injection due to differences in electroosmotic flow rates, geometries, and injection time constants of the capillaries. Recently, we developed a correction scheme to address this exact problem and have been able to reduce injection bias to < 5%, which is adequate for DNA sequencing.

ANL

Argonne National Laboratory

SEQUENCING BY HYBRIDIZATION: TOWARDS LARGE-SCALE COMPILATION OF HUMAN cDNA SIGNATURES.

Crkvenjakov, R., Strezoska, Z., Milosavljevic, A., Zeremski, M., Grujic, D., Paunesku, T., Genome Structure Group, Biological and Medical Research Division, Argonne National Laboratory, Argonnne, Illinois

We have proposed and are developing the SBH Format 1 method in which densely arrayed DNA samples in the form of PCR products or M13 phage are consecutively interrogated by groups of short oligomers with a common hexamer to octamer core. Three levels of information can be achieved depending on the number of hybridizations. Mapping information in the form of clone signatures can be obtained with relatively few (100–300) probes. Positioning and identification of genome structural elements (partial sequencing) requires more extensive hybridizations, and complete sequencing requires data from several thousand probes on 3–5 related genomes (Drmanac & Crkvenjakov, 1992, Intl. J. Genome Res., **1**, 59–79).

As a first step, we are trying to find new human genes by compiling oligo sequence signatures of arrayed human brain cDNA libraries. Using the high-density-spotting robot and image analysis software developed for SBH (see accompanying abstract: Drmanac et al.), we are concentrating on a 8-× 12-cm 3456-dot cDNA array format to test i) various libraries prepared in-house or obtained elsewhere, ii) the reproducibility and reliability of sequencing data, and iii) our signature-compiling software. Over 500 hybridizations with informatic septamers on 10,000 arrayed cDNA clones serve as a base from which the technology is being evaluated. Substitution of ^{33}P for ^{32}P label has led to a fourfold increase in resolution, making correspondingly denser arrays practical. In the next year, we expect to reach a throughput of one million probe-clone bytes of information per day and complete signature analysis on 100,000 cDNA clones with 100–150 septamers. This should provide a substantial number of candidates for new genes to be analyzed further by partial SBH, gel sequencing, chromosome mapping, etc. [Work supported by the U.S. Department of Energy, Office of Health and Environmental Research, under Contract No. W-31-109-ENG-38.]

THE FIRST SEQUENCING BY HYBRIDIZATION (SBH) PRODUCTION LINE IN USE FOR GENE SORTING

R. Drmanac, S. Drmanac, A. Gemmel, J. Jarvis, I. Labat, N. Stavropoulos, and A.Vicentic
Integral Genetics Group, Biological and Medical Research Division, Argonne National Laboratory, Argonne, IL 60439

Last year, we demonstrated the feasibility of SBH by determining 340 bp without error in a blinded experiment (R. Drmanac et al., in *Genome Mapping and Sequencing*, Cold Spring Harbor Laboratory, NY, 1992, p. 317). On the basis of that experience, we built a hybridization data production line for scoring 100,000 clones with 10 probes per day. We have developed i) high-throughput PCR of plasmid inserts starting from bacterial cultures (see accompanying abstract); ii) the capability to spot 31,000 to 55,000 DNA samples on 15– × 23–cm membranes using a Biomek1000 with adapted tablet, pins and programs; iii) image analysis software for scoring hybridization signals from dense dot arrays; and iv) a semiautomatic hybridization setup. Currently we are developing automated entry of data into a data base and an expert system for evaluation and normalization of the hybridization results. The first application of the production setup will be the fingerprinting of 100,000 random cDNA clones from brain tissue (library constructed by M.B. Soares, Columbia University) using 100 6- to 8-mer probes. Up to now, 15,000 clones have been amplified by PCR and the whole procedure has been tested by hybridizing 1500 clones by 20 probes. By comparing the signatures of the clones, we are expecting to recognize about 20,000 new genes and a few hundred new gene families. If successful, the experiment will be repeated on over one million clones combined from a few tissues in order to inventory most of the genes and to gain insight into the gene expression pattern. The capacity of the production line is sufficient for 100-bp-resolution mapping of the human chromosomes and for an efficient, simultaneous sequencing of similar bacterial genomes. The phylogenetic sequencing scheme is especially powerful when SBH data are combined with targeted single-read gel sequences (R. Drmanac, in *Genome Mapping and Sequencing,* Cold Spring Harbor Laboratory, NY, 1992, p. 318). By establishing a preparative PCR procedure in 384-well or 864-well plates (to allow efficient spotting by corresponding pin array) and by constructing an automated hybridization machine, inexpensive sequencing of 100 million bp/yr can be achieved (R. Drmanac at al., *Electrophoresis*, 1992, **13**, 566–573). [Work supported by the U.S. Department of Energy, Office of Health and Environmental Research, under Contract No. W-31-109-ENG-38.]

EFFICIENT CLONE ARRAYING AND HIGH-THROUGHPUT PCR OF cDNA INSERTS

S. Drmanac, A. Gemmell, I. Labat, and R. Drmanac
Integral Genetics Group, Biological and Medical Research Division, Argonne National Laboratory, Argonne, IL 60439

To simplify arraying clones in microtiter plates, we have developed a procedure based on dispensing an optimally diluted transformation mixture into the wells, followed by 20 hours of growth. Success is measured by the number of wells giving a single-band product in PCR. The success of 70% (in comparison with 80% by picking single colonies) is obtained reproducibly by dilutions which give 10% empty wells on average. After growth in 96-, 384- or 864-well plates, PCR is started by inoculating the PCR mixture into corresponding plates using matching pin array. To get clean PCR products, the bacterial culture is diluted 25 times in water. PCRs on six plates are performed in parallel using a BioOven with a rotation platform. In two cycles, 1200-10,000 reactions can be prepared per day. To get successful PCR, it is important to optimize the concentration of primers and enzyme in every batch. Furthermore, we have developed a simple automated procedure for removing oil from the PCR reaction. The procedure is based on passive sucking of the oil into a 96-tip magazine (Dynatech Laboratories, Inc.) mounted on Biomek1000 arm and removing the oil by sliding the tips over a paper tissue. Up to now, 15,000 cDNA inserts have been amplified from random cDNA clones with an average insert of 1700 bp (library constructed by M.B. Soares, Columbia University) toward our goal of amplifying 100,000 clones. DNAs are going to be spotted and hybridized by 100 short probes (mainly 7-mers, labeled either by ^{32}P or ^{33}P in order to recognized new genes and gene families. [Work supported by the U.S. Department of Energy, Office of Health and Environmental Research, under Contract No. W-31-109-ENG-38.]

Accessing Integrated Genomic Data Using GRACE

Ross Overbeek and Morgan Price
Mathematics and Computer Science Division
Argonne National Laboratory
Argonne, IL 60439

GRACE is a database which integrates information from a number of existing databases. Enormous work has gone into developing many carefully curated databases that contain information relating to genomic sequences. Now, there are a number of efforts taking place around the world attempting to offer integrated access to this growing body of valuable information. GRACE is one of these projects. Our goal in developing the system is simply to offer more convenient access to the curated data; as such it builds directly on the efforts of many individuals and projects.

GRACE is an object-oriented database. By this, we mean that the user should think of GRACE as containing information about *objects.* Objects have *attributes.* Each object will have a *type* that categorizes the object. *genome* objects represent an entire genome (and relate to the chromosomes, plasmids, etc.) for specific orgainisms. *chromosome* objects represent specific chromosomes for specific organisms. *sequence_fragment* objects represent a section of DNA sequence that has been captured (in Genbank). *enzyme* objects represent an "abstract enzyme" in they can relate to many distinct peptides and genes (from many organisms), and *peptide* objects represent specific peptide sequences (we take this data from the Swiss Enzyme DB and the Swiss Protein DB). *prosite* objects represent the peptide motifs compiled by Amos Bairoch. Objects of type *cds, rRNA, tRNA,* and *misc_RNA* represent specific genes (from specific organisms). *map* objects are used to represent physical or genetic maps (and most of the information one gets from accessing these objects will be through relationships to the objects contained in the map). Objects of type *eco2dbase* capture the data provided by Fred Neidhardt's project to develop data relating to expression of *E.coli* genes. *rebase_entry* objects describe the sites cut by restriction enzymes (from Rich Roberts' DB). *pdb_entry* objects contain data relating to the coordinates of atoms within a specific peptide for which the crystal structure has been determined. *peptide_alignment* objects contain alignments of peptides (most of these supplied by the PIR). *nucleotide_alignment* objects contain alignments of DNA or RNA sequences (including alignments from the Ribosomal Database Project and the Berlin DB). Objects of type *compound* and *reaction* are used to encode the reactions in metabolic pathways. This is only a partial list, and we find that we add new object types frequently, since the body of curated data is expanding so rapidly.

Flexible access is currently provided via expression evaluation within a logic programming environment. A graphical interface is currently under development.

BNL

Brookhaven National Laboratory

PRIMER WALKING WITH STRINGS OF CONTIGUOUS HEXAMERS

F. William Studier, Jan Kieleczawa and John J. Dunn
Biology Department, Brookhaven National Laboratory, Upton, New York 11973

Strings of three or more hexanucleotides whose complements are adjacent in template DNA can prime DNA sequencing reactions specifically and efficiently when the template DNA is saturated with a single-stranded DNA-binding protein (SSB). The SSB suppresses priming by individual hexamers and most pairs of hexamers but stimulates priming by the 3′ hexamer of most strings of three or more contiguous hexamers. Strings of three or four hexamers representing over 200 of the 4096 possible hexamers primed easily readable sequence ladders at more than 75 different sites in single-stranded or denatured double-stranded templates 6.4 kb to 40 kbp long, demonstrating the generality of the phenomenon. About 60-90% of newly selected strings primed useful sequence information, a percentage that should increase as more is learned about how to select the hexamer strings most likely to prime well. A standard hexamer preparation provides enough material to prime thousands of sequencing reactions, and a library of hexamers would allow rapid and economical sequencing by primer walking on templates up to at least cosmid size. Automating this strategy should increase the efficiency of large-scale DNA sequencing more than an order of magnitude over current practice.

LANL

Los Alamos National Laboratory

Chromosome Sorting by Free-Flow Electrophoresis

Rhett L. Affleck
Los Alamos National Laboratory

Flow cytometry is the current method used to sort chromosomes. It represents the slowest step in the process of creating DNA libraries of the human genome, which is the essential first phase of the Human Genome Project. The primary goal of the proposed research is to pursue a separation technique for the sorting of chromosomes as an alternative to flow cytometry. Specifically, preparative free-flow electrophoresis will be studied for this purpose. Unlike flow cytometry, this proposed technique is suitable for large scale purification, and the apparatus on which the research will be performed could operate at 1,000 to 10,000 times the current rate of sorting by a flow cytometer.

Free-flow electrophoresis is a continuous process used for separation of whole cells, subcellular particles, and biomolecules. The technique is based on a system of laminar flow between two glass plates--the electrophoresis chamber--in which both the electrolyte and the sample solutions are continuously admitted. The electrolyte and sample flow in a direction perpendicular to the forces of an applied electric field. In this process, separation of the sample components occurs as a result of the differences in electrophoretic mobility or to the differences in isoelectric point.

The research to be performed can be divided into a few projects. First, conditions must be found in which chromosomes will both retain their stability and be deflected by the electric field. Buffer parameters, voltage, and flow rates must be appropriately designed for effective electrophoretic separation while maintaining the structural integrity of the chromosomes. Second, experiments to enhance the electrophoretic separation of the various chromosomes will be done. Experiments will include the differential removal of protein from the chromosomes, the addition of charged, nonspecific ligands, and tagging with existing and modified chromosome stains. Finally, various modes of electrophoresis and, if necessary, alternate electrophoretic apparatus will be explored.

One of the most encouraging aspects of this proposal is that any amount of enrichment that can be achieved for a single chromosome type in chromosome preparations will help speed up the flow cytometry sorting of chromosomes--and, therefore, shorten the time necessary to create human gene libraries. Of course, maximal success with this research will produce 100% purified electrophoretic preparations of all 24 chromosomes.

Optimization Tools for DNA Fragment Assembly: Algorithm Comparison

Christian Burks,[1,2,6] Michael L. Engle,[1] Stephanie Forrest,[3] Rebecca J. Parsons,[4] Cari A. Soderlund[1], and Paul E. Stolorz[5]

[1]Theoretical Biology and Biophysics Group; T-10, MS K710; Los Alamos National Laboratory; Los Alamos, New Mexico 87545. [2]Center for Human Genome Studies, LANL. [3]Dept. of Computer Science, University of New Mexico. [4]Computer Research and Applications Group, LANL. [5]Santa Fe Institute. [6]Corresponding author.

Almost all large-scale sequencing projects currently include a significant component of random, or "shotgun," sequencing in order to reduce a prohibitively large segment of DNA into numerous smaller, more manageable fragments [1]. These fragments must then be assembled into an accurate consensus of the region being sequenced, a problem that has proven to be computationally difficult.

Though any software tool will ultimately have to be proven useful on "real" experimental data sets, in the early stages of development, defined and broad-ranging data sets with known correct solutions are extremely useful. In order to compare and contrast various algorithms for sequence assembly, and to develop benchmark data sets, we have developed a set of tools, *genfrag*, that generate artificial sequence fragment data sets [2]. These benchmark tools generate a random, double-stranded fragment set from a known DNA sequence. Fragment size, parent coverage, mutation rate, error distribution, and repeat complexity are parameters which can be independently and systematically varied by the user.

We have expanded initial work on sequence assembly based on simulated annealing procedures [3] to examine the feasibility and efficiency of other stochastic search strategies, including relaxation, self-adaptive annealing, and genetic algorithms. These approaches are also compared to greedy algorithms. Self-adaptive variants of annealing make use of information from the current and recent behavior of the annealing algorithms to slow down and speed up the "cooling" schedule at appropriate stages. Genetic algorithms model optimization of the generation and selection of possible solutions on the mutational changes in chromosomal DNA and natural selection. We compare the results generated by these methods over a number of different input data sets, and examine whether or not optimization of assembly strategies makes a significiant difference in sequence assembly.

This work was funded by the DOE Genome Program (ERW-F137, R. Moyzis, P.I.; ERW-F159, R. Keller, P.I.). C.S. was supported by a DOE Human Genome Postdoctoral Fellowship, and R.P. by a LANL Director's Postdoctoral Fellowship.

[1] Hunkapiller, T., Kaiser, R.J., Koop, B.F., and Hood, L. (1991b) Large-scale and automated DNA sequence determination. *Science*, **254**, 59-67.

[2] Engle, M.L. and Burks, C. (1992) Artificially generated data sets for testing DNA sequence assembly algorithms. Submitted manuscript.

[3] Churchill, G., Burks, C., Eggert, M., Engle, M.L., and Waterman, M.S. (1992) Assembling DNA sequence fragments by shuffling and simulated annealing. Submitted manuscript.

Recognizing Genes Workshop

Christian Burks,[1,2,7] Chris Fields,[3] Stephen Henikoff,[4] Deborah Joseph,[5] and Gary Stormo[6]

[1]Theoretical Biology and Biophysics Group; T-10, MS K710; Los Alamos National Laboratory; Los Alamos, New Mexico 87545. [2]Center for Human Genome Studies, LANL. [3]National Institutes of Health. [4]Fred Hutchinson Cancer Research Center. [5]University of Wisconsin. [6]University of Colorado. [7]Corresponding author.

The Aspen Center for Physics (ACP), in Aspen, Colorado, sponsored a three-week workshop from 18 May to 5 June, 1992, with 25 scientists participating. The workshop, entitled *Recognizing Genes and Other Components of Genomic Structure*, was the third (RG-III) in a series hosted by ACP on this topic; the previous workshops (RG-I [1] and RG-II [2]) occured in 1990 and 1991.

The workshop focussed on discussion of current needs and future strategies for developing the ability to identify and predict the presence of complex functional units on sequenced, but otherwise uncharacterized, genomic DNA. We addressed the need for computationally-based, automatic tools for synthesizing available data about individual consensus sequences and local compositional patterns into the composite objects (e.g., genes) that are -- as composite entities -- the true object of interest when scanning DNA sequences. The general background and justification for a workshop on this topic was discussed earlier in the report on RG-I [1]. Of particular interest over the past year has been the maturation of previously described as well as the emergence of several new approaches to predicting the presence and location of genes (or at least protein-coding exons).

The workshop was structured to promote sustained informal contact and exchange of expertise between molecular biologists, computer scientists, and mathematicians. No participant stayed for less than one week, and most attended for two or three weeks. Computers, software, and databases were available for use as "electronic blackboards" and as the basis for collaborative exploration of ideas being discussed and developed at the workshop.

This work was funded by the DOE Genome Program (ERW-F137, R. Moyzis, P.I.), the Aspen Center for Physics, the NSF, and the Theoretical Division at LANL.

[1] Burks, C., Myers, E. and Stormo, G. (1990) *Recognizing Genes and Other Components of Genomic Structure: Workshop Report (Aspen Center for Physics, 28 May - 15 June, 1990).* Unpublished report, Los Alamos National Laboratory, LA-UR-91-1713.

[2] Burks, C., Fields, C. and Myers, E. (1991) *Recognizing Genes and Other Components of Genomic Structure: Workshop Report (Aspen Center for Physics, 23 May - 7 June, 1991).* Unpublished report, Los Alamos National Laboratory, LA-UR-92-645.

[3] Burks, C., Fields, C., Henikoff, S., Joseph, D., and Stormo, G. (1992) *Recognizing Genes and Other Components of Genomic Structure: Workshop Report (Aspen Center for Physics, 18 May - 5 June, 1992).* Unpublished report, Los Alamos National Laboratory.

Genome Map Evaluation, Assembly, and Diagnosis

James W. Fickett and Michael J. Cinkosky

Theoretical Biology and Biophysics Group, and Center for Human Genome Studies
Los Alamos National Laboratory, Los Alamos, NM.

In the construction of a relatively simple map (containing a few tens of elements), an investigator can keep in mind all of the relevant data, and can weigh and resolve conflicting evidence to produce a map which fits all the experimental results reasonably well. Today, however, much more complex maps (containing thousands of elements) are being constructed, and there is increasing dependence on computational tools to keep track of maps and their relationships to the underlying data. Current map assembly algorithms typically are incapable of recognizing and dealing with conflicts or ambiguities that require revision of a tentative map, but clearly the need to balance conflicting objectives is as great or greater for complex maps as for simple ones. Optimization theory provides a natural conceptual framework for genome map assembly and evaluation. We present a first draft of a mathematical formalism for optimization of genome maps. This paves the way for building maps which fit all the data as well as possible, in a well defined mathematical sense. Equally important, it provides a means to measure how well a map, however constructed, fits the available experimental results. We also discuss means to reveal in a meaningful way conflicts in the experimental data on which a map is based.

Estimation of Protein Coding Density in a Corpus of Sequence Data

James W. Fickett and Roderic Guigó

Theoretical Biology and Biophysics Group, and Center for Human Genome Studies
Los Alamos National Laboratory, Los Alamos, NM.

A number of methods have been reported for estimating the number of genes in a genome, or the closely related *coding density* of a genome, defined as the fraction of base pairs in codons. Until recently these methods have been based primarily on either mutation analysis or mRNA transcript mapping. Recently, DNA sequence data representative of the genome as a whole have become available for several organisms, making the problem of estimating coding density amenable to sequence analytic methods. Different investigators have in many cases given widely varying estimates of the number of genes in the same genome, so that a careful analysis of accuracy has become increasingly desirable. We present a computational method in which a "coding statistic" is calculated for a large number of sequence windows, and the distribution of the statistic is decomposed into coding and noncoding fractions. We then analyze the error of the method by testing it on known cases, and show that coding density can be measured to within a few percent. The method is applied to the yeast chromosome 3 sequence and the eight known cosmid sequences from C. elegans. This method can also be applied to fragmentary data, for example a collection of short sequences determined in the course of STS mapping, and so is applicable at the present to a number of other genomes as well.

Assessment of Protein Coding Regions

James W. Fickett and Chang-Shung Tung

Theoretical Biology and Biophysics Group, and Center for Human Genome Studies
Los Alamos National Laboratory, Los Alamos, NM.

A number of methods for recognizing protein coding genes in DNA sequence have been published over the last 13 years, and new, more comprehensive algorithms, drawing on the repertoire of existing techniques, continue to be developed. To optimize continued development, it is valuable to systematically review and evaluate published techniques. At the core of most gene recognition algorithms is one or more *coding measures* -- functions which produce, given any sample window of sequence, a number or vector intended to measure the degree to which a sample sequence resembles a window of "typical" exonic DNA. In this paper we review and synthesize the underlying coding measures from published algorithms. A standardized benchmark is described, and each of the measures is evaluated according to this benchmark. Our main conclusion is that a very simple and obvious measure -- counting oligomers -- is more effective than any of the more sophisticated measures. Different measures contain different information. However there is a great deal of redundancy in the current suite of measures. We show that in future development of gene recognition algorithms, attention can probably be limited to six of the twenty or so measures proposed to date.

CHROMOSOMAL ORGANIZATION OF HUMAN REPETITIVE DNA SEQUENCES.

Jonathan L. Longmire[1], Larry L. Deaven[2], Robert L. Ratliff[1], Nancy C. Brown[1], Robert J. Baker[3], and Carl E. Hildebrand[1].

[1]Genomics and Structural Biology Group, [2]Center for Human Genome Studies, Los Alamos National Laboratory, Los Alamos, NM 87545.
[3]Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409.

We are investigating the chromosomal organization of repetitive sequences, primarily simple-repeats (microsatellites) which are generally highly polymorphic within the human genome. Multiple-representation human chromosome-specific cosmid libraries have been constructed as an effort on the National Laboratory Gene Library Project. We have used a Biomek robot to make approximately 1-fold representation high density gridded arrays of the chromosomes 16 and Y cosmid libraries. The grids have been screened by hybridization to all possible mono-, di-, and trinucleotide repeats, as well as to other more complex repeats such as Alu, L1, THE, and a variety of satellite sequences. A database is being generated that will contain the identity of all clones found to be positive for each repetitive probe. This approach enables comparative studies of the frequency of specific repeats on human chromosomes 16 and Y, and provides information concerning the coincident linkage of different repeats on fragments of DNA approximately 40,000 base pairs in length. Identification of microsatellite-containing clones within the LANL chromosome 16-specific cosmid library links potentially highly informative genetic markers onto the emerging physical map. These results should help to integrate the physical and genetic maps, as well as facilitate the rapid localization of disease loci on human chromosome 16.

Listing of Molecular Biology Databases

Graham W. Redgrave[1] and Christian Burks[1–3]

[1]Theoretical Biology and Biophysics Group; T-10, MS K710; Los Alamos National Laboratory; Los Alamos, New Mexico 87545. [2]Center for Human Genome Studies, LANL. [3]Corresponding author.

The sheer volume of molecular biological data in the past two decades precipitated a move to computer media for collection, maintenance, and analysis of data. However, recent marked improvements in database management systems, network access to remote database implementations, and the flexibility and richness of user interfaces (as well as the power of the systems supporting them) is leading to another, potentially larger wave of migration of molecular biological data onto computer media. This recent wave is also being precipitated by the Human Genome Project's need for improvements in the collection, comparison, and integration of related data sets from many different sites.

However, a lack of wide-spread standards for organizing and accessing molecular biology databases, and sparse implementation of those standards that do exist, has led to a vast majority of databases that currently exist being relatively unique in terms of access protocols and query platforms. Thus, the scientist wanting to access molecular biological data has to: (i) find the relevant databases among a highly decentralized (geographically and institutionally) set; (ii) learn a number of access protocols and query languages for the relevant, identified databases; and (iii) post-process -- most often with a high degree of manual intervention -- the query results on individual databases to answer the exact query that initiated the search.

The LiMB (**Li**sting of **M**olecular **B**iology Databases) database, which was created to help facilitate and simplify this process [1], is a database of databases, or a database directory. It provides a systematic and coordinated approach to identifying, linking, and accessing heterogeneous, distributed databases relevant to molecular biology by providing an overview of these databases and the data sets they cover. LiMB is currently implemented in a relational database management system that allows for complex, multiconditional queries [2]. We present an overview of LiMB, a description of its current implementation, an updated overview of Release 3.0 [3], and information on how to retrieve the data it contains.

This work was funded by Los Alamos National Laboratory institutional research funds and the DOE Genome Program (ERW-F116, J. Fickett, P.I.).

[1] Lawton, J. R., Martinez, F., and Burks, C. (1989) Overview of the LiMB database. *Nucl. Acids Res.*, **17**, 5885-5899.

[2] Keen, G.M., Redgrave, G.W., Lawton, J.R., Cinkosky, M.J., Fickett, J.W., Mishra, S.K. and Burks, C. (1992) Access to molecular biology databases. *Math. Comput. Model.*, **16**, 93-102.

[3] Redgrave, G.W. and Burks, C. (1992) *LiMB Release 3.0*. Unpublished report, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

DNA Map Assembly Workshop

Cari A. Soderlund[1], and Christian Burks,[1–3]

[1]Theoretical Biology and Biophysics Group; T-10, MS K710; Los Alamos National Laboratory; Los Alamos, New Mexico 87545. [2]Center for Human Genome Studies, LANL. [3]Corresponding author.

The assembly of shot-gun DNA sequence fragments into a consensus sequence and the construction of genomic physical maps from clone overlap data present similar formal problems in the context of combinatoric optimization, and can be structured to take advantage of established and developing computer science techniques.

The Santa Fe Institute sponsored a one-day workshop in April 1992 on the topic of theoretical and experimental approaches to assembly of physical maps and sequencing projects. Experimental biologists, theoretical biologists, and computer scientists discussed overlapping interests and potential collaborations. There were about 30 participants from the Santa Fe Institute, University of New Mexico (Dep't of Computer Science; Pathology Dep't, School of Medicine), and Los Alamos National Laboratory (Center for Non-Linear Studies; Research and Applications Group; Cell Growth, Damage, and Repair Group; Genomics and Structural Biology Group; Complex Systems Group; Theoretical Biology and Biophysics Group)

Ten formal presentations were made, each with a strong tutorial component. Subjects included (i) the experimental methodology for sequencing and physical mapping of DNA, (ii) overviews of genetic algorithms, simulated annealing, and other stochastic (and non-stochastic) optimization methods, and (iii) the applied optimization of DNA sequence assembly, DNA physical mapping, and DNA clone hybridization strategies.

This work was funded by the Theoretical Division at LANL, the Santa Fe Institute, and the DOE Genome Program (ERW-F116, J. Fickett, P.I.; ERW-F137, R. Moyzis, P.I.). C.S. was supported by a DOE Human Genome Postdoctoral Fellowship.

LBL

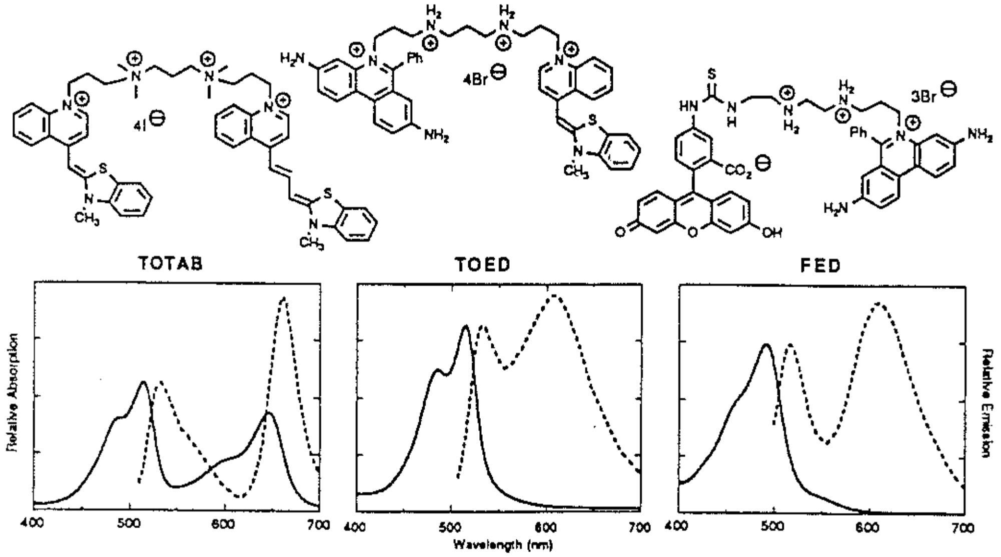
Lawrence Berkeley Laboratory

FLUORESCENT DNA INTERCALATION COMPLEXES WITH HETERODIMERIC DYES DESIGNED FOR ENERGY TRANSFER: PREPARATION AND APPLICATIONS.

Scott C. Benson, Paramjit Singh and Alexander N. Glazer
Division of Biochemistry and Molecular Biology, Department of Molecular and Cell Biology, and the Division of Chemical Biodynamics, Lawrence Berkeley Laboratory, University of California, Berkeley, CA 94720

Double-stranded DNA (dsDNA) forms high affinity fluorescent complexes with homodimeric dyes such as ethidium homodimer (1), thiazole orange homodimer (2), and oxazole orange homodimer (2). Such complexes are stable to electrophoresis and in conjunction with confocal detection of laser excited fluorescence (3) allow detection of dsDNA on gels with picogram sensitivity. These complexes have been applied to the multiplex detection, sizing, and quantitation of restriction fragments, products of the polymerase chain reaction, and DNA-protein complexes (4, 5).

We report here the preparation of a new class of fluorescent reagents designed to bind tightly to DNA and share a common excitation wavelength but to differ in the emission wavelength and permit simultaneous fluorescence detection of multiple targets with high sensitivity. The reagents are heterodimeric polycationic dyes with one dye moiety which serves as an energy donor and a second which serves as an energy acceptor and emits fluorescence. Structures of three such heterodimers are shown below together with the absorption spectra and emission spectra of their complexes with dsDNA at 100 bp:dye. In the complexes with dsDNA, the fluorescence emission of the donor is quenched by approximately 90% and the fluorescence of the acceptor is greatly enhanced.



TOTAB forms complexes with dsDNA stable to electrophoresis and has been successfully applied to the multiplex detection of restriction fragments and DNA-protein complexes. (Supported by DOE under contract DE-FG-91ER61125)

(1) Glazer, A.N., Peck, K., and Mathies, R.A. (1990) Proc. Natl. Acad. Sci. USA 87:3851-3855; (2) Rye, H.S., Yue, S., Wemmer, D.E., Quesada, M.A., Haugland, R.P., Mathies, R.A., and Glazer, A.N. (1992) Nucleic Acids Res.20:2803-2812; (3) Mathies, R.A. and Huang, X.C. (1992) Nature 359:167-169 (1992); (4) Glazer, A.N. and Rye, H.S. (1992) Nature 359:859-861; (5) Rye, H.S. and Glazer, A.N. (1993) Abstracts, Human Genome Workshop, February 7-11, 1993, Santa Fe, NM.

Sequencing by Hybridization:
Methods to Generate Large Arrays of Oligonucleotides

Thomas Brennan
Genetics Dept., Stanford University & Lawrence Berkeley Lab
Stanford, CA 94305

The goal of this project is to develop methods to produce very large, high density arrays of oligonucleotides for use in hybridization sequencing. These arrays of oligonucleotides will be synthesized in parallel chemical reactions on glass plates. Standard amidite coupling chemistry is employed, but the synthetic reactions will be carried out on a picoliter scale. The specific nucleotides will be delivered to each array element by arrays of piezoelectric pumps similar to an ink jet printer. In effect, oligonucleotide synthesis becomes a four-color printing application. A stable hydroxyalkyl or aminoalkyl group bound to the plate acts as the hydrophilic 5'-surrogate on which to initiate strand synthesis. This linker arm enables removal of purine and pyrimidine blocking groups without cleavage of the oligonucleotide from the support.

Each element in the array is separated from its nearest neighbors by a surface tension barrier. Monolayers of perfluoroalkanes are not wet by acetonitrile, and droplets of reagents in adjacent elements can thus be prevented from mixing. Three methods for creating the patterned reaction surfaces have been demonstrated. 1) An array of gold dots is laid down on glass by sputtering through a mask, the region between the dots derivatized with a fluoroalkyl silane, the gold etched with KI, and the newly exposed spot derivatized with the synthesis linker arm. 2) A thick film of polymerized perfluoroalkyl vinyl siloxane is prepared, and the hydrophobic organic surface is directly oxidized with a medium pressure oxygen plasma discharge through a gold or ceramic mask. The resulting silanol groups are then derivatized with the synthesis linker arm. 3) An aminosilane surface is complete blocked as the o-nitrobenzyl carbamate. The nitrobenzyl group is then removed by photolysis through a quartz mask, and the exposed regions are made hydrophobic either as the perfluorooctyl amide, or as the perfluorooctyl sulfonamide derivative. A second photolysis cleaves the remaining nitrobenzyl blocking groups, and exposes the patterned aminoalkyl synthesis linker arm regions.

Array surfaces with 100 μ synthesis elements separated by 25 μ hydrophobic barriers are well behaved through 4 cycles of nucleotide coupling (not by droplet delivery yet, but by flooding entire plate with reagents). The oligonucleotide density is about 0.4 pmol / mm^2. Prototype pizeoelectric pumps have been fabricated in both silicon and glass, and the drop delivery and environmental control chambers have been designed and are now being constructed.

PROGRESS AND NEEDS IN MOLECULAR CYTOGENETICS.

J.W. Gray[1,2], H.-U. Weier[2], W.-L. Kuo[2], F. Waldman[2], M. Pallavicini[1,2], D. Sudar[2], M. Sakamoto[1], G. Gingrich[1], L. Bolund[2,3], and D. Pinkel[1,2]. [1]Lawrence Berkeley Laboratory, Berkeley, CA. [2]University of California, San Francisco. [3]Danish Center for Human Genome Research, Aahrus, Denmark

This presentation will summarize our recent progress in the development of improved molecular cytogenetic techniques, the application of these techniques to problems of clinical and biomedical importance and requirements for further advancement in the field.

Technical developments include: (a) Fluorescence in situ hybridization (FISH) for detection and characterization of disease linked aberrations in individual cells. (b) PRimed IN Situ labeling (PRINS) for rapid hybridization and the potential for near single base detection sensitivity. (c) Comparative Genomic Hybridization (CGH) for rapid detection and mapping of regions of altered gene copy number anywhere in the genome.

Applications where molecular cytogenetics is playing a prominent role include: (a) Probe mapping. Mapping has been almost fully automated so that mapping speed is limited by the rate at which high quality metaphase preparations can be found. Mapping precision (sem) is typically 1-2 Mb. (b) Assessment of exposure to chemical and physical agents. Agents such as radiation may induce aneuploidy and structural chromosome rearrangements. FISH with centromeric repetitive probes is used effectively to detect aneuploidy while FISH with whole chromosome probes allows rapid detection of stable structural aberrations in metaphase spreads. Aberrations can be recognized quickly by human observers and computer assisted microscopy is being developed to automate the scoring process completely. (c) Prenatal and neonatal diagnosis. Trisomies of chromosomes 13, 18, 21, and aneusomies involving the sex chromosomes can be found using FISH to interphase amniocytes or chorionic villus cells. In addition, identification and molecular cytogenetic analysis of rare of fetal cells in the maternal circulation appears feasible. (d) Analysis of human malignancies. Cancers develop and progress through the accumulation of genetic abnormalities at critical loci. FISH, PRINS and CGH are useful for detection and characterization of many of these abnormalities thereby enabling assessment of the tumor genotype, analysis of genetic heterogeneity, and detection of rare malignant cells. These unique capabilities may now enable development of genetically defined tumor grading systems for improved diagnosis, prognostication and treatment.

Although substantial progress has been made, molecular cytogenetic studies are still limited by the lack of informative probes optimized for use in clinical samples and by the labor intensive nature of the analysis process. Development of such probes and improved multicolor computer assisted fluorescence microscopy to facilitate molecular cytogenetic studies will be discussed.

This work was supported by the Office of Health and Environmental Research under Contract no. DE-AC-03-76SF00098, Imgenetics and USPHS grants CA45919, CA44768, CA47537 and HD17665.

Robust Shotgun DNA Sequence Assembly

Frank Olken[*]
Computer Science Research & Development Dept.
Information and Computing Sciences Div.
Lawrence Berkeley Laboratory
1 Cyclotron Road, Berkeley, CA 94720

John Kececioglu[†]
Computer Science Department
University of California at Davis
Davis, CA 95616

We are developing algorithms for DNA sequence assembly from shotgun data that address common sources of difficulty for current systems: repetitive DNA sequences, chimeric clones, and nonuniform error rates along fragments. Our work builds upon the four-phase approach of Kececioglu and Myers [Kec91], which decomposes sequence assembly into overlap detection, fragment orientation, fragment layout, and consensus sequence determination.

Current sequence assembly codes are confounded by repetitive DNA sequences, such as Alu and L1 repeats. Whereas these codes tend to stack or arbitrarily permute fragments that sample copies of a repeat, we intend to separate such fragments by screening against databases of prototypical sequence repeats. Preliminary analysis indicates that by multiple sequence alignment against representative consensus sequences of repeat families, fragments that sample different copies of a repeat may be distiguished, even when the error rate for approximate repeats approaches the sequencing error rate.

Chimeric clones that combine dispersed fragments can cause incorrect melding of regions, and so prevent closure of contigs. We are investigating three approaches for screening chimeric clones: interrupted fragment alignments, likelihood tests on overlap graph conformations, and local violations of interval graph axioms.

Current methods for fragment orientation and layout either use heuristics, which allow these phases to be combined, or use optimization methods, which force the phases to be considered in isolation. We are implementing an orientation and layout algorithm based on graph matchings that combines both phases while applying optimization methods. A two-pass strategy is being tested in conjuction with this algorithm. The first pass solves orientation and layout using only highly reliable overlaps with the low-error ends of fragments not containing known repeats. The second pass closes the resulting contigs by incorporating less reliable overlaps.

Improved consensus sequence determination is based on new algorithms for multiple sequence alignment [Kec92], combined with maximum likelihood "voting" methods based on position- and sequence-dependent error models being developed by of T. Hunkapiller, W.-Q. Chen, G. Alexander and T. Speed.

Acknowledgements. Our thanks to T. Hunkapiller for sequencing project data, and J. Jurka for the prototypical repetitive DNA sequence database.

References

[Kec91] John Kececioglu. *Exact and Approximation Algorithms for DNA Sequence Reconstruction.* PhD thesis, Deparment of Computer Science, The University of Arizona, 1991.

[Kec92] John Kececioglu. The maximum weight trace problem in multiple sequence alignment. Technical report, Computer Science Department, University of California at Davis, 1992.

---

[*] This work was partially supported by the Laboratory Directed Research and Development Funds, Lawrence Berkeley Laboratory supported by the U.S. Department of Energy under Contract DE-AC03-76SF00098. E-mail address: `olken@lbl.gov`

[†] Research supported by a postdoctoral fellowship from the NSF Program in Mathematics and Molecular Biology under Grant DMS-8720208. E-mail address: `kece@cs.ucdavis.edu`

MULTI-COLOR DIGITAL FLUORESCENCE MICROSCOPY FOR MOLECULAR CYTOGENETICS. D. Pinkel[1,2], D. Sudar[2], D. Peters[2], L. Mascio[4], W. L. Kuo[2], M. Sakamoto[2], A. Kallioniemi[2], O. Kallioniemi[2], J. Piper[2,3], D. Rutovitz[3], F. Waldman, J. Gray[1,2]. [1]Lawrence Berkeley Laboratory. [2]University of California San Francisco. [3]MRC Edinburgh. [4]Lawrence Livermore National Laboratory.

We are developing a multi-color digital fluorescence microscopy system for molecular cytogenetic applications, including probe mapping, genotypic and phenotypic analysis of individual cells, and comparative genomic hybridization (CGH). This Quantitative Image Processing System (QUIPS) is built from commercial components. QUIPS employs a multi-band beam splitter and emission filter built to our specification by Chroma Technology (Brattleboro, VT). Separate single band excitation filters are used to excite individual fluorochromes at approximately 360 nm, 405 nm, 495 nm and 570 nm. Cross correlation measurements of the registration of the different spectral bands of the image indicate that registration shifts are less than 0.1 µm in the object plane. Cross talk due to weak excitation of one fluorochrome by the excitation wavelength intended for another is low enough for effective quantitative use of DAPI, FITC and Texas red/rhodamine.

This system has been used extensively for semi-automated mapping of probes on metaphase chromosomes. Chromosomes and candidate hybridization domains are automatically segmented, and the chromosomal medial axis is determined. These are displayed, and the operator selects those domains judged to represent true signals. The location of the intensity-weighted center of mass of each true hybridization domain is projected onto the medial axis and the fraction of the chromosome length (FL) from the p telomere is determined. Measurement of 10 to 15 hybridization signals allows estimation of the probe location to within ~1 Mb (sem) when using metaphase chromosomes of standard length. FL variation, reported as one standard deviation of the FL distributions, averages 2-3 Mb for phage and cosmid probes. Several probes can be pooled, hybridized, and mapped simultaneously if they are known to be on different chromosomes so that they can be individually identified.

QUIPS has been essential for analysis of CGH, a technique we have recently developed for measurement of DNA sequence copy number throughout the genome of tumor cells. In a typical application of CGH genomic DNA from a tumor and from normal cells are differentially labeled and simultaneously hybridized to normal metaphase chromosomes, detected with different fluorochromes, and images of each fluorochrome are obtained. The chromosomes are segmented by thresholding in the background-corrected DAPI image. The intensities of the hybridized tumor and normal DNAs along each chromosome are then calculated by integrating across the chromosome width, correcting for the local background. The intensity ratio at each position is proportional to the ratio of the copy numbers of the sequences that bind there in the tumor and normal genomes. Thus "copy number karyotypes" of the genome are obtained. Deletions and duplications of chromosomal segments longer than 10 Mb, and greater than 7 fold amplification of regions of several hundred kb can be detected.

This work was supported by the Office of Health and Environmental Research under Contract no. DE-AC-03-76SF00098, Imgenetics and USPHS grants CA45919, CA44768, CA47537 and HD17665.

FLUORESCENCE GEL MOBILITY-SHIFT ASSAY: RADIOISOTOPIC-SENSITIVITY DETECTION OF DNA-BINDING PROTEIN INTERACTIONS WITH STABLE DYE-DNA INTERCALATION COMPLEXES.

Hays S. Rye, Becky L. Drees, Hillary C. M. Nelson, and Alexander N. Glazer, Division of Biochemistry and Molecular Biology, Department of Molecular and Cell Biology, and the Division of Chemical Biodynamics, Lawrence Berkeley Laboratory, University of California, Berkeley, CA 94720.

We have previously demonstrated that certain polyfunctional dyes, such as ethidium homodimer (EthD) and thiazole orange homodimer (TOTO), form highly fluorescent complexes with double-stranded DNA (dsDNA) that are stable to electrophoresis. These complexes can be detected in gels with very high sensitivity using a two-color confocal fluorescence gel scanner (1). We have extended the application of these fluorophores to the examination of DNA-binding proteins by the gel mobility-shift assay. This assay relies upon the retardation of migration of a DNA fragment in a polyacrylamide gel upon binding of a protein to the fragment. We have found that we can label a DNA-protein complex with a polyfunctional DNA-binding dye and readily detect mobility shifts of the target DNA by fluorescence. Using the *E. coli* mismatch repair protein mutS in collaboration with J. Rine and A. Lishanskaya (2), we can easily detect the preferential binding of the mutS protein to heteroduplex DNA by labeling target DNA fragments with TOTO. We have also analyzed the interactions of the trimeric yeast heat shock transcription factor (HSF) with DNA (3). Labeling of an HSF-DNA complex with either EthD, TOTO, or the new energy transfer dye thiazole orange-thiazole blue heterodimer (TOTAB) (4), allows the detection of HSF-induced mobility shifts well within the detection limits of ^{32}P end-labeled-DNA autoradiographic methods. Using either EthD, TOTO or TOTAB, as little as 15-30 femtomoles of a 500 bp target DNA fragment can be used to easily detect HSF induced mobility shifts. The complexes formed between DNA, protein and dye produce band-shift patterns that are indistinguishable from those obtained by radioactivity. By fluorescently derivatizing the HSF protein and using the labeled protein in an assay with target DNA stained with TOTB, we have been able to conduct two-color mobility-shift experiments to establish the stoichiometry of DNA-protein complexes.

(Supported by DOE under contract DE-FG-91ER61125)

(1) Glazer, A.N. and Rye, H. S. (1992) Nature 359:859-861; (2) Lishanskaya, A. and Rine, J. (1993) Abstracts, Human Genome Workshop, February 7-11, 1993, Santa Fe, NM. (3) Sorger, P. K. and Nelson, H. C. M. (1989) Cell 59:807-813; (4) Benson, S. C., Singh, P. and Glazer, A. N. (1993) Abstracts, Human Genome Workshop, February 7-11, 1993, Santa Fe, NM.

ORNL

Oak Ridge National Laboratory

Scanning Probe Microscopy of Complete Plasmids

D. P. Allison, T. G. Thundat, T. L. Ferrell, M. J. Doktycz,
K. Bruce Jacobson, and R. J. Warmack

Oak Ridge National Laboratory, Oak Ridge Tennessee 37831

Over the past ten years since its inception, the scanning tunneling microscope (STM) and the atomic force microscope (AFM) have proved to be revolutionary techniques for examining structures at the atomic level. Because these instruments can perform not only in vacuum but also in air and under fluids, their application to biological imaging is obvious. They are the only instruments in which atomic and molecular resolution can be attained in biologically favorable environments. Unfortunately, the interaction of the scanning tip with adsorbed molecules, which are often poor conductors of the tunneling current in STM, has frustrated the application of scanning probes to the detailed examination of such bio-structures. Our primary research goal is therefore to understand and develop new sample preparation techniques compatible with scanning probe examination of DNA. A secondary goal is to develop spectroscopic tools, either electronic, mechanical, or optical, to help analyze the base sequence or to identify labels. Thus, each technique provides, in addition to topographic imaging, new data such as tactile measurements, conductivity profiles, or optical spectroscopic data to be collected on biomolecules. Future hybrid instruments will be even more powerful and yield complementary data simultaneously.

During the past year we discovered techniques to successfully mount DNA for scanning probe investigation. The mounting surface must be extremely flat since topographic contrast is normally used to locate these extremely fine (2 nm) molecules. A number of atomically flat surfaces were tested but were all found to lack sites with adequate binding to DNA. Only very sparsely located and often unstable molecules were observed. Attempts to immobilize DNA on surfaces using *ex situ* or even *in situ* electrochemical deposition proved only partly successful. The new approach for STM mounting involves allowing a thiol (—SH) group at one end of an organic moiety to bind to a specially prepared flat gold surface and chemically polarizable head groups at the other end to attract the DNA. Using radiolabeled DNA and monitoring the uptake onto various types of treated surfaces we developed a rapid prescreening technique to evaluate the efficacy of each treatment. Choosing a promising binder molecule and applying a DNA solution we were able to obtain stable images of DNA at the surface concentrations applied. These are the *first* STM images of complete molecules of a genetically functional DNA.

Using atomic force microscopy, we have studied adsorption of DNA and RNA on chemically modified surfaces. The substrates used include mica modified by chemical treatment with transition metal ions [Mg(II), Zr(IV), Cr(III), Co(II), Ba(II), or La(III)], epitaxial gold films treated with 2-dimethylaminoethanethiol, or silylated mica. Biopolymers air dried on the surface show coiling and 'superwinding' due to surface tension while washed samples show stable isolated strands. To reduce the effect of tip geometry on images specially sharpened tips made in a scanning electron microscope were used. Topographic and spectroscopic images of isolated nucleic acid molecules obtained both in air and under solution will be presented.

The techniques were developed in 1991 by a multi-disciplinary effort by members of Health and Safety Research Division, Biology Division, Chemistry Division, and visiting collaborators. The success of this research is due to an *essential* interdisciplinary mix of personnel from areas of physics, chemistry, and biology.

LUMINESCENT LANTHANIDE ION COMPLEXES AS LABELS FOR DNA SEQUENCING AND MAPPING

Gilbert M. Brown,* Martha L. Garrity,* Jeffrey E. Elbert,* Richard A. Sachleben,* Frederick V. Sloop,** Mitchel J. Doktycz,** and K. Bruce Jacobson,**
Chemistry Division* and Biology Division,** Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831

Luminescence of the lanthanide ions Sm(III), Eu(III), Tb(III), and Dy(III) has several advantages over fluorescent organic compounds as labels for DNA sequencing and mapping. Emission from Ln(III) ions have narrower bandwidths, and the long lifetimes allow detection with a lower background. A derivative of the macrocylic chelating agent, 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA), is used to attach the Ln(III) ions to oligonucleotides. DOTA is an ideal chelating agent for lanthanide ions since it forms stable, kinetically inert complexes. A high yield, general synthetic route was developed for the synthesis of derivatives of DOTA with an aryl substituent at one of the ethylene carbon atoms. The methyl ester of the aryl substituted alanine is the starting material for the synthesis of substituted DOTA ligands. Tosyl protection of the amine followed by reaction with ethylenediamine, further protection, and reduction yield the tosyl protected diethylenetriamine with an aryl substituent. Ring formation was accomplished by treatment with cesium carbonate and 2,2'-(tosylimino)bisethyl ditosylate (Richman-Atkins conditions). Deprotection was accomplished by reduction with lithium tetrahydroaluminate and the acetate arms were attached by treatment with the t-butylester of bromoacetic acid. Acid hydrolysis of the t-butyl groups yielded the tetraacetic acid of the aryl substituted tetraazacyclododecane. Nitration of a benzyl substituent provided a point of attachment of the ligand to DNA. The nitro group was reduced to an amine and reacted with thiophosgene to produce the benzylisothiocyanate derivative of DOTA. Reaction of the isothiocyanate with a hexylamine linker arm on a 17-mer oligonucleotide has been accomplished. The function of the Eu(III) labeled oligo as a primer for the Sanger sequencing procedure and for PCR will be reported.

Detection sensitivity for luminescence from Ln(III) ions can be greatly enhanced if excitation is to a ligand based state having a long lived triplet electronic state. Energy transfer from the ligand triplet to the Ln(III) excited states then results in emission from the narrow bandwidth f-f states. A substituted DOTA ligand having a naphthylmethyl group on the ethylene carbon was synthesized. Naphthalene has a triplet state with energy appropriate to transfer energy to the four Ln(III) ions of interest. The naphthyl group of naphthylmethyl-DOTA thus serves as an antenna to funnel excitation energy to complexed Ln(III) ions. Sensitization of the luminescence via a naphthyl group allows excitation to occur at a single wavelength for simultaneous detection of multiple labels. Sensitized emission was observed from Nap-DOTA complexes and from complexes with a naphthyl derivative of the closely related ligand diethylenetriaminepentaacetic acid. The dynamics of energy transfer from the naphthalene triplet states to the Ln(III) ions will also be reported.

This work is sponsored by the Office of Health and Environmental Research, U. S. Department of Energy, under contract No. DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.

DNA Sequencing by Laser Mass Spectrometry

C. H. Chen, K. Tang, S. L. Allman, R. G. Jones, and K. B. Jacobson

Photophysics Group, Health and Safety Research Division
Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6378

Gel electrophoresis for DNA size analysis is very similar to a time-of-flight (TOF) mass spectrometer to measure molecular weight of different molecules. During the past year, we have used laser mass spectrometry to analyze DNA segments. A Nd-YAG laser or an excimer laser was used to desorb DNA segments which were mixed with small organic molecules served as matrices. A time-of-flight (TOF) mass spectrometer was used to measure the desorbed ions. Oligomers with sizes up to 100 bases for poly-T and 60 bases for mixed base oligomers have been successfully detected by both positive and negative ion spectrum. The kinetics of production and destruction of these oligomer ions has also been studied by a specially designed TOF. It was discovered that negative oligomer ions have a finite lifetime. These parent ions can also be quenched by collision with residual gases. The velocity distributions of desorbed oligmers were also measured, and these results provide very important information of constructing a high-resolution TOF mass spectrometer for fast DNA sequencing.

In addition to matrix-assisted laser desorption/ionization, substrate-assisted laser desorption was also pursued. However, only small oligomers were successfully detected. Post ionization of laser desorbed oligomers have also been pursued. From experimental results, it seems the possibility is high to achieve fast DNA sequencing by mass spectrometry within the next three years, since the speed for mass spectrometry for sequencing compared to gel electrophoresis is much faster. Details on the results and the future plans will be presented.

Research sponsored by the Office of Health and Environmental Research,
U. S. Department of Energy under contract DE-AC05-84OR21400
with Martin Marietta Energy Systems, Inc.

Development of a Real-time, Direct Scanner-reader for DNA Sequencing Films

J. B. Davidson
Oak Ridge National Laboratory*
P.O. Box 2008
Oak Ridge, Tennessee 37831

A scanner and reader for sequencing films is being developed. The goal is to use the original film directly without the need to transfer the image to a computer for analysis. It is based in part on the principle of "same-side" densitometry as exemplified in the film viewer, the "Undimmer", which we demonstrated two years ago.

In the prototype to be described, four novel source-detectors are spaced on the scanning arm of a flat bed recorder so as to span four lanes of a sequencing film. As the four lanes are scanned simultaneously, signal conditioning and logic circuits determine the base-call and provide the sequence directly to a printer or computer.

Progress to date will be presented.

---

*Managed by Martin Marietta Energy Systems, Inc., for the U.S. Department of Energy under Contract DE-AC05-84OR21400

Modification of Electrophoresis Conditions for Optimizing Analysis of Stable Isotope Labeled DNA. Mitchel J. Doktycz[1], Johanna L. Doyle[1], William A. Gibson[2], Heinrich F. Arlinghaus[2], Robert C. Allen[3], and K. Bruce Jacobson[1]. [1]Biology Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-8077, [2]Atom Sciences, 114 Ridgeway Center, Oak Ridge, TN 37830, [3]Medical University of South Carolina, Charleston, South Carolina, 29425.

The SIRIS/LARIS (sputter-initiated and laser-atomization resonance ionization spectroscopy (RIS)) procedure will allow high speed, high resolution analysis of stable isotope-labeled DNA after electrophoresis (See report by Jacobson et al.). To take full advantage of these benefits, a suitable gel electrophoresis system must be developed. To this end we have been investigating ultra-thin, open-faced gel electrophoresis and discontinuous buffer systems. These gels are supported by a plastic backing and can be cast in virtually any thickness. The compatibility of this gel system with RIS was assessed by analysis of Sn-labeled DNA primers and PCR products electrophoresed at varying concentrations. The effect of gel thickness on SIRIS/LARIS detection of Sn-labeled DNAs will be presented. A distinct advantage of this gel system is the ease of incorporation of buffer systems and additives. Several discontinuous buffer systems which enhance the resolution and alter the mobility of PCR or radioisotopically labeled Sanger sequencing products over the conventional TBE buffer system have been assessed. Polyhydroxy alcohols and sugars modify the mobility of DNA when used with discontinuous buffer systems employing borate as a trailing ion. A portion of the effect of the polyhydroxy compounds can be explained by formation of a borate complex which alters the conductivity of the trailing ion. Electrophoresis gels from less than 30 μm to 360 μm have been used. The 30 μm (and thinner) gels were prepared by spin coating on a glass surface in a humid, anaerobic chamber. Simultaneous analysis of multiple, enriched tin isotopes attached to DNA and electrophoresed on these gels will be presented.

(Research sponsored by the OHER, U.S. DOE, under contract DEAC-05-84OR-21400 with the Martin Marietta Energy Systems, Inc. by U.S. DOE under contracts DE-FG05-91ER81235 and DE-FG05-90ER81048 to Atom Sciences Inc., and by NIH under contract 1R43CA54627-01 to Atom Sciences, Inc.)

Oligonucleotide Arrays for DNA Analysis

R. S. Foote, J. B. Davidson[1], R. J. Mural, R. A. Sachleben[2], K.-P. Stengele, and E. C. Uberbacher[3]
Biology Division, [1]Instrumentation and Controls Division, [2]Chemistry Division, and [3]Engineering Physics and Mathematics Division
Oak Ridge National Laboratory, Oak Ridge, TN 37831

Solid-state arrays of oligonucleotides promise to provide a rapid means of mapping, sequencing and characterizing DNA by sequence-specific hybridization to target molecules. A photolithographic method of synthesizing such arrays has been developed at ORNL. Eight-mer oligonucleotides could be synthesized with a high degree of efficiency and fidelity by this method. Prototype arrays are being used to optimize parameters for mismatch-free hybridization and for detection of hybridized sites. In concert with these basic studies, arrays of selected sequences are being designed for specific diagnostic applications. One such application currently under study is the discrimination of coding and non-coding DNA based on the frequency of specific short sequences. Such arrays would allow rapid screening of genomic DNA for exon content. Probe sequences with similar GC content (and uniform hybrid T_m's) may be selected from a large number of oligonucleotides having a high bias for coding DNA. Our experience in the preparation and hybridization of small-scale arrays supports the feasibility of preparing large compact arrays of 8-mers or longer sequences for implementation of the Sequencing by Hybridization (SBH) concept, pattern recognition of genes, analysis of mutations, etc. The photolithographic approach offers a means of achieving array densities of greater than 10^6 probe sites per cm^2.

(Research sponsored by the Office of Health and Environmental Research, United States Department of Energy, under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.)

MATRIX-ASSISTED LASER DESORPTION MASS SPECTROMETRY FOR THE STRUCTURAL CHARACTERIZATION OF OLIGONUCLEOTIDES*

Robert L. Hettich, Greg B. Hurst and Michelle V. Buchanan
Analytical Chemistry Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831-6120

Techniques based upon matrix-assisted laser desorption (MALDI) are being developed for the characterization of normal and modified nucleosides, nucleotides and oligonucleotides. Fourier transform mass spectrometry (FTMS) is a major tool in this study, which allows detailed structural information to be obtained on normal and modified oligonucleotides at the low picomole level. The FTMS permits a wide range of ion manipulation processes to be employed not only to determine the complete sequences of small oligonucleotides, but also to identify any adducts which may be attached to these biomolecules. The MALDI-FTMS technique has been successfully used to characterize a variety of nucleic acid constituents which are substituted with alkyl and polycyclic aromatic hydrocarbons. Collision induced dissociation (MS/MS) and selective ion-molecule reactions which can be performed on ions trapped within the FTMS cell have been developed to identify the adduct, determine the site of substitution, and establish the site of modification in the oligomer sequence. Similar techniques have been used to identify modifications arising from UV-induced damage to small oligonucleotides. A time-of-flight (TOF) instrument has recently been constructed to aid in studying modified oligonucleotides with MALDI. This instrument will allow a direct comparison of MALDI on the FTMS and TOF instrument. A variety of other experiments are also being conducted to probe the fundamental characteristics of the MALDI experiment, to allow the conditions which provide soft ionization for large biomolecules to be better defined.

* Research sponsored by the Office of Health and Environmental Research, U.S. Department of Energy under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.

"The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-84OR21400. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes."

Analysis of Iron, Tin and Lanthanide Labeled DNA by Resonance Ionization Spectroscopy (RIS) for Genome Mapping and Sequencing

K. B. Jacobson, R. A. Sachleben, G. M. Brown, F. V. Sloop , M. L. Garrity, M. J. Doktycz,
H. F. Arlinghaus*, R. S. Foote , F. W. Larimer, R. P. Woychik, and N. Thonnard*
Oak Ridge Natational Laboratory and *Atom Sciences, Inc., Oak Ridge, TN.

Multiple stable isotopes offer promise to provide a basis for very rapid rates of DNA analysis after gel electrophoresis because of 1) the multiplexing possibilities and 2) the rapid repetition rate of the RIS instrument. Mapping and sequencing applications are being considered. Progress is reported for the three main areas of this project: chemistry (see also the report of G. M. Brown et al.), comparison of two modes of RIS (report of H. F. Arlinghaus et al.), and different modes of gel electrophoresis (report of M. J. Doktycz et al.). Three new DNA labels have been synthesized so that enriched isotopes of iron, tin and the lanthanides (rare earths) can be incorporated. For incorporating the four isotopes of iron, ferrocene was synthesized; for incorporating the ten isotopes of tin triethylstannylpropanoic acid was synthesized; and for accommodating the isotopes of any of the rare earths 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA) has been synthesized in a multi-step process and further modifications are being made by G. M. Brown et al. The analysis of all of these isotopic labels is sensitively and selectively accomplished by either of two forms of RIS, sputter initiated RIS and laser atomization RIS. Multiple bands of oligonucelotides labeled with different tin isotopes have been successfully identified after gel electrophoresis and the signal is directly correlated to the amount of the oligonucleotide applied to the gel. A set of bands from a sequencing gel have also been identified. Certain limitations of the analysis occur far above the sensitivity limits of either form of RIS and these are associated with chemical contamination and with the matrix of the polyacrylamide gel. Reduction of these background levels may be accomplished by several strategies that will be discussed. Continuing evolution of laser technology suggests ever faster analytical possibilities. Improvements in the rates of preparation and electrophoresis of DNA will be needed to accomplish such rates. Modification of polyacrylamide gel electrophoresis has been developed by making gels of any thickness from 360 μm to < 30μm and using discontinuous buffer systems and mobility modifiers to resolve differentially DNAs of sizes from 50 to 5000 nucleotides. The separation of PCR products of defined sizes is currently being studied to define the resolution limits of these gels and to develop gels specifically adapted to analysis by RIS.

(Research sponsored by the OHER of the US DOE under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc. and contracts DE-AC05-89ER80735, DE-FG05-91ER81235, and DE-FG05- 90ER81048 with Atom Sciences, Inc. and contract 1R43CA54627-01 from NIH with Atom Sciences, Inc.)

Human Genome Management Information System

Betty K. Mansfield, Anne E. Adamson, Denise K. Casey, K. Alicia Davidson, Rose T. Haas, Sheryl A. Martin, Elizabeth T. Owens, Donna B. Stinnett, **John S. Wassom**, Judy M. Wyrick, and Laura N. Yust
Biomedical and Environmental Information Analysis Section, Health and Safety Research Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6050
615/576-6669, Fax 615/574-9888, Internet: *bkq@ornl.gov*, BITNET: *bkq@ornlstc*

The Human Genome Management Information System (HGMIS) provides communication and information services to the DOE Office of Health and Environmental Research (OHER) Human Genome Program Task Group. HGMIS is charged with (1) helping to communicate genome-related matters and research to contractors, grantees, and other publications and (2) providing a forum for information exchange among investigators in the Human Genome Project.

To fulfill these communication goals, HGMIS produces the bimonthly newsletter *Human Genome News* (cosponsored by OHER and the NIH National Center for Human Genome Research), DOE Human Genome Program reports, contractor-grantee workshop and other technical reports, a traveling exhibit on the DOE genome program, and a textual database of genome-related material. In addition to disseminating information to individual requestors and referring them to other sources, HGMIS supplies support on demand to project leaders in preparing meeting minutes, conducting information searches, writing, and editing.

Human Genome News also serves as a primary Human Genome Project source for discipline-specific publications that extract or reprint information from the newsletter. Some of these secondary distributors are *Bioinformatics;* the ethics journal *Eubios;* and the newsletters of genome centers, biotechnology companies *(BT Catalyst)*, chromosome-specific support groups (*The Chromosome 18 Communique'*), high school biology teachers (*The Genetic Messenger)*, Student Pugwash USA (*Tough Questions*), and the National Society of Genetic Counselors (*Perspectives in Genetic Counseling*).

Human Genome News includes technical and general interest articles, meeting reports, national and international project news, features on informatics and resources for facilitating research, genome event and training calendars, and grant and fellowship announcements. The newsletter mailing list has grown steadily from 800 subscribers in April 1989 to 8132 in May 1992, an average of 600 additions each quarter. Later in 1992, the list expanded 63% with the incorporation of 5156 Genome Data Base (GDB) users. Subscribers now number over 13,000 and include genome and basic researchers at national laboratories, universities, and other research institutions; professors and teachers; industry representatives; legal personnel; ethicists; students; genetic counselors; physicians; science writers; and other interested individuals.

Through the newsletter, readers can ask for data, services, and other publications. By 1992 over half the subscribers had requested some type of information or document, including program and workshop reports and the DOE-NIH 5-year plan. In addition, numerous copies of full or partial documents have been distributed for educational purposes. *The Primer on Molecular Genetics*, originally an appendix to the program reports and now published separately, has proved extremely popular as a resource for teachers, genetic counselors, and educational organizations and as handouts for genome centers. Over 4000 copies have been distributed at public lectures, high school biology teacher meetings, genetic counselor workshops, and in secondary and college biology classrooms.

A HGMIS-initiated text database, designed to support DOE managers and investigators in the Human Genome Project, allows e-mail communication among users and makes *Human Genome News*, bibliographic and research abstracts, and other documents and compilations available electronically. Implemented in Information Dimensions, Inc., BASISplus, the database resides on a VAX 3500/3100 cluster and is accessible via modem and Internet.

Displaying the DOE Human Genome Project traveling exhibit at major scientific conferences and genome-related meetings allows HGMIS staff to exchange ideas with investigators and others interested in the project.

HGMIS invites comments and suggestions about its documents and services, which are available upon request and without charge.

This work is sponsored by the Office of Health and Environmental Research, U.S. Department of Energy, under contract No. DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.

Informatics Support for Mapping in Mouse-Human Homology Regions

Sergey Petrov[1], Manesh Shah[1], Lisa Stubbs[2], Richard Mural[2], and Edward Uberbacher[1]
Engineering Physics and Mathematics Division[1] and Biology Division[2], Oak Ridge National Laboratory, Oak Ridge, TN 37831-6364 615/574-6134,
Fax 615/574-7860, Internet:"UBER@ubersun.epm.ornl.gov".

We have addressed several immediate priorities of the Mouse-Human Mapping Project, including the construction of a mapping database for the project, tools for management and archiving of cDNAs and other probes used in the laboratory, and analysis tools for mapping, inter-specific backcross, and other needs. The ORNL mouse-human mapping effort is in its initial start-up phase and our initial effort has involved purchasing, installing, and debugging the current SYBASE versions, and in assembling the necessary hardware. To maximize efficiency with a small effort and reduce start-up time, we have borrowed existing approaches, standards, and technologies where possible from other DOE and NIH centers.

An initial relational database has been constructed with SYBASE (because SYBASE has become the standard at DOE centers), and using a database schema based on the one implemented at the LLNL center. We chose this approach because of the available documentation for the LLNL system and to maximize compatibility with the human chromosome 19 mapping (major homologies exist between human chromosome 19 and mouse chromosome 7 - the initial focus of the ORNL work). 36 tables are incorporated into the current database, corresponding to the central core of the LLNL system. This framework provides a mechanism for handling the most basic types of information: (1) Laboratory information: Inventory and tracking of samples and probes, experimental and analysis results, etc. (2) Mapping Information: Physical and genetic map information, loci, genes, synteny to human chromosomes, etc. (3) Ancillary information: comments, references, etc.

The description for the logical schema of the system is stored in a "metabase" also implemented in SYBASE. This metabase has been constructed to facilitate schema modifications through use of an interface for creating, modifying, and displaying the logical schema. This interface enforces consistency of the schema and includes a X-based graphical browser which allows visualization of the database tables, their attributes, and interrelationships. This browser will be extended to view actual data entries in the database.

User access to the system is being provided by forms-based and graphical interfaces. Access to the system is provided from either workstations or Macintosh (using the Exodus or Mac-X X-window servers on Macintosh). This provides relatively uniform-looking graphics for users for both database and analysis tools (and other applications such as the "Encyclopedia of the Mouse"), while simplifying development to a single standard environment (UNIX workstation). Specific tools for archiving and tracking of cDNAs and other mapping probes, the analysis of inter-specific backcross data, and restriction mapping of YACs have been implemented.

We would like to acknowledge use of ideas from the LLNL and LBL Human Genome Centers, and database design ideas from Tom Marr of the Cold Spring Harbor Laboratory. (Research sponsored by the Office of Health and Environmental Research, U.S. Department of Energy, under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.)

GENETIC AND PHYSICAL MAPPING WITHIN MOUSE CHROMOSOME 7: Foundations of a chromosome-wide comparative physical map, and detailed analysis of selected regions with special biological interest

Lisa Stubbs, Cymbeline Culiat, Estela Generoso, Dabney Johnson,
and Eugene M. Rinchik
Biology Division, Oak Ridge National Laboratory
PO Box 2009, Oak Ridge, TN. 37831-8077

Over the past year, we have concentrated our efforts on developing tools and resources required for the generation of comparative genetic and physical maps throughout mouse chromosome 7 (MMU7). As part of this effort, we have focused upon collecting conserved human markers derived from human chromosomal regions such as 19q13, 11p15, 11q13, 15q11-13, and 15q23-25, that have known homologies to MMU7. A large number of human markers from each of these chromosomal regions has already been mapped in mouse; these collected markers form the foundation of a system within which new, unmapped sequences may be localized rapidly and with a high degree of precision.
Resources now well at hand include two large *Mus spretus- Mus musculus* interspecies backcrosses, including one that is already typed for a large number of MMU7 markers. This typed backcross allows immediate fine-structure genetic mapping of homologs of any human locus which does map to chromosome 7. A second large backcross is currently being typed for each of the 20 mouse chromosomes so that markers from known or suspected human homology segments that do not map to MMU7 can be included in the genome-wide comparative map.

The detailed comparative genetic map of MMU7, which includes our data as well as that generated by other groups, is intended to serve as a framework upon which a series of detailed physical maps may be constructed. Several regions of the chromosome, however, are already well-enough marked to allow physical map construction to begin. As a first step toward chromosome-wide physical map construction, we have chosen to focus upon moderately-sized regions of MMU7 (1-5 cM) with well-defined human homologies and special biological interest. Our detailed maps include a 5 cM region surrounding the murine *p* locus which is homologous to the Prader-Willi/Angelman Syndrome region of HSA 15q11-q13, and three scattered regions with mixed homologies to HSA 11p15 and 11q13-14. We will discuss our progress toward construction of detailed comparative maps within these regions, and genetic mapping of homologs of conserved human markers throughout the chromosome. We will also outline our strategies to use the physical and related genetic and mutation maps of these regions to predict functional units within corresponding human regions. [Research sponsored by the Office of Health and Environmental Research, U.S. Department of Energy under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.]

GENE RECOGNITION AND ASSEMBLY IN THE **GRAIL** SYSTEM

Edward Uberbacher[1], J. Ralph Einstein[1], Xiaojun Guan[1], Donna Buley[2], and Richard J. Mural[2].
[1]Engineering Physics and Mathematics, and [2]Biology Divisions, Oak Ridge National Laboratory,
Oak Ridge, TN 37831-6364. e-mail:GRAILMAIL@ornl.gov

GRAIL is a modular expert system being constructed to analyze and characterize the genetic structure of DNA sequences. Sufficient components of **GRAIL** have been constructed to allow the evaluation of the basic methodology for the recognition and modeling of genes, and to provide the basis for an e-mail server system for coding region localization.

GRAIL E-mail Server and Feature Recognition: **GRAIL** provides an on-line e-mail service for locating the protein coding regions of DNA sequences. This interface utilizes a multiple sensor-neural network (Uberbacher and Mural, 1991, PNAS 88:11261-11265) to find coding regions and a rule based interpreter to reduce this output to a table. **GRAIL** can analyze up to 100 kbp of sequence at a time and several sequences may be included in a single e-mail message. The analysis is performed on both strands of the input sequence and the interpreter assigns the putative coding region to a preferred strand and reading frame. In our experience the strand and reading-frame assignments are accurate about 95% of the time. GRAIL finds 90% of coding exons over 100 bases in length with a very low false positive rate. The **GRAIL** e-mail server includes an option for translating the preferred reading frame for each potential protein coding region and having the predicted amino acid sequence searched against the SwissProt Data bank on an Intel iPSC/860 parallel computer. The results of these searches are returned with the GRAIL output.

The **GRAIL** e-mail server has been in operation for over a year, currently has about 500 users, and processes 2 megabases of sequence per month. User feedback has been a significant source of information for ways to improve performance. To improve the performance of the system for certain classes of proteins which were not well recognized (such as Zinc finger genes), larger and more comprehensive training sets have been constructed which include additional examples of poorly recognized gene classes. Several new sensors for the system have been designed and evaluated, including statistical algorithms based on "complexity", local and regional composition or isochore indicators, and high-order non-stationary Markov models. These have contributed to the performance of the system. In collaboration with Jim Fickett at LANL, we will be testing several new input algorithms which were suggested by his recent analysis of protein coding sequences.

Model Organism GRAILs: A number of investigators have expressed an interest in versions of **GRAIL** for model organisms, and we have established a working group to support the construction of additional systems and development of the overall technology for computer-based gene recognition and sequence annotation. Several versions of **GRAIL** specific for model organisms are in development. A version of the coding recognition portion of **GRAIL** for *E. coli* has been created by Jude Shavlik, et al. at the University of Wisconsin, and has been combined with the rule-based portion at ORNL and made available as an e-mail option. This system has excellent performance rivaling the human version. A version of **GRAIL** for *S. pombe* is being constructed with the help of Tom Marr, et al. of Cold Spring Harbor Laboratory, and a version for *C. elegans* is being constructed with Chris Fields of TIGR. Chris Fields is also interested in contributing to the technology for incorporating homology information into the gene assembly process. Preliminary work on versions for *Drosophila* and dicotyledonous plants has been carried out at ORNL.

Gene Modeling and Stand-Alone GRAIL: In addition to the current coding region recognition capabilities based on a multiple sensor-neural network and rule base, modules for the recognition of features such as splice junctions, transcription and translation start and stop, and control regions are being constructed and incorporated into an expert system for reliable computer-based assembly of model genes. A gene assembly program (GAP) is being developed which combines the outputs from the various feature recognition modules and attempts to predict the sequence of the spliced mRNA from the genomic DNA sequence. If error conditions are detected by the GAP rule-base, then the program attempts to correct them by the insertion and/or deletion of one or more coding exons. These actions result in a net improvement in gene-model prediction, particularly in the recognition and characterization of very short coding regions.

Preliminary results for gene assembly using the full **GRAIL** system are encouraging. The current **GRAIL** recognition tools and assembly system are very successful in determining the correct strand for a gene, and incorporate, on average, about 90% of the correct message in the best predicted solution. Only 7% false sequence is included in the predicted message, and 81% of the correct splice junctions are found. Logic to handle false positive or unrecognized exons is being tested.

Although these results represent significant progress toward the computer-based assembly of human genes, a number of additional challenges remain before anonymous regions containing multiple genes on both DNA strands can be accurately described from sequence analysis alone. The potential to analyze sequences with multiple genes poses new problems which will require improved ability to accurately assign recognized exons to either the forward or reverse strand (currently this capability stands at about 90%-95% correct), and distinguish parts of one gene from parts of another.

A stand-alone version of GRAIL which provides graphic output, is being developed in an X-windows environment on a Sun Workstation. Currently we are integrating the gene assembly modules into this system.

(This research was supported by the Office of Health and Environmental Research, United States Department of Energy, under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.)

Genome-wide Insertional Mutagenesis and the Molecular Analysis and Mapping of Disease-related Genes in Humans and Mice

Woychik, R.P., Beatty, B.R., Bultman, S., Michaud, E.J., Moyer, J., Kwon, H., and Klebig, M.L. Biology Division, Oak Ridge National Laboratory, P.O. Box 2009, Oak Ridge, TN 37831-8077

Physical mapping and nucleotide sequencing of genomic regions and cDNAs is an important first step in understanding the function of genes in humans and model organisms. This information by itself, however, is of limited value for establishing the relationship between a gene and a specific disease or to reveal the function of any given gene in the context of the developing organism. For this reason we have been conducting a large-scale insertional mutagenesis effort in transgenic mice in an attempt to markedly increase the number of mutant stocks that have important developmental and disease-related phenotypes. It is our experience that insertional mutations are particularly attractive because they contain a marker sequence that can be used as a probe to directly clone and characterize any associated genes. This program is a genome-wide effort and covers the synteny regions of the mouse genome that are highly homologous to individual sections of the human genome. As just one example of the approach we are taking, we have generated a mutation that gives rise to recessive polycystic kidney disease. The phenotype of this mutation in the mouse has a remarkably similar pathology to the autosomal recessive polycystic kidney disease condition in humans. We have cloned and mapped a gene that is directly associated with the mutant locus in these animals, and utilizing a cross-species hybridization approach, we have also cloned and sequenced a gene from humans that is extremely highly conserved at the nucleotide and protein sequence levels with the mouse. We believe that this approach of identifying genes that are associated with phenotypes in mutant mice will prove to be extremely useful for assigning disease-related and whole organism functions to genes on the human genome. The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-84OR21400. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

PNL

Pacific Northwest Laboratory

GnomeView: Graphical Integration of GDB and GenBank Databases

Richard J. Douthart*
Gregory S. Thomas
JoAnne E. Pelkey

Pacific Northwest Laboratory
Richland, Washington 99352

GnomeView is a tool for exploring information generated by the Human Genome Project. GnomeView provides both graphical and textual styles of data presentation; employs an intuitive window-based graphical query interface; and integrates underlying genome databases in a way that allows users to navigate smoothly across databases and between different levels of information.

The first distributed version of GnomeView utilizes the Genome Data Base (GDB) for chromosome map information and GenBank for sequence map information. Map-objects found by query are mapped to stylized representation or chromosomes initially as color density maps. Numerous lists and information windows are available by simple interaction with the graphical representation. Sequence loci are available from GenBank. A sequence map is defined as a graphical representation of a GenBank sequence locus with descriptive features from the database header automatically mapped to it.

Queries can initially be entered at the chromosome level (GDB) or at the sequence level (GenBank). Smooth transition between databases occurs via database provided cross reference tables. Comparisons of similar entries from the chromosome level and the sequence level provides, amongst other things, indications of database inconsistencies.

*Contact for further information:
e-mail: dick@gnome.pnl.gov
Ph: (509) 375-2653
Fax: (509) 375-6821

Work supported by the U.S. Department of Energy under Contract Number DE-AC06-76RLO 1830

A New Mass Spectrometric Approach to High Speed DNA Sequencing

Richard D. Smith, Charles G. Edmonds, Xueheng Cheng, Steven A. Hofstadler, James E. Bruce, and Brian Winger

Chemical Sciences Department
Pacific Northwest Laboratory
Richland, WA 99352

We are investigating a new approach having the potential for high speed DNA sequencing based upon mass spectrometry. In our approach. large single stranded segments (1-20 Kbase) are transferred to the gas phase by an electrospray process as highly charged molecular ions. The ions are then trapped in an ion cyclotron resonance (ICR) cell in a high magnetic field where a single molecular ion can potentially be trapped, isolated and nondestructively detected with high mass measurement accuracy. The aims of this project are to develop (1) the instrumental methods necessary for such accurate mass determination and (2) physical or chemical methods for sequentially cleaving individual nucleotide bases from one terminus of the DNA segment. An accurate and high speed ICR measurement after each step would therefore, alter determination of the oligonucleotide sequence.

In this presentation, initial results obtained under this new project will be presented. These include the demonstration that synthetic and biological DNA and RNA oligomers across a wide range of molecular mass can be efficiently transferred to the gas phase by the electrospray process. New ICR instrumentation utilizing a 7-tesla superconducting magnet has also been developed, and has demonstrated the capability for obtaining sufficiently high resolution and high mass measurement accuracy for use in our sequencing scheme. These initial results will be summarized and concepts for the required cleavage processes will be discussed.

Human Genome Research at Other Institutions

ESTs as a Physiological Tool for Gene Identification and Gene Expression Analysis

Mark D. Adams, Anthony R. Kerlavage, M. Bento Soares, Ruben Moreno, Chris Fields, and J. Craig Venter

The Institute for Genomic Research, 932 Clopper Road, Gaithersburg, MD, 20878

Analysis of over 8000 cDNA clones from three cDNA libraries has resulted in the identification of over 500 different genes that are expressed in the human brain. The three libraries vary greatly in the percentage of clones that can were sequenced in a protein-coding region — from eighteen percent (hippocampus) to 65% (infant brain). The increased coding content of the infant brain library resulted in a large increase in the number of clones that could be identified by searches against the public databases. One interesting observation was the large number of new members of gene families that was found. About ten percent of the ESTs that matched human genes were non-identical matches, thus defining or adding to a gene family. Proteins from organisms as evolutionarily distant as yeast, *E. coli*, and rice contributed to the putative identification or gene family placement of ESTs. In fact, one quarter of the yeast proteins in PIR matched an EST; about one-third of the *Drosophila* proteins in PIR matched an EST. A wealth of new gene information has recently been reported from genomic and EST projects in yeast, *C. elegans*, and humans. Most of the open reading frames in these sequences are unannotated and thus do not appear in the standard database searches. Several methods were devised to compare human ESTs with each other and with unannotated open reading frames from other sources. Over twenty new gene families were found consisting of non-exact matches of EST-encoded proteins to one another in the absence of other matches to the protein database. Several cross-species matches of unknown open reading frames were also found. *C. elegans*, yeast, and human EST and genomic sequencing projects all result in remarkably similar rates of gene identification: about half of the of protein-coding regions matched a known protein in the database. These matches illustrate the value of pursuing genome projects in many species for providing viable starting points to elucidate the function of newly-described, related genes. Over four hundred ESTs have been localized to chromosomes using PCR screening of somatic-cell hybrid DNA panels or fluorescence *in situ* hybridization. Eventually, these sequences will be used as confirmation of coding region in genomic sequence and to further our understanding of the expressed gene complement of the brain.

A statistical approach to the detection of overlap and alignment of DNA fragments.

G.E. Alexander[1] and T.P. Speed[2]

[1] Department of Mathematics & Statistics, American University, Washington, D.C.
[2] Department of Statistics, University of California, Berkeley.

This study was motivated by a desire to make use of available information concerning the nature and distribution of DNA sequencing errors in the computational process known as fragment assembly.

Data kindly made available to us by T. Hunkapillar allows us to observe mis-called bases and insertions and deletions by comparing sequenced fragments to a final consensus sequence, assumed to be correct. The task then is to develop a simple probability model for sequencing errors based on these observations which permits a statistical approach to the detection of overlap and the alignment of sequenced random fragments of a larger DNA sequence. This model can also be used to reconstruct the larger DNA sequence from the sequenced fragments. Of primary interest will be a comparison of the performance of our statistical approach with one of the standard non-statistical approaches to overlap detection, alignment and consensus determination. We are also interested in comparing statistical approaches to sequence comparison with the more common computer algorithms which compute scores summing penalties for mismatches. Where possible, we will try to find the penalty parameters which correspond to error probabilities in our model.

This work, which is currently in its preliminary stage, builds on and attempts to integrate research of Waterman and Gordon (1990)[a] on the relationship between penalty parameters and the distribution of the score of the best matching segments, Churchill and Waterman (1992)[b] on the accuracy of DNA sequences, and Thorne, Kishino and Felsenstein (1991, 1992)[c] on the likelihood approach to sequence alignment.

[a] In *Computers & DNA*, eds. G.I. Bell and T.G. Marr, Addison-Wesley 1990 pp127-135.

[b] *Genomics* **14** (1992) 89-98.

[c] *J. Mol. Evol.* **33** (1991) 114-124; *ibid* **34** (1992) 3-16.

Development of Larger Scale Systems for the Genome Project

Norman G. Anderson and N. Leigh Anderson
Large Scale Biology Corporation
Rockville, MD

A systems analysis of the Genome Project divides the effort into a series of unit processes, each of which exists at present at a bench-level scale. If further scale up is to be done by reduplication of bench-scale systems, then integration of data and samples from multiple sources will become a major effort. If, in contrast, integrated large scale operations involving larger unit process machines are contemplated, then a new set of problems emerge. While new bench-scale systems may be developed and evaluated in almost any operational HG laboratory, large scale (LS) unit process systems are difficult to develop and test in the absence of a complete set of systems operating at the same scale. It is difficult to foresee how separate LS systems for cloning, picking, hybridization, short and long term storage, sample preparation, primer synthesis, PCR amplification etc., data acquisition and analysis, and for system scheduling can be developed in isolation, one from the rest. To break this impasse, we proposed to chose those elements which might have other uses which would justify the cost of development and also provide an alternative means for technical evaluation. These include cloning, colony or plaque identification by hybridization, sample storage, and oligonucleotide synthesis.

The Clonepick System:- The Clonepick system uses 70 mm sprocketed motion picture film as both a support, transport, and storage medium for cells or vectors. Two porous spacers are attached along the film inboard of the sprockets to give a space 5 cm wide and 3,048 cm (100 feet) long having a total growth area of 15,240 cm^2. Agar adheres well to the gelatin on photographic film, hence we have used commercial film processed to remove dyes and silver in an automatic continuous film processor. We have built machines to automatically slit and attach the plastic foam spacers, and to continuously cast agar media of constant thickness the entire length of the film. The film is co-wound with a strengthened cellulose blotting strip to absorb any moisture which may condense during incubation. Methods under development for maintaining sterility, for scanning the film strips, for picking and transferring clones additional strips, and for printing the strips onto PVDF or nylon filter strips for probing will be described. The alternative use for this system is screening for new antibiotics.

PCOS:- The Production-Scale Centrifugal Oligonucleotide Synthesizer is being developed to allow large single batches of DNA for use as standard probes and primers to be synthesized. Solid phase synthesis is done in a centrifugal filed using a zonal centrifuge rotor with a transparent glass end window and fluid line seals on both the upper and lower shafts to allow reagents in the synthesis cycle to be run in to either the rotor center or the rotor edge. Density differences between reagents, magnified by centrifugal force, keep the reagents separate, and insure ideal flow through the bed. The alternative use for this system is production of antisense oligonucleotides for pharmaceutical studies.

Sample Storage:- A system for storing and distribution large numbers of very small DNA samples on film strips with barcode identification will be described. In the completed system an entire reel may be scanned, indexed, and searched. Samples are punched out so that film integrity is preserved. We believe the system will be useful for storing and distribution large numbers of STS sequences, and identified fragments of genomic DNA. The alternative use of this system is for the storage and distribution of large numbers of human DNA samples from individuals with defined ethnic backgrounds, genetic diseases, or susceptibilities, and from organisms from different species in all phyla for comparative sequence studies.

Resonance Ionization Spectroscopy: A New Analytical Technique for Genome Mapping and Sequencing

H.F. Arlinghaus , M.T. Spaar, and N. Thonnard

Atom Sciences, Inc., 114 Ridgeway Center, Oak Ridge, TN 37830

M.J.Doktycz and K.B. Jacobson

Oak Ridge National Laboratory, Oak Ridge, TN, 37831.

Resonance Ionization Spectroscopy (RIS) is becoming recognized as an analytical technique for a wide range of applications. The extremely high element specificity and sensitivity of the RIS process is especially valuable for ultra-trace element analysis in polymeric materials where the complexity of the matrix is frequently a serious source of interferences. RIS utilizes precisely tuned lasers to ionize a chosen element efficiently with virtually no interference from the immense background of other constituents. RIS, when combined with ion sputtering, Sputter Initiated RIS (SIRIS), or, Laser Atomization RIS (LARIS), leads to an exceptionally efficient analytical technique having the ability to localize, with micrometer spatial resolution and virtually no matrix effects, ultra-trace concentrations of a selected element with ultrahigh sensitivity and selectivity. The ionized atoms are counted in a mass spectrometer. If a time-of-flight mass spectrometer is used, all isotopes of an element can be detected simultaneously. At the Oak Ridge National Laboratory, methods have been developed for synthesizing tin labels for oligonucleotides, with other element labels being developed (see report by Jacobson et al.). At Atom Sciences, we have compared both the SIRIS and LARIS technique to determine their characteristics to localize and quantify Sn-labeled DNA. We will present data showing (a) the differences between SIRIS and LARIS response as a function of atomization parameters, substrate and analyte, (b) the detection of subattomole quantities of Sn-labeled DNA, (c) the detection of positively hybridized and unhybridized sites on a DNA sequencing matrix, (d) the utilization of a mass spectrometer with SIRIS/LARIS to detect multiple, enriched tin isotope-labeled DNA after electrophoresis, and (e) the DNA concentration as a function of depth in polyacrylamide gel and nylon membrane. The prospects of using the SIRIS/LARIS technique to perform much faster DNA sequencing and mapping will be discussed.

This work was supported in part by the United States Department of Energy under contract under contract No. DE-AC05-89ER80735, No. DE-FG05-90ER81015, No. DE-FG05-91ER81235 and No. DEFG05-90ER81048, NIH No. 1R43CA54627-01 to Atom Sciences, Inc., No. DEAC-05-84OR-21400 to Martin Marietta Energy Systems, Inc.

Rapid; High-throughput DNA Sequencing Using Confocal Fluorescence Imaging of Capillary Arrays

J. S. Bashkin, D. Roach, E. Rosengaus, D. L. Barker
Molecular Dynamics, Inc.
880 E. Arques Ave.
Sunnyvale, CA 94086

Abstract: The goals of the Human Genome Project require techniques for DNA sequencing that increase throughput over current methods by an order of magnitude or more. Current automated fluorescence sequencing instruments require 5-10 hours for a single gel run and accommodate a maximum of 36 reaction sets. Capillary gels can be run at higher voltages to yield separation of 300-500 bases in 1-2 hours, but current fluorescence detection methods are limited to analyzing one capillary at a time. Confocal fluorescence imaging has the capacity to produce great sensitivity of detection with the geometrical advantage that excitation light is delivered through the same lens that collects fluorescence emission. R. Mathies and colleagues (Analytical Chemistry (1992) 64, 967) have shown that a single scanning confocal fluorescence detector can detect DNA fragments in a parallel array of capillaries.

Based on Mathies' work, we have built a functional breadboard of a confocal scanning system capable of detecting fluorescence data from a large array of capillaries. We have assembled arrays of acrylamide gel-filled capillaries and built an injection manifold capable of simultaneously loading different sequencing samples into each capillary. We have also made progress on the software necessary for calling the DNA sequence on-line, based on the observed fluorescence signal intensities.

Pathways to Genetic Screening: Patient Knowledges, Patient Practices

Diane Beeson, Ph.D., Robert Yamashita, Ph.D. and Troy Duster, Ph.D.
Institute for the Study of Social Change
2420 Bowditch Street
University of California, Berkeley
Berkeley, CA

Our specific purpose is to clarify the processes by which genetic screening and genetic concepts of health and illness penetrate two contrasting communities and become integrated into the family lives of high-risk individuals. In order to clarify the role of culture in integrating genetic explanations and interventions into lived experience we focus on two autosomal-recessive and potentially fatal disorders, cystic fibrosis and sickle cell disease. Since each is primarily associated with a different ethnic and racial group (CF in European Americans and SS in African Americans) each with differing economic and political resources, significant differences can be expected in social adaptation and meaning systems. We are in the early stages of conducting 800 interviews of individuals in families in which there are known to be affected individuals or carriers, and who by virtue of this, have a personal reason to have considered seeking genetic screening.

Data collection consists primarily of focused interviews to explore constructions of the following topics: direct personal experience with the target disorders; the meaning of the target disorder for the lives of affected individuals and family members; health care and insurance issues; prevention and screening; family communication; communication among friends concerning these disorders; and the nature and character of the penetration of these issues into the larger community.

Our interviews to date suggest the following emerging patterns:

1) Health care and potential loss of health care coverage are major issues in the lives of high-risk families.

2) Women carry the primary burden of communicating information about genetic risks, but are sometimes successful in enlisting male support. Grandmothers are often pivotal figures in communication among extended family members.

3) Genetic explanations for disease coexist with other beliefs about why particular individuals are ill.

4) Contrasting patterns of response to genetic disease seem to be emerging. Among the CF families we have interviewed, there appears to be a greater readiness to discuss their attitudes and experiences in relation to being at risk and to engage or support the engagement of their family members as participants in the research project. At the same time there is greater focus on survival and day-to-day care-giving among the poorer sickle-cell families as well as more acceptance of the disability associated with genetic disease.

5) Concepts of genetic disease become relevant primarily as a result of the birth of a child with a disorder.

6) Confusion in communication is common, as well as taboos and prohibitions on open communication among family members regarding risk and inheritance patterns for both disorders.

Capillary electrophoresis with replaceable linear polyacrylamide for DNA sequencing and restriction analysis.

Authors: Jan Berka, Marie C. Ruiz, Frantisek Foret and Barry L. Karger

Barnett Institute, Northeastern University, Boston, MA.

The solutions of linear (noncrosslinked) polyacrylamide have been used as a sieving medium for the DNA analysis by capillary electrophoresis. The advantage of linear polyacrylamide when compared to crosslinked gels stems from its advantageous physicochemical properties. While it provides separation power comparable to crosslinked gel it still maintains properties of a fluid enabling to replace the separation matrix after each analysis. Thus, the most laborious and time consuming steps of capillary gel electrophoresis connected with the preparation and alignment of the separation column are reduced to a simple reloading of the separation matrix from the stock solution. When used with stable coated capillary, this procedure not only provides excellent reproducibility of the analysis but also enables fast optimization of the separation since the composition of the sieving matrix can be easily changed when reqired. In this work the properties of linear polyacrylamide and new type of stable capillary coating were studied for use in both DNA sequencing and restriction fragment mapping. In the first example, problems assotiated with the short column lifetime, gel instability and DNA sample purification, which remain the main limitations of the use of capillary electrophoresis for DNA sequencing, were bypassed by the use of replaceable linear polyacrylamide. DNA sequencing strategy utilized Sanger dideoxy-termination chemistry, two dye-labeled primers and two peak-height ratios to code for all four bases. Complete DNA sequencing information of at least 300 nucleotides long was routinely obtained in less than 30 minutes with the high reproducibility and low error rate. In the second cxample, separations of DNA restriction digests and synthetic DNA ladders, i.e. double stranded DNA fragments in the size range from 70 to 20,000 bp were achieved in short analysis time (less than 20 minutes) using replaceable linear polyacrylamide matrixes. High sensitivity was accomplished using fluorescently active intercalating dyes.

GENOME COMPOSITION AND HUMAN CHROMOSOME 22: ANALYSIS OF 96 MAPPED FOSMIDS

Bruce W. Birren[1], Yoshiaki Tachi-iri[2], Ung-Jin Kim[1], Hiroaki Shizuya[1], Julie R. Korenberg[3], and Melvin I. Simon[1].
1 Division of Biology 147-75, California Institute of Technology, Pasadena, CA 91125
2 Hamamatsu Photonics K.K., Hamakita Research Park, Japan.
3 Ahmansom Dept. of Pediatrics, Cedars-Sinai Medical Center, Los Angeles CA

We have constructed a Fosmid library from a hybrid cell line containing human chromosome 22 as the only human DNA. The Fosmid vector permits packaging of cosmid sized DNA fragments in a single-copy, F-based vector, for stable maintenance in *E. coli.* We have analyzed in detail 96 unique, randomly picked human chromosome 22 Fosmid clones from this library. All clones were assigned to bins of approximately 8 Mb by fluorescent *in situ* hybridization, and we have ordered some using two-color FISH. For each clone, the frequency of sites for several rare cutting enzymes was determined and revealed a non-random distribution with respect to position on the chromosome. In addition, all clones were tested for the presence of the major families of interspersed repeated sequences, including the different sub-families of Alu, the 5' and 3' ends of L1, THE-1, the CG dinucleotide, and various centromeric repeats. Combined, these clones span chromosome 22 and represent nearly 4 Mb. These Fosmids are now being used to isolate corresponding BAC clones. These data suggest tentative conclusions about the overall structure and evolution of the chromosome.

"**Methods for normalization of cDNA libraries and for isolation of chromosome-specific cDNAs**". M. Fatima Bonaldo[1], M-T Yu[2], S. Brown[3], L. Su[1], A. Efstratiadis[2], D. Warburton[2] & M.Bento Soares[1]
Departments of Psychiatry[1], Genetics & Development[2] and Obstetrics and Gynecology[3], Columbia University.

We have developed a method for normalization of directionally cloned cDNA libraries constructed in phagemid vectors, which involves priming of the library in the form of single-stranded circles with a Not I-oligo (dT) primer and controlled extensions with Klenow in the presence of dNTPs and ddNTPs. After purification of the partial duplexes over HAP, melting and reannealing to a moderate Cot, unhybridized (normalized) single-stranded circles are purified by HAP and electroporated into bacteria, generating a normalized library. We have followed this protocol to normalize a human infant brain cDNA library. The extent of normalization was determined by a series of screenings with probes that represent mRNAs from the 3 frequency classes.

We have also developed a method for selection of chromosome-specific cDNAs and successfully utilized it to identify a number of genes on human chromosome 13. A normalized cDNA library, in the form of single-stranded circles, is hybridized to a filter immobilized chromosome-specific phage library and cDNA/λ pairs are visualized by subsequent hybridization with a vector probe. Both members of a cDNA/λpair are then simultaneously isolated: the λclone is purified by standard procedures, and the cDNA circles are eluted off the filter and electroporated into bacteria. Verification of correct cDNA selection is done by back hybridization with the corresponding λ clone. While the λ clone is used as a probe to position the gene to a cytogenetic band by in situ hybridization, the corresponding cDNA is sequenced from both ends for tentative gene function identification. We are currently optimizing this cDNA selection procedure for the identification of transcribed sequences directly from YAC clones.

An Automated Liquid Handling/Gel Loading System Capable of Pipetting Sample Volumes Less Than One Microliter.

Brumley, R. L., Jr., Buxton, E. C., Boville, B. M. and Smith, L. M.
University of Wisconsin, Department of Chemistry, Madison, WI 53706

We have recently developed an automated DNA sequencing instrument based on Horizontal Ultrathin Gel Electrophoresis(1). An integral component of this automated sequencing system is the development of a robotic arm for delivering large numbers of samples quickly and accurately to the electrophoresis cell. Since the HUGE/CCD instrument does not require most of the sample volume produced by conventional sequencing reactions, we have designed a system capable of delivering samples as small as 150 nanoliters. Preliminary data show that this automated system can pipette samples in the range of 0.15-1.0 microliters with less than 10.0% error. In addition, the linear positioning tables that are used will travel at speeds up to 10 in/sec with accuracies of $\pm 7.4 \times 10^{-8}$in and repeatabilities of $\pm 6.2 \times 10^{-9}$in. Since the design is composed of interchangeable modules, the working envelope can be easily modified by substituting different sizes of linear positioners. Thus, the robot design is flexible and amenable to future expansion or modification.

1. Kostichka, A. J., Marchbanks, M. L., Brumley, R. L., Jr., Drossman, H. and Smith, L. M. (1992) Bio/Tech 10, pp.78-81.

CONSTRUCTION AND USE OF HIGH RESOLUTION CYTOGENETIC AND GENETIC MAPS AND HUMAN CHROMOSOME 16

D.F. Callen[1], N.A. Doggett[2], R.L. Stallings[2,3], C.E. Hildebrand[2], J.C. Mulley[1], R.I. Richards[1], G.R. Sutherland[1]

[1]Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital, North Adelaide 5006, Australia; [2]Life Sciences Division, Los Alamos National Laboratory, Los Alamos NM 87545, U.S.A.; [3]Department of Human Genetics, University of Pittsburgh, Pittsburgh PA 15261, U.S.A.

A cytogenetic-based physical map of human chromosome 16 allows the euchromatin of this chromosome to be divided into 59 intervals. This corresponds to an average resolution of less than 1.5 mb. The intervals of the physical map are defined by constitutional and culture-induced chromosome 16 breakpoints isolated in human/mouse hybrid cells and by the use of fragile sites. Mapped DNA markers include STSs and cosmids from cosmid contigs, polymorphic probes - both classical RFLPs and microsatellite repeats, genes - many as STSs, and other anonymous DNA segments. This physical map has been integrated with a detailed genetic map containing classical RFLPs and highly informative microsatellite markers.

Several procedures for the specific recognition of cDNA sequences from chromosome 16 have been evaluated but found to be unsuitable for large scale use. This aspect is undergoing further development.

The following strategy is now being used to clone DNA in the region of disease genes, fragile sites and cancer breakpoints on chromosome 16. For a disease gene the first step involves selection of polymorphic probes for linkage analysis based on the cytogenetic-based physical map. The resolution obtained in defining a region depends on the extent of family material. For fragile sites and cancer breakpoints the region can be defined by FISH. Once an interval is determined, genes mapping to this region are potential candidates. Cloned DNA in the region can be rapidly obtained from cosmid contigs already mapped to the region or by use of other mapped DNA markers to isolate contigs. YACs isolated from the CEPH or Los Alamos library can be rapidly converted to cosmids by use of high density gridded cosmid assays. Examples of these approaches will be presented for cloning the gene for juvenile Batten disease (in collaboration with Dr. R.M. Gardiner, London), the breakpoint of the inv(16) in acute myelomonocytic leukaemia (in collaboration with Dr. C.L. Willman, Albuquerque), and familial Mediterranean fever (FMF) (in collaboration with Dr. D.L. Kastner, NIH).

National Study Conference on Genetics, Religion, and Ethics

C. Thomas Caskey, J. Robert Nelson,[†*] and Hessel Bouma III*
Baylor College of Medicine and the *Institute of Religion, Texas Medical Center, Houston, TX 77225
[†]713/797-0600, Fax 713/797-9199

The objective of this project is the elucidation of four main issues raised by data produced in the Human Genome Project and related genetic research and applied technology. These issues are detailed below.

1. How does emerging genome information challenge traditional religious and philosophical concepts of human nature, behavior, moral freedom, and moral responsibility? How are the doctrines of creation, human life, health, and purpose of living treated similarly or differently in various religious communities? How do religious insights contribute to the understanding by researchers and medical practitioners of the value and inviolability of human life? In light of these views, how should we regard the current practice of positive eugenics and the ideology of eugenics? What principles govern decisions concerning the genetic modification of human beings?

2. What philosophical and theological concepts are given to justify the protecting of personal privacy, patient autonomy, and family integrity against disclosures in respect to individuals who submit to genetic screening, testing, and counseling?

3. What additional challenges might the increasingly available genetic information present in the inevitable reproductive choices that arise from preconception testing and prenatal diagnosis? What aids to moral understanding and guidance are forthcoming and agreeable to counselor and client?

4. What suggestions or guidelines may be helpful to physicians, genetic counselors, and pastoral counselors in educating patients and the public in the meaning, problems, and promises of molecular genetics? What factors constitute the full range of information needed for effective counseling, treatment, and care in varying kinds of cases?

The distinctive character of this project results from the inclusion of theologians, ethicists, and representative leaders of religious bodies, along with molecular genetics researchers, clinical physicians, and social policymakers and students. Its expected value will be to enable people of diverse disciplines, professions, viewpoints, and religious beliefs to discover common, practical ways of understanding certain implications of Human Genome Project data.

The aim of this project is to reveal theological and ethical insights, both traditional and evolving, that may be of practical relevance to the findings and medical applications of genetic science. Scientists will present emerging human genome data that may in turn indicate the need for modification of traditional beliefs about human life and behavior. Cultural and environmental influences will also be considered in analyzing the causes of aberrant behavior and disease.

MOUSE GENOME MAPPING IN AN INTERSPECIFIC CROSS USING RESTRICTION LANDMARK GENOMIC SCANNING (RLGS)

V.M. Chapman[1], S. Hirotusne[2], Y. Okazaki[2], I. Hatada[2], T. Mukai[2], J. Kawai[3], T. Hirasawa[3], Y. Nishitani[3], S. Watanabe[3], T. Shiroishi[4], K.Moriwaki[4], Y.Matsuda[5], K.Manly[1], R.Elliott[1], and Y.Hayashizaki[6]
[1]Molecular and Cellular Biology Dept., Roswell Park Cancer Institute, Elm and Carlton Sts., Buffalo, NY 14263; [2]Dept. of Bioscience, Natl. Cardiovascular Center Res. Inst., 5-7-1 Fujishirodai Suita Osaka 565, Japan; [3]Shionogi Research Laboratories, Shionogi and Co., Ltd., Fukushima-ku, Osaka 553, Japan; [4]Dept. of Cell Genetics, Natl. Inst. of Genetics, Mishima Shizuokaken 411, Japan; [5]Div. of Genetics, Natl. Inst. of Radiological Sciences, Chiba-shi 260, Japan; [6]Tsukuba Life Science Center, RIKEN, Tsukuba Science City, Japan

The RLGS method has been used to identify variation between M. spretus and the laboratory mouse C57BL/6. End-labelled NotI sites were subsequently cleaved with PvuII and PstI for these analyses. Seventy two progeny from an interspecific backcross between (C57BL/6 x M. spretus)F_1 females mated with M. spretus (BSS) were used to analyze the variant fragments. These backcrosses were karyotypically analyzed for the segregation of the centromeres for each chromosome which was used as the primary anchor marker for each chromosome. More than 380 loci were identified. The segregation and linkage relationships of these loci relative to the centromere of each chromosome were established. Initial linkage relationships were determined using a 2x2 analysis at the 0.01% level of significance. Map orders of loci within chromosomes were established with the aid of the Map Manager program that minimized frequency of recombination and that also tested maximum likelihood estimates of gene order for groups of three loci. We have used the recently developed method of primer extension preamplification (PEP) with random 15-mers to expand the available genomic DNA from these backcross resources to analyze simple sequence repeat (SSR) loci. We plan to add 240 of the SSR loci to the genetic maps that are derived from the backcrosses analyzed for RLGS loci. The added SSR loci will integrate the RLGS loci into the mouse genetic maps that are also characterized for functional genes. An additional 37 backcross progeny of a separate series (C3H x M. spretus)F_1 x C3H (CSC) were analyzed for M. spretus-specific loci using NotI, followed by a six-base and a four-base recognition enzyme. The latter system produced a larger number of fragments than were observed in the NotI-PvuII-PstI protocol but only 300 loci could be reliably identified. Linkage relationships of these loci with each other and separately analyzed anchor loci were characterized in a manner similar to the BSS cross. The RLGS Not I landmarks are characterized by specific restriction fragment sizes that may be identifiable in genomic DNA clones. The recent development of Not I linking and boundary libraries using a restriction trapper affinity chromatography isolation of Not I sites will enhance the cloning of the Not I genomic landmarks (Hayashizaki et al.,1992 Genomics 14,733). It is possible to use the RLGS methods to screen the Not I linking libraries by multidimensional sampling of arrayed clone libraries. Thus, there are substantial prospects for using RLGS to make significant expansions of the mouse genetic map and for integrating these maps with other markers in these maps and for applying these methods to the physical mapping of the mouse genome.

A Physical Mapping Database for the Macintosh: the GENOME NOTEBOOK

Stephen P. Clark and Glen A. Evans
Molecular Genetics Lab, The Salk Institute for Biological Studies,
La Jolla, California

An important issue in the genome mapping project is the collection and analysis of experimental results of all members of the research group. It is useful for each researcher to be able to review the results of the other members of the team and to be able to get an up-to-date summary or overview so that the overall status of the project can be determined at any time. To accomplish this we have developed a database program that runs on Apple Macintosh computers, called the GENOME NOTEBOOK, using the relational database system 4th Dimension. We chose this route because of the tools 4th Dimension provides for implementing a standard Macintosh interface with which lab workers can quickly become proficient. The database is maintained on a file-server so that it can be accessed simultaneously from all the Macintoshes in the lab. Every detail of the mapping project is recorded, including the size of a clone, what it overlaps with, what sequences were generated from it, the results of sequence analysis, PCR primers made to the sequences and conditions for efficient PCR, species cross-hybridization with a PCR primer-pair, and, of course, chromosomal mapping position. Because it can record the results from any number of experiments, and extrapolate results from some experiments, it is very easy to assess the quality of the data and conflicting results can be easily identified.

Our main concern is that the data can be entered and accessed quickly, easily and accurately. Three approaches have been used towards this goal: Exploitation of the normal Macintosh interface, database programming and importation of data from files that have been created by other programs. For reviewing the data, the GENOME NOTEBOOK takes advantage of the very powerful and easy-to-use tools for searching and sorting records between related tables that 4th Dimension provides when generating reports. Three kinds of reports are available: lists, graphical displays and ascii files. Lists are available for all the database tables and often show values from related tables such as map position. Lists can also be printed for a permanent record. The ascii file reports can have one of three functions - a more compact and flexible summary than can be generated as a list; a report that can be imported into a word-processor or spreadsheet; or for exporting data to the GDB or for colleagues at remote sites with their own copies of the GENOME NOTEBOOK. The graphical report displays the selected loci next to a chromosomal ideogram and information for each locus can be displayed on the screen by double-clicking its name. The drawing can be refined by hand if necessary using the built-in drawing tools and the output is suitable for publication when printing on a laser printer.

The GENOME NOTEBOOK is being used to keep track of experimental results for the human chromosome 11 and *Giardia lamblia* genome projects in our lab. By changing two configuration values, the database can be used for other chromosomes or species.

HUMAN GENOME PROJECT GENETICS EDUCATION WORKSHOPS FOR MIDDLE AND SECONDARY SCIENCE TEACHERS, Collins, Debra L.[1], Segebrecht, Linda[2], and Schimke, R. Neil[1].

[1] University of Kansas Medical Center, 3901 Rainbow Blvd., 4023 Wescoe, Kansas City, KS 66160-7318
[2] Science Pioneers, 425 Volker, Kansas City, MO 64111

Public awareness of the Human Genome Project - especially the legal, ethical, social and technical, implications - will be increased through summer education workshops for teachers in Kansas City. Middle and secondary teachers (150) from throughout the United States will participate.

Each participant will complete four parts over 1-2 years:

I	first week-long workshop (summer) - speakers and hands-on laboratories
II	utilization of information and materials with students (during school year)
III	second week-long workshop (summer) - review experiences, updates, prepare for peer teaching
IV	peer teaching with colleagues (summer or school year)

Topics will be taught using hands-on labs, interactive talks, and panels of individuals with genetic conditions. Areas covered include ethical, legal and social implications of the HGP, basic human genetics, The Human Genome Project, biotechnology [electrophoresis, polymerase chain reaction, restriction fragment length polymorphisms, DNA sequencing], genetic counseling process, population screening, linkage analysis, alternative non-Mendelian inheritance mechanisms, careers in human genetics, science fair projects in genetics, and genetic resources.

Teaching materials include new DOE/BSCS curriculum "Mapping and Sequencing the Human Genome: Science, Ethics and Public Policy".

Area biology teachers, Midwest Bioethics Center, consumer representative, Science Pioneers, and the University of Kansas Medical Center are planning the workshops.

For teacher applications and Genetic Resource Professional applications contact:

Genetics Education
University of Kansas Medical Center
3901 Rainbow Blvd., 4023 Wescoe
Kansas City, KS 66160-7318

Phone (913) 588-6043
FAX (913) 588-3995

An Interactive Object Oriented System for Retrieval, Manipulation and Analysis of Genomic Data.

Steven Cozza, Elizabeth Cuddihy, Jacqueline Salit, E. Corprew Reed, William Chang and Thomas Marr, Cold Spring Harbor Laboratory.

There is a growing tendency within the genome community to develop databases centered on specific organisms. Now that molecular characterization of the genomes of human and many model organisms is underway in earnest, it is important to be able to link these databases together in a unified manner. It is important to do so because people studying gene function exploit evolutionary conservation extensively in their investigations. We have developed an object oriented computer system which uses a largely organism-independent underlying data model to accomplish unification of genomic data across eukaryotic organisms. We developed a system architecture that naturally operates in client-server mode, using an object oriented database (GemStone, Servio Corp.) as the server. However, the client portion can readily be used in stand-alone mode. In addition, both the client and server components have the ability to execute *ad hoc* and pre-programmed (via graphical user interface) SQL queries to the relational versions of the Genome Database, GenBank, and the CSHL Fission Yeast database. The system software is programmed using the Samlltalk-80 programming language and runs with the same look-and-feel on most UNIX workstations, IBM PC-type computers running Ms Windows, and Macintosh computers running the Mac OS. Other genomic databases, such as Standford Yeast database, the Jackson Lab Mouse database the MRC ACeDB, the Harvard *Drosophila* database, and the NCBI non-redundant sequence database will be added as they become available and as we have resources to do so. The system is highly interactive using intuitive graphical displays, icons, pull-down menus and drag-and-drop paradigms for user actions, allowing the user to retrieve reference data, create new data using object-based editors, or enhance the annotations of existing data, by making links between independently derived data for example. We are currently extending the system to support sequence similarity searching (on a network based, socket-accessible, non-redundant sequence database whose structure is optimized for searching) and tools to examine conserved elements in groups of functionally related protein sequences.

High Capacity Semi-Automated DNA Sequencing

Ronald W. Davis

The Stanford sequencing group has been conducting a feasibility study to determine the rate, cost and accuracy that can be achieved in an effort to complete the sequence of the entire yeast genome. For this feasibility study we are sequencing 200 kb from the end of chromosome 5. We have taken a semi-automated approach, making use of the Beckman Biomek 1000 with side loader. ABI 800 robots, and the ABI 373 sequencer. We have focused on increasing the capacity of the instruments, decreasing labor time and cost, and decreasing material costs.

In a collaboration with Beckman Instruments we have developed a high throughput DNA template procedure, and can produce approximately 400 sequencing templates per day. We have also developed new software to integrate the Biomek 1000 with the robotic arm. Using this equipment and microtitre plates with filter bottoms, we are testing a template preparation procedure that is more automated and is capable of producing approximately 1,000 templates per day.

In collaboration with ABI we have developed new software that allows the ABI 800 catalyst to work in a 96 well format. All sequencing reactions are carried out in this instrument. Also in collaboration with ABI, we have made modifications to the 373 to improve sequencing. This modified instrument contains a higher output laser that gives improved signals for the T reactions. We have made other modifications to the instrument and are conducting studied to determine if we can increase its capacity.

SINGLE MOLECULE DETECTION USING CHARGE-COUPLED DEVICE ARRAY TECHNOLOGY

M. BONNER DENTON
DEPARTMENT OF CHEMISTRY
UNIVERSITY OF ARIZONA
TUCSON, AZ 85721

COLIN W. EARLE
DEPARTMENT OF CHEMISTRY
UNIVERSITY OF ARIZONA
TUCSON, AZ 85721

Investigations into developing and apply a new technique for the detection of single fluorescent chromophores in a flowing stream will be discussed. These studies are intended to contribute toward improved rapid DNA sequencing schemes currently being developed by Los Alamos National Laboratory. In previous investigations, the detection sensitivity has been limited by the background Raman emission from the water solvent.

By employing a Charge-Coupled Device (CCD) detector operated in a time-delay and integration (TDI) mode we are able to enhance the discrimination between fluorescence from a single molecule and the background Raman scattering from the solvent. In this novel mode of operation, the register shifts between rows on the CCD are synchronized with the sample flow velocity such that the fluorescence from a single molecule is collected in a single moving charge packet occupying an area approaching that of a single pixel while the background is spread evenly among a large number of pixels.

Additionally, this approach should contribute to more rapid sequencing since more than one species can be simultaneously present in the observed flow cell.

To date we have investigated the system with fluorescent latex spheres ranging in size from 10 microns down to 0.1 microns. The TDI method has proven to be very effective in enhancing the signal to noise ratio of the fluorescence emission. Current data suggests that single molecule detection should be achieved with an excellent signal to noise ratio.

HIGH SPEED SEPARATION OF DNA SEQUENCING FRAGMENTS BY CAPILLARY GEL ELECTROPHORESIS

Norman J. Dovichi
Department of Chemistry, University of Alberta

DNA sequencing fragments were separated in 20-µm ID capillaries at electric fields ranging from 200 to 1000 V/cm for 3, 4, 5, and 6% LongRanger gels. Three distinct regimes are noted in the data. Small fragments tend to be poorly resolved, particularly at low gel concentrations. Intermediate-size fragments have constant peak spacing. Large fragments co-elute. The first separation regime is related to Ogston sieving and the latter regime is related to reptation with stretching.

We have developed a simple phenomenological model that describes the separation of sequencing fragments across a wide range of fragment size, electric field, and gel composition. The retention time of the fragments is given by

$$T_r = \frac{T_0}{E^2} + \frac{\Delta T}{E^2} \times \mathrm{erf}\left(\frac{N}{N_r/E} - \beta\right)$$

where E is electric field, T_0 is related to the retention time of a very short fragment, ΔT is related to the peak spacing for mid-sized fragments, erf is the error function, N is the fragment length, N_r is related to the largest size fragment that can be separated, and β is a constant related to the midpoint of the transition between sieving and biased reptation. T_0, ΔT, and N_r have been normalized to unit electric field. β is independent of electric field and equals 0.78 for 4% gels. This equation describes the sequencing data with part-per-thousand accuracy for all experimental conditions investigated.

This model reflects several important trade-offs between sequencing rate and maximum size fragment that can be sequenced. The sequencing rate, defined as the inverse of the average peak spacing, scales with the square of electric field. At electric fields of 200 V/cm, the separation rate is about 200 bases/hour in a 35 cm long capillary filled with 4% LongRanger and operated at room temperature. At an electric field of 1000 V/cm, the sequencing rate increases to 4,500 bases/hour ofr 4% gels and 7,700 bases/hour for 3% gels. Extremely rapid separation of sequencing fragments is possible at high electric field.

The onset of reptation with stretching scales inversely with electric field. At 200 V/cm, reptation with stretching is significant for fragments 1000 bases long. At 1000 V/cm, reptation with stretching occurs for fragments 150 bases long. Extremely rapid separation of long sequencing fragments is not possible at high electric field.

However, as methods are developed for the rapid and inexpensive production of primers for walking experiments, it becomes interesting to consider the use of very high electric fields for separation of sequencing fragments. For example, an electric field of 400 V/cm allows separation of fragments of 475 bases in length in about 30 minutes.

Gels may be reused many times if they are prepared in 20-µm capillaries. In our first experiment, we obtained eight runs at 1000 V/cm with 4% Long-Ranger gels. Study is underway to define further the stability of these gels. An extremely sensitive fluorescence detector is required for operation with the 20–µm capillaries because only a minute amount of DNA can be loaded onto the capillary.

Last, we have developed a system for the simultaneous operation of five capillaries. This system is being expanded for simultaneous detection of 32 capillaries. With 32 capillaries operating at 400 V/cm, we can determine sequence for 15,000 bases in 30 minutes. When combined with fast, low cost synthesis of sequencing primers, primer walking applications will allow sequence determination of thousands of cosmids per week.

Working Group on Ethics and the Use of Genetic Information

Betsy Fader, Executive Director
Student Pugwash USA,
1638 R Street, NW Suite 32
Washington, DC 20009
Tel:(202) 328-6555 Fax:(202) 797-4664

Student Pugwash USA, a national, educational, non-profit organization, helps young people of diverse academic and ethnic backgrounds gain a better understanding of social and ethical issues raised by science and technology in the following key areas: health and bioethics, environment and energy issues, peace and security, population and development, and computer technologies. Student Pugwash USA's interactive, educational programs and conferences bring international, interdisciplinary groups of motivated students together with scientists, policy makers, members of academe, and industry leaders for examination of critical issues at the juncture of science, technology and society.

In June 1992, Student Pugwash USA conducted a week-long international conference focusing on science, technology and ethical responsibility. The Conference, entitled "Visions for a Sustainable World", brought together approximately 90 college and university students and 65 eminent professionals from 25 different countries for intensive discussions on the role of science and technology in world affairs.

Six different "working groups" were assembled for the week-long event held at Emory University in Atlanta, Georgia, with one working group focusing on the ethical, legal and social implications of the Human Genome Project. Plenary sessions of the conference also examined critical science and technology issues, one of which was entitled, "The Human Genome Initiative: Human Disease Prevention at What Cost?" Both the working group and plenary session raised the following key questions relating to the Human Genome Project:

- What effects will the genetic knowledge derived from the Human Genome Project have on access to insurance, jobs, and civil rights?
- What long-term impacts will the technologies associated with genetic research have upon society, public health, the global gene pool, and bioengineering?
- Is the Human Genome Initiative the most effective and equitable use of scarce scientific resources?

The Department of Energy (DoE)-supported working group, "Ethics and the Use of Genetic Information," included 17 students representing a range of academic experience (10 graduate/ medical, 7 undergraduate) and international backgrounds (11 U.S, 1 Brasil, 1 Canada, 1 Germany, 1 Greece, 1 India, 1 Trinidad & Tobago). The student participants were similarly joined by professionals representing a range of experiences and perspectives. These "seniors" or resource persons served as moderators for the discussions, and provided professional insight where appropriate. The 8 professional participants of this working group included genetic researchers, physicians, ethicists, legal experts, and the director of a non-governmental organization addressing social aspects of the Human Genome Project.

In advance of the International Conference, the student participants prepared original research papers on ethics and the use of genetic information, which then served as the focal point for discussion in the working group throughout the conference week. Upon the conclusion of the week, students presented a working group statement to the full conference, including policy recommendations for the scientific community.

Student Pugwash USA is currently working with a variety of scientific and academic journals to publish the student papers so that they will be available to a broader, international audience. Thus far, two journals (Science, Technology and Human Values, and Technology in Society) have indicated interest in publishing the students' research.

The GDB Human Genome Data Base Anno 1992

Kenneth H. Fasman, Robert J. Robbins, and Peter L. Pearson

Welch Medical Library, 1830 E. Monument St., 3rd Floor, Baltimore MD 21205

The GDB Human Genome Data Base is a computer repository for the gene mapping information obtained as part of the worldwide Human Genome Project. The database contains information on genomic loci, mutations and polymorphisms, mapping reagents, genomic maps, phenotypic data, and all associated references. GDB is an international collaboration hosted by The Johns Hopkins University School of Medicine and the William H. Welch Medical Library. The Genome Data Base is jointly funded by the U.S. Department of Energy and the National Institutes of Health. Negotiations are underway with other countries to provide additional funds for GDB development and the establishment of a worldwide network of remote sites.

GDB is currently implemented as a relational database with a client-server architecture using the Sybase database management system. In a client-server configuration, the database is divided into one or more "front end" client programs which provide the user's interface to the data, and a single "back end" server which handles data requests from the clients. The main Genome Data Base server is located at Johns Hopkins in Baltimore, with read-only copies of the database at various sites around the world. GDB client software can be run from any Sun workstation connected directly to the Internet.

GDB's efforts in the last eighteen months have focussed on the development of a revised database where referential integrity and data validation is completely encapsulated in the back end server. Previously, much of this checking was performed in the associated front end application, with the result that third party software could not be used with the database. We have developed a Database Access Toolkit (DAT) for GDB Version 5 which incorporates all the procedures necessary for syntactic and semantic validation of data. The DAT has been implemented as a collection of stored procedures, triggers, custom data types, rules, defaults, and external subroutines. Any application which accesses the database through this toolkit can be assured of appropriate data validation and integrity.

Building upon the new DAT, a more powerful and efficient front end application has also been developed for GDB Version 5. New features for handling data in groups have been added to accommodate the rapidly growing volume of data in the database. Enhanced output capabilities have been added. The new front end also provides limited X Windows capabilities over local area networks.

A major effort is underway at GDB to revise the way in which genomic maps and supporting mapping data are represented in the database. The current schema, though capable of representing such fundamental mapping concepts as order and distance, is cumbersome and inadequate. We propose the implementation of a three tiered organization of map information, made up of consensus maps, individual observed maps, and the underlying low-level map data. We also propose to develop, in collaboration with a number of third parties, graphical interfaces for the display and manipulation of GDB map data. Other efforts in the coming year will include full support for GDB accession numbers, arbitrary coordinate systems, and direct electronic data submission from genome centers and other end users.

Identification of genes in anonymous DNA sequences

Chris Fields, Granger Sutton, and Owen White

The Institute for Genomic Research
932 Clopper Rd.
Gaithersburg, MD 20878

301-869-9056 (voice); 301-869-9423 (fax)

Results obtained in the last year by anonymous genomic sequencing projects in both humans (Martin-Gallardo et al., *Nature Genetics* 1: 34-39; McCombie et al., *Nature Genetics* 1: 348-353) and *Caenorhabditis* (Sulston et al., *Nature* 356: 37-41) indicate that genes in large anonymous sequences can be located efficiently by a combination of similarity searches of sequence databases, gene-prediction software, and PCR amplification from cDNA libraries using primers selected from predicted exons. As the number of available expressed sequence tag (EST) sequences increases, the fraction of genes that can be located by similarity searching of an EST database will rapidly approach unity. A "complete" EST database, for the purposes of gene identification, need contain only a single member of each BLAST-identifiable gene family; hence isolation and identification of all cDNAs in an organism is not a prerequisite for ESTs to substantially solve the gene localization problem. The challenges for sequence analysis have, therefore, shifted from finding genes in anonymous sequences to 1) integrating gene prediction more closely with sequence data generation, and 2) predicting gene structure and expression.

The success of large-scale EST projects in identifying new genes by similarity searches shows that single-run sequences of moderate accuracy ($\geq 98\%$) can be analyzed directly, before assembly into large contigs. Integrated software systems for automatically analyzing such sequences by similarity searching and coding-region prediction have been developed. These tools are available for integration into analysis systems designed for genomic DNA. Pre-assembly analysis of genomic sequences fragments to identify corresponding ESTs would allow the identification and full-length sequencing of cDNA clones from which the matching ESTs were derived as a marginal addition to a genomic sequencing project. Sequencing of these cDNAs would, given a complete EST database, define the exon-intron structures of all genes contained in the genomic sequence.

The availability of significant genomic sequence data from a variety of phyla allows statistical methods to be used to characterize both functional sites and bulk compositional features of genes. Our results to date suggest that the mechanisms of exon identification by the spliceosome may depend on features including intron and exon length and sequence composition as well as the structure of the 5' and 3' splice sites. While the structures of splice sites themselves vary only slightly between phyla, the compositional properties of both exons and introns differ significantly. An understanding of these differences should enhance our ability to predict rare alternative splicing patterns from genomic sequence data.

Basic Matrices for the Matrix Assisted Laser Desorption/Ionization Mass Spectrometry of Oligodeoxynucleotides

Michael C. Fitzgerald, Gary R. Parr, and Lloyd M. Smith
Department of Chemistry, University of Wisconsin, Madison, WI 53706

Matrix Assisted Laser Desorption/Ionization (MALDI) Mass Spectrometry is a powerful new analytical technique that permits the mass analysis of high molecular weight biopolymers (up to 300,000 Dalton). The matrix, a small organic molecule, is a key part of the desorption and ionization step in this mass spectral method. By isolating analyte molecules in an appropriate matrix and irradiating the sample with a high intensity, pulsed laser beam it is possible to generate intact, gas phase molecular ions of the analyte. The role of the matrix is threefold: (1) a molar excess of the matrix co-crystallizes with the analyte; (2) absorption of the energy from each laser pulse results in the ejection of both matrix and analyte molecules into the gas phase; and (3) the matrix may play a role in the ionization of analyte molecules through gas phase ion/molecule reactions.

Several matrix materials including nicotinic acid, sinapinic acid, and 2,5-dihydroxybenzoic acid have proven to be very effective for the MALDI analysis of a wide variety of protiens. However, results in our laboratory suggest that these acidic matrices are not generally applicable to the analysis of nucleic acids. In order to determine if pH is an important factor in the MALDI analysis of nucleic acids we screened a number of basic compounds as potential matrices. Here we report a new class of matrix compounds that permits the preparation and analysis of MALDI samples at basic pH's. A summary of the results obtained from using over 25 basic matrices in the MALDI analysis of both proteins and nucleic acids will be shown.

OLIGONUCLEOTIDE ARRAYS FOR HYBRIDIZATION ANALYSIS

Stephen Fodor*, Xiaohua Huang*, Robert Lipshutz#, Martin Diggelmann*, Mark Chee*^ and Ann Pease*. *Affymax Research Institute, Palo Alto, CA, #Wagner Associates, Sunnyvale CA, and ^Dept. of Biochemistry, Stanford University, Stanford, CA..

Light-directed combinatorial chemical synthesis is used to fabricate high density arrays of oligonucleotide probes. Photolabile 5'-protected deoxynucleoside phosphoramidites, surface linker chemistry, versatile combinatorial synthesis strategies, and fluorescence detection schemes were developed for this technology. Matrices of spatially defined oligonucleotide probes were generated, and the availability of the probes on the array demonstrated by hybridizing fluorescent labeled targets. Examination of the hybridization behavior reveals a high degree of sequence specificity that enables identification of oligonucleotide targets. Limited sequence reconstruction from the hybridization data will be presented illustrating the progress in using these arrays for sequencing by hybridization.

Assessing Genetics Risks: Issues and Implications for Health (formerly Predicting Future Disease: Issues in the Development, Application and Use of Tests for Genetic Disorders) **(DE-FG05-91ER61115 and NIH N01-HG-0-001)**

Study Director: Jane E. Fullarton, Division of Health Sciences Policy, Institute of Medicine/National Academy of Sciences, 2101 Constitution Avenue NW (FO3016), Washington, DC 20418, 202-334-3913 (PHONE) 202-334-1385 (FAX)

The Institute of Medicine (IOM), through its Board on Health Sciences Policy, established a panel of experts to evaluate scientific, ethical, legal and social issues in the development, application, and use of tests for genetic disorders. The study has addressed a variety of issues presented by the rapid proliferation of genetic tests capable of identifying and/or predicting disease or genetic predispositions to disease. The study is attempting to develop responsible approaches to resolving current and future problems presented by the rapid development and application of genetic tests. Our understanding of human molecular genetics and technologies applicable to the field of diagnostics and testing has expanded enormously in the past decade. The process of mapping and sequencing the human genome will accelerate the rate of new test development and will substantially expand the potential test population. Sophisticated techniques used to explore the genetic bases of disease are providing tools for identifying genetically inherited diseases generally before treatment is available, as well as tools for identifying genetic predisposition to diseases often long before patients show symptoms.

In particular, the study has focused on the following areas:

- The availability of adequately trained personnel—including genetic counselors, genetic and non-genetic laboratory personnel, and physicians—to administer and interpret tests;
- Quality control and integrity in testing and approval for wide-scale use of genetic tests;
- Cost-effectiveness of testing (focusing on recommendations for additional research, especially methodological research in this area);
- Access to test results (specifically by insurors);
- Ethical dilemmas related to the issues of autonomy and privacy; and
- A research and policy agenda for the present and future.

A 20-member committee with expertise in human genetics, law, health education, economics, ethics, medicine, insurance, psychology, and regulation was appointed to conduct the study, with the advice of liaison panel members. Three workshops and a public meeting were convened to discuss critical issues deemed most important by the sponsoring agencies and the IOM. Papers and testimony from the workshops and public forum will be printed and disseminated. In addition, a comprehensive report including findings and recommendations will be delivered at the end of the 30-month study, with final report release scheduled for Spring 1993. It is anticipated that this report will address issues of interest to a wide audience ranging from the genetics community, government agencies responsible for research funding, regulation, and reimbursement for genetic tests, commercial developers of tests, Congress, state legislatures, medical educators, and disease-specific foundations and advocacy groups.

Prototype Automated Instrumentation for the Human Genome Project

Skip Garner (General Atomics) and Glen Evans (The Salk Institute)

The Human Genome Project (HGP) is advancing much faster than anticipated. Driven by technological advances, new biological methods, international competition, and integration of automation, the physical mapping of the genome is being done on a time scale of a year instead of 5 years as originally estimated. In addition, new approaches for large scale DNA sequencing are promising to accelerate that portion of the project. Key to the continuation of this pace is the rapid development of automation and informatics systems capable of handling the tremendous number of samples to be processed and archiving and analyzing the data.

The Biosciences Division at General Atomics in conjunction with the Human Genome Center at the Salk Institute has been developing hardware and software to target many of the research components required for physical mapping and automated sequencing. Our objective is to develop high-throughput robotic and informatics tools that address known bottlenecks - extraction of DNA from bacteria and yeast, preparation of sequencing templates, automated DNA amplification (PCR) and detection, automatic local processing of raw sequence or mapping data.

Several systems have been completed - a centrifuge based DNA preparation device (Dr. Prepper) to process 48 culture samples per hour, a yeast (YAC) screening laboratory equipped to automatically process greater than 10,000 samples per day by PCR and a variety of support software. Under development are - a large next generation system that automates the processing of biological samples from storage to data base entry, a transputer based local parallel processing system to analyzed sequence data automatically using existing homology search codes and post process the data using an expert system and finally, a variety of tools, plasticware and software to move to even higher sample throughputs.

AUTOMATION OF A LARGE-SCALE MULTIPLEX SEQUENCING PROCESS.

R.F. Gesteland[1,2] and R.B. Weiss[1]. Utah Center for Human Genome Research[1] and [2] Howard Hughes Medical Institute, University of Utah, Salt Lake City, UT 84112.

A multiplex physical mapping/sequencing process for converting large-insert cloned DNA into quality sequence ladders is in an initial test phase using human cosmid and YAC contigs. The core molecular biology procedure has recently succeeded in determining 101 Kb of contiguous sequence from the human NF1 locus (GB:L03723). The sequence was generated from a minimal set of transposon-based multiplex priming sites, and the dideoxy-ladders separated via direct blotting electrophoresis. As the magnitude of the project expands, the repetitive tasks of colony picking, DNA preparation, gel electrophoresis and blotting, hybridization and base-calling are being automated. The blotting and probing steps involved in multiplexing have been improved using a horizontal agarose direct blotter (HADB), a conventional direct sequence blotter and an automated probe chamber. Map positions of transposon-based multiplex priming sites are detected by multiple probings of high-resolution Southern transfers, generated by the HADB. DNA is prepared from primary clones on a Biomek workstation, which also assembles mapping pools, resulting in sample handling reductions of 20 to 50-fold. Digitized mapping films are interpreted within an X-window application that records fragment sizes and predicts minimal spanning sets of transposons required to complete both strands of the target sequence with minimal redundancy. Plasmid templates are converted to single strands following attachment to a solid support through a biotin-streptavidin linkage, and sequenced using Mn^{++}-Sequenase conditions. Our direct sequence blotters are capable of resolving multiplex sequence ladders of > 600 nucleotides in length in 20 individually loaded lane sets using an 80 cm 4% polyacrylamide gel. Multiplex blotted membranes from the HADB or direct blotter are hybridized and washed in an automated probe chamber, which is computer controlled for cycle kinematics, fluid delivery, and temperature. Digitized images of multiplex sequence ladders are analyzed by a new base-calling algorithm that uses communication and signal processing theory.

Lawful Uses of Knowledge from the Human Genome Project

Frank Grad, Neil Holtzman,* Dorothy Warburton,** and Ilise Feitshans
Legislative Drafting Research Fund, Columbia University Law School, New York NY 10027
212-854-2685, Fax 212-854-7946
*Johns Hopkins School of Medicine, Baltimore, MD 21205
**School of Physicians and Surgeons, Columbia University, New York NY 10032

A substantial portion of Part I of the project which deals with the Right To Know or Not To Know about Personal Genomic Information" is close to completion. The study of available legal protections of the rights of confidentiality has thus far examined the stigmatizing effect of the disclosure of genomic information, and the substantial impact of such disclosures on employment and insurance availability. Examining the rationale for the legal protection of privacy, our preliminary conclusion is that privacy is not merely a personal right but a public good, which is deserving of the legal protection and the costs which it involves. Preliminary findings show that protection of privacy is currently very limited, at best partial, and at worst non-existent. While there has been a great deal of writing on the subject of privacy protection, in truth there is very little of it. The examination of current legal methods of protection has included the limited common law protections against disclosure of medical information, as well as the extent to which constitutional protections of privacy are available to safeguard genomic information. There are some protections of personal medical and health information generally under the federal Freedom of Information Act and under the Computer Matching and Privacy Protection Act of 1988. A later effort is the Human Genome Privacy Act introduced in 1991 which was limited in scope and failed of enactment. Another act which primarily deals with protection against discrimination is the Americans with Disabilities Act of 1990 which fails to protect persons with genetic disabilities, except insofar as they have been manifested. There is clear evidence that failure to protect privacy leads to discrimination both by employers and insurance companies. A small number of states have enacted limited legislation to address discrimination in health and life insurance availability.

The second phase of our study, which addresses the Uses of Genomic Information in Public Health Planning, has collected significant information on available public health and therapeutic services and on program development in the genomic area. Significant issues in public health planning involve not only questions of organization of genetic services, but also significant issues relating to the voluntary nature of genetic screening and testing and the possible implications of making available genetic screening and testing when followup of a therapeutic nature may be impossible or unavailable. Other problems, which also relate to the voluntariness of available screening and testing services, involve the proper targeting of at-risk groups and individuals, and the organization of health services in the states and nationally to provide for reliable testing services and for a full range of counselling and follow-up services.

Pooling Strategies for Top-Down Library Ordering: A Test on an *S. Pombe* Cosmid Library

Dietmar Grothues[1], Terrence Speed[2], Charles R. Cantor[3] and Cassandra L. Smith[3]
EMBL, Heidelberg[1]; University of California, Berkeley[2]; Center for Advanced Biotechnology, Boston University[3]

Construction of dense ordered libraries covering single chromosomes with clones suitable for transcript mapping and DNA sequencing remains an important priority of genome research. It is very tedious to construct such libraries by bottom-up fingerprinting approaches. We have been testing a top-down approach which appears to offer a number of advantages for efficient cosmid ordering.

The genome of the fission yeast *Schizosaccharomyces pombe* contains three chromosomes totalling about 13 Mb of DNA. Two filters containing an arrayed cosmid library of *S. pombe* were prepared, consisting of a total of 1728 clones or about five-fold genome coverage. These filters were probed with a series of complex radiolabeled DNA probes including PFG-separated *S. pombe* chromosomes, individual or mixtures of *S. pombe* macro-restriction fragments, and pools of smaller *S. pombe* DNA probes. Autoradiographs of the results of these hybridizations were digitized with a CCD camera, and the resulting images were analyzed to determine spot position and intensity and then normalized to constant amounts of vector density to compensate for differences in cosmid copy numbers and colony growth. Cut off intensities were developed so that each hybridization could be scored as positive or negative. Overlaps between clones were established by a likelihood analysis considering all potentially relevant hybridizations. Finally, these sets of overlapping clones were assembled into contigs. The potential power of the approach is that relatively few experiments can be analyzed to give a great deal of information about cosmid order. For example, just those hybridizations, with intact *S. pombe* chromosomal DNAs, assigned 86% of all the cosmid clones to a unique chromosome.

The procedures we have developed should find considerable use in breaking down yeast chromosome-sized DNAs to ordered sets of smaller clones. The power of the method will be increased if a library with inserts larger than cosmids but smaller than yeast chromosomes is analyzed in parallel.

These rapid library ordering methods can be applied to more complex genomes. Here, inter-*Alu* probes generated from PFG-fractionated *Not* I fragments obtained from a monosomic hybrid cell line DNA would be used to order a library. In the case of human chromosome 21, one would be taking advantage of the ordered *Not* I restriction fragments and mega YACs. This method might, in fact, be used with total human DNA to order a complete genomic library.

A DIRECT SELECTION VECTOR FOR SHOTGUN CLONING AND SEQUENCING IN BACTERIOPHAGE M13

Richard A. Guilfoyle, Jim Uzgiris, Doug Kolner and Lloyd Smith
University of Wisconsin, Deptartment of Chemistry
1101 University Avenue, Madison WI 53706

A new M13-based sequencing vector will be described which allows direct selection of recombinants in a shotgun sequencing library. This vector system has demonstrated significant improvements over the widely used M13mp18/19 system where the discrimination of parentals from clones containing inserts is done by means of a blue-white color selection. (J. Messing, (1983) in Meth. Enzymol. 101, (Part C), p20-78). The mp system, although powerful and convenient, is nonetheless laborious, costly, variable in efficiency, and refractory to effective automation. Because of its "direct" or "positive" selection capability, our vector system has yielded improvements in each of these problem areas and include:

(a) observed ratios of up to 200:1 recombinant: parental plaques (99.5 %). This represents a reduction of 20-100 fold in non-recombinants (including false positives) when compared to the background generated by the mp blue/white color selection system. The theoretical background, based on levels of repression observed, ranges from 10,000:1 (basal level) to 5 million:1 (with induction) using phage infections or RF DNA transfections;

(b) the alleviation of the need for protocols requiring alkaline phosphatase, X-gal and IPTG, (the latter when not inducing), providing savings in time, cost, and labor;

(c) greater amenability to automation of the production of single-stranded DNA template preparations as required for large scale sequencing projects. Some preliminary data will be presented regarding a promising flow-cytometry and sorting based strategy for this capability.

Construction of a human genomic library as double-minute chromosomes in cultured mouse cells,

Peter J. Hahn, SUNY HSC, Syracuse, NY
John Hozier, FIT, Melbourne, Fla.

We are developing technolongy to create human genomic libraries composed of megabase-sized acentric chromosome fragments as double-minute chromosomes (DMs) in cultured mouse cells. We are working on three phases of this project: 1 Library construction, 2, Library screening, and 3. Recovery and physical mapping of DMs. Our goal is to be able construct and screen an entire library without resorting to mass culturing of the individual members. Our strategy to accomplish this goal is to first transfect selective markers into human cells at a high rate so that there is a selective marker within a few megabases of all regions of the genome in a single population of human cells. These cells are then expanded as a single population, lethally X-rayed and fused to mouse cells to rescue the labelled fragmented chromosomes as DMs. The resulting population is screened by PCR for microsatellite markers specific for the region of interest. Our progress to date is that we can now achieve transfection rates that yield 10^5 independent integrations of the plasmids with the selective markers using electroporation techniques. We can also achieve fusion rates nearly that high. Therefore, we can create libraries as populations with sufficient compexity to cover the entire genome with multiple redundancy assuming 5 mb/fragment. We haved also been working on manipulating the cells as colonies for screening the cells analogous to colony screening bacteria, and PCR techniques for identifying hybrid cells using CA repeat microsatellite markers. We can easily detect the presence of a microsatellite marker in a culture of as few as 1000 cells. This should allow us to identify specific genetic markers in a library population by screening the cells at the single colony stage. We have also developed techniques for recovering and physically mapping DM DNA. We have been using DMs from cells with highly amplified natural DHFR genes on DMs of approximately 1 million base-pairs. We have been able to isolate the DM using PFGE and to construct a long range map of this DM combining ethidium bromide staining and Southern hybridization of PFGE of digestions of the DM DNA.

SYNTHESIS OF FLUORESCENT DNA FOR SEQUENCING BY SINGLE MOLECULE DETECTION.

John Harding, Alberto Haces, A. John Hughes and Roger Lasken.

Corporate Research and Molecular Biology Research and Development,
GIBCO BRL, Life Technologies Inc. Gaithersburg MD, 20898.

We and scientists at the Los Alamos National Laboratory are jointly developing an advanced DNA sequencing technology potentially capable of sequencing DNA at rates of 100 - 1000 bases per second. The technique involves synthesizing a fluorescent copy of a DNA target, capturing a single molecule of fluorescent DNA on a solid support, suspending the solid support in a flowing sample stream as in a flow cytometer, digesting the labelled nucleotides sequentially from one end of the DNA molecule using a nuclease and identifying individual nucleotides by laser-induced fluorescence (J. Jett et al, LANL; U.S. Patent 4,962,037). This joint project is performed under the auspices of Cooperative Research and Development Agreement LANL-C-91-001.

A critical step in the sequencing protocol is the synthesis of DNA several kb in size containing fluorescent nucleotides. To determine the feasibility of this step, we have: 1). Synthesized 15 different fluorescent nucleotides containing either rhodamine or fluorescein attached to the base moiety of dCTP, dUTP or dATP, respectively, by linkers of different length and chemical composition. 2). Tested the ability of 6 different DNA polymerases and 3 different reverse transcriptases to synthesize full length copies of M13mp19 DNA (7.25 kb) when one of the normal dNTPs is fully replaced by a cognate fluorescent dNTP. The various enzymes differ in their ability to synthesize fluorescent DNAs. T5(exo-) and T7(exo-) DNA polymerases are the best enzymes thus far tested for this purpose. DNAs greater than 7 kb in size are synthesized with certain nucleotides. 3). Synthesized DNA greater than 4 kb in size containing two fluorescent dNTPs totally replacing the cognate normal dNTPs. 4). Shown that, for some polymerases, synthesis can be enhanced by the presence of auxiliary proteins, such as E coli single-strand binding protein, in the reaction.

These results suggest that it will be possible to synthesize fluorescent DNA for sequencing by single molecule detection in flowing sample streams.

Experiments examining the fidelity of incorporation of fluorescent nucleotides by various polymerases and nuclease digestion of fluorescent DNA bound to solid supports are currently in progress.

Low-Cost Massively Parallel Neurocomputing for Pattern Recognition in Macromolecular Sequences

John R. Hartman* and **Dianne M. Marsh**

Computational Biosciences, Inc.
P.O. Box 2090
Ann Arbor, MI 48106

*e-mail: **john@cbi.com**
Tel: (313)426-9050
Fax: (313)426-5311

Connectionist (neural network) approaches to pattern recognition and classification have attracted great interest because of their flexibility, ability to learn by example, and ability to self-organize such that hidden patterns and relationships in the input data set can be discovered. Neural networks are inherently parallel computational structures, and thus potentially excellent candidates for implementation on massively parallel computers. Particularly in the training stage, serial implementations of connectionist models are often limited with respect to either the size of the network that may feasibly be simulated or the number of training trials that are practical. Exponential growth in the size of public macromolecular sequence databases (e.g., GenBank) have created a need for robust pattern analysis solutions with cost-performance characteristics superior to those currently available.

We describe here a working neural network implementation on an inexpensive SIMD parallel computer with 512 processing elements (PEs). Algorithms are presented for the components of a basic multilayer, feed-forward network model; these include a *linear combiner* for the multiplication of connection weight matrices with input and error vectors, a sigmoidal *activation function* to introduce non-linearity in neuron behavior and stability to learning, and a parallel *back-propagation* (generalized delta rule) training algorithm. Figures quantifying the performance of an early version of the network (not available when this abstract was prepared) on a prototype sequence analysis problem will also be presented.

Ultimately, this work will be generalized as an object-oriented parallel neural network toolkit, and will be employed to augment the capabilities of CBI's **X/Gene**^{TM} distributed/parallel sequence analysis software (currently under development).

The work described is funded through grant# DE-FG02-92ER81390 from the U.S. Department of Energy.

X/Gene^{TM}: Massively Parallel and Distributed Sequence Analysis on Heterogenous Unix Networks

John Hartman*, David Solomon, Clifton Flynt, and David Steinhoff

Computational Biosciences, Inc.
P.O, Box 2090
Ann Arbor, MI 48106

*email: **john@cbi.com**
Tel: (313)426-9050
Fax: (313)426-5311

Progress on the **X/Gene**^{TM} sequence analysis software package is described. **X/Gene**^{TM} is a graphically-oriented collection of programs designed to accommodate the rapidly growing demands of genomic analysis and database management. Salient features include:

- An intuitive and informative *graphical user interface* (GUI) based on the X Window system, designed to facilitate ease of use in both directed and explorative work with minimal reference to printed manuals;
- Support for *distributed processing* (spreading compute-intensive operations over idle CPUs on a local network), facilitated by a complete separation of the user interface from analytical modules and adherance to a client-server model;
- Support for inexpensive *massively parallel processing* as a network resource to accelerate similarity searches, large sequence comparisons and alignments, contig assembly, neural networks, and other compute-intensive operations;
- Robust and high-performance genome analysis and database searching capabilities; and
- *Object-oriented* design and implementation (in C++) for reliability, maintainability, and future extensibility.

The overall design of the **X/Gene**^{TM} package is a response to exponential growth of macromolecular sequence databases that will only accelerate in coming years with the Human Genome Project. It also recognizes the recent revolution in computing technology, with its trend toward increasingly powerful, inexpensive, and interconnected desktop-scale machines and away from large, monolithic, centralized facilities. By carefully isolating hardware-dependent program code and adhering closely to a client-server paradigm, we will be able to better adapt to the needs of heterogenous network environments and take quick advantage of expected opportunities in low-cost, high-performance computational hardware.

The **X/Gene**^{TM} front-end currently runs under OpenLook^{TM} on Sun SPARCstations, and will soon support optional SIMD massively parallel processors from Applied Intelligent Systems and MasPar. Ports are expected to a variety of hardware platforms and graphical environments in the future. Presently under development, **X/Gene**^{TM} is scheduled for commercial introduction during the third quarter of 1993.

This work was funded partially through grant # DE-FG02-90ER80902 from the U.S. Department of Energy.

COMPARISON OF SPECIAL PURPOSE COMPUTER HARDWARE AND STANDARD SOFTWARE METHODS FOR SEQUENCE COMPARISON

G. Herrmannsfeldt and T. Hunkapiller, Division of Biology, California Institute of Technology, Pasadena, CA.

Our research has been focused on redefining the dynamic programming paradigm such that not only could these methods be implemented effectively in silicon, but also provide, in that context, the flexibility required of a robust biological tool. Our biological information signal processor (BISP) represents a systolic implementation of a dynamic programming algorithm based on that of Smith and Waterman, optimizing its ability to define local similarities at extremely high speeds (Chow et al, Proc. Intl. Conf. Supercomp., 1991). The first generation BISP system is now functional and undergoing functional and performance tests. Particularly, we are using members of the immunoglobulin gene superfamily as paradigm sequences for comparative searches of the published databases. This is probably the most frequently represented and diverse group of homologous sequences in the eukaryotic collections. This diversity has made such sequences problematic when using standard database comparison techniques. We have compared the results of the BISP array for issues of sensitivity and selectivity with those of FASTA and BLAST. Performance questions have also been addressed.

ACCESSORY PROTEINS OF BACTERIOPHAGE T7: *ESCHERICHIA COLI* THIOREDOXIN, T7 HELICASE, AND T7 DNA BINDING PROTEIN

Jeff Himawan, Steven M. Notarnicola, Stanley Tabor, and Charles C. Richardson
Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School,
Boston, MA 02115

Bacteriophage T7 DNA polymerase modified to reduce or eliminate its 3' to 5' exonuclease activity has properties that are advantageous for DNA sequence analysis. Some of these properties derive from the fact that T7 DNA polymerase is the enzyme responsible for the replication of the T7 chromosome. As such, it physically interacts with several accessory proteins for the acquisition of properties such as processivity and helicase activity. An understanding of these interactions should facilitate the development of novel DNA polymerases and new sequencing and amplification strategies.

T7 DNA polymerase (gene 5 protein) has low processivity, dissociating from the primer-template after catalyzing the incorporation of 1-50 nucleotides. *E. coli* thioredoxin binds tightly (K_m = 5 nM) to T7 DNA polymerase in a 1:1 stoichiometry, stabilizing the binding of T7 DNA polymerase to a primer-template by 20- to 80-fold and increasing the processivity of polymerization by 1000-fold. To understand how thioredoxin confers processivity, we have been studying the interaction between T7 DNA polymerase and *E. coli* thioredoxin by genetic methods. We have used a thioredoxin mutant (Gly-74 to Asp) that is unable to support the growth of wild-type T7 phage to select for T7 revertant phage that suppress this thioredoxin defect. Three suppressor mutations are located within the 3'-to-5' exonucleolytic domain of gene 5 protein, and three are located within the polymerization domain of gene 5 protein. We are currently purifying the mutant proteins and examining their biochemical properties.

The gene 5 protein/thioredoxin complex, although highly processive on single-stranded DNA, is unable to polymerize nucleotides through duplex regions of DNA. In a reaction coupled to the hydrolysis of a nucleoside 5'-triphosphate the 56-kDa gene 4 protein, the T7 helicase, translocates 5' to 3' on a DNA strand and catalyzes the unwinding of duplex DNA. T7 also induces the synthesis of a 63-kDa protein, the T7 primase, that catalyzes a template directed synthesis of tetraribonucleotides that are used as primers by T7 DNA polymerase. The primase is identical to the helicase except for the presence of an additional 63 amino acids residues at the amino terminus. The smaller helicase is produced via an internal in-frame translation initiation codon within gene 4. Both proteins contain an abridged "A" nucleotide-binding motif. Mutant helicase or primase proteins in which two conserved residues within this site have been altered do not catalyze DNA-dependent nucleotide hydrolysis or translocate on DNA. Interestingly, wild-type helicase can provide translocation activity to a mutant primase thus restoring its ability to find and copy a primase recognition site. Evidence for protein-protein interactions also comes from the finding that mutant gene 4 proteins are dominant lethal when introduced into wild-type T7-infected cells.

T7 also encodes a single-stranded DNA binding protein, the product of gene 2.5. Gene 2.5 mutants of phage T7 catalyze little, if any, DNA synthesis and are defective in recombination. Gene 2.5 protein forms a complex with T7 DNA polymerase in a one-to-one stoichiometry and stimulates the activity and processivity of the polymerase. The purified gene 2.5 protein also stimulates greatly the renaturation of two complementary strands of DNA, increasing the second-order rate constant for annealing approximately 5000-fold over that obtained in the absence of protein at 37 °C. The renaturation reaction requires a saturating amount of DNA binding protein, and is unaffected by a 500-fold excess of heterologous DNA. The ability of T7 single-stranded DNA binding protein to facilitate renaturation could be of general use in procedures that require hybridization of complementary single-stranded DNA.

GRM: A Genome Reconstruction Environment to Support Large-Scale DNA Sequencing

Sandra Honda, N. Wayne Parrott, Thomas Flood, Xiaoqiu Huang[1], Gene Myers[2] and Charles Lawrence

Computational Molecular Biology Group, Department of Cell Biology, Baylor College of Medicine, Houston, TX; [1]Department of Computer Science, Michigan Technical University; and [2]Department of Computer Science, University of Arizona

We are developing a comprehensive environment, the Genome Reconstruction Manager or **GRM**, to support large-scale DNA sequencing. The environment includes:

- Project and Object Management
- Automated preprocessing of primary sequence data
- Multiple fragment assembly kernels
- Constraint enforcement interface for closing mapped gaps
- Suppport for random, directed and mixed sequencing strategies
- Contig Editor
- ABI 373A data file display
- Engineered graphical interface for human interaction and information display

The project is also a test of object-oriented software development technologies. Object-oriented analysis and design have been applied from the beginning of the project. A working prototype for **GRM** has been implemented in ParcPlace Smalltalk 80, an integrated object-oriented programming language. Object persistence is provided by the Gemstone ODBMS.

A key result of the project is an object data model for genome map information at all levels of resolution (Genome Map Model or **GeMM**). Map classes in **GeMM** are used extensively in GRM to represent complex objects in the application domain such as primary sequence data files, sequence contigs, constraints, and high-level organization of discontinuous sequence information. **GeMM** also permits seamless integration of **GRM** with sequence analysis applications and with higher-level genome map browsers.

LARGE-SCALE DNA SEQUENCING

L. Hood, L. Rowen, B. Koop and T. Hunkapiller

We have focused on two areas over the past year: systems integration and technology development, and production line sequencing.

Systems Integration/Technology Development: We have emphasized two areas. First, we have diversified our DNA sequencing strategies. Cycle sequencing has been introduced on the 96 well thermocycler as well as with a robotic DNA sequencer, the catalyst. We have demonstrated the overall pattern of error is similar in both the cycle sequencing and the sequenase sequencing protocols. We have also shown that the two strategies complement one another in that their errors tend to be in different segments of the DNA sequence and accordingly cancel one another out. As a result, we are employing both strategies for each cosmid we sequence. Second, we have distinctly improved the assembly and editing strategies. This effort includes programs that have enabled us to do batch editing rapidly on the Macintosh system, and to examine directly the original chromatographic data a we edit our contigs. As a result of these improvements, we have been able to keep the assembly process in pace with the generation of sequence data.

Our future plans include 1) examining the double-stranded cycle sequencing and dye-terminator protocols for their read length and accuracy in the interest of further diversifying our sequencing strategy and improving our ability to achieve contig closure: 2) switching to larger numbers of well lanes run per gel (32 or 36, whichever proves more successful) and 3) the development of a second generation high throughput fluorescent DNA sequencer.

Production Line Sequencing: Over the past year, we have sequenced approximately 3 megabases of raw DNA sequence data and generated more than 250kb of finished DNA sequence. Thus, the production goal of 250kb of sequence in this year was met. The goal of submitting 200kb of data to GenBank was also met.

The specific regions we focused on were: 1) the complete DNA sequence analysis of the human T-cell receptor α / δ locus from the C_δ to the C_α genes; 2) the analysis of a 35kb cosmid from the junctional region between the cluster of mouse V_α gene segments on the one hand and V_δ gene segments on the other hand; and 3) the data collection, assembly and preliminary analysis on 250kb of human T-cell receptor V_β locus. This locus will be particularly interesting because it contains at least one V_β gene segment that may predispose to multiple solerosis. These data have presented a series of striking new opportunities for exploring the biology of mouse and human T-cell receptor loci.

Xiaohua C. Huang, Mark A. Quesada and Richard A. Mathies*
Department of Chemistry, University of California, Berkeley, CA 94720

Capillary Array Electrophoresis Using Confocal Fluorescence Detection: An Approach to High-Speed, High-Throughput DNA Sequencing

A laser-excited, confocal-fluorescence scanner is used for high-sensitivity, on-column detection of electrophoresis performed using an array of capillaries. A linear array or ribbon of capillaries is assembled on a scan stage and then translated past a laser excitation and confocal-fluorescence detection system. The small depth of focus of the excitation coupled with confocal detection through a spatial filter reduces the background due to scattering, stray fluorescence, and reflections from the capillary walls, while the high numerical aperture objective provides efficient collection of the fluorescence. Because we can run hundreds of capillaries in parallel, capillary array electrophoresis (CAE) is an important new high-throughput analytical technique. To illustrate the power of CAE we have applied this method to the development of high-speed, high-throughput DNA sequencing. First, electromigration data are presented for fluorescently-labeled T-fragments separated on an array of gel-filled capillaries and detected with a one-color system. Data will be presented using 24 capillaries run in parallel, demonstrating the feasibility of CAE. To sequence DNA with capillary electrophoresis it is desirable to detect all four sets of DNA sequencing fragments on the same capillary. Thus, we have developed a sequencing method that utilizes capillary arrays, two-color fluorescence detection, and a two-dye labeling protocol. Sanger DNA sequencing fragments are separated on an array of capillaries and then distinguished by using a binary coding scheme that employs only two different fluorescently-labeled dye primers to identify four sets of fragments. DNA sequencing results will be presented using a 25 capillary array. This apparatus has the capability of sequencing DNA at a rate of 20,000 bases/hour.

Characterization of Radiation Hybrids Produced from a Marked Human Chromosome 19

Cynthia L. Jackson, Hon Fong L. Mark, Deborah E. Britt,
Kathleen Santoro and Kathleen Walsh
Department of Pathology,
Rhode Island Hospital, Brown University,
593 Eddy Street, Providence, RI 02903

Radiation hybrid cell lines have proven to be excellent resources for the physical mapping and the isolation of markers from specific chromosomal regions. We have produced radiation hybrids from a monochromosomal microcell hybrid containing chromosome 19. The human chromosome has been marked with a retroviral vector containing a dominant selectable marker. The hybrids were produced using doses of radiation ranging from 1000 to 8000 rads and selection for the exogenous marker.

A panel of approximately 90 hybrids are being characterized using fluorescent in situ hybridization (FISH) and marker analysis. These methods should allow us to identify the number of chromosome fragments and the region of the chromosome contained in the hybrids. The panel has been tested for over twenty chromosome 19 markers from known locations on the chromosome using the polymerase chain reaction and Southern blotting. We will continue to test the hybrid panel for additional markers. Hybrids have been isolated which appear to contain only 1 or 2 markers from the short arm of the chromosome where the selectable marker is inserted.

In situ hybridization is the most direct method to visualize the number of chromosome fragments in a hybrid. We have utilized fluorescent in situ hybridization (FISH) to label human chromatin material in the hybrids using two protocols. A number of hybrids were characterized using biotinylated total human DNA as a probe hybridized to metaphase psreads prepared from the hybrids. We have also optimized the protocol using biotinylated Alu-PCR products from the hybrids as a probe hybridized to human metaphase spreads. A high percentage of hybrids appear to contain a single fragment when analyzed by FISH. In summary, the hybrids that we have produced will aid in the physical mapping of chromosome 19 and region specific cloning.

Computer-assisted studies on human repetitive DNA

Jerzy Jurka, Linus Pauling Institute of Science and Medicine, 440 Page Mill Road, Palo Alto, CA 94306

Most of the human repetitive families represent medium reiteration frequency (MER) repeats of unknown origin with copy numbers per genome ranging from hundreds to thousands. These repeats represent a fossil record of evolutionary processes shaping the human genome and affecting the chromosomal stability of contemporary human individuals. Therefore, studies of repetitive DNA are of both biological and medical importance.

Our long-term goal is to discover and study novel families of repetitive elements using computer-assisted DNA sequence analyses. This approach proved to be far more successful than its experimental alternatives. It led to the discovery of nearly 30 new families in the available human genome sequences, of which nine will be reported during the meeting (1).

Recently, we compiled and made available electronically a collection of 53 prototypic sequences representing known families of repetitive elements from the human genome (2). This collection will be updated in the near future. Reference collections of repeats are essential for any routine studies of newly sequenced DNA. We also developed software for quick identification of repeats in newly sequenced DNA. A version of this software can be run via electronic mail (2).

We have characterized the second largest family of human interspersed repeats which appears to be universal in mammals (3).

Finally we have developed an algorithm for identification and analysis of so called "simple repeats" (4).

1. Jurka, J., Kaplan, D.J., Duncan, C.H., Walichiewicz, J., Milosavljevic, A., Murali, G., Solus, J.F. (1992) Identification and characterization of new human medium reiteration frequency repeats *(submitted)*.
2. Jurka, J., Walichiewicz, J, & Milosavljevic, A. Prototypic sequences for human repetitive DNA. *J. Mol. Evol.* **35**, 286-291 (1992).
3. Jurka, J. et al. A large family of short mammalian interspersed repesats *(in prepration)*
4. Milosavljevic, A. & Jurka, J. Discovering patterns in genetic sequences by Minimal Length Encoding *(submitted)*.

CHROMOSOME MICRODISSECTION AND PCR-MEDIATED MICROCLONING IN HUMAN GENOME AND GENETIC DISEASE ANALYSIS

Fa-Ten Kao and Jingwei Yu
Eleanor Roosevelt Institute for Cancer Research, and University of Colorado Health Sciences Center, Denver, CO 80206

High resolution physical mapping and cloning of disease-related genes from specific chromosomal regions require large numbers of DNA probes from the region of interest. Among different strategies, a direct method is to use microdissection to physically remove the crucial chromosomal region, followed by microcloning of the dissected DNA sequences to construct a genomic library, as demonstrated first in Drosophila polytene chromosomes. We developed an efficient method for microdissection of human chromosomes followed by PCR amplification of the dissected sequences using MboI linker-adaptor (PNAS 88, 1844-1848, 1991). In this procedure, 20 fragments are dissected, digested with proteinase K, extracted with phenol, and ligated to an MboI linker-adaptor. All these steps are performed in nanoliter volumes under the microscope. The ligated sequences are transferred to an Eppendrof tube for PCR and the PCR products are cloned into pUC19. Large microdissection libraries have been constructed with tens to hundreds of thousands of microclones in each library. The inserts in the microclones averaged 300-400 bp and 50-60% of the clones contained unique sequences. The libraries have been shown to derive from the dissected regions by using FISH in chromosome painting.

We have constructed microdissection libraries for human chromosome 21, 21q21 and 21q22 for physical mapping and molecular analysis of Down syndrome. In addition, we have constructed 4 region-specific libraries for human chromosome 2, including 2q35-q37, 2q33-q35, 2p23-p25, and 2p21-p23 (Genomics, in press). Characaterization of these libraries by Southern blot hybridization and the FISH analysis have shown that they are of good quality and can be used efficiently not only for fine structure physical mapping of these regions, but for the analysis and cloning of disease-related genes mapped to these regions, e.g. Alveolar rhabdomyosarcoma (a pediatric soft tissue tumor mapped to 2q35), aniridia-1 (a dominant disorder of the ocular abnormality mapped to 2p25), holoprosencephaly (an abnormal forebrain and midface development mapped to 2p21-p22), and others.

As we have demonstrated previously (Am. J. Hum. Genet. 51, 263-272, 1992), the microclones can be used efficiently as probes to isolate (i) corresponding YAC clones with large inserts for physical mapping and contig construction of the dissected region, and (ii) region-specific cDNA clones as candidate genes for diseases assigned to the region. This chromosome microtechnology will become even more important as the more accessible areas of the genome are characterized and leave some large gaps, and no relevant cell hybrids are available for those gap regions. Microdissection is ideally suited because it can be targeted anywhere in the genome and no selectable marker is needed.

Construction of High Resolution Contig Map of Human Chromosome 22

Ung-Jin Kim, Hiroaki Shizuya, Bruce Birren, Jeffrey Garnes*,
Pieter deJong*,and Melvin Simon, Division of Biology 147-75,
California Institute of Technology, Pasadena, CA 91125, and
*Human Genome Center, L-452, Lawrence Livermore National
Laboratory, Livermore, CA 94550

We have previously demonstrated the stability of human DNA inserts cloned in pFOS1, an F factor based vector. We have constructed a human chromosome 22 specific Fosmid library with 8X coverage by picking human positive clones prepared from human chromosome 22-hamster hybrid cells, and also from flow sorted chromosome 22. We have also prepared a total human BAC library with larger inserts. Approximately 6X (128 microtiter plates) of the chromosome 22 specific Fosmid library and 1.5X (324 microtiter plates) of the total human BAC library were gridded onto filters at high density. We are currently developing methods to quickly identify subsets of clones using large regional probes, such as YACs and BACs that have already been localized. In addition, filters are being hybridized with STS and cDNA probes and Fosmid clones which have been localized by FISH. Subsets of clones identified by various approaches are being ordered into contigs by restriction digestion and fingerprinting.

Site-Specific Endonucleases for Human Genome Mapping

Kimberly Knoche, George Golumbeski, Lydia Hung and Ray Bandziulis
Promega Corporation, 2800 Woods Hollow Road, Madison, WI 53711
608/274-4330, Fax 608/273-6967

Current genome mapping methodology suffers from a lack of tools for generating specific DNA fragments in the megabase size range. While several multidimensional approaches are underway for cleaving mammalian DNA in this range, there is currently no single step procedure to generate specific DNA fragments of this size. Promega is developing a family of site-specific endonucleases capable of generating DNA fagments in the 2-100 megabase range in a single step. For this project, we are focusing on a recently discovered class of enzymes, the very-rare cutting intron-encoded enconucleases. These enzymes are similar to bacterial type II restriction enzymes; however, the intron-encoded enzymes have much larger recognition sequences (15-39 base pairs) making them much less frequent cutters of DNA.

To generate the family of megabase cutters, we will use two approaches. One approach is to relax the specificity of I-Ppo, an intron-encoded endonuclease with a 15 base pair recognition sequence. This enzyme was chosen because it has excellent catalytic ability, high stability and will cut methylated DNA (this will allow reproducible digestion patterns to be obtained from the DNA of individuals and different species whose methylation pattern may vary). Our second approach is to develop other intron-encoded endonucleases.

The first approach involves three areas of research. (1) We are characterizing the 15 base pair recognition sequence by studying the tolerance for single base pair substitutions. This work is nearly complete. Using this information, we will attempt to establish reaction conditions (pH, ionic strength, enzyme concentration, dielectric constant and metal cofactor) and chemical modifications of I-Ppo that result in relaxed specificity (i.e. more frequent cutting). (2) Structural studies of the enzyme also are underway in order to identify DNA binding domains of the protein. This information will allow us to use site-directed mutagenesis in an attempt to isolate mutant I-Ppo enzymes with relaxed specificity. (3) We are assessing I-Ppo's ability to cut chromosomal DNA in agarose plugs. To date we have found that I-Ppo cuts chromosome XII of the yeast genome and completely excises the human 28S rDNA repeats. Any conditions and/or modified enzymes that result in relaxed specificity will be tested with complex genomic DNA. Both the cutting frequency and enzyme performance will be evaluated.

For the second approach we are currently developing two additional intron-encoded endonucleases. Both enzymes have been expressed in *E. coli*. Currently, we are purifying these proteins and characterizing their specificity.

Gene Sequencing by Scanning Molecular Excitation Microscopy: A Progress Report

Raoul Kopelman, John Langmore,* Bradford Orr,** Zhong-You Shi, Steven Smith, Weihong Tan, Eric Monson,* and Gregory Merritt
Department of Chemistry, *Biophysics Research Division, and **Department of Physics, The University of Michigan, Ann Arbor, MI 48109-1055
313/764-7541, Fax 313/747-4865, Internet: "usergb2q@um.cc.umich.edu"

Ideally, DNA sequencing is performed by rapid direct imaging of very long strands, starting from a known location and reading the sequence quickly, with no error, until another known location is reached. Our proposed scanning molecular excitation microscopy (MEM) was conceived to achieve this ideal. The scanning MEM rapidly locates one end of a DNA molecule. From there it moves quickly along the backbone, reading tens of thousands of bases until it reaches another designated location. The fast read-out is performed with high fidelity and without any perturbation or alteration of the sample. The sample's sequence can be read repeatedly, with high fidelity. The critical features of MEM that make it more promising than TEM, STM, or AFM are: 1) the ability to scan at low resolution in order to automatically and very quickly locate the DNA molecules, 2) the ability to use a wide variety of specimen supports, 3) the ability to non-invasively image non-conductive molecules, and 4) the large, high resolution, base-specific contrast. Furthermore, this method can be *combined* with AFM. We have at this point improved the resolution of near-field-optics/exciton microscopy from about 5000 Å to about 150 Å. Probes made of optical/exciton tips and supertips have been produced with high sensitivity and durability. They now also have the capability of Atomic Force feedback using a new principle of lateral (dithering) force microscopy. For spatial calibration we employ gratings made for electron-microscopy. Preliminary images have also been obtained from various samples, including porous membranes and DNA. A scanning-tip optical spectroscopy has also been developed, with a *spatial* resolution of 400 Å. It has been applied to molecular films containing fluorescent molecular probes. We have experimentally demonstrated the principle of *tip-sample* energy transfer (Förster) as well as the *interfacial* Kasha (external heavy atom) effect. The Förster and Kasha effects are the mechanisms proposed by us for the *in-situ* DNA sequencing. Our very successful nanofabrication of optical excitation probes has been demonstrated by the manufacture and use of fiber optics biochemical sensors that are a thousand times faster and can analyze samples with a billion times fewer molecules, compared to the best present state of the art. Based on our experimental demonstrations we estimate that with the help of this technique a single (connected) Mega-base strand could be sequenced in about 10 min., *i.e.*, better than 1000 bases/sec. We are currently working on proper sample preparation techniques in air and in saline solution and on the development of single molecule supertips. We have also demonstrated that a single molecule can have the required quantum efficiency, stability and sensitivity. We still need to combine these properties via molecular engineering. Several schemes are under investigation for making such a "supermolecule" with a 5-10 Å active group.

Partial Bibliography

R. Kopelman and W. Tan, "Near-Field Optical Microscopy, Spectroscopy and Sensors," in *Spectroscopic and Microscopic Imaging of the Chemical State*, ed. M. Morris, Marcel Dekker (in press).

R. Kopelman, W. Tan, and Z-Y. Shi, "Nanometer Optical Fiber pH Sensor," *Soc. Photographic and Imaging Engineers Proceedings* **1796** (1992).

W. Tan, Z-Y. Shi, and R. Kopelman, "Development of Submicron Chemical Fiber Optic Sensors," *Analyt. Chem.* **64** (1992).

W. Tan, Z-Y. Shi, S. Smith, D. Birnbaum, and R. Kopelman, "Submicrometer Intracellular Chemical Optical Fiber Sensors," *Science* **258**, 778 (1992).

Human cDNA Mapping Using Fluorescence In Situ Hybridization. J. R. Korenberg[1], X. N. Chen[1], J. Gatewood[2], M. Adams[3], J. C. Venter[3]. Ahmanson Dept. of Pediatrics, Div. of Genetics, Cedars-Sinai Medical Center., UCLA[1], Life Sciences Div., Los Alamos Laboratory, NM [2], NIH Receptor Biochemistry and Molecular Biology Section, Bethesda MD[3].

The human genome is estimated to consist of 50,000-100,000 genes. With current technology, it was likely that the majority of human genes would remain unknown for at least the next decade. Therefore, the recent "cDNA strategy", to sequence only transcribed DNA was designed and has been successfully applied to the cloning and partial sequencing of over 10,000 cDNAs by numerous groups. The rapid application of these expressed sequence tags (ESTs) to genome mapping and as disease markers requires a knowledge of their genomic location. Our laboratory has now developed novel banding techniques and hybridization technologies to rapidly localize human cDNAs to single bands of human metaphase chromosomes. Moreover, we have demonstrated that our techniques preferentially identify the transcribed locus in a multi pseudogene system. Using these techniques, we have mapped cDNAs in the range of 0.8 - 4.3 kb to chromosomal regions of about 2-4 megabases. The results of these endeavors have yielded information on the accuracy, sensitivity and speed of fluorescence in situ hybridization of cDNA mapping. It is found that 70% of both random and known cDNAs in the size range of 3.5 kb may be assigned to single human chromosome bands with high accuracy. Using these techniques combined with current technology of multicolor fluorescence in situ hybridization and moderate image analysis capabilities, should allow for the mapping of 3,000 cDNAs in the above size range to single human chromosome bands at 2-4 Mb resolution within two years using only 6 individuals. The immediate mapping of cDNAs provides rapid evaluation of these genes as candidates for genetic disease in humans and, by comparison with known regions of homology, for disease models in the mouse. Such EST's may also serve as the basis for the rapid creation of overlapping large fragment maps that complement other methods to facilitate the completion of the human genome map and its applications to human biology.

Non-Random DNA Rearrangement Events in DNA Sequences Flanking Inserted Retroviral Sequences Following Whole Cell Fusion

M.J. Lane, S. Mante, L. Cherath, S.M. Peshick and D.J. Hanlon*
Departments of Medicine and Microbiology/Immunology, SUNY-Health Science Center at Syracuse, Syracuse, NY 13210;
*Department of Neuropathology, Yale University Medical School, New Haven, CT 06511

We have been involved in a collaborative effort to create a library of "cloned" intact megabase scale human DNA segments, flanking markers introduced into the human genome, by fusion of lethally X-irradiated donor cells to mouse EMT-6 recipient cells, which have shown a high frequency of retention of markers and flanking DNA as double minute chromosomes. The mechanism biasing EMT-6 recipient cells toward retention of DNA as double minute chromosomes (DM's) is not understood. This procedure for cloning large human DNA segments is a relatively new one and given the chimera problems associated with early YAC libraries we were interested in demonstrating that DNA "cloned" in such a fashion truly represented the structure of the human DNA surrounding the introduced marker in the donor cells.

To investigate this question we initially employed a single human chromosome containing human:mouse microcell hybrid, in which the human chromosome had been marked with a single copy of the defective retrovirus pZipNeo. This cell line was lethally irradiated and then fused to mouse EMT-6 cells. Viable fusion cell lines were then selected on G418 containing media followed by cloning and propagation. To our surprise evaluation of the size of the Not I fragment encoding the neomycin gene revealled that none of four daughter fusion lines analyzed retained "parental" restriction fragment size. This Not I fragment comparison prompted us to examine restriction sites in the immediate (5-10 kb) vicinity of the viral insert in a larger sample of daughter clones and compare these fragment sizes with the appropriate parent restriction fragment sizes. We detected retention of parent restriction fragment sizes in only a small minority of the daughter clones examined. Significantly, the same rearranged restriction fragment sizes were observed in several daughter lines and in some cases the parent restriction fragment size was observed on one side of the retrovirus while on the other side of the virus a rearranged fragment length was encountered. Similar results have been generated with several donor lines. We interpret these observations to mean that a viral reinsertion mechanism is not responsible for the restriction fragment length alterations and that the DNA rearrangement events observed are markedly non-random. This information will be presented.

Further experimentation, involving similar restriction analysis of DNA from more than a hundred independent daughter cell lines, obtained from various donor:recipient fusion combinations, results from which will also be presented, has revealed that: 1) the rearrangements observed do not appear to be a consequence of X-ray induced DNA repair - rearrangements occur following fusion even if the experiment is executed with unirradiated donors; 2) provisionally, that both the retroviral LTR and genomic DNA sequences flanking the viral sequence can independently serve as sites for the rearragement events observed; 3) a pZipNeo construct introduced into a donor by transfection can show similar flanking DNA rearrangement frequency, and; 4) both the donor cell line and the recipient cell line employed in a fusion affect the frequency of DNA rearrangements obtained - suggesting that the effect can, in fact, be overcome, allowing successful completion of the human DM library.

Efficient Algorithms and Data Structures in Support of DNA Mapping and Sequence Analysis

Eugene L. Lawler
Computer Science Division
University of California at Berkeley
Berkeley, CA 94720

Daniel Gusfield
Division of Computer Science
University of California at Davis
Davis, CA 95616

The research objective of this project is to identify computational problems of fundamental importance to molecular biologists engaged in the Human Genome Project, to devise new algorithmic techniques for solving these problems, to program and test the algorithms that are developed, and to make useful computer code available to the biological community. The project also contributes to the training of Ph.D.s in computer science who are qualified to take up careers in Computational Biology.

Most of our work concerns the design and adaptation of data structures and algorithms for solving problems in DNA mapping and sequence analysis. Results obtained under DOE Grant DE-FG03-90ER609999 during the period July 1990 to July 1992 include:

- A new algorithm for sublinear expected time approximate string matching (E. Lawler, W. Chang)
- A faster dynamic programming algorithm for approximate string matching and alignment (W. Chang, J. Lampe)
- Algorithms for construction of phylogenetic trees (E. Lawler, T. Warnow, S. Kannan)
- XPARAL: A software package to efficiently and optimally align sequences using parameterized match, mismatch, indel and gap weights (D. Gusfield, K. Balasubramanian, D. Mayfield, J. Bronder, P. Stelling)
- Mathematic results in parametric sequence optimization (D. Gusfield, K. Balasubramanian, D. Naor)
- Efficient algorithms for multiple sequence alignment with guaranteed error bounds (D. Gusfield)
- A new algorithm for constructing suffix arrays (D. Gusfield)
- An efficient algorithm for the all-pairs suffix-prefix problem (D. Gusfield, G. Landau, B. Scheiber)
- Analysis of solutions to the Probed Partial Digestion Problem (I. Newberg, D. Naor)
- Algorithms for sequence comparison with inversions (J. Kececioglu, D. Sankoff)

A number of these problem areas are under continuing investigation. In particular, XPARAL is currently being adapted to parametric analysis of protein sequences.

Efficient Construction of Large Insert YACs in Recombination-deficient Hosts.

Lo-See Lucy Ling and Donald T. Moir, Department of Human Genetics and Molecular Biology, Collaborative Research, Inc., Waltham, MA 02154

The goal of this research is to construct a human DNA YAC library of clones which will be suitable for both physical mapping and DNA sequencing. To this end, we aim to build a selectable high copy number, large insert YAC library which is low in chimeric YACs and low in YACs with internal deletions. The amplifiable vector will also carry convenient unique restriction sites and tag sequences adjacent to the cloning site to permit affinity capture of YACs, YAC inserts, YAC insert ends, or vector-YAC PCR products.

We are testing the effects of host strain mutations on the frequency of chimeric YACs. So far, we have constructed a library of a few thousand amplifiable YACs averaging ≈400 kb in size in a *rad52* host strain. A similar library is being prepared in a *rad52 rad1* double mutant host strain. We are also increasing the size of ligation products isolated by preparative CHEF gel in order to determine the practical limit for efficient transformation of these strains. The *rad52* and *rad52 rad1* hosts transform about 4- and 8-fold less efficiently with plasmid YRp12 DNA, with 200 kb YAC DNA, and with gel-purified YAC ligation mixtures than does the isogenic parental wild-type strain. The integrity of YACs in these host strains will be analyzed in two ways. First, a large number of YACs from each host will be examined for chimerism by fluorescence *in situ* hybridization to prometaphase spreads under conditions which revealed a 40% chimerism frequency when applied to 90 YACs from the libraries prepared at Washington University (Brownstein et al., 1989) and the CEPH (Albertsen et al., 1990). Second, inter-*Alu* PCR products from the *rad52* and *rad52 rad1* YACs will be used as hybridization probes to isolate Washington University and CEPH library YACs in wild-type hosts from the same chromosomal regions. Restriction maps of the YACs will be generated and compared to determine the effects of the host strains on the frequency of microdeletions or rearrangements.

In collaboration with Dr. Calvin Vary (Maine Medical Center Research Institute), we are testing the feasibility of capturing YACs, YAC ends, and vector-YAC PCR products by affinity of solid-phase bound polypyrimidine third strands to polypurine regions embedded in the vector adjacent to the cloning site. YAC vector modifications in progress include the introduction of 18-base recognition restriction sites flanking the cloning site and the inclusion of lac operator sequences for capture of YACs and inserts by solid-phase bound lac repressor as an alternative to triplex capture.

If these experiments are successful, YACs will provide ample, pure DNA, representative of the human genome, and free of yeast genomic DNA and YAC vector DNA suitable for subcloning into plasmids and determination of the sequence.

Electrophoretic Separation of DNA Fragments in Ultrathin Planar-Format Linear Polyacrylamide

M. T. MacDonell and D. B. Roszak
Ransom Hill Bioscience, Ramona, CA 92065

Linear, or uncross-linked polyacrylamides have been employed successfully in the field of capillary electrophoresis for the separation of nucleic acids. Typical acrylamide concentrations for those applications range from 3% to 14% (w/v), with consistencies ranging from virtually liquid to moderately viscous. Due to the absence of cross-links, and the relatively fluid nature of linear polyacrylamide at typically-employed concentrations, it's use in planar (slab) gel electrophoresis has been overlooked. We describe herein, the application of ultrathin (100 µm) high viscosity slabs of linear polyacrylamide to planar electrophoresis of nucleic acid fragments. The approach we describe is rapid and yields high-resolution separations of nucleic acid fragments in linear polyacrylamide supports. The mobilities of DNA fragments of various lengths in a range of concentrations of linear polymer are compared with those observed for conventional cross-linked gels. The reptative migration of larger DNA fragments in linear polymers is predictable from the models derived from work with cross-linked acrylamide and agarose. The migration of smaller fragments, however, is not entirely predicted by the Ogston model. The relative mobilities observed for very small DNA fragments are approximately half those predicted by the Ogston regime (see Figure).

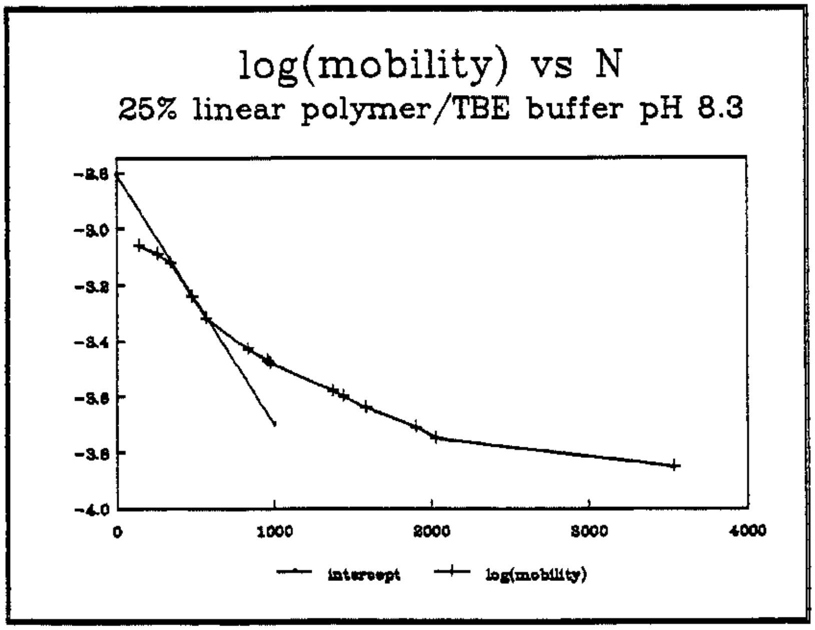


Figure: Plot of log(µ) vs. N for DNA restriction fragments separated in a 25% linear polymer. The straight line indicates the range of fragments migrating by Ogston mobility. Note the mobilities of the smallest fragments deviate from the predicted Ogston mobilities. This is an effect similar to that produced by a field strength gradient

The tendency for smaller fragments to deviate from mobilities predicted by the Ogston model benefits the user since the overall effect is not unlike the employment of a field strength gradient. The smaller bands migrate more slowly than expected. An additional benefit of the use of linear polymers is that, since they are water soluble, quantitative recovery of DNA fragments from these gels requires only that the band be excised and dropped into buffer. The useful range of linear polymer concentrations with respect to planar sequencing applications appears to be 20% to 35% (w/v), appropriate for the electrophoretic separation of DNA fragments in the range of 50 bp to approximately 2500 bp in length.

CHEMILUMINESCENT DNA SEQUENCING WITH 1,2- DIOXETANE SUBSTRATES

Chris S. Martin and Irena Bronstein Tropix, Inc., Bedford, MA 01730

We have developed new methods for non-isotopic DNA sequencing based on chemiluminescent 1,2-dioxetane enzyme substrates. Standard Sanger dideoxy DNA sequencing and multiplex DNA sequencing procedures have been adapted for chemiluminescent signal detection. Coupling multiplex DNA sequencing and chemiluminescence detection is particularly advantageous as the use of large quantities of radioisotopes can be eliminated. Furthermore, the multiplex procedure already requires that the DNA is transferred to a membrane support. While much of our original development work was performed with the 1,2-dioxetane substrate, AMPPD, currently, we have incorporated an improved substrate, CSPD, disodium 3-(4-methoxyspiro[1,2-dioxetane-3,2'-(5'-chloro)tricyclo [3.3.1.1^{3,7}]decan]-4-yl)phenyl phosphate, which shows improved performance in DNA sequencing applications. CSPD exhibits faster kinetics of light emission on nylon membranes and greater DNA band resolution when the results are imaged. Recently, we have investigated an alternative procedure for DNA sequencing which permits an increase in the amount of DNA sequencing information that can be obtained from a round of electrophoresis. This approach, termed-label multiplexing-enables the acquisition of a greater amount of DNA sequence data from a single membrane by imaging overlapping sets of DNA sequencing reactions consecutively using hapten specific enzyme conjugates.

We have further increased the efficiency of chemiluminescent DNA sequencing by utilizing a direct transfer electrophoresis apparatus, Betagen Autotrans 350. With direct transfer electrophoresis, DNA fragments are deposited on a membrane which moves along the bottom of the electrophoresis plates. This enhances band resolution and separation of high molecular weight DNA fragments. Direct transfer electrophoresis has been successfully used by other investigators working with large scale multiplex sequencing projects to accelerate data acquisition.

We have also investigated the feasibility of cooled CCD cameras in digitizing DNA sequence data directly from the chemiluminescent membrane. These cameras enable the acquisition of high quality DNA sequence data which can be further analyzed with available DNA Sequencing software. Limitations of this system include; the size of the CCD array required for adequate resolution of DNA bands and length of the integration periods necessary to generate acceptable signal to noise ratios.

Finally, our efforts in coupling more efficient DNA sequencing methodologies such as label multiplexing with automated processing; direct transfer electrophoresis, CCD imaging and data digitization, indicate that the chemiluminescent detection technique has universal application and is superior to other alternatives.

This work was supported by grants from the US DOE contract No., DE-AC05-89ER80752, DE-FG05-90ER81063, and DE-FG05-92ER81389.

CHEMILUMINESCENT DNA SEQUENCING USING MULTIPLEX LABELING

Chris S. Martin, Corinne E. M. Olesen, and Irena Bronstein Tropix, Inc., Bedford, MA 01730

We have developed a non-isotopic DNA sequencing method utilizing chemiluminescent 1,2-dioxetane substrates for alkaline phosphatase. With this procedure, standard Sanger dideoxy DNA sequencing reactions are initiated with biotinylated primers, separated by gel electrophoresis and transferred to nylon membrane. The chemiluminescent detection procedure is subsequently performed on the membrane which provides a solid support for the DNA fragments during the binding of alkaline phosphatase conjugates and washing steps. The membrane also acts as an enhancer of the chemiluminescent signal. We present here a method for increasing the efficiency of the chemiluminescent detection procedure by maximizing the amount of DNA sequence data that can be obtained from a single membrane. Our strategy is to separate multiple sets of DNA sequencing reactions in each gel lane, transfer the DNA to nylon membrane and subsequently detect each set of sequencing reactions individually. Primers labeled with different haptens at the 5' end are used in separate sequencing reactions, mixed together prior to electrophoresis and the individual templates detected using hapten specific protocols. This enables a significant increase in productivity per gel run without the use of the multiple cloning vectors or hybridization steps required in current multiplex DNA sequencing systems. We have utilized primers labeled with biotin, digoxigenin and fluorescein, each of which is consecutively detected with hapten specific alkaline phosphatase conjugates and CSPD 1,2-dioxetane chemiluminescent substrate. In order to further increase the amount of DNA sequence data which can be imaged from a single membrane, we have utilized a Betagen AutoTrans 350 direct transfer electrophoresis (DTE) apparatus for simultaneous separation of the DNA sequencing reactions and transfer to nylon membrane. The increased separation of the high molecular weight fragments enables the acquisition of 350-450 base pairs of DNA sequence data from each template. By combining the multiplex labeling technology with direct transfer electrophoresis, 21 templates or 6000-8000 bases of sequence information can be analyzed from a single 15 x 40 cm membrane.

This work was supported by the U.S. Dept. of Energy, Grant No. DE-FG05-92ER81389.

Surveys of State Insurance Commissioners and Insurance Company Medical Directors on the Use of Genetic Information in Life Insurance; Review of Related Legislation.

Jean E. McEwen*, **; Katharrine McCarty*; and Philip R. Reilly*

*Eunice Kennedy Shriver Center for Mental Retardation, Division of Social Sciences, Ethics and Law, 200 Trapelo Road, Waltham, MA 02254
**Boston College Law School, 850 Centre Street, Waltham, MA 02159

Rapid advances in our ability to test persons for genetic diseases and genetic traits has generated increasing concern that genetic information will be abused by insurance companies. The potential for abuse may be greatest in the area of life insurance, which unlike health insurance, is generally purchased by individuals. Thus, while most insurers could have some interest in obtaining genetic information, life insurers may have the strongest incentive to use genetic data to rate individual applicants. Thus far, however, reports of genetically based discrimination in life insurance have been based solely on reports from individuals.

We conducted a survey of state insurance commissioners (those responsible for regulating the insurance industry) to determine: (1) the extent of genetic information about applicants that life insurers can legally request, the types of tests they can perform, and the way they can use the data obtained; (2) the extent to which life insurers would be permitted to use predictive tests for various genetic conditions if such tests were available; (3) the extent to which commissioners have received complaints by consumers of genetically based discrimination by life insurers; (4) the extent of insurance commissioners' awareness of and involvement in legislative activities regarding the use of genetic data.

We conducted a companion survey of the medical directors of North American life insurance companies to determine: (1) current practices regarding the accessing and use of genetic information; (2) the extent of their interest in genetic testing; (3) whether they have actuarial data relating to particular conditions; (4) the procedures they use to ensure confidentiality and to communicate test results; (5) their involvement in relevant legislative activities; and (6) their professional backgrounds and level of knowledge about genetics. In both the insurance commissioner and the medical director surveys, we explored some topics using brief case studies. We also reviewed the application forms that a number of life insurers use and analyzed all legislation (both enacted and proposed) that purports to regulate access to and use of genetic data.

The results of our survey of insurance commissioners uncovered little evidence to suggest that these regulators perceive that genetic testing currently poses a significant problem in how life insurers rate applicants or that many consumers are presently filing formal complaints about the use of genetic data by life insurers. However, they also suggest that considerable uncertainty exists among regulators regarding the role of genetic information in life insurance underwriting. In addition, these results, coupled with our own review of the relevant legislation, show that should life insurers increasingly wish to make use of genetic data in the future, they currently have wide latitude to do so. The medical director survey demonstrated that many companies, concerned about the prospect of adverse selection by applicants, already access and make use of much family history and related information, even if they do not actually perform or require genetic testing. This survey also showed that many companies are interested in using genetic testing, at least before issuing policies with high face values, if tests that could determine greatly increased risks of developing particular conditions become available. However, we also found that few companies have relevant actuarial data to support their underwriting decisions, that few have written procedures to govern the dissemination of test results, and that many medical directors' knowledge of genetics is limited.

Rapid Shotgun Cloning Utilizing The Two Base Restriction Endonuclease *CviJ*I

David A. Mead, Chemistry Department, University of Wisconsin, Madison, WI 53706

A new approach has been developed for the rapid fragmentation and fractionation of DNA into a size suitable for shotgun cloning and sequencing. The restriction endonuclease *CviJ*I normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions which alter the specificity of this enzyme (*CviJ*I**) yield a quasi-random distribution of DNA fragments from the small molecule pUC19 (2686 base pairs). To quantitatively evaluate the randomness of this fragmentation strategy, a *CviJ*I** digest of pUC19 was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lacZ minus M13 cloning vector. Sequence analysis of 76 clones showed that *CviJ*I** restricts PyGCPy and PuGCPu, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation. Advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 ug instead of 2-5 ug), fewer steps are involved (no pre-ligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed), and higher cloning efficiencies are obtained (*CviJ*I** digested and column fractionated DNA transforms 3-16 times more efficiently than sonicated, end-repaired, and agarose fractionated DNA).

AUTOMATED MULTIPLEX SEQUENCING, PROTEINS, cDNAs, AND TANDEM REPEATS

Alec Mian, Andrew Link, Nathan Lakey, Hamid Ghazizadeh, Laura Jaehn, Peter Richterich*, Keith Robison, and George M. Church. Harvard Medical School and Howard Hughes Medical Institute, Boston, MA 02115. *Collaborative Research Inc., Waltham, MA.

We have recently integrated direct transfer electrophoresis, automated multiplex hybridizations and automated film reading to sequence E. coli cosmids. Sequence patterns for two cosmids were detected using chemiluminescence with oligonucleotide probes directly conjugated to alkaline phosphatase. Primers for the directed walking and sequence confirmation steps were synthesized with a 15 base tag complimentary to a alkaline phosphatase conjugate for dideoxy sequencing. Sequence patterns for a third cosmid were detected using radiolabled oligonucleotide probes. Film data were automatically read and assembled using the programs REPLICA and GTAC. For the cosmids, 20 gels resulted in 9216 sequences on film. We have developed a program which automatically finds and graphically annotates proteins and ORFs including matches to database sequences. A systematic database of amino terminal protein sequences, pI, MW, abundance checks accuracy of DNA sequences and gene expression modeling. A new method for cDNA sequencing employs T4 RNA ligase to accurately determine the terminal bases. This has been applied to non-polyadenylated transcripts from rDNA tandem promoter repeat regions.

SEQUENCING BY HYBRIDIZATION WITH OLIGONUCLEOTIDE MATRIX

A.Mirzabekov, I.Ivanov, G.Ershov, V.Florentiev, Yu.Lysov, V.Barsky, E.Kreindlin
Engelhard Institute of Molecular Biology, Russian Academy of Sciences, Vavilov str. 32, Moscow 117984

We have further developed the DNA sequencing technique based on its hybridization with an oligonucleotide matrix (SHOM). Sequencing "microchip" consists of a glass plate covered with 20 micron-thick polyacrylamide gel squares (from 30x30 to 100x100 microns) containing individual immobilized octanucleotides. A theory has been developed to explain the experimental data that apparent thermostability of the DNA duplexes with gel-immobilized oligonucleotides increases with the increase in the oligonucleotide concentration and gel thickness. These properties of the chip enable us to adjust concentration of immobilized octanucleotides to equilize the thermostability of A-T- and G-C-rich duplexes. A prototype automatic sequenator has been constructed which consists of microchips, fluorescent microscope, CCD-camera, thermostated plate, special software for image analysis (manufactured in Russia) and computer (imported). The conditions for SHOM have been optimized. We have worked out a technology for manufacturing chips containing tens of immobilized oligos to be used for genetic mutation analysis. We plan to produce chips containing hundreds of immobilised oligos for sequence comparison, mapping and getting sufficient volume of data to build a theory for predicting stabilities of different duplexes. Computer simulations have demonstrated that SHOM with octanucleotide matrix together with continuous stacking hybridization can be effective for sequencing about 3000-5000 nucleotide long DNA.

STUDIES ON GENETIC DISCRIMINATION

Genetics Screening Study Group

Carol Barash, **Marvin Natowicz**, Paul Billings, Joe Alper, Jon Beckwith, Lisa Geller, Andy Ahn, Carol Martin, Catherine Ard and members of the Genetic Screening Study Group; Shriver Center, Waltham, MA 02254

The major objective of the study is to assess the significance of genetic discrimination in our society. Genetic discrimination refers to discrimination directed against an individual and/or members of that individual's family solely because of real or perceived differences in the genetic constitution of that individual. Few studies and limited data exist regarding this issue.

The specific aims of our study are: (1) to determine the particular social institutions that might engage in discriminatory practices such as insurance companies, governmental agencies, employers, educational institutions and the military; (2) to evaluate the nature of the discrimination experienced by individuals and families; and (3): to determine the underlying basis of the discrimination and, in particular, whether it is the result of ignorance or policy.

Questionnaires were mailed to approximately 30,000 persons with single gene disorders, persons at-risk for certain single gene disorders, and asymptomatic heterozygotes. Our sample consists of persons who are carriers for autosomal recessive conditions, e.g. parents of children with mucopolysaccharide disorders, phenylketonuria, and sickle cell disease; persons who are at-risk for Huntington disease, an autosomal dominant condition; and asymptomatic persons with hemochromatosis, an autosomal recessive condition. We are continuing to receive responses and have completed detailed follow-up interviews with heterozygotes for mucopolysaccharidoses and persons at-risk for Huntington disease.

Preliminary results indicate discrimination in a wide variety of contexts including insurance (health, disability, and life), employment, the military, educational institutions and professional training programs, and adoption agencies. Our case analyses also reveal psychological burdens and social stigmas resulting specifically from genetic information. Finally, our results indicate that some persons who were aware of their risk for genetic discrimination took steps to make sure that their condition would not become known or that knowledge of the condition would not harm them.

RAPID IMAGE ANALYSIS OF RADIOISOTOPIC AND CHEMILUMINESCENT SAMPLES

Quan Nguyen, Frank Witney, Dave Heffelfinger, Will Stubblebine and Charles Ragsdale, Genetic Systems Division, Bio-Rad Laboratories, 2000 Alfred Nobel Drive, Hercules, CA 94547

The detection of specific DNA or RNA sequences by genomic Southern or northern hybridization analysis techniques typically employs a radiolabeled probe, which is visualized by autoradiography. The use of a photostimulatable phosphorescent compound-based imaging systems provides an alternative to autoradiography for detection of the ß-particles emitted by these radiolabeled probes. This storage phosphor imaging system has the advantage of being more sensitive to the ß-radiation (10 to 20 fold) with a greater linear dynamic range (over 5 orders of magnitude) than x-ray film.

Recently, a number of non isotopic methods for nucleic acid detection have been developed. However, non isotopic techniques have not gained widespread acceptance due to low sensitivity, high backgrounds, and inconsistent results. We have developed a two-step hybridization method for genomic Southern and northern applications (Biotechniques **13:** 116-123, 1992) which eliminates the major problems associated with standard non isotopic methods. In addition, we have developed a storage phosphor imaging system based upon novel photostimulatable phosphorescent compounds, which can detect the photons emitted by the chemiluminescent substrates (AMPPD and luminol) commonly employed in non isotopic methods. This system exhibits the same advantages with respect to linear dynamic range for both chemiluminescent and ß-particle emitters.

PCR Mapping of Human cDNA Clones

William C. Nierman, Donna R. Maglott, and A. Scott Durkin, American *Type Culture* Collection, 12301 Parklawn Drive, Rockville, MD 20852

The laboratory of J. Craig Venter, NINDS, is identifying genes expressed in the human brain by sequencing portions of cDNA clones (Adams *et al.*, Science 252:1651-1656, 1991; Adams *et al.*, Nature 355:632-634,1992). In collaboration with the Venter group, we are developing methods for rapid localization of newly identified cDNA sequences to human chromosomes. We are using the ABI automated DNA sequencer to analyze fluorescently-tagged PCR products. Primer pairs are designed from the partial cDNA sequence data and tested for specific amplification from human genomic DNA. Primers permitting resolution of human products from rodent products are tested with DNA from somatic cell hybrid cell mapping panels. The presence or absence of specific amplification products in each cell line DNA is determined electrophoretically using the ABI sequencer, and chromosomal assignments are made by discordancy analysis. We are developing methodology for multiplexing the amplification reactions and analysis of the reaction products, to achieve a high throughput with a minimum allocation of resources. This project will determine chromosomal assignments for "Expressed Sequence Tags" (ESTs), provide primer sequence data for subsequent subchromosomal localizations, and generate a broad data set from which to evaluate strategies to identify functional primer sequences from cDNA sequence data.

Analysis of Oligodeoxynucleotides by Matrix Assisted Laser Desorption/Ionization Mass Spectrometry and Its Potential for DNA Sequencing

Gary R. Parr, Michael C. Fitzgerald, Lin Zhu and Lloyd M. Smith
Department of Chemistry, University of Wisconsin, Madison, WI 53706

Analysis of intact, high molecular weight biopolymers by mass spectrometry is now possible due to the recent development of Matrix Assisted Laser Desorption/Ionization (MALDI). In this technique, an analyte and a matrix compound are co-deposited on a mass spectrometer probe tip and irradiated with a high intensity, pulsed laser beam. The energy of the laser pulse is absorbed by the matrix compound, resulting in the ejection, or sublimation, of matrix and analyte molecules into the gas phase. The matrix and laser wavelength are chosen so that little, if any, light energy is deposited in the analyte, resulting in the vaporization of *intact* biopolymer molecules. Subsequent ionization of the analyte probably occurs through gas phase ion/molecule reactions. These ions can be accurately mass analyzed, usually with a time of flight (TOF) mass spectrometer. A complete mass spectrum can be obtained in a few seconds.

The MALDI technique appears to be generally applicable to proteins with no limitations imposed by primary, secondary or tertiary structures. Proteins as large as 275,000 Da have been successfully mass analyzed. Furthermore, complex mixtures of proteins can be readily resolved. If these characteristics (i.e., analysis of mixtures of large, intact nucleic acids irrespective of base composition) can be achieved for nucleic acids, this technique could be a powerful new method for DNA sequence analysis.

Results in our laboratory suggest that matrix materials and/or conditions successfully used for protein analysis are generally useful for oligonucleotides only up to about ten bases in length. For larger oligonucleotides, base composition appears to be an important parameter. For example, oligodeoxythymidylic acids up to 100 bases can be mass analyzed, while oligodeoxyguanylic acids larger than 8 bases cannot be mass analyzed using the same conditions. Mixtures of oligodeoxythymidilic acids of varying length can also be readily analyzed. Thus, the basic requirements for sequence analysis can be met with poly dT. Nonetheless, for this technique to be useful in DNA sequence analysis, dependence on base composition must be eliminated. Toward this end, we are currently investigating new classes of matrix compounds covering a wide range of pH. We are also attempting to understand the nature of oligo/matrix interactions using fluorescent dye labelled oligonucleotides.

Isolation and characterization of expressed sequences from the human X chromosome

Julia E. Parrish and David L. Nelson
Institute for Molecular Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030 713-798-4787, -5386fax, nelson@bcm.tmc.edu

A long term goal of our group is the characterization of the majority of transcribed sequences localized to the human X chromosome. Toward that end, we have begun to analyze previously assigned expressed sequence tags (ESTs) as well as to develop methods for isolating transcribed sequences associated with cloned fragments of the chromosome.

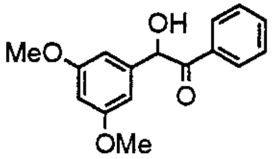
Primers have been synthesized for the 14 ESTs currently assigned to the X in the Genome Data Base, and the corresponding cDNA clones have also been obtained. The primers specific to each locus have been used in PCR of somatic cell hybrids containing fragments of the chromosome providing localization to intervals. These ESTs are also being used to seed contigs of YACs from the Houston, Philadelphia and CEPH YAC libraries, as well as to isolate cosmids from a flow-sorted X library obtained from LLNL. More refined assignment of the clones using FISH mapping will be performed using the cosmids as hybridization probes. EST mapping to new and pre-existing YAC contigs will allow very fine localization.

cDNA clones from which the X-linked ESTs were developed will be used in Southern, northern and in situ RNA analysis to characterize the associated genes in more detail. Additional cDNAs are being sought using several methods, including a variation of an exon-trapping system developed at Baylor, cDNA fishing and direct hybridization. One such clone, H10, isolated with a probe derived from the fragile X locus and found to hybridize to multiple locations on the X is currently serving as a model for characterization. Complete insert sequencing will be carried out for each clone, in order to more fully characterize these genes by comparison to known sequences and for potential identification of additional cDNA clones representing full or nearly full-length sequence.

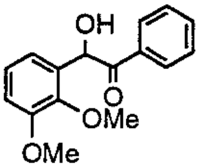
Preparation of Oligonucleotide Arrays for Hybridization Studies

Michael C. Pirrung, Steven W. Shuey, and David C. Lever
P. M. Gross Chemical Laboratory
Department of Chemistry, Duke University
Durham, NC 27708-0346

Photoremovable groups are key to the novel technique of light-directed synthesis, whereby the preparation of large arrays consisting of thousands of biopolymer sequences can be accomplished.[1] Arrays consisting of complete sets of DNA sequences of a given length can be used to sequence DNA by hybridization, using principles enumerated by several groups.[2]



3,5-DMBOH



2,3-DMBOH

The purpose of this project is to develop the method of photochemical DNA synthesis for use in light-directed synthesis. A photoremovable protecting group is required for the phosphotriester method of DNA synthesis. We relied on earlier work of Sheehan concerning benzoin and 3',5'-dimethoxybenzoin (DMB) esters as photoremovable carboxylate protecting groups[3] to develop two new optically-active benzoins to protect phosphotriesters. Since benzoin phosphotriesters are stereogenic at phosphorous and in the benzoin, they can exist as mixtures of diastereomers, which causes difficulties in the chemical characterization of the materials. When these esters are derived from alcohols such as nucleosides that themselves bear stereogenic centers, a complicated mixture results. An asymmetric synthesis of the desired benzoins has been accomplished via the cyanohydrins (method of Oguni[4]) that minimizes the number of diastereomers formed in protection of nucleosides. The compounds (+)-3,5-DMBOH and (+)-2,3-DMBOH were selected on the basis of an extensive investigation of substituted benzoin acetates for yield (both are quantitative) and rate of deprotection. Each was also converted to its diethyl phosphate and irradiated in a Rayonet reactor at 350 nm for 30 min. They show the following half-lives: 2,3-DMBOP, 4.18 min; 3,5-DMBOP, 3.78 min; 3,5-DMBOAc, 4.18 min. All four bases were protected with a 3'-DMB-phosphate, deprotected by irradiation, and coupled to form dinucleotides. These results lay the groundwork for the synthesis of DNA arrays photochemically using a lithographic synthesis instrument now under construction.

References

1. Fodor, S.P.A.; Read, J.L.; Pirrung, M.C.; Stryer, L.; Liu, A.T.; Solas, D. *Science* **1991**, *251*, 767.
2. Bains, W.; Smith, G. C. *J. Theor. Biol.* **1988**, *135*, 303-307. Drmanac, R.; Labat, I.; Brukner, I.; Crkvenjakov, R. *Genomics* **1989**, *4*, 114-128. Khrapko, K. R.; Lysov, Y. P.; Khorlyn, A. A.; Shick, V. V.; Florent'ev, V. L.; Mirzabekov, A. D. *FEBS Lett.* **1989**, *256*, 118-122. Lysov, Y. P.; Florent'ev, V. L.; Khorlyn, A. A.; Khrapko, K. R.; Shick, V. V.; Mirzabekov, A. D. *Dokl. Akad. Nauk. SSSR* **1989**, *303*, 1508-11.
3. Sheehan, J.L.; Wilson, R.M.; Oxford, A.W. *J. Am. Chem. Soc.* **1971**, *93*, 7222-7227.
4. Hayashi, M.; Matsuda, T.; Oguni, N. *J. Chem. Soc., Chem. Commun.* **1990**, 1364.

CANCER AND THE GENOME EXPOSURE REACTION

Theodore T. Puck, Alphonse Krystosek, Kim Marcom, Robert Johnson, Patricia Webb, and Patricia Griffin
Eleanor Roosevelt Institute for Cancer Research and University of Colorado Cancer Center, Denver, CO 80206

Genome exposure is defined as a reaction which causes specific loci in the mammalian cell nucleus to become differentially susceptible to interaction with reacting molecules in the fluid medium. It is measured by increased sensitivity of specific DNA regions to hydrolysis by DNase I added to isolated nuclei or to cells fixed on microscope slides. The location of exposed DNA can be visualized by the nick translation reaction in which labeled bases are incorporated by DNA polymerase into the discontinuities produced by DNase I. In all normal cells examined a region of exposed DNA exists around the nuclear periphery. However, this region is absent or greatly diminished in each of sixteen different cancers so far studied. In addition, a region of exposed DNA is found around the nucleoli in both normal and cancer cells. The difference in pattern of genome exposure between normal and cancer cells can be displayed by direct biopsy cells as well as in tissue culture.

Because of the loss of nuclear, peripheral, exposed DNA in cancer cells, it is postulated that this exposed DNA represents DNA organizational patterns specific to particular differentiation states. The nucleolar exposed DNA presumably represents active DNA sequences common to normal and malignant cells.

The position of the region of nuclear exposed DNA coincides approximately with the location of the lamins A, B, and C and a fraction of topoisomerase II but not topoisomerase I in the nucleus. The cytoskeleton is essential for genome exposure since treatment of cells with agents like colcemid or cytochalasin B, prevents the exposure reaction. Steroids and lipoids like hydrocortisone, prostaglandin E2, retinoic acid and dexamethasone, and phosphorylation modifiers like herbimycin A also can affect the distribution pattern of exposed DNA in particular cells. Fluorescence-in situ-hybridization with DNA probes in interphase cells can be used to demonstrate the location of specific chromosomal loci within the exposed region of particular nuclei.

We hypothesize that genome exposure is the mechanism underlying specific differentiation of mammalian cells, and is regulated by agonists attaching either to membrane receptors which change the cytoskeleton or to cytoplasmic receptors which act on the nucleus directly.

Physical Mapping, Marker Saturation and Identification of Transcribed Sequences From a YAC in Xq28.

Orly Reiner[1], Manfred Wehnert [1], and C. Thomas Caskey[1,2].

[1] Institute for Molecular Genetics, Baylor College of Medicine, Houston, Texas, USA

[2] HHMI, Baylor College of Medicine, Houston, Texas, USA

YAC XY845 covers a region of 500kb, that include several well known and commonly used polymorphic markers, DXS15, DXS52, DXS134. Many disease loci in Xq28 have shown high LOD scores to these markers, especially to DXS52 which is highly polymorphic. We have composed a physical map of this YAC and identified several CpG islands. The YAC has been subcloned into lambda and cosmid vectors and a contig encompassing most of the YAC was generated using *Alu* PCR and fingerprinting of clones containing STRs. STRs were used as physical and genetic markers. The oligo fingerprinting revealed phage clones containing nine different STRs. Seven STRs have been subcloned and sequenced.

Four characterized highly polymorphic STRs provide a marker saturation in a physically defined region of 100 kb between DXS52 and DXS15 useful for fine mapping studies at Xq28. The same strategy will be employed to generate a physically defined marker saturation in the DNA region between F8 and DXS15/DXS52, that can be used for fine mapping of disease loci like Barth syndrome (X-linked myotubular myopathy), X-linked Hydrocephalus (HSAS), Adreno-leukodystrophy (ALD), Emery-Dreifuss muscular dystrophy (EMD) and others.

A minimal coverage of the YAC was obtained by twelve phages and six cosmids. These cosmids and phages were digested and hybridized to zoo blots to identify conserved sequences. Specific conserved fragments were used against Northern blots, and as probes for cDNA libraries. Other cDNAs that hybridize to the YAC were isolated using degenerate PCR probe. Further characterization of these clones may provide the molecular basis for one or more disease loci in this region.

The Use of Modified Streptavidins in Genome Mapping, Sequencing, and Other Sensitive Analyses

Takeshi Sano[1], Takashi Ito[2] Cassandra L. Smith[1] and Charles R. Cantor[1]
Center for Advanced Biotechnology, Boston University[1] and University of Tokyo[2]

The great specificity and remarkable affinity of streptavidin for binding biotin has made this system a centerpiece for a number of sensitive bioassays. We have greatly expanded the capabilities of the system by the creation of a series of genetically engineered streptavidins. Cloned core streptavidin, expressed in *E. coli*, using Studier's tightly regulated T7 system, can be purified to homogeneity by affinity chromatography on 2-iminobiotin. It is a homogeneous product, unlike the protein normally used, and it is very soluble. We have also made core streptavidin containing five terminal cysteine residues, to facilitate attachment to solid supports. In addition, several chimeric streptavidins have been made, including a fusion of streptavidin and metallothionein and a fusion of streptavidin with two domains of staphylococcal protein A. The former allows labeling of biotinylated targets such as DNA with large numbers of metals. The latter allows direct conjugates to be made between biotinylated DNA and immunoglobin G's.

Several applications of these new streptavidins will be illustrated. We have been testing a variant of sequencing by hybridization in which a constant length of duplex DNA containing an overhanging single-stranded probe region is used to detect the end of a target DNA by stacking hybridization followed by DNA ligation. Here, streptavidin allows a very convenient way to position the DNA reagent on a magnetic microbead, for current test purposes. Ultimately the same mode of attachment could be used for membrane surfaces. We have developed a technique for capturing a specific target DNA sequence during electrophoresis. This employs a gel insert with biotinylated DNA probes attached to streptavidin beads. Finally, we have demonstrated one potential advantage of assay methods that combine DNA and immunoglobulins by using the streptavidin-protein A chimera to label a monoclonal antibody with DNA and then performing PCR on the attached DNA to detect an immobilized antigen. This new technique, which we have called immuno-PCR, has 10^5 times the sensitivity of conventional ELISA assays.

OPTIMIZATION OF FIELD INVERSION GEL ELECTROPHORESIS FOR SEPARATION BELOW 150 KB.

Peter A. Schad and Paul W. Zoller. Genetic Systems Division, Bio-Rad Laboratories Inc., 2000 Alfred Nobel Drive, Hercules, CA 94547.

We have developed a pulsed field gel electrophoresis unit and protocols utilizing field inversion gel electrophoresis (FIGE) in order to optimize the separation of DNA fragments up to 150 kb in size. FIGE has several advantages when routine separation of small fragments (< 150 kb) is desired such as: the use of standard agarose electrophoresis gel boxes, room temperature runs, and reduced number of program variables. In addition, previous data has shown that asymmetric voltage FIGE has proven to be the preferred method of choice for separation of DNA from 3 to 50 kb, when compared to other field inversion methods such as asymmetric switch times and CHEF with 120° pulsed field angles. The ability to modify asymmetric voltages, switch times, or both simultaneously, and non-linear ramping provides versatility in optimizing ideal field inversion separation protocols.

The purpose of the present study was to develop conditions that optimize DNA separations up to 150 kb. The key feature is the inclusion of ten defined programs which cover the full separation range along with various subsets ranges. The data presented in this study will show separations achieved with the defined programs along with the effects of non-linear switch time ramps on increasing the separation linearity of various regions. Current experiments are focused on investigating the use of asymmetric voltages simultaneously coupled with asymmetric switch times. In addition data is presented showing that FIGE is useful in clinical applications such as bacterial RFLP analysis in nosocomial outbreaks.

TIME-OF-FLIGHT MASS SPECTROMETRY OF DNA FOR ACCELERATED SEQUENCING

David Schieltz, Chau-Wen Chou, Cong-Wen Luo, David Dogruel, Robert M. Thomas and Peter Williams

Dept. of Chemistry, Arizona State University, Tempe, AZ 85287-1604

We are working to accelerate DNA sequencing technology by developing a capability to size-sort Sanger sequence fragments using time-of-flight mass spectrometry. A key difficulty is the requirement for gas phase molecules; even quite small DNA molecules cannot be volatilized thermally without decomposition. We have shown that DNA molecules are ejected intact into the gas phase when a thin frozen film of a buffered aqueous DNA solution is ablated from an oxidized copper substrate by a focussed 8 ns pulse from a dye laser operating in the visible. Absorption of ~ 10 J/cm^2 in the copper oxide surface produces a high-pressure plasma which drives a shock wave through the ice film sufficient to explosively ablate the ice. The result is a supersonic water vapor plume which propels the DNA molecules into the gas phase and, through expansion, also cools and stabilizes them. Double-stranded DNA in excess of 600 base pairs (> 400 kilodaltons) has been ablated intact in this way. Fragmentation-free ionization appears to occur by attachment of sodium ions from the buffer solution or copper ions from the substrate to the ablated DNA molecules. A mass spectrum of the resulting molecular ions is obtained in a time-of-flight mass spectrometer. Early results show that ablation of mixtures of single-stranded DNA containing oligomers up to 60 nucleotides in length can produce simple mass spectra dominated by a singly-charged molecular ion peak for each DNA segment (Figure 1). The prospects for accelerated sequencing using this approach will be discussed.

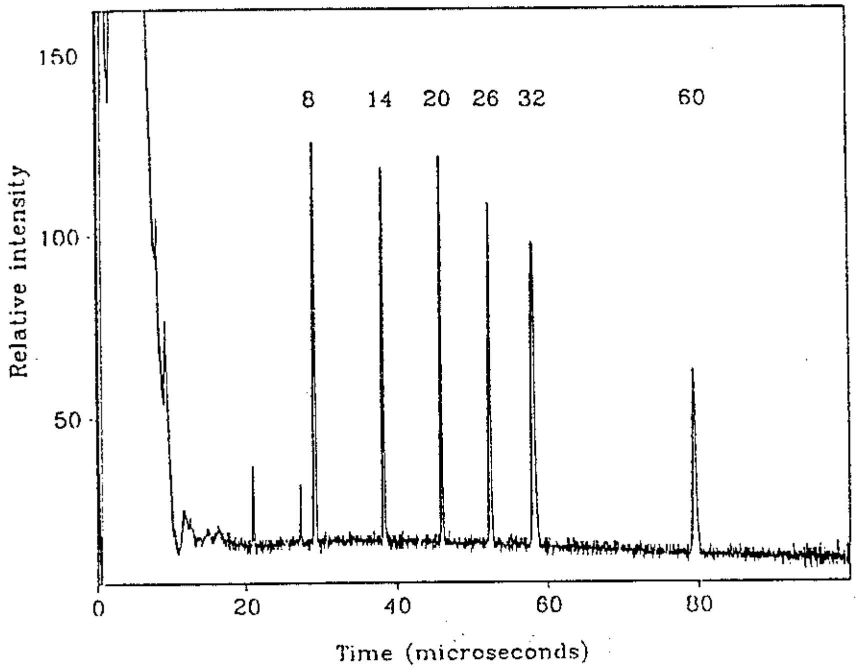


Fig. 1. Time-of-flight mass spectrum of 6-component mixture of single-stranded DNA oligomers: 8, 14, 20, 26, 32, 60 nucleotides. Laser wavelength 578 nm; pulse length ~ 8 ns; pulse energy ~ 10 J/cm^2 @ 1 Hz. 9-shot average. The small peaks ~ 21 and 27 μs appear to be impurities.

Matrix-Assisted Laser Desorption Mass Spectrometry of Oligonucleotides

Klaus Schneider and Brian T. Chait
The Rockefeller University, 1230 York Avenue, New York, N.Y. 10021

Matrix assisted laser desorption mass spectrometry (LDMS) is a new technique that enables the rapid, accurate analysis of proteins in the mass range extending up to a few hundred thousand daltons. If the method can be extended, with similar success, to the analysis of DNA fragments, then LDMS may provide a rapid alternative to gel electrophoresis for DNA sequence analysis. Several groups have recently reported the use of LDMS to analyze relatively short stretches of single stranded DNA (generally < 60 nucleotides in length). However, a general finding of these experiments is that the mass spectrometric response falls off precipitously for DNA fragments longer than approximately 20 nucleotides in length.

During the past year we have carried out a series of experiments designed to understand this falloff in the mass spectrometric response as a function of oligonucleotide length. For these first experiments, we chose to investigate a series of simple nucleotide homopolymers. In agreement with previous experiments of L. Smith and coworkers at the University of Wisconsin, we have found that dT homopolymers yield intense mass spectra for species containing as many as 180 bases (i.e. a molecular weight of 54,000 daltons). By contrast, we found that the mass spectrometric responses to dA, dC, and dG homopolymers were virtually non-existent, except for very short polymers. We discovered 20 different laser desorption matrix compounds that were found to produce mass spectra of dT_{18}. Significantly, none of these matrices produced usable spectra from dG_{18}.

We hypothesize three possible causes for the lack of response to polyguanosine: (i) low ionization efficiency, (ii) high degree of fragmentation, and (iii) unwanted aggregation of polyguanosine chains and/or "non-ideal" interactions of the polyguanosine molecules with the matrix molecules. A series of experiments will be described that explore these different possible sources for the lack of LD mass spectrometric response to polyguanosine:

(i) Measurements on very short stretches of polyguanosine (e.g. dG_7) gave a reasonably intense response, indicating that low efficiency for ionizing polyguanosine is not the primary reason for our failure to observe longer stretches of polyguanosine.

(ii) The mass spectra of dG_7 showed evidence for a substantial degree of fragmentation of the parent ion in the mass spectrometer. If the amount of fragmentation increases as a function of polymer length, such mass spectrometric fragmentation may explain our inability to observe longer stretches of polyguanosine (e.g. dG_{18}). Experiments on chemically modified polynucleotides (e.g. polyinosine) provide independent evidence that the detailed chemical nature of the bases have a profound effect on the amount of fragmentation observed.

(iii) Our working hypothesis is that a good matrix-assisted LD mass spectrometric response requires appropriate interaction of DNA with the matrix molecules, leading to the incorporation of analyte into the matrix materials as isolated molecules. Results of these studies will be presented.

Linguistic Analysis of DNA

David B. Searls   dsearls@cbil.humgen.upenn.edu
Department of Genetics, University of Pennsylvania School of Medicine

"Linguistic" approaches to biological sequence analysis are increasingly in evidence, due in part to the popularity of the common metaphor of DNA as language—a language, that is, describing the structures and processes required (at a minimum) to establish and maintain life. Much of the work in this area deals with analyses of "vocabularies" in the tradition, by and large, of classical linguistics. The author's approach, however, is more concerned with analysis of higher-order structural aspects of DNA, in the sense of Chomsky and the mathematical discipline of formal language theory. This allows for rule-based descriptions of features up to the level of genes with generative *grammars*, and the analysis of sequence data by *parsing*.

Thus the question naturally arises as to where the language of DNA is situated relative to known language classes. Pattern-matching algorithms now used for DNA sequences are largely based on regular expression search; in the Chomsky hierarchy the corresponding *regular* languages (RLs) are at the lowest level of expressive power. Yet, we have shown that such biologically important features as inverted repeats (and the corresponding secondary structures, e.g. stem-and-loop formations) belong to the class of *context-free* (CF) languages and not RLs. In fact, certain important forms of secondary structure are *non-linear, non-deterministic,* and *inherently ambiguous*—all formally-defined language-theoretic properties which have significant consequences for general algorithmic approaches to recognition. Generalized CF grammars have now been written for so-called "orthodox" secondary structure of nucleic acids, as well as specific grammars to recognize important instances of such structure, such as tRNA genes and certain introns, in primary sequence data.

Other features of the language of DNA suggest that even CF grammars may not suffice—for instance, tandem repeats formally belong to the non-CF *copy languages,* and non-orthodox secondary structures called *pseudoknots* are not CF. The secondary and tertiary structure of proteins and the resulting interactions between residues suggests that genes themselves are also CF or greater, as does the non-determinism of gene expression. A class of languages lying between CF and context-sensitive, called *indexed* languages (ILs), appears to adequately handle all these phenomena. Genomic rearrangement may have significant effects on the linguistic complexity of any underlying language, since the CF languages are not closed under the operations of duplication, inversion, or transposition—that is to say, when such evolutionary operations are applied to strings contained in a CF language, there is no guarantee that the resulting language will still be CF. Thus evolution by its nature may provide pressure toward increasing linguistic complexity. Another such pressure may arise from *superposition* of multiple levels of information, e.g. signals for successive steps in gene expression, since CF languages are also not closed under intersection.

We have used the Definite Clause Grammar formalism associated with logic programming as a basis for tools for the practical analysis of sequence information on a large scale. In addition to implementing a formalism for *string variables* that specifies the required elements of ILs for sequence repeats in arbitrary conformations, we have incorporated a number of other domain-specific features to allow for powerful pattern-matching search; these include a variant of chart parsing, imperfect matching, and special control operators for the logic-based search paradigm. The resulting parser has been successfully used for finding high-level patterns in sequences of the scale of yeast chromosome III, mitochondrial genomes, and entire GenBank divisions. The grammars, besides being flexible tools for rapid-prototyping of sophisticated motif discriptors, are also effective frameworks for the application of other algorithms, focussing them on regions of interest, managing parameters and cost thresholds, and assembling results into a hierarchically structured parse tree. An associated X-Windows-based parse visualization tool can be used for interactive search.

Applying Machine Learning Techniques to DNA Sequence Analysis

Jude W. Shavlik
University of Wisconsin

Michiel O. Noordewier
Rutgers University

We are pursuing the following research tasks:

(1) Using machine-learning techniques to create accurate recognizers of genes (and genomic signals, such as promoters, terminators, and ribosome binding sites). The strength of our algorithms is that they refine existing biological knowledge, rather than solely inducing gene recognizers from sample sequences. That is, in our *knowledge-based neural networks* are able to make effective use of prior knowledge about the task being learned.

In addition to considering DNA sequences as strings of individual nucleotide locations, we have been developing extended descriptions of features of biological interest (e.g., dinucleotide frequencies, bending sites, local secondary structure, etc.). We are studying their contribution to classification accuracy, as well as increased understandability of extracted rules (see below).

(2) Evaluating these recognizers on "anonymous" DNA produced by a sequencing laboratory; we are collaborating closely with F. Blattner of Wisconsin's Genetics Department who is sequencing *E. coli*.

(3) Developing sequence-alignment algorithms that compare DNA to known proteins. These algorithms are able to detect frameshift errors. Blattner's group applies our alignment algorithms to all of their *E. coli* contigs.

We are also developing a parallel DNA-to-protein alignment algorithm on the Wisconsin CM-5 64-node supercomputer. This algorithm, unlike BLAST and FASTA, computes the complete alignment matrix, as well as detects frameshift errors.

(4) Creating methods for automatically producing human-comprehensible descriptions of the "rules" learned by our algorithms; this will allow biologists to inspect what it is learned by our computer programs.

(5) Developing computer-based visualization tools for inspecting the information acquired about "anonymous" DNA sequences.

Sample Publications

Towell, G. G. & Shavlik, J. W. (to appear). The extraction of refined rules from knowledge-based neural networks. *Machine Learning*.

Craven, M. W. & Shavlik, J. W. (1993). Training neural networks to predict reading frames in *E. coli* DNA sequences. *Proc. of the 26th Hawaii Intl. Conf. on Systems Science: Biocomputing Track*.

Craven, M. W. & Shavlik, J. W. (1992). Visualizing learning and computation in neural networks. *Intl. Journal on Artificial Intelligence Tools* 1:3.

Shavlik, J. W., Towell, G. G., & Noordewier, M. O. (1992). Using neural networks to refine biological knowledge. *Intl. Journal of Genome Research* 1:1, 81-107

Shavlik, J. W., (1991). Finding genes by case-based reasoning in the presence of noisy case boundaries. *Proc. of the DARPA Cased-Based Reasoning Workshop*, Washington, D. C.: Morgan Kaufmann.

Noordewier, M. O., Towell, G. G., & Shavlik, J. W. (1990). Training knowledge-based neural networks to recognize genes in DNA sequences. *Advances in Neural Information Processing Systems (NIPS)*, Denver: Morgan Kaufmann.

CONSTRUCTION OF A HUMAN DNA LIBRARY IN BACTERIAL ARTIFICIAL CHROMOSOME

Hiroaki Shizuya, Bruce Birren, Ung-Jin Kim, Valeria Mancino, Tatiana Slepak, Yoshiaki Tachi-iri, and Mel Simon, Division of Biology, 147-75, Caltech, Pasadena, California 91125

The BAC vector, a single copy plasmid based on the F-factor, is capable of maintaining human genomic DNA fragments as large as 300 kb with a high degree of structural stability in *E. coli*. We have constructed a human DNA library of approximately 2X coverage with an average molecular weight of 150 kb. The frequency of co-cloning events in the BAC clones appears to be significantly lower than that commonly obtained with YAC clones, as judged by fluorescence *in situ* hybridization. 40,000 BAC clones in the library are stored in 400 microtiter plates and gridded on nylon filters at a high density (2,304 clones per sheet) for hybridization. We are currently evaluating the extent of coverage of the library by probing with various markers, and constructing a sub-library of chromosome 22 by identifying chromosome 22 specific clones with various probes including inter-Alu PCR products.

STSs Derived from hn-cDNA Libraries Made from Hybrid Cells

Michael J. Siciliano, Ph.D.; Department of Molecular Genetics, The University of Texas M.D. Anderson Cancer Center, Houston, TX 77030

The purpose of this research is to develop methods for the isolation of STSs of expressed sequences from specific human chromosomes. The essence of the approach is to use a hybrid cell (5HL9-4) containing the chromosome of interest (the 19) as the only human component. Polyadenylated RNA is isolated and coding sequences enriched by making cDNA using consensus 5′ intron splice sites as primers for first strand synthesis. Optimization of conditions for that process has been recently published (Liu et al., 1992 *Somat. Cell Molec. Genet.* **18**:7-18). A cDNA, or more properly an hn-cDNA, library is made from this material and human transcribed sequences isolated by screening with human repetitive DNA (present in introns). Resultant clones contain small inserts (usually <500 bp) which are sequenced for exon STS identification.

While initial results were successful (Liu et al., 1989, *Science* **246**:813-815) scaling up to produce large numbers of clones resulted in a wide variety of anomalies in the first 50 produced this year -- most were not able to be sequenced by our collaborator (T. Carrano, Livermore) because of stuttering in homopolymeric stretches. Also, several were not human specific, and other contained only repetitive sequences.

These results indicated the need to implement certain alternative strategies -- making hn-cDNA libraries from 5HL9-4 RNA enriched for the heterogeneous form; utilizing PCR for isolating human hn-cDNA inserts.

1. Enrichment for hn-RNA.

As an assay for the relative amounts of hn-RNA vs mature RNA in a preparation, reverse transcriptase/PCR (RT/PCR) species specific primers were designed for the products of the *ERCCI* gene on human chromosome 19 (Liu et al. *op. cit.*). To identify unprocessed message in RNA preparations, one primer was located in an intron and the other in the adjacent exon. To identify mature RNA, both primers were located in the same exon. In poly-A RNA preparations from 5HL9-4 cells, the assay revealed approximately 5x more mature rather than hn-RNA present.

Since our aim is to make cDNA libraries from hn-RNA we sought to enrich for that fraction by isolating RNA from nuclear matrix preparations (to eliminate mature RNA) and then selecting molecules with poly-A tails (to eliminate overwhelming levels of rRNA). Testing this product with the RT/PCR primers described above revealed 5x more hn-RNA than mature RNA in these preparations (approximately a 25 fold enrichment of hn-RNA over mature RNA than present in our original preparations). Additional nuclear matrix RNA poly A-selected has been isolated from 5HL9-4 cells for preparation of an enriched hn-cDNA library.

2. PCR strategies for isolating human hn-cDNA inserts.

Our experience thus far has indicated that the rate limiting step in the isolation of hn-cDNAs from hybrid cells is in the screening of the highly heterogeneous hn-cDNA library filters. Based on highly efficient inter-Alu-PCR primers we designed (Liu et al. in press, *Cancer Genet. Cytogenet.*), Alu-vector arm PCR was conducted on minipreps of clones pooled from each replica plate of the hn-cDNA library. In probing human chromosome 19 minipanels, these PCR products all appear to be human 19-specific but still have some repetitive sequences. This is therefore a quick and effective procedure for isolating human hn-cDNA inserts from the library and will be used to screen the new hn-cDNA library made from nuclear matrix RNA (described above).

SINGLE PASS AND FULL LENGTH SEQUENCING AND PHYSICAL AND GENETIC MAPPING OF HUMAN BRAIN cDNAS. J.M. Sikela*, T.J. Stevens, A.S. Wilcox, M.H. Polymeropoulos‡, R. Berry, A.S. Khan, J.A. Hopkins, and A.K. Orpana. Univ. of Colorado Health Sci. Ctr., Denver, CO 80262 and ‡Laboratory of Biochemical Genetics, NIMH, Washington, DC 20032.

The large scale collection, sequencing and mapping of human brain cDNAs represents a productive approach to the identification of most human brain genes and to the development of a cDNA expression map of the genome. For these reasons we have refined strategies for automated single pass and full length sequencing of human brain cDNAs and for physical mapping of cDNAs to specific locations in the genome. Currently, in collaboration with Charles Auffray, we are applying high throughput approaches to assigning sequenced brain cDNAs to the megabase YACs developed by Daniel Cohen and to mapping resulting cDNA-positive YACs to chromosomal regions by fluorescence *in situ* hybridization. We have also identified a subset of cDNAs that contain polymorphic microsatellite sequences and demonstrate how they can be converted to highly informative (PIC value > 0.7) gene-associated genetic markers. At present, most of the first thousand sequenced cDNAs correspond to potentially new human brain genes, while a significant number of cDNAs appear to represent human homologs of interesting genes found in other species. Prescreening of cDNA libraries with total brain cDNA and selection of non-hybridizing clones resulted in a significantly lower frequency of highly represented cDNAs. We have also explored strategies for identifying and rapidly sequencing anonymous cDNAs that contain complete protein coding regions. Currently in our laboratory automated single pass sequencing is being carried out at a rate of several thousand cDNAs per year. Therefore, coordination of this effort with other laboratories doing similar work should permit the sequence identification of most of the genes expressed in the human brain within the next few years.

HIGH SPEED DNA SEQUENCING BY HORIZONTAL ULTRATHIN GEL ELECTROPHORESIS (HUGE)

Lloyd M. Smith, Robert L. Brumley, Eric Buxton,
Michael Giddings, and Michael Marchbanks

Department of Chemistry
1101 University Ave
University of Wisconsin
Madison, Wisconsin
53706

We have been exploring the utility of capillary electrophoresis for increasing the throughput of fluorescence-based automated DNA sequencing instruments for the past two years. Using this method it is possible to separate and detect fluorescently labeled products of DNA sequencing reactions up to 26 times more rapidly than in conventional electrophoresis. In order to extend this high speed separation to the parallel analysis of multiple samples we have recently developed an apparatus for performing automated fluorescence-based DNA sequencing in ultra thin (50-100 μm) slab gels. The fluorescence detection system employs a charge coupled device (CCD) detector operated in frame transfer mode for the real-time acquisition of fluorescence data. Readable sequence out to 410 bases may be obtained from a 50 minute electrophoresis, and 18 samples may be analyzed in parallel. A variety of improvements are under development to permit the analysis of fifty samples in parallel, an instrument throughput of 26,000 bases raw sequence data per hour.

CONSTRUCTION OF A SEQUENCE-TAGGED SITE MAP FOR HUMAN CHROMOSOME 11 - MANUAL AND AUTOMATED APPROACHES

Michael W. Smith, Stephen P. Clark, Jane S. Hutchinson, Yalin Wei,
Allan C. Churukian, Lori B. Daniels, Karin L. Diggle, Michael W. Gen,
Ying Lin, Anthony J. Romo and Glen A. Evans

Molecular Genetics Laboratory and Human Genome Center,
The Salk Institute for Biological Studies, La Jolla, California

The Genome Project is working towards the completion of low resolution YAC contig maps, high resolution clone maps and ultimately DNA sequence-based maps of the human genome and of model organisms. Our approach has been to utilize a variety of technologies with increasing levels of mapping detail to first complete low resolution maps of chromosome 11 and then begin the assembly of high precision maps as a prelude to complete DNA sequencing. As part of the chromosome 11 effort, we developed techniques that rapidly generated over 330 sequence-tagged sites (STSs). These have been regionally localized using fluorescence *in situ* hybridization and somatic cell hybrid panels, and are being used to construct a low resolution YAC contig map of chromosome 11.

The STSs were generated from sequences of previously characterized chromosome 11 genes found in GenBank, or from newly determined sequences of cosmid clones determined mostly by automated DNA sequencing. Previously existing primer sets used by our laboratory or others for gene identification were also exploited. Selection of primer sets from DNA sequence data was carried out by computer aided analysis. Potential STS primer sets were tested for specificity on human, Chinese hamster, mouse, yeast, and chromosome-11-containing somatic cell hybrid DNAs. Successful STSs were screened against a panel of somatic cell hybrids for chromosomal sublocalization.

The sequences we derived from cosmid ends correspond to about 0.1% of the DNA content of chromosome 11. These sequences were analyzed for repetitive DNA, similarity to known genes or motifs, and potential exons. Even though many of the cosmid end sequences contained repetitive DNA elements, we were able to successfully generate STSs from a high proportion of them. Some cosmids contained simple nucleotide repeats, a few of which are being developed into polymorphic markers. Exons predicted by the GRAIL program were recorded for future reference; a number of authentic genes were identified using BLAST searches of nucleotide and protein databases. Mapping and DNA sequencing information are stored in the GENOME NOTEBOOK, a Macintosh-based relational database for data analysis and presentation.

This set of STSs provides the reagents for rapidly assembling ordered YAC contig maps as a step towards generating high resolution cosmid maps from chromosome-specific libraries. To facilitate the latter process, we are currently developing methods and automated techniques to accelerate high resolution genome mapping and large scale DNA sequencing of the human and other genomes.

HIGH SPEED SAMPLE PREPARATION IN A CAPILLARY BASED INSTRUMENT

Harold Swerdlow*, Kerry Dew-Jager, Carl Wittwer and Raymond Gesteland
University of Utah Human Genome Center
6160 Eccles Genetics Bldg. Univ. of Utah, Salt Lake City, UT 84112

With the advent of the Human Genome Initiative, pressure has been placed upon the scientific community to develop DNA diagnostic and sequencing methods which are simultaneously more rapid, accurate and cost-efficient than current methodologies. Capillary electrophoresis offers numerous advantages in speed, resolution and automation for these endeavors. The acceptance of capillary electrophoresis as an analytical tool has been limited by the lack of applications properly matched to the power of the technique. Samples are normally prepared in large volumes from which 1-10 nanoliters are loaded in a typical run. Efforts to design and implement micro-volume sample preparation techniques for capillary electrophoresis have begun. We are currently exploring such techniques for both PCR and Cycle Sequencing.

The polymerase chain reaction (PCR) has allowed researchers to easily amplify specific DNA sequences from complex mixtures of nucleic acids. Conventional block cycling instruments are far too slow to take advantage of the inherent speed of the reaction. A novel thermal cycler developed by C. Wittwer, was commercialized by Idaho Technologies (1605 Air-Thermo Cycler). It employs thin-walled capillary tubes (500 μm i.d., 1000μm o.d.) with a low thermal mass, and air with its low specific heat, to accomplish very rapid temperature transitions. Consequently, it has been shown that the kinetics of denaturation, annealing, and elongation are much faster than previously suspected for PCR. This allows 5-10 μl PCR reactions to be cycled in only 15-25 minutes. Furthermore, minimizing the time spent at or near optimal annealing temperatures can improve specificity in some amplifications. Fluorescent PCR products cycled in this way have been analyzed in a home-made capillary electrophoresis instrument.

Cycle sequencing requires only a few hundred nanograms of template DNA to produce ample quantities of sequence product for manual or automated, radioactive or fluorescent techniques. Additionally, the method appears to overcome the problems associated with direct sequencing of PCR products. Although cycling provides several advantages over conventional sequencing protocols, it is far slower. To improve the attraction of cycle sequencing a radical reduction in the total cycling time was necessary. To accomplish this, we adapted the cycling protocol to work in the air thermal cycler, reducing the total reaction time to only 25 minutes. Results obtained compare favorably with control reactions performed on block thermal cyclers in 2-3 hours. Radioactive and fluorescent reactions were optimized using both Taq and Vent (exo$^-$) thermostable polymerases. The use of sealed glass capillary tubes in this instrument obviates the need for oil overlays to prevent evaporation.

Genomic Sequencing by Ligation-Mediated PCR.

V. T. Törmänen, G. P. Pfeifer, K.S. Graham, and A. D. Riggs.
Department of Biology Beckman Research Institute of the City of Hope, Duarte, CA 91010

We have continued our efforts to use ligation-mediated PCR (LMPCR) as a method to sequence genomic DNA directly, without any cloning, as may be necessary for portions of the genome that are unclonable or difficult to clone. The method consists of (i) treatment of genomic DNA with base-specific cleavage agents, (ii) primer extension using a gene specific oligonucleotide, (iii) ligation of an oligonuleotide linker to the blunt ends produced in the primer extension step, (iv) exponential PCR using a gene-specific primer and a linker-specific primer, and (v) sequencing gel analysis of the PCR products. Several aspects of the procedure have been studied, including new DNA cleavage agents, but the major improvement so far has come from the use of biotinylated primers. Use of biotinylated primers in the first primer extension step enables streptavidin-coated magnetic beads to be used for the capture and enrichment of extension products. This is not only a step towards automation by eliminating centrifugation, but it also reduces the sequence complexity in the PCR reaction and increases the quality of the sequencing gels obtained. Methods for multiplexing, analyzing longer products, and nonradioactive detection are currently under investigation.

FLUORESCENCE IN SITU HYBRIDIZATION MAPPING IN INTERPHASE CHROMATIN

Barbara Trask[1], Anne Fertitta[2], Mari Christensen[2], Susan Allen[2], Marge Segraves[2], Anne Bergmann[2], Hillary Massa[1], Rainer Sachs[3], and Ger van den Engh[1],
[1]Department of Molecular Biotechnology, University of Washington, Seattle, WA;
[2]Human Genome Center, Lawrence Livermore National Laboratory, Livermore CA;
[3]Dept. of Mathematics, University of California, Berkeley, CA

A strategy has been developed that uses fluorescence in situ hybridization to produce dense maps (100-kbp average spacing) of markers ordered along a chromosome. The strategy is based on (1) our finding that the square of the mean distance between hybridization sites in interphase chromatin is linearly correlated with genomic distance in the 50 kbp to 1-2 Mbp range, and (2) the capability to obtain interphase measurements rapidly. A linear correlation between mean square interphase distance and genomic distance has been observed in several chromosomal regions, 4p16.3, 6p21.3 and Xq28, from 50 kbp to 1-2 Mbp. Chromatin behaves as a polymer over this range and follows random walk model predictions. As a consequence, probe order can be reconstructed from mean interphase distance measurements <1-2 µm. Deviation from the linear relationship occurs at ≈2 Mbp in 4p16.3 and at ≈1 Mbp in 6p21 and Xq28, suggesting that higher order constraints on random chromatin organization may exist in different chromosomal regions. A system has been developed to rapidly obtain distance measurements between fluorescence *in situ* hybridization sites. The system relies on film-based acquisition and storing of images, due to advantages of image resolution, image size, speed of acquisition and redisplay, simplicity, and low cost. The photographic film is projected onto a digitizing board through which the distances between hybridization sites are recorded. Over 5000 measurements can be made with 0.01 µm precision by an individual per day. In a test of the mapping strategy, a map of 13 probes from a 4-Mbp region of chromosome 4 could be reconstructed from a matrix consisting of 56 pair-wise distance measurements. The order and relative spacing of markers in this map was similar to published maps. Given this validation, we tested the practicality of interphase mapping on chromosome 19. To date, 120 cosmids representing genes or contigs have been ordered along this chromosome, at an average density of one marker per 0.5 Mbp.

Impact of Human Genome Initiative-Derived Technology on Genetic Testing, Screening and Counseling: Cultural, Ethical and Legal Issues.

DOE/NIH Grant #DE-FG02-92ER61396

Ralph W. Trottier, Ph.D., J.D. (PI)[*], Lee A. Crandall, Ph.D. (Co-PI)[**], Faye Cobb Hodgin, Ph.D., J.D.[*], Mwalimu Imara, D. Min.[*], Ray E. Moseley, Ph.D.[**], David Phoenix, Dr. P.H.[*], Delores Armotrading (Graduate Student)[**], and Sherrill Lybrook (Graduate Student)[*], Morehouse School of Medicine[*], Atlanta, GA and University of Florida College of Medicine[**], Gainesville, FL.

Genetic medical services provided by the Georgia Division of Public Health in two northern and two central districts are compared to services provided in a district in which a tertiary care facility is located. Genetics outreach public health nurses play key roles in Georgia's system of Children's Health Services Genetics Program, including significant roles as counselors and information sources on special needs social services and support organizations. Unique features of individual health districts,(e.g., the changing face of some rural communities in ethnocultural diversity and socioeconomic character), present new challenges to current and future genetics services delivery. Preparedness as to educational needs of both health professionals and the lay population is of foremost concern in light of the ever expanding knowledge and technology in medical genetics. Perspectives on genetics and an overview of services offered by a local private sector counselor are included for comparison to state supported services. The nature of the interactions which transpire between private and public genetic services resources in Georgia will be described. A special focus of this research includes issues associated with sickle cell disease newborn screening service delivery process in Georgia, with particular attention paid to patient follow-up and transition to primary care. Of particular interest to this focus is the problem of loss to follow-up in the current system. Critical factors in education and counseling of sickle cell patients and the expectations of expanding roles of primary care physicians are discussed. The Florida approach to the delivery of genetic services contrasts to the Georgia model by placing more emphasis on a consultant-specialist team approach. The state's three medical schools house genetics teams that provide specialty satellite services, under contract with the Department of Health and Rehabilitative Services, to 22 sites. Although some aspects of this system are similar to the tertiary care center genetic physician contracts for services in the Georgia health districts, there are no on-site genetics services personnel at the site of clinical services provided in Florida. The Florida services include genetic testing and counseling provided by teams that include medical geneticists and masters degree trained genetic counselors. Florida Children's Medical Services is a program representing a dramatic expansion of federal crippled children's programs. Unique aspects of this structure are discussed and compared to programs in other states within the southeast region. Ethical issues, related to the principle of justice, are discussed in terms of rural-urban differences in access to genetic services and the interrelationship of these differences to concepts of race, ethnicity, variable incidence of genetic diseases and level of genetic predisposition to multifactorial diseases. Legal concerns involve expanding liabilities in the realm of general medical practice, risk communication and issues surrounding the concept of informed consent in genetic medicine and genetic counseling.

FLOW KARYOTYPING AND FLOW INSTRUMENTATION DEVELOPMENT

Ger van den Engh[1], Richard Esposito[2], Mike Vardanega[3], Hillary Massa[1], and Barbara Trask[1], [1]Department of Molecular Biotechnology, University of Washington, Seattle, WA; [2]Human Genome Center, Lawrence Livermore National Laboratory, Livermore CA; [3]Systemix, Palo Alto, CA

Flow karyotyping, or chromosome analysis by flow cytometry, has become an important tool in genetic analysis. The technique determines the DNA content and base pair composition of individual chromosomes by measuring the fluorescence of DNA-specific dyes. We are applying analytical flow karyotyping to a variety of areas related to genomic research and medical diagnostics. In addition, chromosomes purified by flow sorting are used for the production of cloned or PCR-amplified DNA libraries. We have built a new high speed sorter that combines high measurement accuracy and high sorting speed. Electronics with minimal dead-time, non-rectangular sort windows, and digital error checking have been incorporated. Redesign of droplet drive and deflection electronics and the nozzle assembly has resulted in consistent and stable drop deflection. Sorting runs producing 3-5 million chromosomes/day (10 times the rate of commercial instruments) have been achieved. Sorted chromosomes are delivered in smaller volumes, with better error detection and coincidence event rejection, than have been achieved previously. The new flow sorter design has been licensed to industry. Emphasis on ease of operation, ease of sterilization, and use of standard components will facilitate the export of this technology to other genome centers that require chromosome analysis and purification capabilities. The capability for simultaneous sorting in 8 directions has been implemented. Software that facilitates the analysis and comparison of human flow karyotypes has been developed. An increasing understanding of the interactions of DNA-binding dyes and chromatin has led to improved techniques for chromosome preparation and staining. As a result of these developments, quantitative DNA measurement of human chromosomes has become a reproducible technique that can be applied to a variety of genetic studies. Examples are the description of normal chromosome heteromorphism, quantification of deletion size in contiguous gene syndromes to facilitate construction of long-range physical maps, and routine monitoring of somatic cell hybrids.

Development of a Human Virus-Based Genomic Library of 150-200 kb Inserts

Jean-Michel H. Vos[1,2], Subrata Banerjee[1], Zachary Kelleher[3] and Tian-Qiang Sun[2]
1 UNC Lineberger Comprehensive Cancer Center, 2 Department of Biochemistry & Biophysics, and 3 School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7295. (919) 966-6887, Fax (919) 966-3015, Internet: "vos@med.unc.edu".

The objective of this project is to develop a method for cloning large-size DNA exclusively in human cells and propagating these fragments for genomic mapping, gene isolation, and mutation detection. A human-based library of 150-200 kb human DNA inserts would be very useful resource in global mapping/sequencing efforts. The proposed strategy is based on the development of a novel virus-mediated gene-transfer technology. Epstein-Barr virus (EBV), a human lymphotropic herpesvirus is one of the largest human virus with a genome of 172 kb. After EBV infection, its genome is stably maintained as nuclear episome during latent phase, and amplified 500-1000 fold and packaged into virions during lytic phase. A novel EBV-based system is being developed to clone, amplify, package and transfer 150-200 kb genomic DNA library between human cells. Using a mini-EBV (i.e. 16-20 kb) and a helper cell line carrying resident EBV, we have demonstrated that 1) between 150-200 kb DNA was packaged into EBV virions; 2) the engineered DNA was packaged as efficiently as the helper EBV genome; 3) the DNA was not rearranged during amplification and packaging into EBV virions.

Using this EBV cloning system, we have constructed a partial human genomic library covering approximately 25% of the human genome. Detailed analysis of this library allowed the following conclusions: a) miniEBV carrying 150-350 kb human genomic inserts were maintained as circular episomes in the human helper cell line; b) single clones isolated from the library were shown by Alu-PCR fingerprinting to be genetically distinct; c) the episomal inserts were stably maintained with no detectable rearrangements; d) EBV virions were produced which carry human DNA. Current efforts are focused on the preparation of a complete EBV-based human genomic library. Such EBV-based library should provide a unique resource to map, isolate, sequence and assay large genes and other functional human genomic regions.

Partial Bibliography:

Sun T.Q and J-M.H. Vos "Packaging of 200 Kb Engineered DNA as Infectious Epstein-Barr Virus" Int. J. Genome Res. 1, 45-57 (1992).

A Strategy Employing Homologous Recombination and Gene Amplification to Expedite Cloning of Large Genomic Regions in YACs.

Geoffrey M. Wahl.
The Salk Institute, San Diego, CA

We constructed a specialized vector, HARY, to be integrated into predetermined loci in mammalian genomes by homologous recombination. We designed the vector to facilitate the mapping and rescue of large chromosomal regions adjacent to the insertion site in YACs. This strategy should expedite mapping as well as isolation and characterization of disease genes or chromosomal rearrangements.

We first targeted the human HPRT gene to provide a model system for the use of the vector and the HARY strategy. We have introduced the HARY vector at the HPRT locus in the human cell lines HT1080 and 293. We chose this locus because it offered straightforward genetic and physical methods to rapidly identify homologous recombinants. These were identified by three methods; 1) genetic selection (they resist 6-thioguanine and are sensitive to HAT selection), 2) Southern blotting (they display patterns consistent with disruption of the HPRT locus, which has already been mapped), 3) fluorescent in situ hybridization (hybridization of metaphase chromosomes with the HARY vector reveals a single signal on the X-chromosome at the position of the HPRT gene).

To facilitate YAC rescue, a mutated DHFR gene (DHFR*) was included in HARY to allow detection of amplified vector sequences in preference to the endogenous DHFR genes. Flow cytometry detected a significantly higher expression of the cell surface marker IL2R (present in HARY) in the Mtx^{r} 293 cells as compared to the non-amplified cells. Cells with the highest and lowest 10% fluorescence were isolated by FACS. The difference in the fluorescence level correlated with the copy number of the introduced vector, as determined by Southern blots. Finally, FISH analyses of the clones displaying a high level of fluorescence revealed the presence of HARY at high copy number. At present we are working on the rescue of genomic DNA flanking the insertion site in YACs. We have recovered DNA flanking a random insertion site and we are attempting to rescue DNA flanking the HPRT locus in targeted and amplified human 293 cells.

The HARY vector, once integrated into the human genome, contains rare cutter restriction sites that serve as anchor points for long range mapping. By targeting the HARY vector to the HPRT locus, a NotI site present in the vector dissected a 2 MB NotI DNA fragment from the HT1080 parental cell line into two smaller fragments, one of which was 680 kb in size. The vector contains an additional site for the meganuclease I-SceI, which has a unique recognition sequence in the human genome. The enzyme cuts DNA in agarose efficiently, as evidenced by digestion of a 440 kb SrfI fragment containing the human HPRT gene into two smaller fragments detected by PFGE and Southern analyses. These sites, and additional ones not tested on this clone, demonstrate that once the HARY vector is introduced to a specific locus, it provides a landmark useful in determining or refining existing restriction maps.

The successful targeting of HPRT suggests that the HARY insertion vector should be useful for targeting other loci as well. To test the generality of the HARY insertion approach, we constructed an insertion vector to target telomeres and other loci in the human genome.

Isolation of the EPM1 Gene

Janet A. Warrington, Lynn Bernard, Ursula Edmond, Nancy Stone, Jasper Rine*,
Richard M. Myers, and David R. Cox

Department of Biochemistry & Biophysics, University of California, 513 Parnassus Ave.,
San Francisco, CA 94143-0554. *University of California, 225 Barker Hall, #401,
Berkeley, CA 94720.

Recent advances in physical mapping and positional cloning techniques have made the isolation of disease genes possible. However, many problems in positional cloning remain unresolved and existing methods need improvement if these techniques are to find general applicability. The focus of this project is the development of a positional cloning approach that will be applicable to the isolation of any disease gene. The project focuses on improving positional cloning methods using the isolation of a gene involved in an inherited disease as a model system. The disease, myoclonic epilepsy of the Unverricht-Lundborg type (EPM1), is an autosomal recessive disorder and the underlying biochemical defect is unknown. Clinical features of the disease include, incapacitating stimulus sensitive myoclonic and tonic-clonic seizures and a gradual intellectual decline. The age of onset of symptom ranges from 6-15 years and the severity of symptoms and a rate of progression varies between and within families. The long term objective of this project is to improve methods of positional cloning using the isolation f the EPM1 gene as a model system. The specific aims include: 1) Identify markers that most closely flank the disease gene by recognizing informative meiosis using a) somatic cell genetics and b) genomic denaturing gradient gel electrophoresis to make all markers informative. 2) Clone the DNA between the closet flanking markers and, 3) Isolate genes from the cloned DNA using the technique of exon trapping. Using the cloning of the epilepsy gene as our experimental model, we will address the problems confronted in positional cloning and develop a system which will be applicable to the cloning of any disease gene.

Application of arbitrarily-primed PCR (AP-PCR) for construction of chromosome-specific high complexity DNA libraries and isolation of repeat DNA probes.

Heinz-Ulrich Weier[1], Beate M. Miller[1], K. Harry Scherthan[2], Scott Cram[3]
Daniel Polikoff[1], Loh-Chung Yu[1] and Joe W. Gray[1]

[1] Division of Molecular Cytometry MCB 230, Dept. of Laboratory Medicine, University of California, San Francisco, CA 94103-0808, [2] Biology Dept., Universität Kaiserslautern, Kaiserslautern, Germany and [3] Life Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545.

In vitro DNA amplification using the polymerase chain reaction (PCR) with mixed base primers and DNA templates comprised of flow sorted chromosomes has been applied for construction of chromosome–specific high complexity DNA libraries. The DNA templates were comprised of 250-2000 flow sorted human, hamster or mouse chromosomes. We used two different oligonucleotide primers that generate pools of PCR products in the size range of 200-800bp and 300-1200bp, respectively. Two sets of chromosome-specific DNA libraries were generated for all human chromosomes but chromosome 20, for which no flow sorted chromosomes were available at the time. High complexity rodent libraries were made for mouse chromosome 17 and the chinese hamster ovary (CHO) cell line X chromosome. The complexity and chromosome-specificity of the pools were checked by non-isotopical labeling and fluorescence in situ hybridization (FISH). The AP-PCR generated libraries compare favourably with complete digest libraries of flow human chromosomes produced earlier, particularily in regions of the genome that appeared underrepresented in Hind III and Eco R1 libraries. A number of the high complexity PCR product pools have subsequently been cloned into plasmid vectors. This allows easy propagation and distribution of the libraries. Reduction of the library complexity furthermore allowed us to rapidly isolate repeat DNA sequence clones for the CHO X chromosome.

Human Genetics and Genome Analysis: A Practical Workshop for Public Policy Makers and Opinion Leaders

Jan Witkowski [1], David Micklos [2], and Mark Bloom [2]
Banbury Center [1] and DNA Learning Center [2]
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724
516/549-0507 Fax: 516/549-0672

These workshops provide a basic understanding of genetics and molecular genetics for nonscientists who have a special interest in the societal implications of the Human Genome projects. The workshops are joint effort of the Banbury Center and DNA Learning Center at Cold Spring Harbor Laboratory.

We have held three workshops so far, with a total of 72 participants. Each workshop lasts two-and-one half days and has three components: (1) seminars that cover genetics, a modern view of the gene, recombinant DNA techniques, and their applications to human inherited disorders. (2) four seminars by invited speakers. These have included talks on the genome projects; complex diseases; societal issues; cloning genes; and gene therapy. (3) A unique feature of these workshops is that participants perform three experiments: cutting DNA and running gels; bacterial transformation; and DNA fingerprinting by PCR.

There is a maximum of 24 participants for each workshop so we select participants who will be most effective at passing on what they learn. Participants include representatives from action groups for specific disorders (16); educators from high schools and science museums (20); science journalists (10); congressional staff and members of other Washington based groups, e.g. OTA (7); foundations (3); ethicists and theologians (12); and lawyers (4). Participants have come from 28 states.

Mapping and Sequencing the Human Genome: Science, Ethics, and Public Policy

Joseph D. McInerney, Jenny Stricker, and Katherine Winternitz
Biological Sciences Curriculum Study, The Colorado College, Colorado Springs,
CO 80903

Today's high school students will be affected throughout their adult lives by the Human Genome Project (HGP), and the high school biology course is an appropriate mechanism for introducing young adults to the scientific, ethical, and public-policy dimensions of the HGP. The Biological Sciences Curriculum Study (BSCS) and the American Medical Association (AMA), therefore, have developed and distributed a 94-page instructional monograph titled *Mapping and Sequencing the Human Genome: Science, Ethics, and Public Policy.* The module, designed for use with average, first-year students in high school biology, includes 30 pages of background materials for the teacher and 4 inquiry-oriented activities for the classroom. BSCS has distributed one copy of the monograph free of charge to each high school biology teacher in the U.S. (approximately 48,000). Teachers have permission to reproduce the materials for classroom use. The module was designed and written by individuals experienced in human genetics, molecular biology, ethics, and public policy, and was field tested in 34 schools across the country. Evaluation of the field test showed, among other things, that the materials increased student awareness of the role of ethical inquiry in dealing with social issues and student understanding of genotype/environment interaction in the expression of human traits. In addition, the evaluation showed that orientation sessions for teachers increased the overall effectiveness of the module in the high school classroom.

Chromosomal Localization of Active Genes

K. Denison, J. M. Gatewood, J. R. Korenberg, and X-N. Chen
Life Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545
and
Department of Pediatrics, Cedars-Sinai Medical Center, Los Angeles, CA 90048

From a natural but abnormal pregnancy (hydatidiform mole), a directionally cloned cDNA library was constructed. Clones from the library are first characterized by sequencing from the defined 3' end; the sequences are then screened against GENBANK primate sequences for homology matches.

Chromosomal localizations of cDNA clones have been obtained using the fluorescent *in situ* hybridization (FISH) technique of Korenberg and Chen. Briefly, clones that show no homology to known genes are further selected for the absence of hybridization to random-primed genomic DNA (suggesting the absence of highly repetitive DNA), and insert sizes greater than 2 kb. Slides prepared from BrDU-synchronized primary lymphocyte cultures are used in a specially modified FISH reaction with the biotinylated clones. Fluorescein-avidin detection is followed by R-banding, allowing the hybridization signal and banding pattern to be viewed simultaneously.

An alternative approach to mapping that has been investigated involved sequence tagged site (STS) primer generation, followed by hybridization of PCR products to hybrid panels. In-house comparisons of the mapping resolution and efficiency of obtaining positive localizations using FISH strongly suggest this to be the superior technique.

Appendices

Appendix A: Subject Index

Abstracts listed by abstract number.

Mapping
1, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 22, 23, 24, 25, 28, 29, 32, 34, 37, 39, 40, 42, 45, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 62, 63, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 85, 88, 93, 95, 98, 104, 106, 107, 109, 110, 112, 119, 120, 122, 124, 125, 132, 133, 140, 142, 151, 152, 153, 154, 155, 157, 158, 159, 160, 172, 175, 177, 182, 183, 184, 186, 189, 192, 193, 194, 198

Sequencing
8, 11, 27, 28, 32, 37, 38, 39, 41, 45, 46, 48, 52, 64, 65, 67, 77, 78, 81, 83, 84, 86, 87, 91, 92, 94, 96, 97, 98, 99, 100, 101, 102, 103, 104, 108, 110, 111, 115, 116, 118, 128, 129, 130, 134, 135, 138, 141, 143, 144, 146, 147, 148, 149, 150, 156, 159, 161, 162, 163, 165, 166, 167, 169, 171, 173, 178, 179, 180, 181, 184, 185, 187, 188

Resources
1, 2, 3, 4, 5, 9, 13, 17, 18, 19, 22, 23, 32, 33, 35, 36, 38, 42, 49, 50, 51, 52, 57, 59, 61, 63, 65, 66, 70, 71, 77, 88, 89, 93, 102, 105, 107, 108, 109, 110, 112, 119, 120, 125, 132, 133, 135, 137, 140, 141, 142, 143, 147, 148, 151, 153, 155, 157, 158, 159, 160, 162, 163, 170, 172, 173, 176, 182, 184, 191, 192, 193, 195

Instrumentation
2, 10, 11, 12, 14, 21, 26, 27, 28, 30, 31, 39, 40, 41, 44, 45, 58, 60, 76, 79, 82, 91, 95, 96, 97, 99, 100, 101, 103, 104, 111, 114, 115, 116, 118, 121, 128, 129, 130, 134, 137, 138, 149, 150, 156, 161, 166, 169, 171, 177, 178, 179, 185, 187, 191

Informatics
2, 4, 16, 18, 19, 32, 35, 36, 43, 46, 47, 48, 52, 56, 68, 70, 72, 74, 75, 80, 83, 84, 85, 86, 87, 89, 90, 94, 106, 108, 110, 113, 127, 132, 133, 137, 144, 145, 146, 148, 152, 159, 180, 181

Ethical, Legal, and Social Issues (ELSI)
20, 117, 123, 126, 131, 136, 139, 164, 168, 190, 196, 197

Small Business Innovation Research (SBIR)
114, 115, 116, 135, 145, 160, 161, 162, 163

Appendix B: Author Index

First authors are in **bold.**

Adams, Mark D. 112, 157
Adamson, Anne E. 105
Affleck, Rhett L. 82
Aggarwal, A. 32
Ahn, Andy 168
Alegria-Hartman, Michelle 49, 52, 57, 58
Alexander, G. E. 113
Alleman, Jennifer 53, 61
Allen, Robert C. 101
Allen, Susan 50, 52, 189
Allison, D. P. 97
Allman, S. L. 99
Alper, Joe 168
Altherr, Michael R. 1
Ambrose, W. P. 10
Amemiya, Chris T. 49, 51, 52, 57, 63
Anderson, N. Leigh 114
Anderson, Norman G. 114
Anzick, Sarah L. 13
Ard, Catherine 168
Arlinghaus, H. F. 101, 104, 115
Armotrading, Delores 190
Armstrong, R. 48
Ashworth, Linda K. 52, 75
Baker, Robert J. 88
Balding, D. 2
Bandziulis, Ray 155
Banerjee, Subrata 192
Barash, Carol 168
Barber, William M. 4
Barker, D. L. 116
Barlett, R. 52
Barsky, V. 167
Bashkin, J. S. 116
Batzer, Mark A. 49, 52, 53, 57, 58, 63
Beatty, B. R. 109
Beckwith, Jon 168
Beeson, Diane 117
Bellefroid, Eric 51
Ben-Shachar, O. 47
Benner, W. H. 21
Benson, Scott C. 91
Bergmann, Anne 52, 54, 189
Berka, Jan 118
Bernard, Lynn 194
Berry, R. 184
Beugelsdijk, Tony J. 14
Billings, Paul 168
Birren, Bruce W. 49, 119, 154, 182
Blajez, R. 34, 42
Bloom, Mark 196
Bolund, L. 93
Bonaldo, M. Fatima 120
Bouma III, Hessel 123
Boville, B. M. 121
Brandriff, Brigitte F. 50, 52, 54, 55, 59, 60, 69, 73
Branscomb, Elbert W. 52, 54, 56, 68, 69, 72, 74, 75
Brennan, Thomas 92
Bridgers, Michael A. 4
Britt, Deborah E. 151
Bronstein, Irena 162, 163
Brown, Gilbert M. 98, 104
Brown, Nancy C. 5, 88
Brown, S. 120
Bruce, D. 2
Bruce, James E. 111
Brumley, Jr., Robert L. 121, 185
Bruno, W. 2
Buchanan, Michelle V. 103
Buckingham, J. M. 5
Buckler, Alan J. 1
Buley, Donna 108
Bultman, S. 109
Burgin, M. 52
Burks, Christian 19, 83, 84, 89, 90
Buxton, Eric 121, 185
Callen, D. F. 6, 7, 122
Campbell, Evelyn W. 5, 13, 17
Campbell, M. L. 3, 5
Cantor, Charles R. 40, 140, 176
Cantu III, Roy 74
Carrano, Anthony V. 49, 50, 51, 52, 54, 55, 59, 64, 66, 67, 69, 73
Casey, Denise K. 105

Caskey, C. Thomas 123, 175
Chait, Brian T. 179
Chang, William 127
Chapman, V. M. 124
Chee, Mark 135
Chen, C. H. 99
Chen, Chira 49, 52, 57
Chen, I-Min 36
Chen, X. N. 157, 198
Cheng, J.-F. 22, 42
Cheng, Xueheng 111
Cherath, L. 158
Chou, Chau-Wen 178
Christensen, Mari 52, 189
Church, George M. 166
Churukian, Allan C. 186
Cinkosky, Michael J. 4, 6, 85
Clark, Stephen P. 125, 186
Collins, Debra L. 126
Copeland, Alex 52, 58, 69
Cox, David R. 194
Cozza, Steven 127
Cram, Scott 195
Crandall, Lee A. 190
Crkvenjakov, R. 77
Cuddihy, Elizabeth 127
Culiat, Cymbeline 107
Daniels, Lori B. 186
Davidson, J. B. 100, 102
Davidson, K. Alicia 105
Davis, Cheryl A. 37, 38
Davis, Ronald W. 128
Davy, D. 32, 43
de Jong, Pieter J. 49, 51, 52, 53, 57, 58, 61, 63, 71, 154
Deaven, Larry L. 3, 5, 6, 9, 13, 17, 88
Deininger, Prescott L. 53
Denison, K. S. 8, 198
Denton, M. Bonner 129
Devlin, Lorie G. 50, 52, 59
Dew-Jager, Kerry 187
Diggelmann, Martin 135
Diggle, Karin L. 186
Doggett, N. A. 2, 6, 7, 19, 122
Dogruel, David 178
Doktycz, Mitchel J. 97, 98, 101, 104, 115
Douthart, Richard J. 110
Dovichi, Norman J. 130
Doyle, Johanna L. 101
Drees, Becky L. 96
Drmanac, R. 78, 79
Drmanac, S. 78, 79
Duesing, L. A. 7
Dunn, John J. 81
Durkin, A. Scott 170
Duster, Troy 117
Earle, Colin W. 129
Edmond, Ursula 194
Edmonds, Charles G. 111
Efstratiadis, A. 120
Einstein, J. Ralph 108
Elbert, Jeffrey E. 98
Eldarov, M. 12
Elliott, Jeffrey M. 52, 59, 67
Elliott, R. 124
Engle, Michael L. 83
Ericsson, Cheryl 29
Ershov, G. 167
Esposito, Richard J. 52, 60, 191
Evans, Glen A. 125, 137, 186
Eveleth, Jerry 52, 61
Faber, Vance 16
Fader, Betsy 131
Fasman, Kenneth H. 132
Fawcett, J. J. 3, 5
Feitshans, Ilise 139
Ferrell, T. L. 97
Fertitta, Anne 50, 52, 189
Fickett, James W. 4, 6, 85, 86, 87
Fields, Chris 84, 112, 133
Fitzgerald, Michael C. 134, 171
Flood, Thomas 148
Florentiev, V. 167
Flynt, Clifton 145
Fodor, Stephen 135
Foote, R. S. 102, 104
Ford, Amanda A. 2, 16
Foret, Frantisek 118
Forrest, Stephanie 83
Fullarton, Jane E. 136
Garcia, Emilio 50, 52, 67
Garner, Skip 137
Garnes, Jeffrey A. 49, 52, 61, 63, 154
Garrity, Martha L. 98, 104

Gatewood, Joe M. 8, 18, 157, 198
Geller, Lisa 168
Gemmell, A. 78, 79
Gen, Michael W. 186
Generoso, Estela 107
Genetic Screening Study Group 168
Georgescu, A. 52, 58
Georgi, D. 52
Gesteland, Raymond F. 138, 187
Ghazizadeh, Hamid 166
Ghiso, Neil S. 65
Gibson, William A. 101
Giddings, Michael 185
Gingrich, G. 93
Gingrich, Jeffrey C. 23, 24, 25, 42
Giorgi, D. G. 71
Glazer, Alexander N. 91, 96
Goldberg, Mark 16
Golumbeski, George 155
Gonzalez, G. 41
Gonzalgo, Mark A. 51
Goodwin, Edwin H. 15
Goodwin, P. M. 10
Gordon, Laurie 50, 52, 54, 55, 59, 69, 73
Gracia, Emilio 59
Grad, Frank 139
Grady, D. L. 5, 9
Graham, K. S. 188
Granados, G. L. 44
Gray, Joe W. 93, 95, 195
Griffin, Patricia 174
Grothues, Dietmar 140
Grujic, D. 77
Guan, Xiaojun 108
Guigó, Roderic 86
Guilfoyle, Richard A. 141
Gusella, James F. 1
Gusfield, Daniel 159
Haas, Rose T. 105
Haces, Alberto 143
Hahn, Peter J. 142
Hanlon, D. J. 158
Hansen, A. D. A. 26
Hansma, Helen 40
Harding, John D. 11, 143
Hartman, John R. 144, 145
Hatada, I. 124
Hayashizaki, Y. 124
Heffelfinger, Dave 169
Henikoff, Stephen 84
Herrmannsfeldt, G. 146
Hettich, Robert L. 103
Hildebrand, C. E. 6, 7, 19, 88, 122
Himawan, Jeff 147
Hirasawa, T. 124
Hirotusne, S. 124
Hodgin, Faye Cobb 190
Hoffman, S. M. G. 52, 55, 62, 73
Hofstadler, Steven A. 111
Holtzman, Neil 139
Hom, K. 21
Honda, Sandra 148
Hood, L. 149
Hopkins, J. A. 184
Hozier, John 142
Huang, Xiaohua C. 135, 150
Huang, Xiaoqiu 148
Hubbard, Oroon 29
Hugentobler, M. 26
Hughes, A. John 143
Hung, Lydia 155
Hunkapiller, T. 146, 149
Hurst, Greg B. 103
Hutchinson, Jane S. 186
Ijadi, Mohamad 4
Imara, Mwalimu 190
Ioannou, Panayiotis A. 57, 63
Ito, Takashi 176
Ivanov, I. 167
Jackson, Cynthia L. 151
Jacobson, K. Bruce 97, 98, 99, 101, 104, 115
Jaehn, Laura 166
Jaklevic, J. M. 21, 26, 27, 28, 30, 31, 39, 44, 45
Jarvis, J. 78
Jett, J. H. 10
Johnson, Dabney 107
Johnson, M. E. 10
Johnson, Robert 174
Jones, R. G. 99
Joseph, Deborah 84
Jurka, Jerzy 152
Kallioniemi, A. 95
Kallioniemi, O. 95

Kao, Fa-Ten 57, 153
Karger, Barry L. 118
Katz, J. E. 28, 30, 31
Kawai, J. 124
Kececioglu, John 94
Kelleher, Zachary 192
Keller, Richard A. 10, 11
Kerlavage, Anthony R. 112
Khan, A. S. 184
Kieleczawa, Jan 81
Kim, Ung-Jin 49, 119, 154, 182
Kimmerly, William J. 29, 37
Kirk, V. C. 44
Klebig, M. L. 109
Knoche, Kimberly 155
Kolbe, William F. 28, 30, 31, 40
Kolner, Doug 141
Koop, B. 149
Kopelman, Raoul 156
Korenberg, Julie R. 119, 157, 198
Kouprina, N. 12
Kreindlin, E. 167
Kroisel, Peter 57, 63
Krystosek, Alphonse 174
Kuo, W.-L. 93, 95
Kwan, C. 52, 58
Kwon, H. 109
Labat, I. 78, 79
Lakey, Nathan 166
Lamerdin, Jane E. 52, 59, 64, 66, 67
Lane, M. J. 158
Langlois, Richard 60, 61
Langmore, John 156
Larimer, F. W. 104
Larionov, V. 12
Lasken, Roger 143
Lawler, Eugene L. 159
Lawrence, Charles 148
Lemanski, C. L. 8
Lennon, Gregory G. 52, 58, 65, 66
Lever, David C. 173
Lewis, S. E. 32, 39, 43, 46
Lieuallen, Kimberly 52, 65, 66
Lin, Ying 186
Ling, Lo-See Lucy 160
Link, Andrew 166
Lipshutz, Robert 135
Lishanskaya, A. I. 33
Lobb, R. 8
Longmire, Jonathan L. 5, 88
Lowry, Stephen R. 24, 34, 42
Luo, Cong-Wen 178
Lybrook, Sherrill 190
Lysov, Yu 167
MacDonell, M. T. 161
Macken, C. 2
MacLennan, D. H. 71
Maglott, Donna R. 170
Mancino, Valeria 182
Manly, K. 124
Mansfield, Betty K. 105
Mante, S. 158
Marchbanks, Michael 185
Marcom, Kim 174
Mark, Hon Fong L. 151
Markowitz, Victor M. 35, 36, 43, 47
Marr, Thomas 127
Marrone, Babetta L. 10, 13
Marsh, Dianne M. 144
Martin, Carol 168
Martin, Chris S. 162, 163
Martin, Christopher H. 29, 37, 38, 46
Martin, J. C. 10
Martin, Sheryl A. 105
Martin-Gallardo, Antonia 52, 67
Martinez, A. 5
Martinez, E. 3
Mascio, L. 95
Massa, Hillary 61, 189, 191
Mathies, Richard A. 150
Matsuda, Y. 124
Mayeda, Carol A. 37, 38
McCarthy, J. 32, 43, 47
McCarty, Katharrine 164
McCormick, Mary K. 2, 3, 5, 6, 13, 17
McEwen, Jean E. 164
McFarlane, J. 48
McInerney, Joseph D. 197
McNinch, Jennifer S. 49, 52, 57, 61
McPherson, J. D. 9
Mead, David A. 165
Medvick, Patricia A. 14
Meincke, L. J. 5
Meltzer, Paul 57

Meng, J. D. 28, 39
Merritt, Gregory 156
Meyne, Julianne 15
Mian, Alec 166
Michaud, E. J. 109
Micklos, David 196
Miller, Beate M. 195
Milosavljevic, A. 77
Mirzabekov, A. 167
Mohrenweiser, H. W. 52, 55, 62, 73
Moir, Donald T. 160
Monson, Eric 156
Montgomery, M. 52, 64
Moreno, Ruben 112
Moriwaki, K. 124
Moseley, Ray E. 190
Moyer, J. 109
Moyzis, R. K. 3, 5, 6, 7, 9, 12, 17
Mukai, T. 124
Mulley, J. C. 122
Mundt, Mark O. 16
Munk, C. 17
Mural, Richard J. 102, 106, 108
Murray, Matthew N. 40
Myers, Gene 148
Myers, Richard M. 194
Natowicz, Marvin 168
Nelson, David L. 172
Nelson, David O. 52, 68, 69
Nelson, Hillary C. M. 96
Nelson, J. Robert 123
Nguyen, Quan 169
Nierman, William C. 170
Nishitani, Y. 124
Noordewier, Michiel O. 181
Notarnicola, Steven M. 147
Ogletree, D. Frank 40
Okazaki, Y. 124
Olesen, Corinne E. M. 163
Olken, Frank 47, 94
Olsen, Anne S. 50, 52, 55, 58, 69
Orpana, A. K. 184
Orr, Bradford 156
Ostrander, E. A. 34, 42
Overbeek, Ross 80
Ow, David J. 52, 70
Owens, Elizabeth T. 105
Palazzolo, Michael J. 29, 30, 31, 32, 37, 38
Pallavicini, M. 93
Parr, Gary R. 134, 171
Parrish, Julia E. 172
Parrott, N. Wayne 148
Parsons, Rebecca J. 83
Paunesku, T. 77
Pearson, Peter L. 132
Pease, Ann 135
Pecherer, Robert M. 16, 18
Pelkey, JoAnne E. 110
Peshick, S. M. 158
Peters, D. 95
Petrov, Sergey 106
Pfeifer, G. P. 188
Phillips, M. S. 71
Phoenix, David 190
Pinkel, D. 93, 95
Piper, J. 95
Pirrung, Michael C. 173
Pitluck, S. 32, 43, 46
Polikoff, Daniel 195
Pollard, M. J. 30, 41, 44
Polymeropoulos, M. H. 184
Price, Morgan 80
Puck, Theodore T. 174
Quesada, Mark A. 150
Ragsdale, Charles 169
Ratliff, Robert L. 88
Redgrave, Graham W. 89
Reed, E. Corprew 127
Reilly, Philip R. 164
Reiner, Orly 175
Resnick, M. 12
Richards, R. I. 7, 122
Richardson, Charles C. 147
Richterich, Peter 166
Riggs, A. D. 188
Rinchik, Eugene M. 107
Rine, Jasper 33, 34, 42, 194
Rivenburgh, Reid 16
Roach, D. 116
Robbins, Robert J. 132
Robinson, D. L. 5, 9
Robison, Keith 166
Romo, Anthony J. 186
Rose, G. 21

Rosengaus, E. 116
Roszak, D. B. 161
Rouquier, S. P. 52, 71
Rowen, L. 149
Rubin, Gerald M. 29
Ruiz, Marie C. 118
Rutovitz, D. 95
Rye, Hays S. 96
Sachleben, Richard A. 98, 102, 104
Sachs, Rainer 189
Sakamoto, M. 93, 95
Salit, Jacqueline 127
Salmeron, Miquel 40
Sano, Takeshi 176
Santoro, Kathleen 151
Saunders, E. 17
Schad, Peter A. 177
Scheidecker, L. 52, 64
Scherer, Stewart 25, 42, 48
Scherthan, K. Harry 195
Schieltz, David 178
Schimke, R. Neil 126
Schneider, Klaus 179
Schor, P. L. 3, 5, 17
Scott, D. 22, 42
Searles, W. L. 26, 45
Searls, David B. 180
Segebrecht, Linda 126
Segraves, Marge 52, 189
Shadravan, Farideh 23, 24, 25, 42
Shah, Manesh 106
Shavlik, Jude W. 181
Shera, K. 17
Shi, Zhong-You 156
Shiroishi, T. 124
Shizuya, Hiroaki 49, 119, 154, 182
Shoshani, Arie 35
Shuey, Steven W. 173
Siciliano, Michael J. 183
Sikela, J. M. 184
Simon, Melvin I. 49, 119, 154, 182
Singh, Paramjit 91
Slepak, Tatiana 182
Slezak, Thomas R. 52, 70, 72, 74, 75
Sloop, Frederick V. 98, 104
Smith, Cassandra L. 40, 140, 176
Smith, Lloyd M. 121, 134, 141, 171, 185
Smith, Michael W. 186
Smith, Richard D. 111
Smith, Steven 156
Soares, M. Bento 112, 120
Soderlund, Cari A. 16, 19, 83, 90
Solomon, David 145
Spaar, M. T. 115
Speed, Terrence P. 113, 140
Spengler, Sylvia 40
Stallings, R. L. 6, 7, 122
Stavropoulos, N. 78
Steinhoff, David 145
Stengele, K.-P. 102
Stevens, T. J. 184
Stevko, Victor 29
Stilwagen, S. 64
Stinnett, Donna B. 105
Stolorz, Paul E. 83
Stone, Nancy 194
Stormo, Gary 84
Strathmann, Mike 37
Strezoska, Z. 77
Stricker, Jenny 197
Stubblebine, Will 169
Stubbs, Lisa 106, 107
Studier, F. William 81
Stultz, Karen 29
Su, L. 120
Sudar, D. 93, 95
Sun, Tian-Qiang 192
Sutherland, G. R. 6, 122
Sutton, Granger 133
Swerdlow, Harold 187
Tabor, Stanley 147
Tachi-iri, Yoshiaki 119, 182
Tan, Weihong 156
Tang, K. 99
Taylor, John A. 76
Tebbs, R. 64
Tesmer, J. G. 7
Theil, E. H. 32, 39, 43, 46
Thomas, Gregory S. 110
Thomas, Robert M. 178
Thompson, L. H. 64
Thomson, John 23
Thonnard, N. 104, 115
Thundat, T. G. 97

Törmänen, V. T. 188
Torney, David C. 2, 6, 16
Torok, T. 22, 42
Trask, Barbara J. 50, 60, 61, 189, 191
Trottier, Ralph W. 190
Troup, Charles D. 4
Tsujimoto, S. 52, 73
Tung, Chang-Shung 87
Uber, Don C. 39, 44, 45
Uberbacher, Edward 102, 106, 108
Uzgiris, Jim 141
Van den Engh, Ger 60, 61, 189, 191
Vardanega, Mike 191
Veklerov, E. 43, 46
Venter, J. Craig 112, 157
Vicentic, A. 78
Vos, Jean-Michel H. 192
Wagner, Mark C. 52, 72, 74
Wagner, R. P. 9
Wahl, Geoffrey M. 193
Waldman, F. 93, 95
Walsh, Kathleen 151
Warburton, Dorothy 120, 139
Warmack, R. J. 97
Warrington, Janet A. 194
Wasmuth, J. J. 9
Wassom, John S. 105
Watanabe, S. 124
Webb, Patricia 174
Wehnert, Manfred 175
Wei, Yalin 186
Weier, Heinz-Ulrich 93, 195
Weiss, R. B. 138
White, Owen 133
Whittaker, C. 2
Wilcox, A. S. 184
Williams, Peter 178
Wilson, K. M. 34, 42
Winger, Brian 111
Winternitz, Katherine 197
Witkowski, Jan 196
Witney, Frank 169
Wittwer, Carl 187
Wong, Benjamin S. 49, 52, 58, 61
Woychik, R. P. 104, 109
Wyrick, Judy M. 105
Yamashita, Robert 117
Yeh, T. Mimi 52, 70, 75
Yesley, Michael S. 20
Yeung, Edward S. 76
Yoshida, Kaoru 37
Yoshida, Thomas M. 13
Yu, Jingwei 153
Yu, Loh-Chung 195
Yu, M-T 120
Yust, Laura N. 105
Zeremski, M. 77
Zhu, Lin 171
Zhu, Y. 22, 42
Zoller, Paul W. 177
Zorn, M. D. 39, 43, 47, 48

Appendix C: Anticipated Workshop Attendees

Tentative Attendees List For The DOE Contractor-Grantee Meeting

February 7-10, 1993

Mark D. Adams
Receptor Biochemistry/Mol. Biology
National Institutes of Health, NINDS
Park Bldg., Rm. 405
Bethesda, MD 20892
phone: 301-496-8800
fax: 301-480-8588
E-mail:

Rhett Affleck
Life Sciences, MS M880
Los Alamos National Laboratory
Los Alamos NM 87545
phone: 505-667-2692
fax: 505-665-3024
E-mail:

Michelle Alegria-Hartman
Human Genome Center
Lawrence Livermore National Lab.
P. O. Box 808, L-452
Livermore, CA 94551-9900
phone: 510-422-4098
fax: 510-423-3608
E-mail:

Gregory E. Alexander
Mathematics & Statistics Dept.
University of California Berkeley
327 Evans
Berkeley CA 94720
phone: 510-642-4272
fax:
E-mail:

Michael Allen
Human Genome Center
Lawrence Livermore National Lab.
P.O. Box 808, L-452
Livermore, CA 94551-9900
phone: 510-422-6284
fax: 510-422-2282
E-mail:

Susan Allen
Human Genome Center
Lawrence Livermore National Lab.
P.O. Box 808, L-452
Livermore, CA 94551-9900
phone: 510-422-6284
fax: 510-422-2282
E-mail:

David P. Allison
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, TN 37831
phone: 615-574-5823
fax: 615-574-6210
E-mail:

Michael R. Altherr
Genetics Group
Los Alamos National Laboratory
MS M886
Los Alamos NM 87545
phone: 505-665-4007
fax: 505-665-3024
E-mail:

Chris T. Amemiya
Human Genome Center
Lawrence Livermore National Lab.
P. O. Box 808, L-452
Livermore, CA 94588
phone: 510-423-3634
fax: 510-423-3608
E-mail:

Norman G. Anderson
Large Scale Biology Corporation
9620 Medical Center Drive, St. 201
Rockville, MD 20850-3300
phone: 301-424-5989
fax: 301-762-4892
E-mail:

Sarah Anzick
MS-M888
Los Alamos National Laboratory
P.O. Box 1663
Los Alamos, NM 87544
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Linda Ashworth
Biology & Biotechnology Research
Lawrence Livermore National Lab.
P. O. Box 5507
Livermore, CA 94550
phone: 510-422-5665
fax: 510-423-3608
E-mail: linda@snrp.llnl.gov

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

Susan Bailey
Los Alamos National Laboratory
Los Alamos NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Joseph Balch
Electronics Engineering
Lawrence Livermore National Laboratory
P.O. Box 808, L-156
Livermore, CA 94550
phone:
fax: 510-422-3013
E-mail:

David F. Barker
Dept. of Medical Informatics
University of Utah Medical School
420 Chipeta Way #180, Research Park
Salt Lake City, UT 84086
phone: 801-581-5070
fax:
E-mail: dfbarker@cc.utah.edu

Benjamin J. Barnhart
Health Effects Research Division
U.S. Department of Energy
Office of Health and Energy Research
ER-72 F-201GTN
Washington DC 20585
phone: 301-903-5037, FTS 233
fax: 301-903-5051
E-mail: Barnhart@OERV01.ER.DOE.GOV

Mark Batzer
Human Genome Center
Lawrence Livermore National Laboratory
P.O. Box 808, L-452
Livermore, CA 94550
phone:
fax:
E-mail:

Diane Beeson
Department of Sociology
California State University, Hayward
Hayward CA 94542
phone:
fax:
E-mail:

Henry Benner
Engineering and Human Genome Center
Lawrence Berkeley Laboratory
MS70A-4475A
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-5636
fax: 510-486-5857
E-mail:

Scott Benson
University of California, Berkeley
Dept. of Molecular and Cell Biology
229 Stanley Hall
Berkeley CA 94720
phone:
fax:
E-mail:

Douglas E. Berg
Department of Molecular Microbiology
Washington University Medical Center
724 South Euclid Avenue
St. Louis, MO 63110-1093
phone: 314-362-2772
fax: 314-362-1232
E-mail: berg@borcim.wustl.edu

Claire M. Berg
Prof. of Biology
The University of Connecticut
Box U-131
354 Mansfield Road
Storrs CT 06269-2131
phone: 203-486-2916
fax: 203-486-1936
E-mail: BERG@UCONN.VM

Anne Bergmann
Human Genome Center
Lawrence Livermore National Laboratory
P.O. Box 808, L-452
Livermore, CA 94551-9900
phone:
fax:
E-mail:

Jan Berka
Barnett Institute
341 Mugar Bldg., Northeastern University
360 Huntington Avenue
Boston, MA 02115
phone: 617-437-2867
fax: 617-437-2855
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting

February 7-10, 1993

George Bers
Bio-Rad Labs
15111 San Pablo Avenue
Richmond, CA 94806
phone:
fax:
E-mail:

Anthony J. Beugelsdijk
Mechanical and Electronic Engineering
Los Alamos National Laboratory
MS/J580
Los Alamos NM 87545
phone: 505-667-3169
fax: 505-665-3911
E-mail:

Bruce W. Birren
Division of Biology, 147-75
California Institute of Technology
1201 E. California Blvd.
Pasadena CA 91125
phone: 818-356-4504
fax: 818-796-7066
E-mail:

M. Fatima Bonaldo
Dept. of Psychiatry
Columbia University
722 West 168th Street, Box #41
New York NY 10032
phone: 212-960-2313
fax: 212-795-5886
E-mail:

Edwin M. Bradbury
Life Sciences Division, Center for HG
Los Alamos National Laboratory
P.O. Box 1663
Los Alamos, NM 87545
phone: 505-667-2690
fax: 505-665-3024
E-mail:

Brigitte F. Brandriff
Human Genome Center
Lawrence Livermore National Lab.
P. O. Box 808, L-452
Livermore, CA 94551
phone: 510-423-0758
fax: 510-423-3608
E-mail:

Elbert Branscomb
Biomedical Science Division
Lawrence Livermore National Lab.
P. O. Box 5507, L-452
700 E. Avenue
Livermore, CA 94550
phone: 510-422-5681
fax: 510-422-2282
E-mail: elbert@alu.llnl.gov

Thomas M. Brennan
Department of Genetics
Stanford University
School of Medicine
Stanford CA 94305
phone: 415-725-7423
fax: 415-723-7016
E-mail:

Irena Bronstein
Tropix, Inc.
47 Wiggins Avenue
Bedford, MA 01730
phone: 617-271-0045
fax: 617- 275-8581
E-mail:

Gilbert M. Brown
Chemistry Department
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge TN 37831-6119
phone: 615-576-2756
fax: 615-576-5235
E-mail:

Nancy Brown
Mail Stop A114
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-685-3858
fax: 505-665-3858
E-mail:

David Bruce
Mail Stop A114
Los Alamos National Laboratory
Los Angeles, CA 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

Bob Brumley, Jr.
Department of Chemistry
University of Wisconsin
1101 University Avenue
Madison W I 53706
phone: 608-263-2594
fax: 608-262-0381
E-mail:

William Bruno
T-10, Mail Stop K710
Los Alamos National Laboratory
Los Alamos N M 87545
phone: 505-776-9468
fax:
E-mail:

Michelle V. Buchanan
Chemistry Department, MS 6120
Oak Ridge National Laboratory
P.O. Box 2008
Bldg. 4500 South
Oak Ridge TN 37831-6120
phone: 615-574-4868
fax: 615-576-5235
E-mail:

Donna M. Buley
Biology Division
University of Tennessee
c/o Oak Ridge National Laboratory
P.O. Box 2009,
Oak Ridge, TN 37831-807
phone:
fax:
E-mail:

Christian Burks
Group T-10, MS K710
Los Alamos National Laboratory
Los Alamos, N M 87545
phone: 505-667-6683
fax: 505-665-3493
E-mail: cb@life.lanl.gov

David Callen
Dept. of Cytogenetics & Mol.Genetics
Adelaide Children's Hospital
72 King William Road
North Adelaide S.A. 5006
phone: 618-267-7284
fax: 618-267-7342
E-mail:

Mary Campbell
Center for Human Genome Studies
Los Alamos National Laboratory
Los Alamos N M 87545
phone: 505-665-4438
fax: 505-665-3024
E-mail:

Evelyn Campbell
Center for Human Genome Studies
Los Alamos National Laboratory
Los Alamos N M 87545
phone: 505-665-4438
fax: 505-665-3024
E-mail:

Charles Cantor
Dir., Center for Advanced Res. of Biotech.
Boston Universtiy
36 Cummington, Street
Boston, M A 02215
phone: 617-353-8500
fax: 617-353-8501
E-mail:

Charles Carlson
The Exploratorim
3601 Lyon Street
San Francisco, CA 94123
phone: 415-461-0341
fax: 415-561-0307
E-mail:

Anthony Carrano
Director, Human Genome Center
Lawrence Livermore National Lab.
P. O. Box 808, L-452
Livermore, CA 94550
phone: 510-422-5698
fax: 510-423-3608
E-mail: avc@sts.llnl.gov

Denise K. Casey
Health and Safety Research Division
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge TN 37831-6119
phone: 615-576-6669
fax: 615-574-9888
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

C. Thomas Caskey
Institute for Molecular Genetics
Baylor College of Medicine
Texas Medical Center, T-809
One Baylor Plaza
Houston, TX 77030
phone: 713-798-4774
fax: 713-798-7383
E-mail: e-mail:

Alonso Castro
Center for Human Genome Studies
Los Alamos National Laboratory
MS-D434
Los Alamos NM 87545
phone: 505-667-3228
fax: 505-665-3644
E-mail:

Brian Chait
The Rockefeller University
1230 York Avenue
New York, NY 10021
phone:
fax:
E-mail:

Verne M. Chapman
Molecular & Cell Biology
Roswell Park Cancer Institute
Elmand Carlton Street
Buffalo, NY 14263
phone: 716-845-2300
fax: 716-845-8169
E-mail: e-mail:

C. H. Winston Chen
Health and Safety Research Division
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, TN 37831-6119
phone: 615-574-5895
fax:
E-mail:

Chira Chen
Human Genome Center, Biomecial Sci.
Lawrence Livermore National Laboratory
P.O. Box 5507
Livermore CA 94550
phone: 510-423-4927
fax: 510-423-3608
E-mail:

David Chen
Mail Stop A114
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Jan-Fang Cheng
Human Genome Center
Lawrence Berkeley Laboratory
1 Cyclotron Road
MS74-3110
Berkeley CA 94720
phone: 510-486-6590
fax: 510-486-6816
E-mail: JFCheng@lbl.gov

Xueheng Cheng
Chemical Sciences Department
Pacific Northwest Laboratory
1101 University Avenue
Richland W A 99352
phone: 509-376-0723 or 5665
fax: 509-376-0418
E-mail:

L Cherath
Department of Medicine and
State University of New York-Health Sci.
750 E. Adams Street
Syracuse NY 13210
phone: 315-464-5446
fax: 315-464-8255
E-mail:

Chau-Wen Chou
Deparment of Chemistry
Arizona State University
Tempe, AZ 85287-1604
phone: 602-965-3461
fax: 602-965-2747
E-mail:

Mari Christensen
Lawrence Livermore National Laboratory
P.O. Box 5507, L-452
Livermore, CA 94550
phone:
fax:
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting

February 7-10, 1993

Michael Cinkosky
Theoretical Biology and Biophysics Group
Los Alamos National Laboratory
P.O. Box 1663
Los Alamos NM 87545
phone: 505-665-0840
fax: 505-665-3493
E-mail: michael@genome.lanl.gov

Steven Clark
Molecular Genetics Laboratory
The Salk Institute for Biological Studies
P.O. 85800
La Jolla CA 92037
phone: 619-453-4100
fax: 619-558-9513
E-mail:

Lynn Clark
Mail Stop A114
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Debra L. Collins
Department of Medical Genetics
University of Kansas Medical Center
Rainbow at 39th
Kansas City KS 66103
phone: 913-588-6043
fax: 913-588-3995
E-mail:

Francis S. Collins
Medical Genetics
University of Michigan Medical Center
1150 W. Medical Center Dr., 4570
Ann Arbor, MI 48109-0618
phone: 313-747-3414
fax: 313-763-4692
E-mail:

Alex Copeland
Human Genome Center
Lawrence Livermore National Lab.
P. O. Box 808, L-452
Livermore, CA 94550
phone: 510-422-5665
fax: 510-423-3608
E-mail:

Leilani Corell
Human Genome Center
Lawrence Livermore National Lab.
P. O. Box 808, L-452
Livermore, CA 94550
phone: 510-423-3110
fax: 510-423-3608
E-mail:

David Cox
Neurogenetics Lab
University of California, San Francisco
Dept. of Psychiatry & The Langley Porter
401 Parnassus Avenue
San Francisco CA 94143-0984
phone: 415-476-4212
fax: 415-476-4009
E-mail:

L. Scott Cram
Life Sciences Division
Los Alamos National Laboratory
Los Alamos, CA 87545
phone: (505) 667-2690
fax: (505) 665-3024
E-mail:

Lee A. Crandall
Department of Community Health &
University of Florida
Gainesville FL 32610
phone: 904-392-4321
fax:
E-mail:

Radomir Crkvenjakov
Biological & Medical Res. Division
Argonne National Laboratory
9600 South Cass Avenue
Argonne, IL 60439-4833
phone: 708-972-3161
fax: 708-972-3387
E-mail: crkve@mcs.anl.gov

Jackson B. Davidson
Instrumentation and Controls Division
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, TN 37831-6119
phone: 615-574-5599
fax: 615-574-4058
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

Courtney Davidson
Electronic Engineering
Lawrence Livermore National Laboratory
P.O. Box 808, L-156
Livermore, CA 94550
phone:
fax:
E-mail:

Ronald Davis
Department of Biochemistry
Stanford University
School of Medicine
Beckman Center, B400
Stanford, CA 94305-5307
phone: 415-723-6277
fax: 415-723-6783
E-mail:

Donn Davy
Information and Computing Sciences
Lawrence Berkeley Laboratory
50B-3216
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-5041
fax: 510-486-4004
E-mail:

Pieter De Jong
Human Genome Center
Lawrence Livermore National Lab.
P. O. Box 5507, L-452
Livermore, CA 94551
phone: 510-423-8145
fax: 510-423-3608
E-mail: pieter@pcr.llnl.gov

Larry Deaven
Center for Human Genome Studies, Life
Los Alamos National Laboratory
LS-4, MS M888
Los Alamos NM 87545
phone: 505-667-3114
fax: 505-665-3024
E-mail:

Karen Denison
Mail Stop A114
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

M. Bonner Denton
Department of Chemistry
University of Arizona
Tucson AZ 85721
phone: 602-621-8246
fax: 602-621-8272
E-mail:

Margie Dere
Human Genome Center, MS 1-213
Lawrence Berkeley Laboratory
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-4943
fax: 510-486-5717
E-mail:

Lorie Devlin
Human Genome Center
Lawrence Livermore National Lab.
P. O. Box 808, L-452
Livermore, CA 94551
phone:
fax:
E-mail:

Norman Doggett
Genetics Group
Los Alamos National Laboratory
MS M886
Los Alamos NM 87545
phone: 505-665-4007
fax: 505-665-3024
E-mail: Doggett@Flovax.LANL.Gov

Mitchel J. Doktycz
Biology Division
Oak Ridge National Laboratory
P. O. Box 2009
Oak Ridge, TN 37831
phone: 615-576-2756
fax: 615-576-5235
E-mail:

Peter Domenici
427 Dirksen Bldg.
Washington, DC 20003
phone: 202-224-6621
fax: 202-224-7371
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

Nancy Domenici
427 Dirksen Bldg.
Washington, DC 20003
phone: 202-224-6621
fax: 202-224-7371
E-mail:

Richard J. Douthart
Life Sciences Center
Battelle Pacific Northwest Laboratory
P. O. Box 999
Mail Stop K4-13
Richland, WA 99352
phone: 509-375-2653
fax: 509-375-3649
E-mail: dick@gnome.pnl.gov

Norman J. Dovichi
Department of Chemistry
University of Alberta
Edmonton, Alberta
CANADA T6G 2G2
phone: 403-492-3254 or 2845
fax: 403-492-8231
E-mail:

Dan Drell
DOE Human Genome Program
OHER, U.S. Department of Energy
ER-72 GTN
Washington DC 20585
phone: 301-903-4742
fax: 301-903-5051
E-mail:

Radoje Drmanac
Biological & Medical Res. Div.
Argonne National Laboratory
9700 South Cass Avenue
Argonne, IL 60439-4833
phone: 708-972-3175
fax: 708-972-3387
E-mail:

Snezana Drmanac
Biological & Medical Res. Div.
Argonne National Laboratory
9700 South Cass Avenue
Argonne, IL 60439-4833
phone: 708-972-3175
fax: 708-972-3387
E-mail:

Lynn Duesing
Mail Stop A114
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

John J. Dunn
Biology Department
Brookhaven National Laboratory
Upton, NY 11973
phone: 516-282-2123
fax: 516-282-3407
E-mail:

Troy Duster
Institute for the Study of Social Change
University of California
2420 Bowditch Avenue
Berkeley CA 94720
phone: 510-642-0813
fax: 510-642-8674
E-mail:

Jeffrey M. Elliott
Human Genome Center
Lawrence Livermore National Laboratory
P.O. Box 808, L-452
Livermore, CA 94551-9900
phone:
fax:
E-mail:

Mike Engle
Group T-10, Mailstop K710
Los Alamos National Laboratory
Los Alamos NM 87545
phone:
fax:
E-mail: mle@life.lanl.gov

Gary A. Epling
Dept. of Chemistry
University of Connecticut
U-60
Storrs, CT 06269
phone: 203- 486-3215
fax: 203-486-2981
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

Sandra Escobar
Human Genome Center, MS 1-213
Lawrence Berkeley Laboratory
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-4943
fax: 510-486-5717
E-mail:

Rich Esposito
Human Genome Center
Lawrence Livermore National Laboratory
P.O. Box 808, L-452
Livermore, CA 94551-9900
phone:
fax:
E-mail:

Glen Evans
Molecular Genetics Laboratory
Salk Institute for Biological Studies
P.O. Box 85800
San Diego CA 92037
phone: 619-453-4100 x279
fax: 619-558-9513
E-mail: gevans@salk_sd2.sdsc.edu

Jerry Eveleth
Human Genome Center
Lawrence Livermore National Laboratory
P.O. Box 5507, L-452
Livermore, CA 94551-9900
phone: 510-422-2780
fax: 510-423-3608
E-mail:

Vance Faber
Information and Computing Sciences Div.
Los Alamos National Laboratory
T-10, K710
Los Alamos NM 87545
phone: 505-667-7510
fax: 505-665-3493
E-mail:

Betsy Fader
Student Pugwash USA
1638 R Street, NW Suite 32
Washington, DC 20009
phone: 202-328-6555
fax: 202-797-4664
E-mail:

Kenneth H. Fasman
Genome Data Base
Johns Hopkins University
Welch Medical Library
1830 E. Monument Street
Baltimore, MD 21205
phone: 301-955-9705
fax: 301-955-0054
E-mail:

Elise Feingold
Human Genome Research
National Institutes of Health
Bldg. 38A, Rm. 605
9000 Rockville Pike
Bethesda, MD 20892
phone: 301-496-0844
fax: 301-402-0837
E-mail:

Ilise L. Feitshans
Legislative Drafting Research Fund
Columbia University Law School
435 West 116 Street
New York NY 10027
phone: 212-854-2685
fax: 212-854-7946
E-mail:

Steve Ferris
Bio-Rad Labs
2000 Alfred Noble Drive
Hercules, CA 94547
phone:
fax:
E-mail:

James W. Fickett
Theoretical Biology & Biophysics
Los Alamos National Laboratory
Group T-10, MS K710
Los Alamos, NM 87545
phone: 505-665-5340
fax: 505-665-3493
E-mail: jwf@life.lanl.gov

Chris Fields
The Institute for Genomic Research
932 Clopper Road
Gaithersburg, MD 20878
phone: 301-869-9056
fax: 301-869-9423
E-mail: e-mail: cfields@tigr.org

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

Michael C. Fitzgerald
Department of Chemistry
University of Wisconsin
1101 University Avenue
Madison W I 53706
phone: 608-263-2594
fax: 608-262-0381
E-mail:

Steve Fodor
Affymax Research Institute
4001 Miranda Avenue
Palo Alto, CA 94304
phone:
fax:
E-mail:

Robert S. Foote
Biology Division
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, TN 37831-6119
phone:
fax:
E-mail:

Frantisek Foret
Barnett Institute
341 Mugar Bldg., Northeastern University
360 Huntington Avenue
Boston, MA 02115
phone: 617-437-2867
fax: 617-437-2855
E-mail:

Jane E. Fullarton
Div. of Health Sciences Pooicy
National Academy of Sciences
Institute of Medicine
2101 Constitution Avenue (FO3016)
Washington DC 20418
phone: 202-334-3913
fax: 202-334-1385
E-mail:

David Galas
Associate Director, OHER
U.S. Department of Energy
ER70 GTN
Washington DC 20585
phone: 301-903-3251
fax: 301-903-5051
E-mail:

Emilio Garcia
Human Genome Center
Lawrence Livermore National Lab.
P. O. Box 808, L-452
Livermore, CA 94551-9900
phone: 510-422-8002
fax: 510-423-3608
E-mail:

Skip Garner
Bldg. 2, MS 616
General Atomics
3550 General Atomics Ct.
San Diego, CA 92121
phone: 619-455-3464
fax: 619-455-3233
E-mail:

Jeffrey A. Garnes
Human Genome Center
Lawrence Livermore National Laboratory
P.O. Box 808, L-452
Livermore, CA 94551-9900
phone:
fax:
E-mail:

Joe M. Gatewood
Life Sciences Div., Center for HG Studies
Los Alamos National Laboratory
P.O. Box 1663
Los Alamos, NM 87545
phone: 505-667-2690
fax: 505-665-3024
E-mail:

Raymond F. Gesteland
Department of Human Genetics
University of Utah
6160 Eccles Genetics Bldg.
Salt Lake City, UT 84112
phone: 801-581-5190
fax: 801-585-3910
E-mail: RAYG@UTAHMED

Jeff C. Gingrich
Human Genome Center
Lawrence Berkeley Laboratory
MS74-157
1 Cyclotron Road
Berkeley CA 97420
phone: 510-486-6580
fax: 510-486-6816
E-mail: JCGingrich@lbl.gov

Tentative Attendees List For The
DOE Contractor-Grantee Meeting
February 7-10, 1993

Dominique Giorgi
Human Genome Center
Lawrence Livermore National Laboratory
P.O. Box 808, L-452
Livermore, CA 94551-9900
phone:
fax:
E-mail:

Gerald Goldstein
Physical & Technical Research Division,
U.S. Department of Energy
ER 74-GTN
Washington DC 20585
phone: 301-903-3213
fax: 301-903-5051
E-mail:

George Golumbeski
Promega Corporation
2800 Woods Hollow Road
Madison W I 53711
phone: 608-274-4330 X1252
fax: 608-273-6967
E-mail:

Ed Goodwin
Mail Stop A114
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Frank Grad
Legislative Drafting Research Fund
Columbia University Law School
435 West 116 Street
New York NY 10027
phone: 212-854-2685
fax: 212-854-7946
E-mail:

Deborah L. Grady
Center for HumanGenome Studies
Los Alamos National Laboratory
MS M886
Los Alamos, NM 87545
phone: 505-667-2695
fax: 505-665-3024
E-mail:

Kenneth Graham
Biology Department
Beckman Research Institute of the City of
1450 East Duarte Road
Duarte CA 91010-0269
phone: 818-301-8352
fax: 818-358-7703
E-mail:

Joe Gray
Division of Molecular Cytometry
University of California, San Francisco
Rm. 230
1855 Folsom Street
San Francisco CA 94143-0808
phone: 415-476-3461
fax: 415-476-8218
E-mail: joe_gray@dmcmail.ucsf.edu

Dietmar Grothues
Boston Universtiy
36 Cummington, Street
Boston, MA 02215
phone: 617-353-8500
fax: 617-353-8501
E-mail:

Xiaojun Guan
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, TN 37831-6050
phone: 615-576-6669
fax: 615-574-9888
E-mail:

Roderick Guigo
Mail Stop A114
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Richard A. Guilfoyle
University of Wisconsin
1101 University Avenue
Madison W I 53706
phone: 608-263-2594
fax: 608-262-0381
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

Daniel Gusfield
Dept. of Computer Science
University of California, Davis
Chemistry Annex Bldg.
Davis CA 95616
phone: 916-752-7131
fax: 916-752-4767
E-mail:

Roswitha T. Haas
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, TN 37831-6119
phone: 615-574-5599
fax: 615-574-4058
E-mail:

Peter Hahn
Department of Radiology
State University of New York-Health Sci.
750 E. Adams Street
Syracuse NY 13210
phone: 315-464-5956
fax:
E-mail:

James F. Hainfeld
Brookhaven National Laboratory
Upton NY 11973
phone: 516-282-3372
fax: 516-282-3407
E-mail:

Tony Hansen
Human Genome Center, MS 70A-3363
Lawrence Berkeley Laboratory
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-7158
fax: 510-486-5857
E-mail:

John D. Harding
Corporate Research
Life Technologies, Inc.
P.O. Box 6009
8/17 Grovemont Circle
Gaithersburg, MD 20877
phone: 301-840-8000
fax: 301-948-8977
E-mail:

Fred C. Hartman
Biology Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831-8077
phone: FTS 624-0212
fax: 624-9297
E-mail:

John R. Hartman
Computational Biosciences, Inc.
P. O. Box 2090
Ann Arbor, MI 48106
phone: 313- 426-9050
fax: 313-426-5311
E-mail: e-mail: john@cbi2.cbi.com

Yoshihide Hayashizaki
Tsukuba Life Science Center
Riken Japan
phone:
fax:
E-mail:

G. Herrmannsfeldt
Department of Molecular Biotechnology
University of Washington
MS GJ-10
4909 25th Avenue N.E.
Seattle, WA 98195
phone: 206-685-7387
fax: 206-685-7367
E-mail:

Robert L. Hettich
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, TN 37831-6119
phone: 615-574-5599
fax: 615-574-4058
E-mail:

C. Edgar Hildebrand
Life Sciences Division
Los Alamos National Laboratory
MS M885
Los Alamos, NM 87545
phone: 505-667-2746
fax: 505-665-3024
E-mail: Ceh@telomere.lanl.gov

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

Jeff Himawan
Dept. of Biological Chem. & Mol. Phar.
Harvard Medical School
240 Longwood Avenue
Boston MA 02115
phone: 617-432-3129
fax: 617-432-3362
E-mail:

Susan M. G. Hoffman
Human Genome Center
Lawrence Livermore National Laboratory
P.O. Box 808, L-452
Livermore, CA 94551-9900
phone:
fax:
E-mail:

Robert M. Hollen
Mechanical and Electronic Engineering
Los Alamos National Laboratory
MS/J580
Los Alamos NM
phone: 505-667-3169
fax: 505-665-3911
E-mail:

Linda Holmes
Science and Engineering Education
Oak Ridge Associated Universities
P.O. Box 117
Oak Ridge TN 37831-6119
phone: 615-576-3192
fax: 615-576-0202
E-mail:

Sandra Honda
Baylor College of Medicine
One Baylor Plaza
Houston TX 77030-3498
phone: 713-798-6226
fax: 713-790-1275
E-mail:

Leroy H. Hood
Department of Molecular Biotechnology
University of Washington
MS GJ-10
4909 25th Avenue N.E.
Seattle, WA 98195
phone: 206-685-7367
fax: 206-685-7301
E-mail:

Xiaohua Huang
937 Jackson Street
Mountain View, CA 94043
phone: 415-496-2304
fax:
E-mail:

Li-Chun Huang
Gene Expression Laboratory
The Salk Institute for Biological Studies
P.O. Box 85800
La Jolla, CA 92037
phone: 619-453-4100 x 587
fax: 619-455-1349
E-mail:

Lydia Huang
Promega Corporation
2800 Woods Hollow Road
Madison WI 53711
phone: 608-274-4330
fax: 608-273-6967
E-mail:

Eliezer Huberman
Biological and Medical Res. Div.
Argonne National Laboratory
9700 South Cass Avenue
Argonne, IL 60439-4833
phone: (708) 252-3819
fax: (708) 972-3387
E-mail:

Cynthia L. Jackson
Cytogenetics Laboratory
Rhode Island Hospital
593 Eddy Street
Providence RI 02903
phone: 401-277-4370
fax: 401-277-8514
E-mail:

K. Bruce Jacobson
Biology Division
Oak Ridge National Laboratory
P. O. Box 2009
Oak Ridge, TN 37831
phone: 615-576-2756
fax: 615-576-5235
E-mail:

Tentative Attendees List For The
DOE Contractor-Grantee Meeting
February 7-10, 1993

Joe Jaklevic
Engineering Division and Human
Lawrence Berkeley Laboratory
MS70A-3363 F
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-5647
fax: 510-486-5857
E-mail: JMJaklevic@lbl.gov

James Jett
Center for Human Genome Studies, Life
Los Alamos National Laboratory
LS-4, MS M888
Los Alamos NM 87545
phone: 505-667-3843
fax: 505-665-3024
E-mail: Jett@flovax.lanl.gov

Myrna Jones
Mail Stop A114
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Jerzy Jurka
Linus Pauling Institute of Science and
440 Page Mill Road
Palo Alto, CA 94306
phone: 510-327-4064
fax: 510-327-8564
E-mail: jurek@jmullins.stanford.edu

Fa-Ten Kao
Eleanor Roosevelt Institute
1899 Gaylord Street
Denver, CO 80262
phone: 303-333-4515
fax: 303-333-8423
E-mail:

Joe Katz
Engineering and Human Genome Center
Lawrence Berkeley Laboratory
MS70A-4475A
1 Cyclotron Road
Berkeley, CA 94720
phone: 510-486-5636
fax: 510-486-5857
E-mail:

Richard A. Keller
Center for Human Genome Studies
Los Alamos National Laboratory
MS G738
Los Alamos NM 87545
phone: 505-667-3018
fax:
E-mail:

Jan Kieleczawa
Brookhaven National Laboratory
Upton, Long NY 11973
phone:
fax:
E-mail:

Ung-Jin Kim
Division of Biology
California Institute of Technology
147-75
Pasadena CA 91125
phone: 818-356-3781
fax: 818-796-7066
E-mail:

William J. Kimmerly
Human Genome Center
Lawrence Berkeley Laboratory
MS 74-157
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-5909
fax: 510-486-6816
E-mail:

Jan Kleeczawa
Biology Department
Brookhave National Laboratory
Building 463
Upton, NY 11973
phone:
fax:
E-mail:

Leonard Klevan
Molecular Biology Researh and
Life Technologies, Inc.
P.O. Box 6009
8/17 Grovemont Circle
Gaithersburg, MD 20877
phone: 301-840-8000
fax: 301-948-8977
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

Kimberly Knoche
Promega Corporation
2800 Woods Hollow Road
Madison, W I 53711
phone: 608-274-4330
fax: 608-273-6967
E-mail:

Julie R. Korenberg
Medical-Genetics
Cedars-Sinai Medical Center
ASB 378
8700 Beverly Blvd.
Los Angeles CA 90048-0750
phone: 310-855-6451
fax: 310-967-0112
E-mail:

Chinnie Kwan
Human Genome Center
Lawrence Livermore National Laboratory
P.O. Box 808, L-452
Livermore, CA 94551-9900
phone:
fax:
E-mail:

Michael J. Lane
Department of Medicine and
State University of New York-Health Sci.
750 E. Adams Street
Syracuse NY 13210
phone: 315-464-5446
fax: 315-464-8255
E-mail:

William Kolbe
Human Genome Center
Lawrence Berkeley Laboratory
MS70A-2205
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-7199
fax: 510-486-5857
E-mail: WFKolbe@lbl.gov

Natasha Kouprina
Human Genome Center
Los Alamos National Laboratory
CHGS - MS M885
Los Alamos, NM 87545
phone: 505-667-3912
fax: 505-665-3024
E-mail:

Cheryl Lamanski
Mail Stop A114
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Rich Langlois
Human Genome Center
Lawrence Livermore National Laboratory
P.O. Box 808, L-452
Livermore, CA 94551-9900
phone: 510-423-3841
fax:
E-mail:

Raoul Kopelman
Department of Chemistry
University of Michigan
4744 Chemistry Bldg.
2200 Bonisteele Blvd.
Ann Arbor MI 48109-1065
phone: 313-764-7541
fax: 313-764-7315
E-mail: usergb2q@ub.cc.umich.edu

Wen-Lin Kuo
Division of Molecular Cytology
University of California, San Francisco
Department of Laboratory Medicine
San Francisco CA 94143-0808
phone: 415-476-3461
fax: 415-476-8218
E-mail:

Jane E. Lamerdin
Human Genome Center
Lawrence Livermore National Lab.
P. O. Box 808, L-452
Livermore, CA 94550
phone:
fax:
E-mail:

John Langmore
Biophysics Research Division
University of Michigan
Ann Arbor MI 48109-1065
phone: 313-764-7541
fax: 313-764-7315
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting

February 7-10, 1993

Vladimir L. Larionov
NIEHS
Box 12233
Research NC 27709
phone:
fax:
E-mail:

Charles B. Lawrence
Department of Cell Biology
Baylor College of Medicine
Molecular Biology Information Res.
One Baylor Plaza, Rm. M525
Houston, TX 77030
phone: 713-798-6226
fax: 713-790-1275
E-mail: chas@mbir.bcm.tmc.edu

Gregory G. Lennon
Human Genome Center
Lawrence Livermore National Laboratory
P.O. Box 808, L-452
Livermore CA 94550
phone: 510-422-5711
fax: 510-423-3608
E-mail: greg@mendel.llnl.gov

David Lever
Department of Chemistry
Duke University
Durham, NC 27706
phone: 919-660-1564
fax: 919-660-1591
E-mail:

Suzanna Lewis
Information and Computing Sciences Div
Lawrence BerkeleyLaboratory
MS 50B-3029H
1 Cyclotron Road
Berkeley, CA 94720
phone: 510-486-7370
fax: 510-486-4004
E-mail: selewis@lbl.gov

Kim Lieuallen
Lawrence Livermore National Laboratory
P.O. Box 5507, L-452
Livermore, CA 94550
phone:
fax:
E-mail:

Lo-See Lucy Ling
Collaborative Research, Inc.
1365 Main Street
Waltham, MA 02154
phone: 617-487-7979
fax: 617-891-5062
E-mail:

Alla Lishanskaya
Human Genome Center, MS 74-157
Lawrence Berkeley Laboratory
1 Cyclotron Road
Berkeley, CA 94720
phone: 510-486-7332
fax: 510-486-6816
E-mail:

Beckie Lobb
Mail Stop A114
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Jonathan Longmire
Center for Human Genome Studies, Life
Los Alamos National Laboratory
MS M886
Los Alamos, NM 87545
phone: 505-667-8208
fax: 505-665-3024
E-mail:

Steven Lowry
Human Genome Center, MS 74-157
Lawrence Berkeley Laboratory
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-6549
fax: 510-486-6816
E-mail:

Amanda Lumley
Science & Engineering Education. Div.
Oak Ridge Associated Universities
P.O. Box 117
Oak Ridge TN 37831-6119
phone: 615-576-4811
fax: 615-576-0202
E-mail: lumleya%orau2@cunyvm.cuny.edu

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

Linda MaCamant
Science/Engineering Education Div.
Oak Ridge Inst. for Science and Education
P.O. Box 117
Oak Ridge, TN 37831-0117
phone: 615-576-3192
fax: 615-576-3192
E-mail:

Michael T. MacDonell
Ransom Hill Bioscience
P.O. Box 219
Ramona CA 92065
phone:
fax:
E-mail:

Kathy Macken
Theoretical Biology and Biophysics Group
Los Alamos National Laboratory
Group T-10, MS K710
Los Alamos, NM 87545
phone: 505-665-1970
fax: 505-665-3493
E-mail:

Catherine Mackern
Mail Stop A114
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Donna R. Maglott
American Type Culture Collection
12301 Parklawn Drive
Rockville MD 20852-1176
phone: 301-231-5559
fax: 301-770-1848
E-mail:

Vladimir Makarov
Biophysics Research Div.
University of Michigan
2200 Bonisteel Blvd.
Ann Arbor MI 48109-2099
phone: 313-264-5258
fax: 313-264-3233
E-mail:

Janice Mann
Human Genome Center, MS 1-213
Lawrence Berkeley Laboratory
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-4943
fax: 510-486-5717
E-mail:

Betty K. Mansfield
Health and Safety Research Division
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, TN 37831-6119
phone: 615- 576-6669
fax: 615-574-9888
E-mail: bkq@ornl.gov

Victor M. Markowitz
Data Management Group, ICSD
Lawrence Berkeley Laboratory
50B-3209B
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-6835
fax: 510-486-4004
E-mail: markowitz@lbl.gov

Thomas Marr
The Cold Spring Harbor Laboratory
P.O. Box 100
Cold Spring NY 11724
phone: 516- 367-8393
fax: 516-367-8461
E-mail: marr@cshl.org

Babetta L. Marrone
LS-4, M888
Los Alamos National Laboratory
Los Alamos, NM 87544
phone: 505-667-3279
fax: 505-665-3024
E-mail: Marrone@FLOVAX.LANL.gov

Christopher Martin
Tropix, Inc.
47 Wiggins Avenue
Bedford, MA 01730
phone: 617-271-0045
fax: 617-275-8581
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting

February 7-10, 1993

John C. Martin
Los Alamos National Laboratory
P.O. Box 1663
Los Alamos NM 87545
phone: 505-667-3018
fax: 505-665-3024
E-mail:

Christopher H. Martin
Human Genome Center
Lawrence Berkeley Laboratory
MS 74-157
1 Cyclotron Road
Berkeley, CA 94720
phone: 510-486-5909
fax: 510-486-6816
E-mail: chrism@lbl.gov

Christopher Martin
Tropix, Inc.
47 Wiggins Avenue
Bedford, MA 01730
phone: 617-271-0045
fax: 617-275-8581
E-mail:

Antonia Martin-Gallardo
Human Genome Center
Lawrence Livermore National Laboratory
P.O. Box 808, L-452
Livermore, CA 94551-9900
phone:
fax:
E-mail:

Richard Mathies
Chemistry Dept.
University of California, Berkeley
312 Hildebrand
Berkeley CA 94720
phone: 510-642-4192
fax: 510-642-3599
E-mail:

Kathleen Mavournin
Health and Safety Research Division
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, TN 37831-6119
phone: 615- 576-6669
fax: 615-574-9888
E-mail: bkq@ornl.gov

Carol A. Mayeda
Human Genome Center
Lawrence Berkeley Laboratory
MS 74-157
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-5909
fax: 510-486-6816
E-mail:

Linda McCamant
Science/Engineering Education Division
Oak Ridge Institute for Science and
P.O. Box 117
Oak Ridge, TN 37831-0117
phone: 615-576-1089
fax:
E-mail:

Erin C. McCanlies
Los Alamos National Laboratory
P.O. Box 1663
MS-M880
Los Alamos, NM 87544
phone:
fax:
E-mail:

John McCarthy
Information and Computing Sciences
Lawrence Berkeley Laboratory
50B-3216
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-5041
fax: 510-486-4004
E-mail:

Mary Kay McCormick
Center for Human Genome Studies
Los Alamos National Laboratory
M 886
Los Alamos NM 87545
phone: 505-665-4438
fax: 505-665-3024
E-mail: mkm@flovax.lanl.gov

Jean E. McEwen
Shriver Center
200 Trapelo Road
Waltham, MA 02254
phone: 617-642-0292
fax:
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting

February 7-10, 1993

Joseph D. McInerney
Biological Sciences Curriculum Study
The Colorado College
830 North Tejon, Ste. 405
Colorado Springs CO 80903
phone: 719-578-1136
fax: 719-578-9126
E-mail:

Jennifer McNinch
Human Genome Center
Lawrence Livermore National Laboratory
P.O. Box 808, L-452
Livermore, CA 94551-9900
phone:
fax:
E-mail:

David A. Mead
Department of Chemistry
University of Wisconsin
1101 University Avenue
Madison W I 53706
phone: 608-262-2021
fax: 608-262-0381
E-mail:

Patricia A. Medvick
Mechanical and Electronic Engineering
Los Alamos National Laboratory
MS J580
Los Alamos NM 87545
phone: 505-667-2676
fax: 505-665-3911
E-mail: pm@lanl.gov

Linda Meincke
Center for Human Genome Studies
Los Alamos National Laboratory
Los Alamos NM 87545
phone: 505-665-4438
fax: 505-665-3024
E-mail:

John D. Meng
Engineering Div. and Human Genome
Lawrence Berkeley Laboratory
MS 29-100
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-5117
fax:
E-mail:

Julianne Meyne
Center for Human Genome Studies
Los Alamos National Laboratory
LS-3 MS M886
Los Alamos NM 87545
phone: 505-667-3912
fax: 505-665-3024
E-mail:

Alex Mian
Harvard Medical School
240 Longwood Avenue
Boston MA 02115
phone: 617-432-7561
fax: 617-432-7663
E-mail:

Jerome P. Miksche
Director, Office of Plant Genome Mapping
U.S. Department of Agriculture
Agricultural Research Service
BARC-West, Bldg. 005, Rm. 331C
Beltsville, MD 20705
phone: 301-504-6029
fax: 301-504-6231
E-mail:

Harvey Mohrenweiser
Human Genome Center
Lawrence Livermore National Lab.
P. O. Box 808, L-452
Livermore, CA 94551-9900
phone: 510-423-0534
fax: 510-422-2282
E-mail: harvey@cea.llnl.gov

Donald Moir
Research Director
Collaborative Research Inc.
1365 Main Street
Waltham MA 02154
phone: 617-275-0004
fax: 617-891-5062
E-mail:

Fred A. Morse
Los Alamos National Laboratory
P. O. Box 1663, MS A114
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

Robert K. Moyzis
Director, Human Genome Center
Los Alamos National Laboratory
CHGS - MS M885
Los Alamos, NM 87545
phone: 505-667-3912
fax: 505-665-3024
E-mail: moyzis@flovax.lanl.gov

Mike Mucenski
Biology Division
Oak Ridge National Laboratory
Oak Ridge TN 37831-8077
phone: 615-574-0953
fax: 615-574-1283
E-mail:

Mark Mundt
Information and Computing Sciences Div.
Los Alamos National Laboratory
T-10, K710
Los Alamos NM 87545
phone: 505-667-7510
fax: 505-665-3493
E-mail:

Christine Munk
Mail Stop A114
Los Alamos National Laboratory
Los Alamos NM 87545
phone: 595-667-1600
fax: 505-865-3858
E-mail:

Richard Mural
Biology Division
Oak Ridge National Laboratory
P. O. Box 2009
Oak Ridge, TN 37831-8077
phone: 615-576-2938
fax: 615- 574-1274
E-mail:

Matthew Murray
MS 66-214
Lawrence Berkeley Laboratory
Human Genome Center
1 Cyclotron Road
Berkeley, CA 94720
phone: 510-486-4823
fax:
E-mail:

Cleo Naranjo
Center for Human Genome Studies
Los Alamos National Laboratory
Los Alamos NM 87545
phone: 505-665-4438
fax: 505-665-3024
E-mail:

David L. Nelson
Inst. for Molecular Genetics
Baylor College of Medicine
T809
One Baylor Plaza
Houston, TX 77030
phone: 713-798-4787
fax: 713-798-6370 or 5386
E-mail: nelson@condor.mbir.bcm.tmc.edu

David O. Nelson
Human Genome Center
Lawrence Livermore National Lab.
P. O. Box 808, L-452
Livermore, CA 94551-9900
phone: 510-423-8898
fax: 510-423-3608
E-mail: daven@gauss.llnl.gov

J. Robert Nelson
Institute of Religion
Baylor College of Medicine
One Baylor Plaza
Houston, TX 7725
phone: 713-797-0600
fax: 713-797-9199
E-mail:

Quan Nguyen
Life Science Group
Bio-Rad Laboratories, Inc.
2000 Alfred Nobel Drive
Hercules, CA 94547
phone: 510-741-1000
fax: 510-741-1060
E-mail:

Deborah Nickerson
Department of Molecular Biotechnology
University of Washington
MS GJ-10
4909 25th Avenue N.E.
Seattle, WA 98195
phone: 206-685-7387
fax: 206-685-7367
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

William C. Nierman
American Type Culture Collection
12301 Parklawn Drive
Rockville M D 20852-1776
phone: 301-231-5559
fax: 301-770-1848
E-mail:

Stephen Notarnicola
Harvard Medical School
240 Longwood Avenue
Boston M A 02115
phone: 617-432-1864
fax: 617-432-3362
E-mail:

Frank Olken
Human Genome Center
Lawrence Berkeley Laboratory
MS50B-3220
1 Cylcotron Road
Berkeley CA 94720
phone: 510-486-5891
fax:
E-mail: F_Olken@lbl.gov

Anne Olsen
Human Genome Center
Lawrence Livermore National Lab.
P. O. Box 808, L-452
Livermore, CA 94551-9900
phone: 510-423-4927
fax: :510-423-3608
E-mail: olsen@ecor1.llnl.gov

Maynard V. Olson
Dept. of Molecular Biotechnology, GJ-10
University of Washington
Seattle, W A 98195
phone: 206-685-7346
fax: 206-685-7344
E-mail:

Ross Overbeek
Math and Computer Science Division
Argonne National Laboratory
MCS 221/D236
9700 S. Cass Avenue
Argonne, IL 60439
phone: 708- 252-7856
fax: 708-972-5986
E-mail: overbeek@mcs.anl.gov

David J. Ow
Human Genome Center
Lawrence Livermore National Lab.
P. O. Box 808, L-452
Livermore, CA 94551-9900
phone:
fax:
E-mail:

Elizabeth T. Owens
Human Gen. & Toxicology Grp,MS 6050
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, TN 37831-6050
phone: (615) 574-0601
fax: 615-574-9888
E-mail: TUG@ORNL.Gov

Michael J. Palazzolo
Human Genome Center
Lawrence Berkeley Laboratory
MS 74-157
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-5909
fax: 510-486-6816
E-mail: michaelp@lbl.gov

Gary Parr
Department of Chemistry
University of Wisconsin
1101 University Avenue
Madison W I 53706
phone: 608-263-2594
fax: 608-262-0381
E-mail:

Julia E. Parrish
Institute for Molecular Genetics
Baylor College of Medicine
1 Baylor Plaza
Houston, TX 77030
phone: 713-798-3122
fax: 713-798-5386
E-mail: e-mail: jparrish@bcm.tmc.edu

Wayne Parrott
Cell Biology, Room 453A
Baylor College of Medicine
1 Baylor Plaza
Houston, TX 77030-3498
phone: 713-798-3733
fax: 713-790-1275
E-mail:

Tentative Attendees List For The
DOE Contractor-Grantee Meeting
February 7-10, 1993

Rebecca Parsons
MS K987
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-2655
fax: 505-665-5220
E-mail:

Peter L. Pearson
Genome Data Base
Johns Hopkins University
Welch Medical Library
1830 E. Monument Street
Baltimore MD 21205
phone: 301-955-9705
fax: 301-955-0054
E-mail: PEARSON@WELCH.JHU.EDU

Robert Pecherer
Theoretical Biology and Biophysics Group
Los Alamos National Laboratory
Group T-10, MS K710
Los Alamos, NM 87545
phone: 505-665-1970
fax: 505-665-3493
E-mail: rmp%life@lanl.gov

Joanne E. Pelkey
Life Sciences Center
Battelle Pacific Northwest Laboratory
P. O. Box 999
Mail Stop K7-22
Richland, WA 99352
phone: 509-375-6947
fax: 509-375-3641
E-mail: je_pelkey@pnl.gov

Roger B. Perkins
Los Alamos National Laboratory
P. O. Box 1663, MS A114
Los Alamos, NM 87545
phone: FTS 855-3858
fax: (505) 665-3858
E-mail:

Sergey Petrov
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, TN 37831-6050
phone: 615-576-6669
fax: 615-574-9888
E-mail:

Dan Pinkel
Division of Molecular Cytology
University of California, San Francisco
Department of Laboratory Medicine
San Francisco CA 94143-0808
phone: 415-476-3461
fax: 415-476-8218
E-mail:

Michael C. Pirrung
P.M. Gross Chemical Lab.
Duke University
Box 90346
Durham NC 27708-3046
phone: 919-660-1556
fax: 919-660-1591
E-mail:

Marty Pollard
Engineering Division and Human
Lawrence Berkeley Laboratory
MS70A-3363 F
1 Cyclotron Road
Berkeley, CA 94720
phone: 510-486-5647
fax: 510-486-5857
E-mail:

Mihael Polymeropoulos
Neuro Science Center at St. Elizabeth's
National Institutes of Mental Health
LBG Rm. 120
2700 Martin Luther King Avenue
Washington DC 20032
phone: 202-373-6077
fax: 202-373-6087
E-mail:

Theodore T. Puck
Eleanor Roosevelt Inst. for Cancer
1899 Gaylord
Denver, CO 80206
phone: 303-333-4515
fax: 303-333-8423
E-mail:

Steen Rasmussen
Theoretical Biology and Biophysics Group
Los Alamos National Laboratory
Group T-10, MS K710
Los Alamos, NM 87545
phone: 505-665-1970
fax: 505-665-3493
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

Robert Ratliff
Center for Human Genome Studies
Los Alamos National Laboratory
Los Alamos NM 87545
phone: 505-667-2872
fax: 505-665-3024
E-mail:

Graham W. Redgrave
T-10, Mail Stop K710
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail: grw@life.lanl.gov

David E. Reichle
Environmental, Life & Social Sciences
Oak Ridge National Laboratory
P. O. Box 2008, MS-6253
Oak Ridge, TN 37831-6253
phone: (615) 574-4333
fax: (615) 576-2912
E-mail:

Philip Reilly, JD
President, Chief Executive Officer
Shriver Center
200 Trapelo Road
Waltham MA 02254
phone: 617-642-0222
fax: 617-319-5721
E-mail:

Arthur D. Riggs
Biology Department
Beckman Research Institute of the City of
1450 East Duarte Road
Duarte CA 91010-0269
phone: 818-301-8352
fax: 818-358-7703
E-mail:

Eugene Rinchik
Biology Division
Oak Ridge National Laboratory
Oak Ridge TN 37831-8077
phone: 615-574-0953
fax: 615-574-1283
E-mail:

Jasper Rine
Human Genome Center, MS 401 Barker
University of California, Berekeley
225 Barker Hall
Berkeley, CA 94720
phone: 510-642-7047
fax: 510-642-6420
E-mail:

Robert Robbins
Welch Medical Library
Johns Hopkins University
3400 N. Charles Streets, 316 Garland Hall
Baltimore, MD 21218-269
phone: 410-955-9637
fax: 410-955-0054
E-mail:

Donna Robinson
Center for Human Genome Studies
Los Alamos National Laboratory
Los Alamos NM 87545
phone: 505-665-4438
fax: 505-665-3024
E-mail:

Sylvie Roquier
Human Genome Center
Lawrence Livermore National Lab.
P. O. Box 808, L-452
Livermore, CA 94551-9900
phone:
fax:
E-mail:

Hays S. Rye
Molecular and Cell Biology
University of California, Berkeley
Stanley Hall, Room 229
Berkeley CA 94720
phone: 510-642-4192
fax: 510-642-3599
E-mail:

Jacqueline Salit
The Cold Spring Harbor Laboratory
P.O. Box 100
Cold Spring, NY 11724
phone: 516- 367-8393
fax: 516-367-8461
E-mail: marr@cshl.org

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

Miquel Salmeron
Materials and Chem. Science Division
Lawrence Berkeley Laboratory
MS66
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-6230
fax: 510-486-4495
E-mail:

Gary Salzman
Mail Stop A114
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Takeshi Sano
Boston Universtiy
36 Cummington, Street
Boston, MA 02215
phone: 617-353-8500
fax: 617-353-8501
E-mail:

Liz Saunders
Mail Stop A114
Los Alamos National Laboratory
Los Alamos NM 87545
phone: 595-667-1600
fax: 505-865-3858
E-mail:

Peter Schad
Life Science Group
Bio-Rad Laboratories, Inc.
2000 Alfred Nobel Drive
Hercules, CA 94547
phone: 510-741-1000
fax: 510-741-1060
E-mail:

David Schieltz
Department of Chemistry
Arizona State University
Tempe AZ 852870-1604
phone: 602-965-3461
fax: 602-965-2747
E-mail:

R. Neil Schimke
Department of Medical Genetics
University of Kansas Medical Center
Rainbow at 39th
Kansas, City KS 66103
phone: 913-588-6043
fax: 913-588-3995
E-mail:

David Schlessinger
Molecular Microbiology
Washington University School of
660 S. Euclid
St. Louis MO 63110
phone: 314-362-2744
fax: 314-362-3203
E-mail:

Klaus Schneider
The Rockefeller University
1230 York Avenue
New York, NY 10021
phone:
fax:
E-mail:

David B. Searls
Department of Human Genetics
University of Pennsylvania
P.O. Box 517
Philadelphia PA 19104-6145
phone: 215-574-0953
fax: 215-573-5892
E-mail: dsearls@cis.upenn.edu

Linda Segebrecht
Science Pioneers
425 Volker Blvd.
Kansas City MO 64110
phone:
fax:
E-mail:

Farideh Shadravan
Human Genome Center
Lawrence Berkeley Laboratory
MS74-157
1 Cyclotron Road
Berkeley CA 97420
phone:
fax:
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting

February 7-10, 1993

Manesh Shah
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, TN 37831-6050
phone: 615-576-6669
fax: 615-574-9888
E-mail:

Brooks Shera
Center for Human Genome Studies
Los Alamos National Laboratory
MS-D434
Los Alamos NM 87545
phone: 505-667-3228
fax: 505-665-3644
E-mail:

James M. Sikela
Department of Pharmacology
University of Colorado Health Sciences
Campus Box C-236
4200 East Ninth Avenue
Denver CO 80262
phone: 303-270-8637
fax: 303-270-7097
E-mail: sikela_j%mauia@vaxf.colorado.ed

F. V. Sloop
Oak Ridge National Laboratory
P. O. Box 2009
Oak Ridge, TN 37831-8077
phone: 615-576-2938
fax: 615- 574-1274
E-mail:

Jude W. Shavlik
Department of Computer Sciences
University of Wisconsin
1210 W. Dayton Street
Madison W I 53706
phone: 608-262-7784
fax: 608-262-9777
E-mail: shavlik@cs.wisc.edu

Kathy Shera
Los Alamos National Laboratory
Los Alamos N M 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Paul H. Silverman
President
Beckman Instruments, Inc.
Box 3100
2500 Harbor Blvd.
Fullerton CA 92634
phone: 714-773-7745
fax: 714-773-7617
E-mail:

Cassandra Smith
Boston University
36 Cummington Street
Boston, MA 02215
phone: 617-353-8500
fax: 617-353-8501
E-mail: clsmith@ux5.lbl.gov

R. B. Shelton
Oak Ridge National Laboratory
P. O. Box 2009
Oak Ridge, TN 37831-8077
phone: 615-576-2938
fax: 615- 574-1274
E-mail:

Hiroaki Shizuya
Division of Biology
California Institute of Technology
147-75
Pasadena CA 91125
phone: 818-356-4154
fax: 818-796-7066
E-mail:

Thomas Slezak
Human Genome Center
Lawrence Livermore National Lab.
P.O. Box 808, L-452
Livermore, CA 94550
phone: 510-422-5746
fax: 510-423-3608
E-mail: Tom@yac.llnl.gov

Lloyd Smith
Department of Chemistry
University of Wisconsin
1101 University Avenue
Madison W I 53706
phone: 608-263-2594
fax: 608-262-0381
E-mail: smith@bert.wisc.edu

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

Richard D. Smith
Chemical Sciences Department
Pacific Northwest Laboratory
1101 University Avenue
Richland WA 99352
phone: 509-376-0723 or 5665
fax: 509-376-0418
E-mail:

Michael Smith
Molecular Genetics Laboratory
The Salk Institute for Biological Studies
P.O. 85800
La Jolla CA 92037
phone: 619-453-4100
fax: 619-558-9513
E-mail:

David Smith
Health Effects & Life Sciences, Research
U.S. Department of Energy
ER-72 GTN
Washington DC 20585
phone: 301-903-5468
fax: 301-903-5051
E-mail:

Jay Snoddy
U. S. Department of Energy
ER-70 GTN
Washington, DC 20585
phone: 301-903-3251
fax: 301-903-5051
E-mail:

Marcelo Bento Soares
Dept. of Psychiatry
Columbia University
722 West 168th Street, Box #41
New York NY 10032
phone: 212-960-2313
fax: 212-795-5886
E-mail:

Carol A. Soderlund
Group T-10, Mailstop K710
Los Alamos National Laboratory
Los Alamos NM 87545
phone:
fax:
E-mail: cari@life.lanl.gov

Cari A. Soderlund
Group T-10, Mailstop K710
Los Alamos National Laboratory
Los Alamos NM 87545
phone:
fax:
E-mail: cari@life.lanl.gov

Doug Sorenson
Mail Stop A114
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Terence Speed
Chair, Statistics Dept.
University of California Berkeley
327 Evans
Berkeley CA 94720
phone: 510-642-4272
fax:
E-mail: TERRY@STAT.Berkeley.EDU

Sylvia Spengler
Human Genome Center, MS 1-213
Lawrence Berkeley Laboratory
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-5874
fax: 486-5717
E-mail: sylviaj@violet.berkeley.edu

John Steinkamp
Mail Stop A114
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Marvin Stodolsky
Office of Health & Environmental
U.S. Department of Energy
ER-72 GTN
Washington DC 20585
phone: 301-903-4475
fax: 301-903-5051
E-mail: STODOLSKY@OERV01.ER.DOE.

Tentative Attendees List For The
DOE Contractor-Grantee Meeting
February 7-10, 1993

Michael Strathmann
Human Genome Center
Lawrence Berkeley Laboratory
MS 74-157
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-7330
fax: 510-486-6816
E-mail:

Linda D. Strausbaugh
Dept. of Molecular & Cell Biology
The University of Connecticut
Box U-125
75 N. Eagleville Road
Storrs, CT 06269-3125
phone: 203-486-2693
fax: 203-486-4331
E-mail: molce12@UCONNVM

Zaklina Strezoska
Biological & Medical Res. Div.
Argonne National Laboratory
9700 South Cass Avenue
Argonne, IL 60439-4833
phone: 708-972-3175
fax: 708-972-3387
E-mail:

Gary Strniste
Los Alamos National Laboratory
Los Alamos NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Lisa Stubbs
Biology Division
Oak Ridge National Laboratory
P.O. Box 2009
Oak Ridge TN 37831-8077
phone: 615-574-0865
fax: 615-574-1283
E-mail:

F. William Studier
Biology Department
Brookhaven National Laboratory
Upton, NY 11973
phone: 516-282-3390
fax: 516-282-3407
E-mail:

Damir Sudar
Division of Molecular Cytology
University of California, San Francisco
Department of Laboratory Medicine
San Francisco CA 94143-0808
phone: 415-476-3461
fax: 415-476-8218
E-mail:

Betsy Sutherland
Biology Department
Brookhaven National Laboratory
Upton, NY 11973
phone: 516-282-3380
fax: 516-282-3407
E-mail:

Harold Swerdlow
Department of Human Genetics
University of Utah
6160 Eccles Genetics Building 533
Salt Lake City, UT 84112
phone: 801-581-5163
fax:
E-mail:

Ernest Szeto
Information & Computing Sciences Div.
Lawrence Berkeley Laboratory
MS 46A-1123
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-7565
fax:
E-mail:

Stanley Tabor
Dept. of Biological Chem. & Mol. Phar
Harvard Medical School
240 Longwood Avenue
Boston MA 02115
phone: 617-432-3128
fax: 617-738-0516
E-mail:

Judy Tesmer
Mail Stop A114
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

Ed Theil
Lawrence Berkeley Laboratory
1 Cyclotron Road
MS46A-1120
Berkeley CA 94720
phone: 510-486-7501
fax: 510-486-6940
E-mail: EHTheil@lbl.gov

Gregory S. Thomas
Life Sciences Center
Battelle Pacific Northwest Laboratory
Mail Stop K7-22
Richland, WA 99352
phone: 509-375-6943
fax: 509-375-3641
E-mail: jthomas@gnome.pnl.gov

Sue Thompson
Mail Stop A114
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Linda S. Thompson
MS-M880
Los Alamos National Laboratory
Los Alamos, NM 87544
phone: 505-667-1600
fax:
E-mail:

Thomas G Thundat
Health and Safety Division
Oak Ridge National Laboratory
P.O. Box 2008, MS/6123
Oak Ridge TN 37831-6119
phone: 615-574-6215
fax: 615-574-6210
E-mail:

David Torney
Information and Computing Sciences Div.
Los Alamos National Laboratory
T-10, K710
Los Alamos NM 87545
phone: 505-667-7510
fax: 505-665-3493
E-mail: dct@life.lanl.gov

Barbara Trask
Dept. of Molecular Biotechnology, GJ-10
University of Washington School of
4909 25th Avenue N.E.
Seattle, WA 98195
phone: 206-685-7367
fax: 206-685-7301
E-mail:

Ralph W. Trottier
Dept. of Pharmacology and Toxicology
Morehouse School of Medicine
720 Westview Drive, S.W.
Atlanta GA 30310
phone: 404-752-1711
fax: 404-755-7318
E-mail:

Susan Tsujimoto
Human Genome Center
Lawrence Livermore National Lab.
P. O. Box 808, L-452
Livermore, CA 94551-9900
phone:
fax:
E-mail:

Don Uber
Engineering Division and Human
Lawrence Berkeley Laboratory
MS74-157
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-6378
fax: 510-486-6816
E-mail: uber@petvax.lbl.gov

Edward Uberbacher
Engineering Physics & Mathematics Div.
Oak Ridge National Laboratory
Oak Ridge, TN 37831-6364
phone: 615-574-6134
fax: 615-574-7860
E-mail: ube@stc10.ctd.ornl.gov

Ger van den Engh
Dept. of Molecular Biotechnology, MS
University of Washington School of
4909 25th Avenue N.E.
Seattle, WA 98195
phone: 206-685-7367
fax: 206-685-7301
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

Eugene Veklerov
Information and Computing Sciences
Lawrence Berkeley Laboratory
50B-3216
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-5041
fax: 510-486-4004
E-mail:

Jean-Michel Vos
Department of Biochemistry & Biophysics
University of North Carolina at Chapel Hill
Lineberger Cancer Research Center,
Chapel Hill NC 27599-7295
phone: 919-966-6888
fax: 919-966-3015
E-mail: Vos@Med.UNC.Edu

Mark Wagner
L-156
Lawrence Livermore National Lab.
P. O. Box 808
Livermore, CA 94551
phone: 510-422-2866
fax:
E-mail: mwagner@kooler.llnl.gov

Robert Wagner
MS-M880
Los Alamos National Laboratory
P.O. Box 1663
Los Alamos, NM 87544
phone:
fax:
E-mail:

Wagner P. Wagner
MS-M880
Los Alamos National Laboratory
P.O. Box 1663
Los Alamos, NM 87544
phone:
fax:
E-mail:

Geoffrey M. Wahl
Gene Expression Laboratory
The Salk Institute for Biological Studies
P.O. Box 85800
La Jolla, CA 92037
phone: 619-453-4100 x 587
fax: 619-455-1349
E-mail:

Susan Wallace
The Institute for Genomic Research
932 Clopper Road
Gaithersburg, MD 20878
phone:
fax:
E-mail: e-mail: swallace@tigr.org

Thomas Walter
Boehringer Mannheim Gmbh
Werk Penzberg
Nonnenwald 2
Postfach 1152 D-812 Penzberg/Obb.
phone: 011-8856-60-2147
fax: 011-8856-60-3180
E-mail:

Robert J. Warmack
Oak Ridge National Laboratory
P.O. Box 2008, MS 6123
Oak Ridge TN 37831-6123
phone: 615-574-6215
fax: 615-574-6210
E-mail: rjw@ornlstc.Bitnet

Janet Warrington
Neurogenetics Lab
University of California, San Francisco
401 Parnassus Avenue
San Francisco, CA 94143-0984
phone: 415-476-4212
fax: 415-476-4009
E-mail:

Paul Watkins
Molecular Biology Reseah and
Life Technologies, Inc.
P.O. Box 6009
8/17 Grovemont Circle
Gaithersburg, MD 20877
phone: 301-840-8000
fax: 301-948-8977
E-mail:

Nancy E. Watters
Tropix, Inc.
47 Wiggins Avenue
Bedford, MA 01730
phone:
fax:
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting

February 7-10, 1993

Heinz-Ulrich Weier
Division of Molecular Cytology
University of California, San Francisco
Department of Laboratory Medicine
San Francisco CA 94143-0808
phone: 415-476-3461
fax: 415-476-8218
E-mail:

Robert Weiss
HHMI-Human Genetics Dept.
University of Utah
6160 Eccles Genetics Bldg.
Salt Lake City, UT 84112
phone: 801-581-5190
fax: 801-585-3910
E-mail: WEISS@UTAHMED

Sherman Weissman
Department of Human Genetics
Yale University School of Medicine
336-BCMM
295 Congress Avenue
New Haven, CT 06510
phone: 203-785-2677
fax: 203-785-3033
E-mail:

Burton Wendroff
Los Alamos National Laboratory
B-284
Los Alamos NM 87545
phone: 505-667-6497
fax:
E-mail: bbw@LANL.gov

Thomas Whaley
MS-M880
Los Alamos National Laboratory
P.O. 1663
Los Alamos NM 87545
phone: 505-667-2765
fax:
E-mail:

Bud Whaley
Mail Stop A114
Los Alamos National Laboratory
Los Alamos, NM 87545
phone: 505-667-1600
fax: 505-665-3858
E-mail:

Mark Wilder
Center for Human Genome Studies, Life
Los Alamos National Laboratory
LS-4, M888
Los Alamos NM 87545
phone: 505-667-2750
fax:
E-mail: wilder@flovax.lanl.gov

Peter Williams
Dept. of Chemistry
Arizona State University
Tempe AZ 85287-1604
phone: 602-965-4107
fax: 602-965-2747
E-mail:

Julie S. Wilson
MS-M880
Los Alamos National Laboratory
P.O. Box 1663
Los Alamos, NM 87544
phone:
fax:
E-mail:

Jan A. Witkowski
Banbury Center
Cold Spring Harbor Laboratory
P.O. Box 534
Cold Spring NY 11724
phone: 516-549-0507
fax: 516-549-0672
E-mail:

Frank Witney
Life Science Group
Bio-Rad Laboratories, Inc.
2000 Alfred Nobel Drive
Hercules, CA 94547
phone: 510-741-1000
fax: 510-741-1060
E-mail:

A. K. Wolfe
Oak Ridge National Laboratory
P. O. Box 2009
Oak Ridge, TN 37831-8077
phone: 615-576-2938
fax: 615-574-1274
E-mail:

Tentative Attendees List For The DOE Contractor-Grantee Meeting
February 7-10, 1993

John Wooley
Office of Health and Environmental
U.S. Department of Energy
ER71, GTN
Washington DC 20585
phone: 301-903-3153
fax: 301-903-5051
E-mail:

Richard P. Woychik
Mammalian Genetics, Biology Division
Oak Ridge National Laboratory
P.O. Box 2009
Oak Ridge TN 37831-8077
phone: 615-574-0865
fax: 615-574-1283
E-mail:

Judy M. Wyrick
Health and Safety Research Division
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, TN 37831-6050
phone: 615-576-6669
fax: 615-574-9888
E-mail:

Mimi Yeh
Lawrence Livermore National Laboratory
P.O. Box 5507, L-452
Livermore, CA 94550
phone:
fax:
E-mail:

Michael S. Yesley
MS A187
Los Alamos National Lab.
P.O. Box 1663
Los Alamos NM 87545
phone: 505-665-2523
fax: 505-665-4424
E-mail: yesley_michael_s@ofvax.lanl.gov

Edward S. Yeung
Department of Chemistry
Iowa State University
Ames Laboratory
Ames IA 50011
phone: 515-294-8062
fax: 515-294-0266
E-mail:

Kaoru Yoshida
Lawrence Berkeley Laboratory
529 Stanley Hall
1 Cyclotron Road
Berkeley CA 94720
phone: 510-642-5841
fax: 510-642-1188
E-mail:

Jing-Wei Yu
Eleanor Roosevelt Institute
1899 Gaylord Street
Denver, CO 80262
phone: 303-333-4515
fax: 303-333-8423
E-mail:

Yiwen Zhu
Human Genome Center, MS 74-362
Lawrence Berkeley Laboratory
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-7278
fax:
E-mail:

Manfred Zorn
Information and Computing Sciences
Lawrence Berkeley Laboratory
50B-3216
1 Cyclotron Road
Berkeley CA 94720
phone: 510-486-5041
fax: 510-486-4004
E-mail: