Annual Genome Conference
Montreal, Quebec, October 16-18, 1994

Final Report
for Period September 1, 1994 - August 31, 1995

Katheleen Gardiner

Eleanor Roosevelt Institute
1899 Gaylord Street
Denver, CO 80206

November 1995

Prepared for
THE U.S. DEPARTMENT OF ENERGY
AGREEMENT NO. DE-FG03-93ER61851

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

30 final

at

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product or process disclosed or represents that its use would not infringe privately-owned rights.

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

GRANT #: DE-FG03-93ER61851
TITLE: Annual Genome Conference
P.I.: Katheleen Gardiner, Ph.D.
Eleanor Roosevelt Institute for Cancer Research, Inc.
1899 Gaylord Street
Denver, CO 80206
GRANT DATES: 09/01/94-08/31/95

Abstract

Tremendous progress has been made in the construction of physical and genetic maps of the human chromosomes. The next step in the solving of disease related problems, and in understanding the human genome as a whole, is the systematic isolation of transcribed sequences. Many investigators have already embarked upon comprehensive gene searches, and many more are considering the best strategies for undertaking such searches. Because these are likely to be costly and time consuming endeavors, it is important to determine the most efficient approaches. As a result, it is critical that investigators involved in the construction of transcriptional maps have the opportunity to discuss their experiences and their successes with both old and new technologies.

Final Report

The Fourth Annual Workshop on the Identification of Transcribed Sequences was held in Montreal, Quebec, October 16-18, 1994 and had 74 attendees. Topics discussed included: i) progress in the application of the hybridization based approaches (e.g., cDNA hybrid selection and direct cDNA screening) and exon trapping, ii) progress in transcriptional map construction of selected genomic regions, iii) computer assisted analysis of genomic and protein coding sequences and additional new approaches, iv) sequencing and mapping of random cDNAs.

Enclosed is the workshop notebook, containing the agenda, abstracts presented and list of attendees, as well as a copy of the proceedings of the workshop.

Submitted by:

Katheleen Gardiner, Ph.D.
Institute Fellow
Principal Investigator

12/4/95
Date

Approved by:

David Patterson, Ph.D.
President

12/5/95
Date

The 4th International Workshop on the Identification of Transcribed Sequences

*Montreal, Quebec
October 16-18, 1994*

Sponsored by

*Canadian Genome Analysis and Technology Program
and
The United States Department of Energy*

**4th International Workshop on the
Identification of Transcribed Sequences**

Table of Contents

	Page
Meeting Agenda	1-6
 Abstracts (<i>arranged by presentation order</i>)	
Techniques	7-30
Exon Trapping	31-34
Regional Maps	35-53
cDNA Sequencing and Mapping	54-58
Informatics	59-62
 Alphabetical List of Abstract Authors	 63-64
 1994 Roster of Attendees	 65-72

4TH INTERNATIONAL WORKSHOP ON THE IDENTIFICATION OF TRANSCRIBED SEQUENCES

October 16-18, 1994 Montreal, Quebec

AGENDA

Saturday, October 15, 1994

5:30 pm - 7:00 pm

Welcoming Reception

.....

Sunday, October 16, 1994

7:30 am - 7:55 am

Continental Breakfast - Viger A, B,C

7:55 am - 8:00 am

Welcome/Opening Remarks

8:00 am - 6:00 pm

TECHNIQUES

Chairs: Sherman Weissman, Anthony Brookes and
Bento Soares

8:00 am - 12:30 pm

Morning Session

Monica Justice, Ph.D.

Using retroviral tagging to study mouse B-cell lymphomas.

Roger Brent, Ph.D.

Characterizing genes and genomes by interaction mating.

Srinivasa Rao, Ph.D.

Identification of potential target genes for the transcription
factor c-jun.

Vasi Sundaresan, Ph.D.

Isolation of transcribed sequences by YAC transfection and
differential display.

Marcia Budarf, Ph.D.

Identification and isolation of genes involved in
DGS/VCFS: A sequence-based approach.

B. Rajendra Krishnan, Ph.D.

Identifying genes within the HLA-C linked DNA region by
"Feature Mapping": sequencing from nested deletions.

Christian Neri, Ph.D.	A pilot study for obtaining a catalog of human brain cDNAs containing CAG or CGG repeats potentially involved in neurodegenerative pathologies.
Yuri Lebedev, Ph.D.	Construction of chromosome-specific EST libraries. Mapping of Chr19 hn-cDNA EST's.
David Kurnit, M.D., Ph.D.	Transcriptional analysis of genes on human chromosome 21.
Daniela Toniolo, Ph.D.	Identification of genes expressed during development of the central nervous system.
Susan Ackerman, Ph.D.	Identification of novel human neuron-specific transcripts.
Catherine Nguyen, Ph.D.	A cDNA project aimed at identifying new genes involved in processes of immunological interest using expression analysis and genome mapping in mouse and man.
12:30 pm - 2:00 pm	Lunch - L'Escapade
2:00 pm - 6:00 pm	<u>Afternoon Session</u>
Genevieve Pietu, Ph.D.	Analysis of gene expression in human skeletal muscle by screening an arrayed cDNA library on high density filters and <i>in situ</i> hybridization.
Shinji Hadano, Ph.D.	High-resolution mapping and sequence analysis of 597 cDNA clones transcribed from the 1 Mb region in human chromosome 4p16.3 containing Huntington disease gene.
Christian Gruber, Ph.D.	The identification of uniquely expressed or highly induced genes from fibroblast growth factor stimulated BC3H1 cells by subtractive hybridization.
Paul Drew, Ph.D.	Rapid cloning and characterization of transcription factors expressed in human brain by signature sequencing analysis.
Sherman Weissman, M.D.	Supplementary approaches for screening cDNA libraries.

Marcelo Bento Soares, Ph.D.	Identification of transcribed sequences in the SMA critical region by direct selection of cDNA clones from normalized libraries.
Patanjali Sankhavaram, Ph.D.	Sensitive and specific transcript mapping on YACs by cDNA selection using total yeast DNA: A study of Ataxia Telangiectasia, Kit receptor and MHC regions.
Anthony Bench, Ph.D.	Isolation of new genes by a modified direct selection technique.
Cynthia Jackson, Ph.D.	Identification of transcribed sequences in human chromosome 9p by an efficient cDNA selection strategy using somatic cell hybrids.
Anand Swaroop, Ph.D.	A sandwich-hybridization method for cDNA selection from large genomic regions: Search for candidate genes for XLRP locus.
Anthony Brookes, Ph.D.	Gene identification by 'End-ligation coincident sequence cloning' (EL-CSC).
Teresa Liang	Comparative evaluation of techniques for capturing transcribed sequences.

.....

Monday, October 17, 1994

7:30 am - 8:00 am	Continental Breakfast - Viger A, B, C
8:00 am - 10:15 am	EXON TRAPPING
	Chair: Johan den Dunnen
Johan den Dunnen, Ph.D.	Exon trapping
Greg Landes, Ph.D.	Utilization of exon trapping to generate a transcriptional map of the PKD1 locus.
Yun-Fai Chris Lau, Ph.D.	Towards an exon map of the human Y chromosome.
Paul Nisson, Ph.D.	Trapping internal and 3'-terminal exons from YACs.

10:30 am - 6:00 pm

REGIONAL MAPS

Chairs: Johanna Rommens and Daniela Toniolo

10:30 am - 12:30 pm

Morning Session

Wufang Fan, Ph.D.

Discovering, mapping and characterization of 39 new genes in the human MHC class I region by cDNA selection.

Phillipe Sanseau, Ph.D.

Identification of new coding sequences in large genomic regions using a combination of different methods.

Carissa Smith

Identification of transcripts in YAC contig located in 11q13 between PGA and PYGM.

Anat Bar-Shira, Ph.D.

A transcriptional map of the Ataxia-Telangiectasia locus on chromosome 11q22-23.

Stanislav Volik, Ph.D.

Mapping of a sequence homologous to LTR of endogenous human retrovirus on human chromosome 19.

Johanna Rommens, Ph.D.

Progress towards a transcriptional map of the q21-22 region of chromosome 7.

12:30 pm - 1:45 pm

Lunch - L'Escapade

1:45 pm - 6:00 pm

Afternoon Session

Greg Lennon, Ph.D.

Hybrid selection of human chromosome 19 cDNAs.

Hongxia Xu, M.D., Ph.D.

Large regions of DNA from chromosome 21q21 may be transcriptionally silent.

Katheleen Gardiner, Ph.D.

Transcriptional mapping of chromosome 21 by cDNA hybridization selection from YAC clones.

Michel Fontes, Ph.D.

cDNA selection in the human Xq12-Xq21 region.

Daniela Toniolo, Ph.D.

Identification of genes and construction of a transcriptional map in Xq28.

Ida Biunno, Ph.D.	Functional analysis of expressed sequences isolated from human Xq27-28.
Bernard Korn, Ph.D.	Transcriptional mapping by direct cDNA selection: From Xq28 to the entire X chromosome.
Weilong Gong, M.D.	Identification and isolation of genes involved in DGS/VCFS: A cDNA selection-based approach.
Howard Sirotkin, Ph.D.	Isolation of cDNAs from 22q11; the region commonly deleted in DiGeorge and Velo-cardio-facial syndromes.
Catherine Lambert, Ph.D.	Isolation of transcribed sequences in the reeler region of mouse chromosome 5.
Sally Camper, Ph.D.	Construction of a three megabase contig on mouse chromosome 11 for positional cloning of the Ames dwarf locus.
Miklos Peterfy, Ph.D.	Identification of transcribed sequences from distal mouse chromosome 11.
Miriam Meisler, Ph.D.	Transcript identification in the insertion site region of the transgene induced neurological mutant motor endplate disease (MED).

.....

Tuesday, October 18, 1994

8:00 am - 8:30 am	Continental Breakfast - Caf' Conc'
8:30 am - 10:00 am	cDNA SEQUENCING AND MAPPING
	Chair: Greg Lennon
Kevin Brady, Ph.D.	Mapping mouse expressed sequences using single-strand conformation polymorphism analysis.
Michael Jones, Ph.D.	Construction of a human adult testis expression map.
Donna R. Maglott, Ph.D.	Characterizing human brain ESTs of unknown function: Expression patterns, mapping, and additional sequencing.

Donald Moir, Ph.D. Human cDNA mapping by hybridization to high-density megaYAC dot blots.

Greg Lennon, Ph.D. The IMAGE consortium: A collaborative effort to characterize arrayed cDNA libraries.

10:15 am - 12:00 pm

INFORMATICS

Chairman: Richard Mural

Martin Ringwald, Ph.D.

Gene expression information resource: Phase 1.

Jean-Michel Claverie, Ph.D.

Entering the vast territory of uncharted ESTs.

Mark Boguski, M.D., Ph.D.

New feature of dbEST - NCBI's public repository of "Single Pass" sequences and mapping data.

Richard Mural, Ph.D.

Integrated gene recognition, modeling and database searching in the Grail and Genquest environments.

12:00 pm - 1:15 pm

Lunch - Meeting Room

1:15 pm - 2:45 pm

DISCUSSION GROUPS

2:45 pm - 4:00 pm

SUMMARY

WORKSHOP ORGANIZERS

Katheleen Gardiner, Ph.D.
Eleanor Roosevelt Institute
1899 Gaylord Street
Denver, CO 80206

Richard J. Mural, Ph.D.
Oak Ridge National Lab
P.O. Box 2009
Oak Ridge, TN 37831

Johanna Rommens, Ph.D.
The Hospital for Sick Children
555 University Avenue
Toronto, Ontario M5G 1X8
Canada

Techniques

Using retroviral tagging to study mouse B-cell lymphomas.

M. J. Justice, G. M. Hansen, and R. L. Gross. Division of Biology, 312 Ackert Hall, Kansas State University, Manhattan, KS 66506 USA

Retroviruses and other transposable elements are often found as causative agents in mouse cancers. One way these mobile elements can cause cancer is to integrate near, and alter the expression of normal cellular proto-oncogenes. The retroviruses act as insertional mutagens and serve as very useful molecular tags to clone new genes associated with tumorigenesis. Some AKXD recombinant inbred strains of mice have a high incidence of B-cell lymphomas caused by the expression and subsequent integration of murine leukemia retroviruses. The proviruses in the AKXD B-cell lymphomas are useful molecular tags to clone common integration sites that represent putative proto-oncogenes.

One novel locus on mouse Chr 18 that was identified by molecular tagging, called ecotropic viral integration site-3 (*Evi3*), is specifically involved in B-cell lymphoma development. Proviral integrations at *Evi3* that have occurred in multiple independent B-cell lymphomas are tightly clustered within a CpG island. At least two genes flank the CpG island at *Evi3*, and both have properties consistent with their role as proto-oncogenes. One, designated *Evi3B*, is expressed in mouse B-cells, and the other, *Evi3A*, is expressed during mouse embryogenesis. The genomic organization of the two genes is complex; consequently, genomic sequence analysis using the GRAIL computer program, and transcriptional analysis using RT-PCR have been crucial in the molecular resolution of the locus. In humans, *EVI3* maps to a region of Chr 18 that is frequently altered by translocation breakpoints in synovial sarcomas. Thus, either or both of the genes identified at *EVI3* may be involved in human cancer. Because of the cell- and time-specific expression of the two genes, the identification of human transcribed sequences is difficult. Therefore, we are cloning and sequencing genomic regions that are well-conserved between mouse and human in an attempt to understand the human locus and its role in human cancer.

We are using retroviral tagging to clone additional novel common sites of retroviral integration from the AKXD B-cell lymphomas. An ongoing challenge is the identification of expressed sequences from the genomic DNA isolated from the retroviral integration site. The study of the relationships and interactions of these novel genes with other proto-oncogenes will provide additional information about the multistep process of B-cell lymphomagenesis.

This work is supported by a grant from the Alfred P. Sloan Foundation and by NASA grant NAGW 2328 to MJJ.

Characterizing genes and genomes by interaction mating

Russ Finley and Roger Brent

Department of Molecular Biology, Massachusetts General Hospital, 50 Blossom Street, Boston, Massachusetts 02114 and Department of Genetics, Harvard Medical School

We have been using a two hybrid system called the interaction trap to study protein interactions among cell cycle regulatory proteins and to isolate new ones (see for example Gyuris et al., Cell 75, 791 (1993)). We have recently modified the interaction trap to create a rapid and simple assay for interactions between a large number of proteins.

In an interaction mating, we mate a collection of yeast strains, each of which contains a different "bait" protein fused to the DNA-binding protein LexA, to a test strain of the opposite mating type that contains a protein fused to a transcription activation domain. We collect diploids from the mating. In those in which the bait and activation tagged protein interact, LexA-operator containing reporter genes are activated. The results, which represent the collection of assayed binary protein interactions, are displayed in two dimensional arrays called interaction matrices.

We used these techniques to isolate and characterize a set of cell cycle regulatory proteins that may govern the development of *Drosophila melanogaster*. The new cell cycle regulators include a D-type cyclin, E-type cyclin, and a novel cyclin. Examination of the pattern of individual binary protein-protein interactions indicates that some of these proteins function in ternary complexes. We are using similar analysis of interaction matrices to detect other multiprotein complexes, and to establish connectivity between individual complex members. Similar analysis may help identify domains involved in protein-protein interactions, and aid the dissection of genetic regulatory networks. It may also help assign function to new proteins identified by genome characterization efforts, and to position them within ordered genetic pathways.

Identification of Potential Target Genes for the Transcription Factor c-jun.

Srinivasa K. Rao¹, S. Acharya², Jon R. Cohen¹ and William A. Goddard III³

¹Long Island Jewish Medical Center, Department of Surgery, New Hyde Park, NY 11042, ²Department of Medicine, Bronx, NY 10461, Albert Einstein College of Medicine, ³Beckman Institute, California Institute of Technology, Pasadena, CA 921125.

Transcription factors are essential components in the regulation of gene expression. The protooncogene, c-jun has been shown to bind to its target DNA as a homodimer as well as heterodimer (jun-fos). c-jun plays an important role in the normal cell regulation through the PKC pathway and transformation of cells. Although c-jun binding domains have been identified for some genes, they represent a small fraction of c-jun-regulated genes present in the human genome. In order to determine possible roles of the c-jun protooncogene, it is essential to identify all the genes regulated by c-jun.

The classic approach to identify c-jun-binding sites has been to isolate and sequence the target gene and to analyze its 5'-flanking region. We will describe a novel strategy for identifying all genes regulated by c-jun in human genome. In this approach the c-jun DNA binding domain peptide sequence will be cross-linked to a solid support and packed in a column. Human genomic DNA will be fragmented into approximately 1.5kb units by sonication and loaded onto column. The buffer conditions will be similar to the ones established in our laboratory (WAG) (50µg/ml BSA, 10% (v/v) Glycerol, 20 mM Tris HCl pH 7.5, 4mM KCl, 2mM MgCl₂). Under these conditions, the genomic DNA fragments containing the c-jun-binding sequences will be retained on the column which will be eluted using higher salt concentrations. The eluted DNA fragments will be cloned into a suitable vector and sequenced using the automated sequencing techniques.

We expect to sequence the 5'-flanking region and a part of coding region of several c-jun target genes. The sequences will be compared with the DNA sequence data bases and with cDNA sequence data banks in public and private domains. This approach is expected to expedite the process of identifying all genes regulated by the protooncogene, c-jun.

This approach can be used later with appropriate modification to identify the target genes for other suitable transcription factors.

Isolation of transcribed sequences by YAC transfection and differential display

V. Sundaresan¹, A.C. Heppell-Parton¹, A. Graham¹, D. Ogilvie², M. Lerman³, P.H. Rabbitts¹

¹Clinical Oncology & Radiotherapeutics Unit, MRC Centre, Hills Road, Cambridge, CB2 2QH, ENGLAND, UK.

²Biotechnology Department, Zeneca Pharmaceuticals, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK.

³Laboratory of Immunobiology, National Cancer Institute, Frederick Cancer Research Facility, Maryland 21701, USA.

The aim of this study is to isolate genes from the region 3p11-13, chosen because it is the location of a homozygous deletion identified in a small cell lung cancer cell line, U2020, and therefore likely to include a tumour suppressor gene (TSG) important in lung cancer development. In previous reports of YAC transfection into mammalian cells, successful YAC transfer and expression of marker genes was identified when the recipient cells' genetic deficiency was complemented. These observations suggest that the transfer of large genomic fragments of DNA (cloned in YACs) into mammalian cells, followed by comparative analysis of expression patterns may provide a means of identifying transcribed sequences (ie. cDNA subtraction and or differential display (DD)). U2020 deletion related transcripts may be differentially expressed in the transfected cells when compared with the parent cell line, for example by DD. These can then be confirmed to be deletion related, by use as probes on Northern and Southern blots.

Fifty six YACs, representing 10 megabases of DNA, which span part of this deletion have already been isolated and partly characterised. YACs known to be good candidates for harbouring genes (on the basis of CpG islands), have been retrofitted by homologous recombination with yeast insertional plasmids, bearing the "neo" gene.

A variety of techniques, including spheroplast fusion, lipofection and electroporation have been used to introduce retrofitted YACs into mouse LMTK- cells and the U2020 cell line with variable success. Spheroplast fusion was successful in transferring YAC and yeast DNA into LMTK- cells, as confirmed by fluorescent in situ hybridisation, Southern blot and PCR analysis. YAC transfection into the human U2020 cells was not successful.

The transfection of 56 separate YACs (of average size 350 kb) into mouse fibroblast cell line, and the subsequent analysis of these cell lines by DD is a formidable task. As a strategy to reduce the number of transfections and the subsequent analysis, the region of interest with the putative TSG is being further narrowed by deletion mapping of tumours showing allele loss in the 3p11-13 region. In addition, larger YAC clones spanning CpG islands, and therefore good candidates for harbouring genes are presently being isolated for transfection from the CEPH Mega YAC library.

Identification And Isolation Of Genes Involved In DGS/VCFS: A Sequence-Based Approach

Marcia L. Budarf^{1,2}, Bruce Roe³, Zhili Wang³, Weilong Gong^{1,2}, Jennifer Seigert^{1,2}, Joelle Collins¹, Elizabeth Goldmuntz¹, and Beverly S. Emanuel^{1,2}.

¹The Children's Hospital of Philadelphia, and ²Human Genome Center for Chromosome 22, Philadelphia, Pennsylvania, ³University of Oklahoma, Norman, Oklahoma.

Velo-cardio-facial syndrome (VCFS) is a common, autosomal dominant disorder. Patients with VCFS have a spectrum of defects including cleft palate, characteristic craniofacial dysmorphism, conotruncal cardiac malformations and learning disabilities. The majority of VCFS patients (>85%) studied have microdeletions of 22q11.2. The minimal region of overlap between numerous VCFS patient deletions is large, and overlaps with the deletions found in DiGeorge Syndrome patients (i.e., the DGCR). These findings suggest that the loss of function of more than a single gene is required for the pathogenesis of this disorder.

We have identified several genes within the DGCR. To date we have sequence of five cosmids, providing 204,010 bp of contiguous sequence around N25 and 44,580 bp of sequence from the N41 cosmid, a more distal locus. Genomic sequence from the cosmid in the distal region of the DGCR was searched against the existing databases (GENBANK). Two genes were identified by homology to known rodent genes. An additional match was found with a previously unmapped human expressed sequence tag (EST). The EST matched the 3' end of a gene we had already independently isolated by screening a cDNA library. Comparison of this 4.5kb cDNA sequence to the GRAIL predicted exons demonstrated that GRAIL correctly identified 10 out of the 13 exons. Further, only two additional open reading frames (ORFs), not found in the 4.5 kb transcript, were predicted by GRAIL in the 30 kb region containing this gene.

Results from our analysis of sequence from cosmids in the proximal region of the DGCR is similar. Database searches identified 6 potential genes by homology to genes from other mammalian species and another previously unmapped EST. Some additional ORFs were predicted by GRAIL which had no homology to known genes. Regarding the 6 genes identified by homology: We had previously isolated one of the 6 "genes" by probe hybridization to a fetal brain cDNA library. GRAIL correctly identified 11 of the 12 ORFs contained within the genomic region corresponding to this gene. We verified two additional genes by RT PCR using primers from predicted ORFs. We were unable to detect transcripts for a fourth predicted gene and the last two potential genes were identified very recently. Regarding the GRAIL predicted ORFs which did not detect homologies in the database: PCR primers to multiple exons were designed for RT-PCR screening for two GRAIL predicted genes. One gene was successfully verified using this approach. We were unable to detect a transcript for the second GRAIL predicted gene. In summary using genomic sequence we have been able to identify 9 novel human genes in 240 kb of sequence for an average of ~1 gene/30 kb.

**Identifying genes within the HLA-C linked DNA region by
"Feature Mapping": sequencing from nested deletions.**

B.R. Krishnan¹ and D.D. Chaplin^{1,2}

¹Washington University School of Medicine and ²Howard Hughes Medical Institute, St. Louis, MO 63110.

To enable systematic analysis of less well characterized portions of HLA, we are defining significant "features" of these DNA regions, particularly, locations of putative genes by prediction of exons (GRAIL analysis) and by comparison of sequence data for homology to the database (Blastx analysis). We cloned HLA-C linked DNA segments ranging in size from 5 to 35 kb into a transposon $\gamma\delta$ -based cosmid vector designed for generating nested deletions *in vivo* (Wang et al., Proc. Natl. Acad. Sci., USA **90**: 7874-7878, 1993). Numerous nested deletions from each of these clones were obtained, and HLA-specific sequence adjacent to each deletion endpoint was determined by fluorescent-automated technology. These high density islands of DNA sequences constitute the foundation for "feature" maps of these subcloned DNAs. Our data (i) verified the locations of exons of the previously described keratinocyte-specific S gene, (ii) localized the cell growth regulated SC1 gene, previously cytologically mapped to the 6p21-6p22 region, (iii) located additional sequences (200-300 nt) with high homology to sequences in the database, (iv) located several otherwise undefined putative exons within the HLA DNA and (v) located *Alu* elements. These observations suggest that the centromeric portion of the Class I region of the HLA complex may be as gene dense as its telomeric portion.

A pilot study for the obtention of a catalog of human brain cDNAs containing CAG or CGG repeats potentially involved in neurodegenerative pathologies

C. Néri¹, S. Meier-Ewert², P. Millasseau¹, H. Lehrach², J. Weissenbach³, D. Cohen¹, & H. Cann¹

¹Fondation Jean-Dausset-CEPH, Paris, France; ²ICRF Genome Analysis, London, UK; ³Genethon, Evry, France.

The aim of this study is to establish an interface between genome analysis and molecular neurobiology, and to provide the scientific community with a catalog of human brain cDNAs containing polymorphic (CAG)_n or (CGG)_n sequences potentially involved in inherited neurodegenerative disorders. The expansion of such repeats have been associated with several different inherited neurodegenerative pathologies, and may constitute a general mechanism critical to several of the other neurodegenerative diseases observed in humans. As a pilot study, we have currently screened 60,000 gridded clones from an oligodT-primed, human foetal brain cDNA library (S. Meier-Ewert, ICRF Genome Analysis). To select the longer repeats most likely to be polymorphic, we have compared 2 screening conditions using (CAG)_n or (CGG)_n oligonucleotide probes where n=6 or 12. 355 CAG-positive (0.6%), and 462 CGG-positive (0.8%) cDNAs were selected using 6-repeat probes, including 180 CAG-positive cDNAs (0.3%) and 260 CGG-positive (0.4%) also detected by 12-repeat probes. These cDNAs are being uniquely identified using automated sequencing. Fingerprinting analysis is also being used for clone to clone comparisons (ICRF Genome Analysis). Sequences for the 5' end, 3' end, and trinucleotide repeat in each cDNA are being determined. The cDNAs confirmed to contain long repeats will be PCR-mapped at the subchromosomal level. The polymorphism of repeats in CEPH reference pedigrees, and inter-species conservation will also be determined. Our data will be available to the scientific community as they are generated (CEPH server & RLDB database). Data on the efficiency of our approach in selecting long (CAG)_n or (CGG)_n in cDNAs and on the progression of our study will be presented.

Construction of chromosome-specific FST libraries. Mapping of Chr19 hn-cDNA ESTs.

Yu. Lebedev¹, L. Nikolaev¹, S. Volik¹, T. Vinogradova¹, G. Monastyrskaya¹, G. Lennon², E. Sverdlov¹.

¹Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia; ²Human Genome Center, LLNL, CA, USA

The goal of this investigation is to isolate, sequence and map transcribed sequences (hn-cDNA ESTs) and nuclear matrix associated sequences (MAR STSs) for development of functional sequence tags (FST) maps of human chromosomes 19, 20 and 21. To extract and specifically clone human hn-cDNA from hybrid cell lines containing chromosomes 19, 20 and 21, arrayed hn-cDNA libraries containing about 2000 independent clones each were constructed using improved protocol of Alu repeat-primed cDNA synthesis and nested PCR strategy. GL-12-42 and UV5HL9- 5B cell lines were used for Chr19 hn-cDNA libraries. About 200 clones were sequenced and characterized. To prove human origin of the clones we used PCR amplification of DNA from parental and hybrid cell lines with primers corresponding to unique part of hn-cDNA inserts. More than 30 clones were hybridized to high-density colony filters made from the flow-sorted Chr19 cosmid library. The presence of the hn-cDNA sequences in contigs containing hybridized cosmids was verified by PCR-amplification with primers used to check the human origin of the clones. 27 hn-cDNA clones were considered to be mapped to human chr 19. PCR analysis of 96 clones from the libraries proved the human origin of more than 95% of clones. Proportion of chromosome-specific clones in the libraries depended on quality of monosomic hybrid cell lines. All human-specific clones from hn-cDNA libraries made from UV5HL9-5B, A9(neo 20)-3 and WAV17 cell lines were found to be specific for Chr19, Chr20 and Chr21, respectively. As another source of FSTs, a library of short human chr19 sequences able to specifically bind *in vitro* to nuclear matrices from human cells was constructed. The library contains about 400 independent clones, about 30 of them were sequenced and partly characterized. Current work is focused on sequencing and mapping new FSTs using the constructed libraries as a source

Transcriptional analysis of genes on human chromosome 21.

Pei-Wen Chiang^{1,2}, Jan-Fang Cheng⁴, Margaret L. Van Keuren^{1,2}, Monika Buraczynska², Yonghui Jiang², Yan-Zhen Pan², Wei Lu³, Kathryn Whitley², Eric Crombez² and David M. Kurnit^{1,2,3}

¹Howard Hughes Medical Institute, ²Departments of Pediatrics and ³Human Genetics, University of Michigan, Ann Arbor MI, ⁴Lawrence Berkeley Laboratory, Berkeley CA.

We utilized the polymerase chain reaction (PCR) to construct, screen and isolate longer transcripts from cDNA libraries.

Construction of cDNA libraries. RNA was isolated from human embryos and complementary double-stranded DNA synthesized. Since the resultant material was too limited for direct cDNA library construction, linkers were attached, PCR amplification was performed and cDNA libraries were then constructed. Libraries were constructed from human embryonic tissues (53 days - 78 days of gestation), including adrenal, brain, eye, gut, heart, kidney, liver, lung, skeletal muscle and spleen. Although this yielded complex cDNA libraries, the insert size was small. To circumvent this problem, we have synthesized longer cDNA molecules using a cocktail of an exonuclease⁺ thermal-resistant polymerase and an exonuclease⁻ thermal-resistant polymerase. We are constructing a new set of human embryonic cDNA libraries using this improved methodology.

Determination of transcription in a given library. In this technique, 2 oligonucleotides taken from the gene of interest are used to PCR amplify 10 μ l of phages comprising 10^7 - 10^8 pfu of a cDNA library. The presence of a PCR-generated fragment at the expected molecular weight indicates that the sequence of interest is transcribed in a given cDNA library; the presence of additional bands indicates alternative splicing and/or transcription of related genes.

This technique was first used to demonstrate that huntingtin is expressed ubiquitously in 41 cDNA libraries encompassing human life. Subsequently, 99 pairs of oligonucleotide primers on chromosome 21q22.2→21q22.3 were applied to analyze 41 different cDNA libraries. Many of these primers were obtained following techniques to isolate expressed fragments, *viz.* hybrid selection and exon amplification. Seventy seven of the 99 chromosome 21 primers were positive with at least one library screening, indicating that the majority of primers derived from transcribed sequences. We examined embryonic (*vide supra*), fetal and adult human cDNA libraries. Even with our most complex human fetal brain library, we detected only *ca.* 60% of transcribed sequences, underscoring the need to screen multiple human cDNA libraries to determine if transcription occurred. Since only 4/99 gene sequences were present in only one cDNA library, the vast majority of transcribed sequences are present in more than one tissue. This result is in accord with *ca.* 50 genes on other chromosomes that we have examined with these techniques. Because this study involves random genes rather than genes selected because of their import, this demonstrates that only few anonymous genes are transcribed in a single tissue type. Expression of genes in different cDNA libraries aids the comparison between genotype and phenotype.

Isolation of longer transcripts. We developed a PCR-based technology that retrieves the largest sequences present in a cDNA library. In this method, we gel purified the largest molecules resulting from combinations of gene-specific and vector-specific primers used to achieve amplification. These larger inserts are cloned into plasmids and sequenced.

Identification of genes expressed during development of the Central nervous System

M. Penichet* , G. Consalez°, C. Stayton°, S. Nesti**, E. Boncinelli°, C. Sala° and D. Toniolo*

*Istituto di Genetica Biochimica ed Evoluzionistica, CNR, Pavia and ° DIBIT-HSR, Milano, Italy.

Random sequencing of cDNAs from different tissues will soon identify most transcripts of the mammalian genome. An approach complementary to random sequencing is the identification of transcript specific of tissues or cell types which may lead to the identification of rare transcripts and of genes with specific regulatory function.

Our work is aimed to identify developmental and district specific transcripts in the central nervous system (CNS) of the early mouse embryo to obtain information on genes expressed at specific developmental points and in specific regions of the CNS. The mapping of the mouse genes and the isolation of the corresponding human genes will provide candidates for genetic disorders of the CNS. To this purpose we have prepared cDNA libraries from RNA of total CNS of 10 days mouse embryos. Clones were picked and ordered in microtiter plates, and replicas of high density filters have been made.

To identify differentially expressed transcripts, comparative screenings are being made using different approaches. RNA extracted from adult mouse CNS, mouse embryo CNS and from different CNS districts is used to generate radiolabeled probes made from reverse transcription of the RNA using anchored oligodT primers. cDNAs synthesized with anchored oligodT will correspond to different sub population of mRNAs and will detect mainly abundantly expressed transcripts. To identify less abundant transcripts, radiolabeled probes will be made using the anchored oligodT primers in conjunction with a random decamer oligodeoxynucleotide for subsequent PCR amplification (2,3). Analysis by polyacrilamide gel electrophoresis of the RT-PCR products has demonstrated a differential pattern of bands using different RNAs as template. Hybridizations are in progress and the results will be presented

Identification of novel human neuron-specific transcripts.

S. L. Ackerman¹, P. W. Andrews², and B.B. Knowles¹.

¹The Jackson Laboratory, Bar Harbor, ME 04609; ²University of Sheffield, Sheffield, UK S10 2TN.

The clonal human teratocarcinoma-derived cell line N-TERA2 cl. D1 (NT2) differentiates into several cell types, including stable postmitotic neurons, when treated with retinoic acid (RA). This and the fact that such neurons can be produced in pure populations large enough to do molecular and biochemical studies make NT2 cells an ideal model system to study genes involved in neuronal differentiation and function. We constructed and analyzed cDNA libraries from the NT2 stem cells, a homogeneous population of NT2-derived neurons, and a population of RA-induced, non-neuronal cells. These libraries are large and representative of the cells from which they were derived. Transcripts encoding P- and E-cadherin and α E-catenin are down-regulated whereas the prion protein (PrP) transcript is up-regulated during NT2 neuronal differentiation. We have constructed a subtraction library and have used differential display analysis of these cDNA libraries to isolate partial cDNAs corresponding to genes expressed specifically in NT2 neurons. The libraries were also homology-mined using degenerate oligonucleotides corresponding to the C₂H₂ family of zinc finger genes, and new family members differentially expressed during NT2 neuronal differentiation identified. The mouse homologues of these human neuron-transcribed sequences will be genetically mapped to determine whether any previously described mutants suggest their function.

A cDNA PROJECT AIMED AT IDENTIFYING NEW GENES INVOLVED IN PROCESSES OF IMMUNOLOGICAL INTEREST USING EXPRESSION ANALYSIS AND GENOME MAPPING IN MOUSE AND MAN

Catherine NGUYEN, Dominique ROCHA, Karine BERNARD, Samuel GRANJEAUD, Robert NGUYEN and Bertrand R. JORDAN, Centre d'Immunologie INSERM/CNRS Marseille, France

We attempt to interface genome research and "conventional" biology in the framework of an "immunologically oriented cDNA project" in which we select sets of a few hundred murine cDNA clones for preferential expression in a given cell type (thymic epithelium in our case) and proceed to obtain quickly and efficiently molecular data and map position for large numbers of them.

In collaboration with the Genome Analysis Laboratory (Hans LEHRACH, ICRF, London), we have constructed a mouse thymus cDNA library (8×10^5 clones, average insert size 1-1.5 kb) in the pcDNA1 vector that allows for expression in *cos* cells, and 10^5 clones have been organized in 96- and 384-well microtiter plates. 16-fold high density colony filters prepared with a BIOMEK robot are hybridized with labelled cDNA probes made from total cellular mRNA. Signals (indicating the abundance of the corresponding sequence in the probe) are detected and quantitated over a range of 10^3 using an imaging plate system (FUJI) and specially adapted software running on an UNIX workstation, that provides individual spot detection and quantitation as well as assignment of clone designation. We have verified the linearity and reproducibility of the system; the detection threshold is approximately 0.01% abundance. Analysis of a high-density filter is fast and essentially free of manual editing. Successive hybridizations with probes from various cell types provide a series of signatures for each clone (see Figure); those that fall within a given n-dimensional window (defined according to biological criteria) are rearrayed for further studies.

1536 THYMUS cDNA CLONES HYBRIDIZED WITH COMPLEX PROBES FROM FOUR CELL TYPES AND TISSUES

Tag sequencing of these selected clones is performed at Genexpress (Charles AUFFRAY, Genethon, PARIS) to identify "new" genes; Northern and Southern blotting are done in "spaghetti" format to check expression specificity and prepare for localization. Mapping in the mouse genome using the European backcross facility, and localization in the human genome with somatic hybrids, YACs from the physical map and *in situ* hybridization are expected to indicate correlations with known mouse mutants or human genetic diseases. Other experiments aimed at functional characterization include introduction in *cos* cells of pooled cDNAs selected for preferential expression in thymic epithelial cells and assay with an extensive set of locally derived monoclonal antibodies (P NAQUET *et al*, CIML, Marseille), as well as localization studies on tissue sections.

This project implements in a "conventional" setting instruments, techniques and concepts derived from Genome laboratories and demonstrates an effective interface between the Genome Project and Immunology.

Analysis of gene expression in human skeletal muscle by screening an arrayed cDNA library on high density filters and *in situ* hybridization.

**Geneviève Piétu, Nancy Fréchin, Valérie Guichard and Charles Auffray.
Genexpress, Généthon, BP60, 91002 Evry and CNRS UPR 420, BP 8, 94801
Villejuif, France.**

Identifying genes implicated in genetic diseases, the corresponding molecular defects and their functional impact would benefit from in advance basic knowledge on the site and level of their expression.

Analysis of arrays of cDNA clones on high density filters is a useful approach to obtain such information on a large number of genes. We have performed large scale and systematic analysis of gene expression in human skeletal muscle using hybridization on high density filters. A pseudo-normalized selection of 1339 unique cDNA clones inserts from a human skeletal muscle cDNA library have been spotted as PCR products on high density nylon membranes. Filters were hybridized with ^{33}P -radiolabeled probes transcribed from total cellular mRNA of various tissues. The intensity of hybridization of a particular clone with these complex probes is proportional to the relative abundance of the complementary species in the probe and thus to the level of expression of the corresponding gene. Signal identification has been performed with the Molecular Dynamics imaging plate system. Quantitative analysis of the amount of hybridized probe on each cDNA spot was obtained using a software specifically designed for this application by Cose (Le Bourget, France). Hybridization signatures are edited in a worksheet format and analysed with graphing and population analysis software.

Among the 1339 cDNA clones, 10-15 % were detected with cDNA probes transcribed from human skeletal muscle mRNA preparations and could be classified in strong, medium and low signal intensity. These results were correlated with the analysis of the partial sequences derived from the ends of the cDNA inserts which identified known, related or new genes.

36 cDNA clones were analyzed by *in situ* hybridization. They were used as riboprobes and hybridized on sections of various tissues from rat or mouse embryo. Among 8 clones corresponding to new genes, six were positive on sections of skeletal muscle, heart and intercostal muscle of mouse embryo, and two other labelled brain cells.

This work was supported by AFM, CNRS, INSERM, MRES and the EC BIOMED Program.

High-resolution mapping and sequence analysis of 597 cDNA clones transcribed from the 1 Mb region in human chromosome 4p16.3 containing Huntington disease gene.

S. Hadano¹, Y. Ishida¹, H. Tomiyasu^{1,2}, W. Muramatsu¹, K. Wakasa¹, G.P. Bates⁴ and J.-E. Ikeda^{1,3}

¹Ikeda GenoSPHERE Project/ERATO, ²Dept. of Neurology, ³The Institute of Medical Science, Tokai Univ. School of Med., Bohseidai, Isehara, Kanagawa 259-11, Japan, ⁴Division of Medical and Molecular Genetics, Guy's Hospital, London SE1 9RT, UK.

To complete a transcription map of the 1 Mb region in human chromosome 4p16.3 containing Huntington disease(HD) gene, the isolation of cDNA clones are being done throughout. Our method relies on a direct screening of the cDNA libraries probed with single copy microclones from 3 YAC clones spanning 1 Mbp of HD gene region. YAC-DNAs were isolated by a preparative pulsed-field gel electrophoresis, amplified by both a single unique primer (SUP)-PCR (*Gent Anal Techn Appl* 10:105-108,1993.) and a linker ligation PCR, and 6 microclone-DNA libraries were generated. Then, 8,640 microclones from these libraries were independently amplified by PCR, and arrayed onto the membranes. 800~900 microclones that were not cross-hybridized with total human and yeast genomic DNA, YAC vector DNA, and ribosomal cDNA on a dot hybridization (putatively carrying single copy sequences) were pooled to make 9 probe pools. A total of $\sim 1.8 \times 10^7$ plaques from three human brain cDNA libraries was screened with 9 pool-probes, and then 672 positive cDNA clones were obtained. So far, 597 cDNA clones were defined and arrayed onto a map of the 1 Mbp of the HD gene region by hybridization with HD region specific cosmid contig and YAC clones. Further characterization including a DNA sequencing and Northern blot analysis is currently underway.

The Identification of Uniquely Expressed or Highly Induced Genes From Fibroblast Growth Factor Stimulated BC3H1 Cells by Subtractive Hybridization

Authors: Gruber, C.E., Li, W.-B., Lin, J.-J., Lim, R., D'Alessio, J.M. and Jessee, J.A.

Subtractive hybridization technology is used to remove common sequences between two closely related nucleic acid populations and to enrich for differentially expressed sequences that are present in the population of interest. We report an improved and simplified subtractive hybridization procedure based on the hybridization of a single-stranded phagemid cDNA library (target) to biotinylated RNA (driver). Using this method, we created subtracted cDNA libraries from fibroblast growth factor-induced (target) and uninduced (driver) BC3H1 cDNA libraries. After hybridization at a high C_0t value ($C_0t = 1000$), common cDNAs were subtracted up to 231-fold, whereas cDNAs representing several highly induced genes were enriched from 2-15-fold. Interestingly, moderately induced genes were subtracted even at a low C_0t value ($C_0t = 50$). This data suggests that subtractive hybridization tended to equalize the uninduced and moderately induced common sequences within target populations regardless of the C_0t value used or the abundance of the gene species. Therefore, subtractive hybridization should only be used to identify target genes that are either uniquely expressed or highly induced.

Rapid cloning and characterization of transcription factors expressed in human brain by signature sequencing analysis

P.D. Drew¹, J. D. Nagle², R. D. Canning¹, W. E. Biddison¹, K. Ozato³, and K. G. Becker¹. Neuroimmunology Branch¹ and Neurogenetics Branch², NINDS; and Laboratory of Molecular Growth Regulation, NICHD³, NIH, Bethesda, MD 20892

Transcription factors generally exist as families containing conserved DNA binding domains. We have isolated novel cDNA clones by screening brain libraries with probes prepared against conserved domains of transcription factors followed by single-run EST automated sequencing. We term this directed EST protocol-Signature Sequencing Analysis. A variety of potential transcription factors have been cloned and characterized including over 120 novel C2H2-type zinc finger proteins. We have also isolated a variety of clones encoding potential transcription factors including clones containing glutamine (CAG) repeats, and the forkhead family of developmental control genes. Individual clones are being analyzed for 1) levels of mRNA expression in different brain cell types, 2) full-length sequencing, 3) chromosomal sublocalization, and 4) DNA binding specificity.

Supplementary Approaches for Screening cDNA Libraries

H-X Xu¹, N. Baskaran¹, S. Patanjali¹, R. Sivakamasundari¹, H. Wei¹, W-F Fan¹,
S. Nalabolu¹, G. Nallur¹, F. Hisama¹, and Sherman M. Weissman¹.

¹Yale University School of Medicine, New Haven, CT 06510

We have applied cDNA selection to a number of genomic areas, and extensively studied the products obtained from a contig of YACs spanning much of the human MHC. With very few exceptions we have detected all known genes in each region and often found additional transcribed sequences even in rather extensively studied areas. In the MHC we have identified approximately 50 new cDNAs (genes) including a group of olfactory receptor genes, several products that are preferentially expressed in T cells, a B cell specific diubiquitin-like molecule, additional transcription factors, etc. Deriving long cDNA clones from selected short fragments is a relatively time-consuming and expensive process. To simplify this we are attempting to use the assemblage of selected cDNA fragments from a YAC to directly screen long insert cDNA libraries (either standard or normalized). This has the potential to bypass purification of individual cDNA inserts and should be particularly useful in conjunction with arrayed libraries. We will review our progress in this and associated efforts to obtain true 5' end cDNA libraries.

Affinity capture is a useful approach to prepare libraries enriched for particular sequence features such as simple repeats, mRNA destabilization motifs, etc. We have continued to work with affinity capture using both conventional oligonucleotides and, in collaboration with J. Coule of Millipore, peptide nucleic acids (PNAs). Current results suggest that PNAs may be useful in situations where it is impractical to use ordinary oligonucleotides for selection.

IDENTIFICATION OF TRANSCRIBED SEQUENCES IN THE SMA CRITICAL REGION BY DIRECT SELECTION OF cDNA CLONES FROM NORMALIZED LIBRARIES

M. Fatima Bonaldo^{1,3}, L. Lawton^{1,3}, T. Carter^{1,2,3}, B. Ross^{1,2,3}, T.C. Gilliam^{1,2,3} & M. Bento Soares^{1,3}
Departments of Psychiatry¹ and Genetics & Development², Columbia University, and The New York State Psychiatric Institute³, New York, NY 10032.

We have recently reported a method [Hum. Mol. Genet. 3(9), 1994] to isolate transcribed sequences from genomic DNA which involves hybridization of filter immobilized YAC DNA with an oligo-dT primed normalized directionally cloned cDNA library, in the form of single-stranded circles. The hybridized cDNAs are then eluted and directly electroporated into bacteria. We have successfully applied this method to isolate a number of genes that map to the Spinal Muscular Atrophy critical region at 5q13. In one such experiment, DNA from two overlapping YACs encompassing approximately 700 kb were filter immobilized and hybridized to a human fetal liver and spleen normalized cDNA library. A total of 770 colonies were obtained, 237 of which have already been analyzed. Plasmid DNA was prepared and sequentially hybridized with two probes: (a) human Cot1 DNA, and (b) labeled YAC DNA. All clones that hybridized with the latter but not with the former probe were subsequently verified to map to the critical region by Southern analysis. 24 of these 237 cDNAs (10%) hybridized back to the critical region, and the hybridization was shown not to have occurred through repetitive sequences. Sequence analysis revealed that these 24 cDNAs correspond to 14 different genes. The remaining 533 clones are currently being analyzed.

Sensitive and specific transcript mapping on YACs by cDNA selection using total yeast DNA : A study of Ataxia Telangiectasia, Kit receptor and MHC regions.

S.R. Patanjali¹, G. Rotman², H. Xu¹, K. Savitski², F. M. Hisama¹, Y. Shiloh² and S. M. Weissman¹.

¹Department of Genetics, Yale University, New Haven CT 06510, USA

²Department of Human Genetics, Tel Aviv University, Tel Aviv 69978, Israel

Reports of cDNA selection as described earlier (1,2) involved isolation of YAC by pulse field gel electrophoresis. Often, YAC clones bearing inserts of > 1 Mb are difficult to purify from the yeast chromosomes. In addition, the process is labor intensive and results in poor recoveries of DNA from the gel. Hence we obviated the need to purify YAC and reported cDNA selection using whole yeast genomic DNA containing the YAC clone of interest (3,4) by membrane bound target approach ("solid phase"). The sensitivity and specificity were retained by the use of specific quenchers for human, yeast and E.coli ribosomal DNA. In addition Alu and LINE element quenchers were used to suppress repeats. Typically 10-30% of the selected clones are ribosomal, which can be screened by filter hybridization.

Using this approach we performed cDNA selection on four non-chimeric YACs (AT16, AT67, AT71 and AT400) from the Ataxia Telangiectasia region.. Over 500 fragments of selected cDNA were analysed by hybridization and over 50 clones were sequenced and 18 unique sequences were identified. Specificity of the selected clones was verified by Southern blots of the DNA from the YAC contig as well as human genomic DNA. Specificity was confirmed by the appearance of cognate bands in the YAC of origin, overlapping YAC(s) and human genomic DNA. In addition, the unique clones were hybridized to a somatic cell hybrid panel of Chromosome 11 to confirm the specificity. 17 of the 18 clones (94%) tested map back to the region of interest. Isolation of full length clones is in progress. Collaborative efforts are underway to identify the AT gene from these clones.

We also performed cDNA selection on a YAC (950 kb) containing Kit proto-oncogene, a hematopoietic stem cell specific receptor. Kit cDNA was present in only two of the six libraries used as a pool to perform selection at a frequency of less than 1 in 250, 000. Such a system ideally provided conditions to test the specificity of the selection. About 10% of the selected clones were Kit and about 80% of the clones were ribosomal DNA. Large number of ribosomal clones in the selected DNA may be due to an aberrant ribosomal DNA gene or only a fraction of the yeast clones contained YAC DNA resulting in decreased availability of the target. Such a large enrichment of even very rare cDNAs by using whole yeast genomic DNA as target is comparable to the selections performed on purified targets as reported earlier. In a similar study, a putative olfactory receptor was observed in the MHC region whose abundance in the input library was less than 1 in 500, 000 (5).

Thus we conclude that direct cDNA selection using immobilized total yeast genomic DNA as target is at least as sensitive and selective as the earlier reported methods. Hence, it can be used as an effective tool to map transcripts from genomic regions of interest.

References: (1) Parimoo, S. et.al., Proc.Natl.Acad.Sci. U.S.A. (1991)88, 9623 (2)Lovett, M. Current Protocols in Human Genetics, Vol I, 6.3.8 (3) Parimoo, S. et.al. Nucleic Acids Res. (1993) 18, 4422 (4) Patanjali, S.R. et. al. Proc. of 3rd International Workshop on Transcribed Sequences (1994) (in press) (5) Fan, W. and Weissman, S.M. (1994) (Manuscript in preparation)

Isolation of New Genes by a Modified Direct Selection Technique

A.J.Bench¹, A.P.Cahn¹, H.Lehrach², G.Zehetner², P.H.Rabbitts¹

¹MRC Clinical Oncology and Radiotherapeutics Unit, MRC Centre
Hills Road, Cambridge, CB2 2QH

²Imperial Cancer Research Fund, P.O. Box 123, 44 Lincoln's Inn Fields,
London WC2A 3PX

Direct selection techniques aim to specifically isolate cDNAs encoded by a genomic sequence (e.g. a YAC) from a background of other cDNAs. The end result is a library containing an increased proportion of the selected cDNA. Considerable time and effort is then required to identify those cDNAs which are derived from the YAC concerned usually by hybridising cDNAs back onto the YAC. An alternative to this is to screen the selected library with the YAC itself. This has an advantage over direct YAC screening of an unselected library in that the proportion of putative positive clones will be much higher.

Accordingly, a YAC, code name ICRFy900D0561, (a chimaeric YAC from the ICRF library containing the thioredoxin pseudogene) and RNA from the cell line Ludlul (known to express thioredoxin) were used for direct selection. This YAC was of extra interest for two reasons. Firstly, it contained a *Not*I site, a good indicator of coding sequence. Secondly, direct hybridisation of the YAC to genomic DNA, under suppression conditions, identified a number of unique fragments, several of which were not detected when the thioredoxin cDNA was used as a probe.

After two rounds of selection, the proportion of thioredoxin sequences in the cDNA population was shown to have increased considerably. The selected cDNAs were cloned and further analysis indicated that 0.2% of clones were thioredoxin positive. One of these clones was sequenced; this sequence was in agreement with that published for human thioredoxin cDNA.

Hybridisation of the 2nd round selected cDNAs back onto the YAC identified non-thioredoxin bands indicative of the presence of additional genes within this YAC. In order to locate the genes, the 2nd round cDNA library was screened, under suppression conditions, with the YAC. Duplicated positives were picked, rescreened with the YAC and then with human ribosomal DNA probes. Those clones which were YAC positive and rDNA negative were used as probes and many were shown to map back to YAC and human DNA. Several clones were used for Northern hybridisation, giving unique bands. Some of these clones have been sequenced and open reading frames identified. At present, these genes remain unidentified.

This study shows that it is possible to isolate genes by a combination of direct selection and YAC hybridisation with no *a priori* knowledge of the sequences concerned.

Identification of transcribed sequences in human chromosome 9p by an efficient cDNA selection strategy using somatic cell hybrids

J. Sang and C.L. Jackson

Dept. of Pathology, RI Hospital/Brown University, Providence, RI 02903

In an effort to create a transcriptional map for human chromosome 9p along with the construction of a physical map, we have developed an efficient cDNA selection strategy using complex genomic DNA from somatic cell hybrids. In contrast to other cDNA selection techniques, this strategy reduces the complexity of cDNA sources by subdividing the library into fractions and selecting cDNAs from each fraction separately. This approach increases hybridization efficiency because of the increased ratio between target cDNAs and their genomic counterparts as well as the increased concentration of individual cDNA clones. It allows cDNAs to be selected from larger genomic regions, which will aid in the construction and verification of physical maps. The cDNA inserts amplified by vector primers from individual fractions were hybridized in solution to genomic DNA of somatic hybrid containing chromosome 9. The non-hybridized cDNAs were separated from the genomic DNA and the hybridized cDNAs. These retained cDNAs were then amplified by vector primers and the selected cDNAs are directly used as probes to screen the parental fraction of the cDNA library. This identifies the original cDNA phage clone without the need for subcloning thus eliminating the duplication of original clones associated with subcloning. A human fetal brain cDNA phage library was used in our initial studies. cDNAs obtained by this approach have been used as probes in Southern hybridization and shown to be positive in both total human and chromosome 9 specific hybrid DNA, but not hamster DNA. Preliminary sequence analysis indicates some cDNAs are ESTs with unknown chromosome locations and are now assigned to chromosome 9. A panel of somatic hybrids containing portions of chromosome 9 is being used to determine the regional localization of these cDNAs. Further work to complete the isolation of 9p cDNAs from all fractions of the cDNA library and the use of additional cDNA libraries is in progress. We have demonstrated that chromosome-specific cDNAs can be selected by very complex genomic DNAs of somatic hybrids using this protocol, which should also be applicable to pools of YAC or cosmid clones. This strategy should increase the efficiency of creating a transcriptional map of chromosome 9 and facilitate the isolation of disease genes as a complementary method to other techniques.

A sandwich-hybridization method for cDNA selection from large genomic regions: Search for candidate genes for XLRP locus.

D. Yan¹, R. Fujita¹, C. McHenry¹, A. Swaroop^{1,2}

¹Department of Ophthalmology and ²Human Genetics, University of Michigan, Ann Arbor, MI 48105

We have developed a novel hybridization-based selection method for isolating cDNAs from large genomic regions. This approach consists of: 1. hybridization of single-stranded circular cDNAs to the corresponding tagged genomic RNA; 2. retention of "genomic RNA" - cDNA hybrids on an avidin-matrix through a biotinylated RNA complementary to the tag; 3. elution of specific cDNAs from the avidin-matrix with ribonuclease A. We have previously demonstrated the feasibility of the method in model experiments, by selecting a specific retinal cDNA from complex mixtures using a sub-library from its corresponding 5 kb genomic clone. To specifically identify candidate genes for the RP3 locus, we are now applying this selection strategy to select cDNAs from human adult retina, fetal eye and fetal brain libraries, with the "tagged genomic RNA" derived from two 450 kb MluI fragments isolated from OTC-C and 55B YAC clones spanning the region of X-linked retinis pigmentosa (XLRP) locus RP3 at Xp21.1. The enrichment of TCTEX-1L gene that maps within the 55B YAC region is being used for monitoring the efficiency and specificity of the method. Of the 25 selected cDNA clones so far isolated, two cDNAs appear to map to the specific cosmid clones derived from the 55B YAC. They are now being characterized further. This hybridization-based method allows the isolation of full-length cDNAs, and should be suitable for selecting cDNAs from genomic source of any complexity.

Gene Identification by 'End-Ligation Coincident Sequence Cloning' (EL-CSC)

A.J. Brookes¹, E.M. Slorach¹, K.E. Morrison², S.J. Qureshi¹, S. Anderson¹, K.L. Evans¹, D. Blake³, K.E. Davies³ and D.J. Porteous¹. ¹MRC Human Genetics Unit, Edinburgh, UK., ²Neurosciences and ³Molecular Genetics Groups, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford, UK.

Methods for Coincident Sequence Cloning (CSC) attempt to isolate sequences that are shared between pairs of complex and largely distinct DNA mixtures. This general experimental strategy could be applied to many and various DNA resources, enabling the facile cloning of genomic DNA or cDNA fragments of interest. To be of practical utility a CSC method needs i) to be able to handle mixtures of genome complexity, ii) to exclude the recovery of repetitive elements that may be represented in both input DNA resources, and iii) furnish a very high degree of selective enrichment for coincident DNAs. We have now developed a simple generalised procedure that fulfils these criteria, selectively enriching coincident DNAs by over 10^6 fold. The method, termed End-Ligation CSC (EL-CSC), involves the formation of inter-resource DNA duplexes between a pair of input mixtures, with one resource having been previously biotinylated. Only genuinely coincident elements will form duplexes that have fully end matched regions. A high stringency ligation reaction is used to selectively join these perfectly base paired ends to synthetic oligonucleotides (capture oligonucleotides). Finally, after physical capture upon Dynal magnetic beads, the coincident DNAs are selectively amplified by the Polymerase Chain Reaction (PCR) using primers complementary to the capture oligonucleotide sequences.

Of its many possible uses, EL-CSC is particularly valuable for the identification of transcribed sequences within extensive tracts of genomic DNA. This involves the reaction of genomic DNA with an appropriate cDNA resource. Sequences coincident between these DNAs will be the expressed genes encoded within the genomic DNA. Primary cDNA is optimal because its overall complexity, and therefore the number of genes available for recovery, will be greater than for pooled clones from a harvested cDNA library. The tissue source of the cDNA is guided by the expected expression pattern of the genes one wishes to identify. Pooling several tissues has obvious advantages.

To date we have successfully applied the EL-CSC procedure to genomic DNAs, of up to 1.3 Mbp in length, cloned in either phage, cosmid and Yeast Artificial Chromosome (YAC) vectors. A representative experiment involved the use of 260kb of human genomic DNA available as cosmid recombinants. These clones were derived from YACs that spanned markers previously linked to the autosomal Spinal Muscular Atrophy (SMA) locus on chromosome 5. By reacting these cosmids against foetal brain cDNA we produced material that appeared as distinct bands when examined by agarose gel electrophoresis. Cloning and subsequent analysis of this material has so far identified 14 transcribed gene fragments encoded by the input DNA. Many of these fragments were represented in the input cDNA at levels of only 10^{-5} - 10^{-6} . Summed together these gene sequences represent approximately 75 % of the derived PCR product, reflecting the great specificity of the EL-CSC procedure. The low level at which (potentially) false sequences were recovered allowed their exclusion to be accomplished with minimal time and effort. EL-CSC has thus been demonstrated to be an effective and sensitive approach to the identification of transcribed sequences.

Comparative evaluation of techniques for capturing transcribed sequences.

T. Liang¹, N. Udar¹, S. Chiplunkar¹, S. Xu¹, S. Dandekar¹, N. Patel¹, H. Yang² and R. Gatti¹.

¹Department of Pathology, UCLA School of Medicine, Los Angeles, CA 90024 and

²Department of Medical Genetics, Peking Union Medical College, Beijing, China.

In an effort to clone an ataxia-telangiectasia gene we have simultaneously compared several techniques, namely, exon amplification (Buckler *et al* PNAS 88: 4005, 1991), direct cDNA selection (Parimoo *et al* PNAS 88: 9623, 1991), magnetic beads (Korn *et al* Hum Mol Genet 1: 235, 1992) and splice site consensus sequence screening (Melmer *et al* Hum Mol Genet 1: 433, 1992) using the same DNA starting material. These techniques were also compared to the classical library screening procedure. Our data substantiate the following conclusions: 1) exon amplification is very specific and easy to perform but further screening and characterization of the product is time consuming and laborious; 2) cDNA selection is relatively specific but not sensitive enough and is also time consuming; 3) magnetic beads have the least specificity and 4) splice site consensus sequence screening is most sensitive but not very specific. All the hybridization techniques need a fairly high level of cDNA representation in the screened libraries. Exon amplification is an especially useful adjunctive because it does not involve hybridization or tissue specificity. We conclude thusfar that the most reliable approach would be to combine several or all of these techniques.

Exon trapping

Den Dunnen, J.T., Dauwerse, J.G., Datson, N.A. and Van Ommen, G.J.B.
Department of Human Genetics, Leiden University, Leiden, the Netherlands.

We have used several techniques to identify transcribed sequences in cloned DNA, including direct cDNA-selection and exon trapping. Exon trapping was performed with several newly developed vector-systems. pETV-SD2 contains a splice donor site upstream of a multiple cloning site and exons are identified by the functional detection of a splice acceptor site in the cloned insert. Using the same RNA-sample, pETV-SD2 allows trapping of both internal and 3' terminal exons. Isolation of internal exons is facilitated by removal of a LacZ α marker gene in spliced clones and by easy cloning of the RT-PCR products requiring a circularization step only. Identification of 3' exons is performed by a 3' RACE reaction and turns out to be highly specific since clones are selected both by splicing and polyadenylation. Furthermore, the trapped 3' exons have a larger mean length and they serve as a landmark for the gene, defining its 3' end.

One disadvantage of exon trapping is that usually only one exon is trapped per clone. Consequently, trapped exons need to be hybridized to cDNA libraries both to obtain the remainder of the gene and to order all individually trapped exons. If the gene of interest is not or hardly expressed in the tissue(s) screened, these hybridizations do not result in the isolation of positive cDNA-clones. Alternatively, one may have problems because only homologous (highly expressed) cDNA's are isolated. To allow trapping of more exons from one clone we have developed a new vector. This vector, a cosmid, contains a human growth hormone gene (hGH) cloned behind a highly active eukaryotic promoter. A multiple cloning site, located within one of the gene's introns, allows easy construction of cosmid libraries with insert sizes of up to 40 kb. RNA isolated from cosmid-transfected COS-cells should allow, depending on the choice of primers for RT-PCR, the isolation of all types of exons. Internal exons are trapped by hGH exon-exon PCR, 3' terminal exons by 3' RACE and first exons by 5' RACE. Clones containing exon-less inserts can be identified by a high expression of the hGH-gene. We have constructed total genomic and YAC-derived cosmid libraries and specifically cloned multi-exonic gene fragments in the new vector and are currently investigating their behaviour in the system.

Utilization Of Exon Trapping To Generate A Transcriptional Map Of The PKD1 Locus.

G. Landes¹, T. Burn¹, T. Connors¹, W. Dackowski¹, J. Millholland¹, L. Petry¹, H. Shukla¹, R. Hakim², N. Doggett³, K. Klinger¹. ¹Integrated Genetics, Framingham, MA, ²Vanderbilt University, Nashville, TN, ³Los Alamos National Laboratories, Los Alamos, NM

Autosomal dominant polycystic kidney disease (APKD) is one of the most common heritable diseases with a frequency of 1 per 1000, and accounts for approximately 10% of all cases of end-stage kidney disease. The PKD1 gene, whose mutation is responsible for at least 90% of APKD cases, has been mapped to a 750 kb CpG rich region within chromosome 16p13.3. This region also contains the disease gene for tuberous sclerosis (TSC2) and possibly the gene responsible for familial Mediterranean fever (MEF). Our group has cloned the PKD1 interval as a series of 17 overlapping P1 clones and proceeded to use exon trapping and cDNA selection to clone and identify coding sequences from the APKD locus. For exon trapping, individual P1 clones from the APKD locus were subcloned into the pSPL3 splicing vector and transfected into COS-7 cells. Trapped exons were amplified by RT-PCR and subcloned. Approximately 100 putative exons, from 750 kb of genomic sequence, have been identified to date. Depending on the P1 clone, we have obtained as few as four to as many as twenty trapped exons per P1 clone. In many cases, we have found multiple exons which have been spliced and trapped together. Exon trapping has also allowed us to identify a novel alternative splice in the coding region of the TSC2 gene. Trapped exons have been characterized using southern and northern analysis and their sequence determined using Pharmacia ALF automated DNA sequencers. In greater than 95% of cases, the trapped exons contain nonrepetitive non-vector derived sequences which hybridize specifically to the parental P1 clones. We have also used some of the trapped exons successfully to screen cDNA libraries. In summary, we have found exon trapping to be a rapid and powerful technique for the rapid identification and cloning of coding regions from genomic DNA. The trapped exons are currently being used as probes to screen for full length cDNA clones in order to assemble a transcriptional map of the PKD1 locus.

Towards an Exon Map of the Human Y Chromosome

Yun-Fai Chris Lau¹, Rongxian Ding¹ and David B. Krizman²

¹University of California, San Francisco, CA 94121

and ²National Center for Human Genome Research, NIH, Bethesda, MD 20892

The human Y chromosome is one of the smallest chromosomes of the human genome and contains approximately 2% of the total human haploid DNA. Recent advances in physical mapping have generated a YAC-based contig map of the entire euchromatin, thereby providing an excellent opportunity for the construction of a transcription map of this chromosome. Currently, only a few Y-genes have been identified, most of them are located on the pseudoautosomal region. Further, little information is available regarding the possible expression patterns of the other, yet-to-be identified, genes. Hence, in order to construct a transcription map, a positional cloning strategy has to be developed for a large scale collection of functional sequences from this chromosome.

As a preliminary study to evaluate the 3' exon trapping technique as the basic approach to achieve the above objective, we have generated a 3000-colony 3' exon library using the 3' terminal exon trapping vector, pTAG4, and total DNA from a pool of 4,600 cosmids of a flow-sorted human Y chromosome library. These cosmids represent about 3 genome equivalence of this chromosome. The plasmid, pTAG4, harbors an exon trapping modular gene that consists of the SV40 early promoter, two adenovirus exons interrupted by two introns. The human DNA was inserted in the multiple cloning site at the last intron and was transfected into the mammalian COS-7 cells. If a functional 3' terminal exon is present in the human DNA, a hybrid transcript will be generated and will consist of the adenovirus exons, the trapped 3' terminal exon(s) and a poly (A) tail derived from the polyadenylation signal of the trapped terminal exon. The hybrid transcripts can be specifically selected using reverse transcription, PCR amplification and various combinations of specific primers. Since this technique requires the 3' terminal exon from each gene to be present for a successful exon trapping, multiple trappings of the same genes from a particular chromosome can be easily identified. The 3' exons, therefore, are ideal sequences for representing their respective genes in a chromosome-specific transcription map. To evaluate the quality of this Y chromosome exon library, 576 colonies were arrayed in 6 microtiter plates and hybridized with two specific repeat genes, *TSPY* and *YRRM*, from the human Y chromosome. Our results indicated that 35 and 39 colonies hybridized positively with *TSPY* and *YRRM* probes respectively. Random sequencing of 54 clones indicated that 21 clones (39%) were derived from plasmid or *E. coli* sequences. Of the remaining clones, 5 (9.3%) were derived from exons of *TSPY* and *YRRM* genes and 28 clones (52%) show either some homologies to other sequences (apart from vector and *E. coli*) or no homology to sequences in the GenBank. The latter group, therefore, represent potential Y chromosome exons and can be used probes for the isolation of the respective cDNAs from the appropriate cDNA libraries and for physical mapping on the YAC contig of the Y chromosome. Our results indicate that 1) multiple Y chromosome equivalence can be used for a large-scale 3' terminal exon trapping; 2) >60% of trapped exons are derived from either known genes or potentially unknown genes of this chromosome; 3) the trapped exons can be useful probes for cDNA isolation and exon mapping of the human Y chromosome.

Trapping internal and 3'-terminal exons from YACs

P. Nisson¹, A. Ally¹, R. Bollo¹, D. Krizman², N. Sacchi³, and P. Watkins⁴

¹Life Technologies, Inc, Gaithersburg, MD; ²National Center for Human Genome Research, Bethesda, MD; ³University of Milan School of Medicine, Milan, Italy;

⁴Sequana Therapeutics, La Jolla, CA.

Exon trapping has proven to be an integral part of many positional cloning projects including the discoveries of the Huntington's Disease gene and the gene for Neurofibromatous Type 2. Until recently, several exon trapping vectors have been available but they all share the common feature of targeting internal exons, that is exons that are flanked by splice donor and acceptor sequences. One of the most commonly used internal exon trapping vectors, pSPL3, contains the SV40 early promoter, origin of replication, splice donor and acceptor sequences from the HIV-tat gene. Cloned genomic sequence is screened for exons by subcloning it into the HIV-tat intron, transfecting the pSPL3 subclones into COS-7 cells to allow RNA expression, followed by RT-PCR and cloning. An exon is selected or trapped by the COS-7 splicing apparatus by virtue of the splicing signals that flank it. 3'-Terminal exons can be trapped by targeting the poly-A tail and the upstream splice acceptor sequence using the vector pTAG4. This method is similar to internal exon trapping except that an oligo-dT adapter primer is used to initiate first strand cDNA synthesis. PCR is subsequently used to amplify the trapped 3'exon terminal. To date, a majority of exon trapping has been done with bacteriophage or cosmid clones. With the advent of improved physical maps of the human and mouse genomes using YACs, it would be useful to exon trap directly from YACs. We report progress in the development of internal and 3'-exon trapping from a human YAC and general improvements in the 3'-terminal exon trapping method.

Regional Maps

Discovering, mapping and characterization of 39 new genes in the human MHC class I region by cDNA selection

Wufang Fan*, Satish Parimoo., and Sherman Weissman

Department of Genetics, Yale University, New Haven, CT 06510

*Human Genome Center, LLNL, CA 94550

The human MHC is located in the short arm of chromosome six covering 3.5-4.0 megabases of DNA. Genes in this complex can be subdivided into three classes according to their structures and functions. Recently, more than 40 new genes have been discovered within the class II and class III regions, but the class I region, covering 2 megabases of genomic DNA, has not been extensively studied.

In the study presented here, a novel strategy, cDNA hybridization selection, has been applied to investigate the gene content of three yeast artificial chromosomes (YAC) covering about 1 megabase of DNA within the class I region. The results demonstrate that this method was very specific and efficient. It can be used to isolate fragments of cDNA encoded by a given YAC with little or no contamination from ribosomal sequences, repetitive sequences or unique copies of genomic sequences not present in the YAC.

Using this strategy, more than 39 new non-class I genes were discovered in the class I region. DNA sequence analysis showed no obvious homology to any known sequences available in the GenEMBL database. Existence of these new genes were further confirmed by isolation of the full length cDNAs from 3 different cDNA libraries; by genomic DNA hybridization, Northern blotting and physical mapping. All of the new genes were mapped to their own loci.

To explore the patterns of gene expression, Northern hybridization was performed with poly-A mRNA from different tissues or cell types. Some of the new genes were specifically expressed in T lymphocytes, some of them only expressed in B lymphocytes, and others strongly expressed either in testis tissues or in Hela cells. These results imply different functions or common features. The B or T lymphocyte specific cDNAs may play some unknown roles in the immune system.

Interestingly, several olfactory receptor-like genes have been identified in the investigated region. One of full-length cDNA was sequenced. The result shows a striking homology, i.e., more than 60% identity in both nucleic acid and amino acid levels, to the mouse olfactory receptor genes. It has been observed for a long time that mice, and perhaps men, exhibit a preference for mating with individuals of disparate MHC haplotypes. In mice this discrimination among MHC haplotypes is believed to be mediated by olfactory detection of substances in the urine, so called odortypes. These effects are also seen in outbred strains of mice indicating that the relevant genes are located within the MHC. It is very interesting to see whether or not the olfactory receptor genes encoded in MHC class I region are indeed involved in the observed phenomenon.

Identification of new coding sequences in large genomic regions using a combination of different methods.

P.Sanseau¹, A.Jackson², K. Bankhead¹, J.Riley¹, M.L. Yaspo², R.Sherrington³, M. North², L.Hosking², B.Korn⁴, J. Trowsdale², S. Beck², I. Purvis¹, P. St-George Hyslop³, and C. Dykes¹.

¹Glaxo Research and Development, Molecular Pathology Dpt, Greenford Road, Greenford, UB6 0HE, U.K; ²Imperial Cancer Research Fund, 44, Lincoln's Inn Fields, London WC2A 3PX, U.K.; ³Centre For Research in Neurodegenerative Diseases, 6 Queen's Park, Toronto M5S 1A8, Canada; ⁴DNKFZ 280, 6900 Heidelberg, Germany.

The identification of new transcribed sequences is one of the major goals of the Human Genome Project. Recently, several methods have been described to uncover new transcripts. Since each method has its advantages and drawbacks, we have used a combination of these techniques on a well characterised region of the human genome: the class II region of the MHC located on chromosome 6. Cosmids covering the region have been the starting material. A physical map has been established to localise CpG islands and the clones have been used directly as probes on cDNA libraries, northern or zoblots to find genes such as TAPs, LMPs and IPP2 (1,2). Cosmids have also been subcloned to initiate the complete sequencing of the locus. So far more than 200 kb have been sequenced and analysed with the Grail program (3). This genomic sequence allows the understanding of the organisation from a large region and the finding of potential new coding sequences. Exon amplification and cDNA selection have been performed on the cosmids. Known genes and two new conserved sequences, under current analysis, have been found with these techniques (4). In our hands this combination approach was very powerful to uncover new potential transcripts and to characterise in detail an already well-known region of the human genome.

The same strategy is being used on chromosome 14 at q24.3 where a linkage has been described between early onset pedigrees for Alzheimer's Disease and DNA markers.

1. Trowsdale et al., (1993) Immunology Today, 14:349
2. Sanseau et al., (1994) Mammalian Genome, in press.
3. Beck et al., (1992) Journal of Molecular Biology, 228:433
4. North et al., (1993) Mammalian Genome, 4:466

**Identification of transcripts in YAC contig located in 11q13
between PGA and PYGM.**

D.S. Gerhard, C.M. Smith, E. Lawrence, T. Kamarasova, J. Wu¹

Dept. of Genetics and ¹Surgery, Washington University School of Medicine, St. Louis, Mo.
63110

We have previously isolated a large number of bacteriophage clones specific for human chromosome 11 and mapped them with a new series of somatic cell hybrids. One hundred and forty-one clones map to 11q12-q14, a region harboring a number of human disease genes including atopy, McArdle's syndrome, B-cell lymphomas, Best's macular degeneration, Bardet-Biedl syndrome and multiple endocrine neoplasia type 1.

Eight of these bacteriophage clones that map into the segregation group located between PGA and PYGM were converted into sequence tagged sites (STSs) and used to screen the Washington University YAC library. All STSs found corresponding YAC clones. A contig of approximately 600 kb containing all 8 STS within 14 YACs was achieved. Four non-chimeric YACs were digested with Sac II, Mlu I, Not I and Sfi I. This region of the genome has a high frequency of restriction enzymes as evidenced by long range restriction mapping of this contig. Within this 600 kb, we identified 7 Sac II sites, 7 Sfi I sites, one which appears to be polymorphic and 3 Mlu I and Not I sites. At one end of this contig is a cluster of 2 Mlu I sites and one Sac II site.

Of the eight bacteriophage clones that map to this region, two clones cross-hybridize with rodent DNA. This cross-species homology suggests the presence of a gene or genes within this region. Based on the STS and restriction mapping information, these two clones are approximately 200 kb apart. We are currently pursuing two independent strategies to identify the transcribed sequences of this YAC contig. One, we are initiating the use of the exon-amplification system with pSLP3 to obtain random exons from isolated YAC DNAs; and two, we are screening subclones of the bacteriophage clones that cross-hybridizes to rodent genomic DNA against a few cDNA libraries. We shall report our progress on this project.

A Transcriptional Map of the Ataxia-Telangiectasia Locus on Chromosome 11q22-23

A. Bar-Shira¹, K. Savitski¹, S. Gilad¹, S.R. Patanjali², G. Rotman¹, Y. Ziv¹, L. Vanagaitel¹, A. Simmons³, G. Clines³, D. Tagle⁴, F.S. Collins⁴, S. Weissman², M. Lovett³, Y. Shiloh¹. ¹Dept. Human Genetics, Tel Aviv University, Tel Aviv 69978, Israel; ²Boyer Center for Molecular Medicine, Yale University School of Medicine, New Haven, CT 06536; ³Dept. Biochemistry and McDermott Center, University of Texas Southwestern Medical Center, Dallas, TX 75235, and ⁴National Center for Human Genome Research, N.I.H., Bethesda, MD 20892.

Ataxia-telangiectasia (A-T) is an autosomal recessive disorder featuring cerebellar degeneration, immunodeficiency, chromosomal instability, cancer susceptibility, and radiosensitivity. Four complementation groups have been observed in A-T. The two major groups, A and C, were localized to chromosome 11q22-23, and the other two, D and E, may map to the same chromosomal region. Genetic analysis recently confined the A-T locus to a 2 Mb region, between D11S1819 and D11S1960. We constructed YAC and cosmid contigs spanning this region, and applied the following methods to identify transcribed sequences: direct selection of cDNAs by magnetic bead capture (MBC), using YAC and cosmid clones; direct selection using YACs immobilized on a solid matrix; exon amplification; 3' exon trapping; and, screening of cDNA libraries with cosmid clones. The first two methods usually yield small (100-600 bp) cDNA fragments. YACs proved to be no less efficient than cosmids as starting material for MBC, with more than 50% of the products mapping back to the genomic clones. However, when we used a small panel of radiation hybrids spanning the A-T region to test the cDNA fragments, we found that some clones obtained by MBC or immobilized YACs, that hybridized back to YACs and cosmids, were not derived from this region. This pitfall probably stems from limited homology between certain portions of different genes, and points up the necessity to use radiation hybrid mapping when testing the authenticity of the captured sequences, and not to rely solely on cloned DNA for this purpose. Homology searches in sequence databases showed that only one of the first 60 cDNA fragments that we mapped to the A-T region was homologous to a sequence previously deposited in one of the databases, as an EST. Sequence analysis and physical mapping showed that direct selection methods and exon amplification are complementary in identifying transcribed sequences. The availability of a deep cosmid contig enables rapid and precise physical localization of the cDNA fragments and captured exons, leading to a detailed transcriptional map of the region. These sequences are being used as anchor points in isolating full-length clones from a variety of cDNA libraries. Initial screening of the libraries by PCR, using primer sets derived from individual cDNA fragments or exons, helps identify the libraries most likely to yield corresponding cDNA clones. The density of our transcriptional map indicates that the A-T region is rich in transcribed sequences. The involvement of these genes in A-T is being investigated by searching for mutations in patients, and by testing the ability of their cDNAs to complement the radiomimetic sensitivity of A-T cells.

Mapping of a Sequence Homologous to LTR of Endogenous Human Retrovirus on Human Chromosome 19.

Volik, S. V.¹, Lebedev, Yu. B.¹, Obradovic, D.¹, Ermolaeva, O. D.¹, Lennon, G.², Sverdlov, E. D.¹

¹Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow; ²Human Genome Center, LLNL, USA.

The human genome contains multiple copies of sequences related to one of the families of endogenous retroviruses, HERV-K homologous to the B-type mouse mammary tumour virus. A DNA fragment closely resembling a HERV-K LTR has been detected in a library of hncDNA clones enriched with sequences of human chromosome 19. The sites homologous to this sequence have been identified on human chromosome 19 by the hybridization to the arrayed chromosome 19 cosmid libraries and thus the first high resolution map of LTR distribution along an individual human chromosome has been constructed. We estimate the total number of such sites to be at least 170 on human chromosome 19 alone. In at least 4 cases this sequence is located within 1 kb of transcribed *Alu* repeat. In a number of cases the cosmids positive for this probe are located in vicinity of known genes, that could be an indication of possible regulatory functions of this sequence.

Progress towards a transcriptional map of the q21-22 region of chromosome 7.

J. M. Rommens^{1,2}, L. Mar¹, J. McArthur¹, L.-C. Tsui^{1,2}, S. W. Scherer^{1,2}.

¹Department of Genetics, Research Institute, The Hospital for Sick Children and ²Department of Molecular and Medical Genetics, University of Toronto, Toronto, Canada.

In order to understand the defects that cause the genetic conditions that map to the q21-q22 region of human chromosome 7, a detailed integrated map is being constructed. The entire region, which is estimated to contain 30 Megabases of DNA, has been divided into thirteen intervals on the basis of the breakpoints of rearranged chromosome 7 copies contained in somatic cell hybrid lines. Over 250 YAC clones have been assembled into four large contigs that cover most of the region and to date, twenty-six cloned genes have been finely mapped. To facilitate the rapid isolation of additional transcribed sequences, a direct selection procedure is being applied to 100 YAC clones that cover this region with a minimal amount of overlap. Essentially, purified artificial chromosomes are being immobilized onto nylon filters and subjected to two rounds of hybridization with pooled cDNAs of combinations of ten to twelve fetal and adult tissues. The primary sources of cDNA have been prepared such that the generated ends of each cDNA retains identification of tissue source and also instructs the subsequent expansion, amplification and cloning steps. Through a combination of characterizations including overlap screening, physical mapping and RNA hybridization, the cloned cDNAs are being arranged into tentative transcription units. This information provides the preliminary framework for the construction of a detailed transcription map that is fully integrated with the physical map. This map will provide insight into the organization and function of this chromosome region.

Hybrid selection of human chromosome 19 cDNAs

G. Lennon, C. Prange, S. Rouquier, and D. Giorgi

Human Genome Center, L-452, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore CA 94550

We have used a hybrid selection approach to isolate cDNAs from human chromosome 19. Our strategy combines elements that have each been used successfully in other hybrid selection experiments (Lovett et al., 1991; Parimoo et al., 1991; Korn et al., 1992), but uses flow-sorted chromosomes as the initial genomic substrate, and uses the selected material to probe arrayed cDNA library filters. From the 40,000 clones present in the arrays, 600 were initially identified as potentially from chromosome 19. To date, sequences from 121 clones have been obtained. 50% are a combination of false positives and true-but-nonetheless-unwanted positives (from pseudogenes present on ch 19). 5 % derive from already known chromosome 19 genes, and 45% from hitherto unknown genes. All cDNAs within this latter category are unique. Half of these hybridize to chromosome 19 cosmids. PCR reactions with somatic cell hybrid substrates has independently confirmed that at least 3 cDNAs map exclusively to ch 19. Overall, our analysis indicates that in this one hybrid selection experiment we are likely to have isolated cDNAs from over 70 novel chromosome 19 genes.

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

Large Regions of DNA from Chromosome 21q21 may be Transcriptionally Silent.

Xu, H-X¹, Gardiner, Katherine², Weissman, Sherman M.¹

¹Yale University School of Medicine, New Haven, CT 06510; ²Eleanor Roosevelt Institute, Denver, CO 80206.

As part of an effort to detect the transcribed cDNA sequences from chromosome 21q, we performed cDNA selection on 9 YACs (615C9, 449F6, 328D5, 93B1, 746B10, 221B7, 264F9, 193G12 and 745C11) from the dark band at 21q21, using a mixture of six short fragment cDNA libraries. The selected cDNA libraries contained predominantly short fragment clones of ribosomal cDNA, repetitive DNA and multi-primer sequences, and no single copy cDNA inserts. The same results were obtained when one of the YACs from 21q21 was selected with a normalized short fragment cDNA library. Parallel experiments with a range of other YACs yielded cDNA insert libraries that consisted largely of YAC specific single copy cDNA inserts. Approximately 500 plaques from each selected cDNA library were hybridized with a pool probe of ribosomal DNA, total human genomic DNA and multiprimer sequences. In general, over 95% of the clones were positive. Ten negative clones from each library were sequenced and none of them proved to be single copy cDNA. One YAC from 21q21 was chosen for more intensive investigation. EcoR1 fragments from the YAC DNA were cloned in λ -zap and twenty different fragments were sequenced in from both ends. None of the sequences showed even a marginal signal for an exon with the GRAIL program and 19 of the 20 inserts contained repetitive sequences not specific for chromosome 21. Together with the success of the selection methods for finding even very scarce cDNAs in other YACs, these results suggest that several megabases of DNA in this dark band region lack genes. In contrast, single copy cDNAs have been selected from YACs in other 21q dark band YACs.

Transcriptional mapping of chromosome 21 by cDNA hybridization selection from YAC clones

K. Gardiner¹, H. Xu², F. Tassone¹, H. Wade¹ and S. Weissman²

¹Eleanor Roosevelt Institute, Denver, CO 80206; ²Yale University School of Medicine, New Haven, CT 06510.

The technique of cDNA hybridization selection is being applied to a minimum tiling path of non-chimeric YAC clones from human chromosome 21q. Using cDNA from fetal brain, whole fetus, testes, liver, thymus and spleen, selected cDNA libraries have been constructed from 19 YAC clones, together totalling 10 Mb, and comprising >8 Mb of non-overlapping DNA, derived largely from regions characterized as gene-rich.

Ribosomal RNA and repetitive DNA content of all libraries appears low, and of 10 genes known to map within these YACs, all but one have been found at significant levels in the appropriate libraries. In all libraries, between 65% and 100% of unique sequence clones map to the correct YAC and/or chromosomal region.

Characterization of unique clones includes sequencing, followed by Grail and BlastN/X analysis, screening of cDNA libraries and expression studies, and mapping within the parent YAC. While conversion of the short cDNA fragments to full length cDNAs remains difficult and time consuming, estimates of the number and distribution of new genes can be made. It appears that novel genes, often including those with low levels of transcription (based on presence at $1/10^6$ in long insert cDNA libraries), are being obtained from essentially all YACs. In some cases, gene densities appear as high as 1/50 kb, in others 1/300 kb. Continued analysis should yield, not only useful candidates for chromosome 21 associated diseases, but also further information on human genome organization.

cDNA SELECTION IN THE HUMAN Xq12-Xq21 REGION.

Villard L., Colleaux L., Lossi A.M., Passage E., Fontes M. INSERM U406. 27, Bld J. Moulin. 13385 MARSEILLE CEDEX 5 FRANCE.

Genome mapping projects require the development of efficient techniques to build physical and transcriptional maps. The first goal has been achieved using YAC contigs. For the second, we decided to use direct cDNA selection (Parimoo et al., PNAS, 1991, 881, 9623-9627). Usually, total YAC DNA is prepared by PFGE, an often difficult and time-consuming step, especially when using clones with large inserts. We thus have developed an alternative strategy in which YAC DNA is amplified by IRS-PCR from preparations of total yeast DNA using degenerated oligonucleotides for Alu sequences. To test the feasibility of that approach, we hybridised IRS-PCR products from various YAC clones of a test region (Xq13.3) with probes originating from six different genes present in this sub-band. In all cases, the IRS-PCR products were shown to contain at least one exon of the known genes present in the original YAC.

The clones obtained after two rounds of selection were hybridised with probes for human rDNA and human Cot-1 DNA. We found rDNA to be the major contaminant of the selected clones. Depending on the YACs used and the quenching conditions, 10 to 80 % of clones are rDNA. On the other hand, no more than 5 % of the clones contain repetitive elements. We found that it was necessary to increase the quenching time of the amplified cDNAs up to several hours to reduce the number of rDNAs in the selected mini-libraries. Additionally, clones can be positively selected by hybridisation with the IRS-PCR products. They are then mapped back onto YAC panels. As opposed to what has been observed for the cDNA selection using PFGE prepared YAC DNA, the use of IRS-PCR products does not show any normalisation property. Extensive characterization of genes encoded by a given region would therefore probably require a combination of both protocols.

We have constructed several mini-libraries of selected products using adult, fetal and embryonic cDNA libraries and four YAC contigs in Xq13.1, Xq13.2, Xq13.3 and Xq21.1 respectively. We have characterized 15 new genes so far. More clones are under analysis. All isolated clones are being tested for their potential involvement in genetic diseases assigned to that part of the human X chromosome.

Moreover, we now plan to use this technique on cell-sorted single chromosome to extend this preliminary transcriptional map to the whole human X chromosome.

Identification of genes and construction of a transcriptional map in Xq28

E.Maestrini, S.Rivella, S.Bione, F.Tamanini, C. Tribioli, M.Mancini,
F.LiBergolis, and D.Toniolo.
Istituto di Genetica Biochimica ed Evoluzionistica, CNR, Pavia.

In the distal portion of Xq28, two gene rich regions have been characterized in detail. Cosmid and YAC contigs have been constructed: one links the FVIII to the G6PD and to the Color Vision (CV) genes, the second links the HCF, L1CAM and ALD genes to the BGN gene. CpG islands were identified and mapped back to the cosmids: cDNAs were isolated and new genes with and without CpG islands were identified. More CpG islands and cDNAs are being isolated and mapped.

In 450 kb of genomic DNA between the G6PD and the CV genes, 17 genes and 3 CpG islands, whose flanking cDNA has not been identified yet, are present. Similarly, in the L1CAM region 7 different transcripts have been characterized in 100 kb of DNA, but several additional CpG islands demonstrate that more genes are to be identified in the remaining 400-500 kb. Based on the orientation of transcription, in tissue distribution and in conservation of gene order in evolution "transcriptional domains" of about 100 kb were identified in the two regions. The significance of such order is being investigated by different approaches. In collaboration with Dr.G.Herman, we are characterizing the corresponding region in the mouse. We have determined that all the human genes are present in the mouse, with the same order and the same orientation of transcription. The pattern of expression of genes in the same transcriptional domain is being investigated to look for a common transcriptional pattern during mouse development.

FUNCTIONAL ANALYSIS OF EXPRESSED SEQUENCES ISOLATED FROM HUMAN Xq27-28.

Appierto V., Sacchi M.C., Pergolizzi R., De Bellis G., Biunno I.

Istituto Tecnologie Biomediche Avanzate, CNR , Via Ampere 56
20131 Milano ,Italy

In order to contribute to the isolation and analysis of genes, in the human X chromosome we have concentrated our efforts to the study of the genic content in three cosmid clones from the q27-28 regions (26H7,14B7, 2G2). The X chromosome is the most widely studied of all the chromosomes due to the high number of clinically important genes. The sub-telomeric portion (q28) of this chromosome has been the target of intense mapping studies, due to the high number of inherited clinical disorders. The region proximal to FRA X (q27) has instead been poorly studied because lack of genic signals such as HTF islands. We have performed direct cDNA selection (1) to screen cDNA libraries from human fetal, liver and muscle. We have isolated cDNAs pertaining to the three cosmids of interest, analysis of their sequences showed lack of total identity to previously identified genes. While the gene contained in the cosmid on q28 (2G2) region gave immediate analysing results on Northern blots, for cosmids on q27 (26H7,14B7) further manipulations were required including isolation of overlapping clones and extensive genomic sequencing. The gene on Xq28 (denominated Xib) (2), showed a striking conservatin in evolution, DNA fragments were observed even in fish and bird DNA. On Northern blots, Xib resolved two transcripts of 2.0 and 2.5 Kb only in some fetal tissues (kidney, lung and liver). A single 2.0 mRNA species was instead observed in fetal and heart tissues and adult skeletal muscle, moreover in these tissues, over-expression was also observed. Protein data bank analysis for Xib, showed homology to a deoxyribonucleic complex. For the genes contained on the q27 region (3) conserved fragments were observed even in mouse DNA, but the expression profile indicated very low and temporal specific espression. For one genomic clone (26H7) complete sequence was performed and analysis demostrated 89% identity to human and mouse U7snRNP. U7snRNP is one of the several factors involved in the 3' end maturation of histone premessage RNA's (4).

References:

- 1) Korn et al. (1992) Human Molecular Genetics 1N°4 235-242
- 2) Pergolizzi et al. (1994) Submitted Genomics
- 3) Mostardini et al. (1994) Submitted Genomics
- 4) Scharl et al. (1994) EMBO J. 13:2432-2440

Transcriptional mapping by direct cDNA selection: From Xq28 to the entire X chromosome

B. Korn and A. Poustka, Deutsches Krebsforschungszentrum, Heidelberg, Germany

The human X chromosome has been studied by genetic and molecular means for a long time. The terminal part of the long arm of the chromosome has been screened for transcripts in very detail by cDNA selection. We were able to re-isolate all known genes from this region out of cDNA sublibraries generated by using genomic clones as templates for the selection technology. Over all we used about 600 cosmids, 2 P1 clones and 7 YACs to generate more than 20 000 cDNA clones, which are analysed by hybridisation and exclusion sequencing. The complexity of our cDNA source (from three human tissues) seems to be very high. In the current state of analysis, we have not identified any clones which are identical. Due to mapping randomly picked clones, we estimate that more than 80 % of the cDNA clones do map back to the chromosomal region used for cDNA selection.

We are now in the process to construct 3'-cDNA sublibraries from the entire X chromosome and other chromosomes, using modified protocols. As cDNA source we currently use cDNA from six different human tissues. The genomic templates are either flow sorted chromosomes or pools of genomic clones. Mapping of the selected cDNA clones is performed on characterised YAC clones of the X chromosome, hybrid cell line panels or by FISH. The rapidly growing transcriptional map does provide candidate genes for genetic diseases which are mapped along the chromosome.

The analysis will be complemented by exon trapping (internal and 3') of specific regions. In combination with other mapping approaches on the X chromosome it will be possible to construct an integrated physical, genetic and transcript map.

Identification and Isolation of Genes Involved in DGS/VCFS: A cDNA Selection-Based Approach

W.Gong^{1,2}, B.S. Emanuel^{1,2}, J. Siegert^{1,2}, J. Collins^{1,2}, E. Goldmuntz¹, D.H. Kim¹ and M.L. Budarf^{1,2}.

¹The Children's Hospital of Philadelphia, and ²Human Genome Center for Chromosome 22, Philadelphia, Pennsylvania.

Isolation of expressed sequences from a defined genomic region of biological interest has been a major step in the identification of genes involved in the etiology of genetic diseases. DiGeorge syndrome is a developmental field defect of the third and fourth pharyngeal pouches resulting in abnormal development of the thymus, parathyroid glands and heart. The majority of patients are deleted for a specific region (DGCR) of 22q11.2. Patients with Velocardiofacial Syndrome also have similar deletions within the DGCR. To isolate genes involved in these disorders we have constructed a cDNA sublibrary specific for a 250 kb area within the DGCR by direct selection of cDNAs. In order to increase the complexity of the starting cDNA, mRNA from a variety of tissues (such as brain, muscle and liver) was used to generate cDNA. Tissue specific linkers were added, and the cDNAs were amplified by PCR. To select for region-specific cDNAs, cosmids spanning the 250 kb region were biotinylated by nick translation and hybridized to the amplified cDNAs in solution. The cosmid/cDNA complexes were captured on streptavidin-coated magnetic beads, and the specific cDNAs were eluted, amplified and subcloned.

The insert size of cDNA clones ranges from 350 bp to 1.2 kb. The initial cDNA sub-library consisted of 570 clones. To date, 432 cDNAs (76% of 570 clones in this sub-library) were identified and grouped by hybridization to cosmids representing this region. Of the remaining 138 clones: 12 appeared to contain only vector, 56 gave multiple bands when PCR amplified and 70 did not hybridize back to the cosmids. The distribution of selected cDNAs is not uniform over the genomic region, ranging from one to 40 clones every 10 kb of genomic DNA. Based on analysis of sequence and hybridization data, 99 of the clones contain regions of homology to several novel cDNAs which have been independently mapped within this region (see abstract by Budarf et al.). One cDNA contig, consisting of 35 cDNA clones, contains expressed sequences encompassing a 2.6 kb genomic region. The gene has 80% of homology to a mouse protein kinase. Another cDNA contig, with 24 cDNA clones represents the sequence of a novel partial cDNA, LAN. The original cDNA was isolated from a human fetal brain cDNA library probed with a cosmid from the proximal portion of the DGCR. To date, several novel transcripts mapping to the DGCR have been identified by this method and further analysis is ongoing. This region-specific cDNA library will be used as a resource to develop a complete transcriptional map and to identify the candidate genes associated with the DiGeorge and Velocardiofacial Syndromes.

Isolation of cDNAs from 22q11; the region commonly deleted in DiGeorge and Velo-cardio-facial syndromes.

Sirotkin, H.¹, Morrow, B.¹, DasGupta, R.¹, Parimoo, S.², Patanjali, S.², Weissman, S.², Kucherlapati, R.¹

¹Albert Einstein College of Medicine, Bronx, NY; ² Yale University, New Haven, CT.

DiGeorge sequence (DGS) and Velo-cardio-facial syndrome (VCFS) are relatively common disorders characterized by cleft palate, cardiac abnormalities, learning disabilities and characteristic facial dysmorphism. There is considerable variability in the spectrum of clinical features of the both disorders. A subset of DGS patients and the majority of VCFS patients have interstitial deletions of 22q11 of 1-2 mb. No correlation has been made between the extent of the deletion and the severity of the phenotype. In an effort to isolate the gene(s) responsible for DGS and VCFS we have utilized a hybrid selection protocol to recover expressed sequences from three nonoverlapping YACs comprising almost 1 mb of the commonly deleted region. Total yeast genomic DNA or isolated YAC DNA was hybridized with a mixture of random primed short fragment cDNA libraries. Six human short fragment libraries derived from total fetus, fetal brain, adult brain, testes, thymus and spleen have been used for the selections.

Short cDNAs shown to originate from the YACs and from chromosome 22 are being used to isolate full length cDNAs. In order to speed screening, PCR primers were designed from the short fragments and used to screen cDNA libraries that have been subdivided into smaller pools of approximately 30,000 clones. For each selection, a previously cloned gene was isolated. Among the genes we have isolated are catechol-O-methyltransferase, tuple1 and clathrin heavy chain. Additionally, a gene related to the murine p120 gene and several novel cDNAs have been recovered. The role of these genes in DGS and VCFS is being investigated.

Isolation of transcribed sequences in the reeler region of mouse chromosome 5.

C.Lambert de Rouvroit, M.-C. Beckers, I.Bar, I.Royaux and A.M. Goffinet.

Department of Physiology, FUNDP School of Medicine, rue de Bruxelles 61, B-5000 Namur, Belgium.

reeler (rl) is a recessive mutation in mice that affects brain development. As the product of the gene is unknown, positional cloning is an obvious strategy to approach the locus, that maps in the centromeric portion of chromosome 5.

Analyses of interspecific crosses showed that *rl* maps within walking distance (0.0 - 1.0 cM) from markers *D5Mit61* and *D5Mit72*. A PCR screen of the MIT/Princeton YAC library yielded 4 YACs positive for *D5Mit61* and 3 positives for *D5Mit72*, and a contig was assembled using cloned YAC termini. Genetic evidence demonstrated that *rl* was contained within the cloned interval defined, on the centromeric side, by two extremities of YACs positive for *D5Mit61*, and distally by the microsatellite *D5Mit72*.

The present work is focused on the detection and characterization of transcription units present within this cloned genomic domain. The first approach is a cDNA enrichment strategy based on the hybridization of cDNA probes to biotinylated genomic DNA in solution, with subsequent capture of the hybrid complexes on streptavidin-coated paramagnetic beads. cDNA libraries from E14 embryo or embryonic brain were used. After the enrichment procedure, cDNA clones that mapped to the original YAC clones were selected for further study. So far, homology searches failed to detect related sequences. Clones were assessed for cross-species homology, using zooblots, and for expression in the mouse embryo using Northern hybridization and/or RT-PCR. Several positive clones representing candidate genes are currently under investigation in order to isolate full-length cDNA clones.

As a second approach, exon trapping was recently initiated. Two YAC clones from the region of interest were subcloned into cosmids. Pools of six cosmids were then used to construct libraries in pSPL3. Putative exons are being sequenced and analyzed using the same criteria as for the cDNA selection method.

CONSTRUCTION OF A THREE MEGABASE CONTIG ON MOUSE CHROMOSOME 11 FOR POSITIONAL CLONING OF THE AMES DWARF LOCUS

Dawn Watkins-Chow, Marion S. Buckwalter, Michelle L. Roller, Frank J. Probst, Matthew M. Newhouse, Amy C. Lossie, and Sally A. Camper*, Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48019-0618 USA

The Ames dwarf gene, *df*, is involved in cell proliferation in the anterior pituitary gland. Homozygotes are one third to one half the size of their normal litter mates and lack thyroid stimulating hormone, growth hormone, and prolactin. Their hypocellular anterior pituitary glands lack three cell types; thyrotropes, somatotropes, and lactotropes. In order to clone the *df* gene, we have defined the map position precisely by segregation analysis and molecular mapping using over 1000 meioses from *M. castaneus* backcross and intercross progeny. Genetic mapping has eliminated several transcription factors expressed during pituitary ontogeny as candidates for the *df* locus, including *Pit-1* and thyrotrope embryonic factor (*Tef*). Genes encoding other pituitary transcription factors are currently being evaluated. The structural genes for the hormones affected in the dwarf mice and the genes encoding the receptors for growth hormone releasing factor (*Grfr*) and thyrotropin releasing hormone (*Trhr*) have also been eliminated. We have undertaken identification of the *df* gene by positional cloning. A chromosome walk was initiated by screening YAC libraries for markers located 0.2 and 1.0 cM proximal and distal to *df*, respectively. YAC ends were obtained by inverse PCR or by homologous recombination plasmid end rescue. Unique ends from the YACs were genetically mapped to confirm linkage to mouse Chr 11 and establish orientation relative to the centromere. End clones were used to rescreen the YAC library, and a contig consisting of over 25 YACs and 6 bacteriophage P1 clones has been assembled. Gene identification within the contig is in progress using pituitary cDNA selection. Although the non-recombinant interval around *df* is approximately 0.1 cM, physical mapping suggests this corresponds to 1.0 Mb. In contrast, the 1.0 cM interval just distal to *df* corresponds to 0.2 Mb, demonstrating variation in recombination rate across this chromosomal region. Genetic analysis of an equivalent number of progeny from back crosses lacking *df* suggests that recombination rates are not influenced by the *df* mutation itself. Synteny homology observed in the *df* region predicts that the human counterpart to *df* is located on human chromosome 5q. The cloning of the *df* gene will identify an important player in the process of pituitary cytodifferentiation and cell proliferation.
(NIH R01 HD30428)

Identification of transcribed sequences from distal mouse chromosome 11.

Miklós Péterfy, Tibor Gyuris, László Takács

Amgen, Inc. , Thousand Oaks, CA 91320

We are interested in the identification of novel genes with hematopoietic functions. A large number of cytokines, growth factors and growth factor receptors have been mapped to mouse chromosome 11. For our analysis we have chosen the chromosomal region proximal to the Tca-3 cytokine cluster (containing Tca-3, Mip1- α , Mip1- β , Sigje). This region shows synteny conservation to the human chromosomal regions 17p13.3 and 17q11-q21 which are also known to contain tumor suppressor genes. We have been collecting clones from a mouse YAC library using microsatellite markers and identifying transcribed sequences from the YAC clones.

TRANSCRIPT IDENTIFICATION IN THE INSERTION SITE REGION OF THE TRANSGENE INDUCED NEUROLOGICAL MUTANT A4. David Kohrman, Nicolas Plummer, Dan Burgess and Miriam Meisler, Department of Human Genetics, University of Michigan, Ann Arbor MI 48109-0618.

Transgenic line A4 demonstrates an autosomal recessive neuromuscular disorder with onset at postnatal day 10, muscle atrophy and death by 3 weeks of age. The insertion site has been mapped to distal mouse chromosome 15. Analysis of the interspecific backcross (C57BL/6J-A4 x CAST/Ei)F₁ X C57BL/6J demonstrated the following order and approximate distances: centromere - *D15Mit34* - 9 cM - transgene - 4 cM - *D15Mit16*, *D15Mit35*. The human homolog of this neuromuscular disease gene is located within the conserved linkage group on human chromosome 12.

In order to isolate the mutated gene responsible for the disorder, we constructed a cosmid library from transgenic DNA. Flanking clones from both transgene junctions were isolated by screening the library with a transgene derived probe. Sequencing revealed transgene insertion into a LINE element. Using primers from flanking sequences, P1 clones from the region have been isolated. Nonrepetitive subclones detecting transcripts on Northern blots have been identified, and several cDNA libraries have been screened. A small deletion (< 20 kb) has been detected at the insertion site. Several putative exons have been recognized by GRAIL-2. Characterization of transcripts from this region of mouse chromosome 15 will be described.

cDNA Sequencing and Mapping

Mapping mouse expressed sequences using single-strand conformation polymorphism analysis.

Kevin P. Brady¹, Todd Telle¹, T.J. Stevens², James Sikela², and David R. Beier¹
1 Division of Genetics, Brigham & Women's Hospital, Harvard Medical School, Boston, MA 02115; 2 University of Colorado Health Sciences Center, Denver, CO 80262

We have initiated a project to map expressed sequences in the mouse genome using the method of single-strand conformation polymorphism analysis (SSCP). Sequences from the 3' untranslated regions or intron regions of expressed mouse genes were PCR amplified and polymorphisms between inbred mouse strains were detected using SSCP analysis. These amplifiable sequences were obtained by accessing the Genbank sequence database or by sequencing individual clones from a directionally cloned mouse brain cDNA library. PCR primers were designed from these sequences to yield a product of optimal length for SSCP analysis. 3' untranslated and intron sequences were chosen because they are less subject to sequence conservation and more likely to exhibit a polymorphism between inbred mouse strains. The function of approximately 70% of the cDNAs sequenced to date was known based upon DNA sequence homology, but the biological function of the remaining 30% was completely unknown. Once a polymorphism between inbred strains was identified we used DNA from the corresponding recombinant inbred mapping panel to determine the strain distribution pattern of the polymorphism in this mapping panel. We then used the RI Manager program to rapidly map the polymorphic expressed sequence.

To date we have analyzed 202 expressed sequence primer pairs using this technique. Greater than 75% of the tested primers were polymorphic between inbred mouse strains. Of these polymorphic sequences approximately 25% exhibited polymorphisms, and were subsequently mapped, on BxD, AKxD, or AKxL crosses. The remaining 75% were polymorphic between B6 and *M. spretus*, and are being mapped using a (B6xSp)F₁ x Sp interspecific backcross generated at the Jackson Laboratory. We shall present our most recent findings and a more complete map of expressed mouse sequences.

Construction of a Human Adult Testis Expression Map

Michael H. Jones, Konstantinos N. Tirosvoutis, Philip M. Davey, and Nabeel A. Affara. University of Cambridge Department of Pathology, Tennis Court Rd., Cambridge, CB2 1QP, England.

The diversity of mRNA expression in the human testis presents a rich resource for the systematic identification of genes. As part of a positional candidate cloning strategy targeted at disorders of male fertility, we are systematically characterizing human adult testis transcripts. Ongoing, more than 600 clones from a 3' non-directional library have been sequence tagged. 70% of these clones show no matches (<60% identity) to sequences on data-bases whereas 3% are identical. Many of the transcripts represent genes with a role in meiosis and, significantly, 3% are zinc fingers. Southern analysis of monochromosomal hybrid panels has been used to map to single chromosomes more than two hundred of these transcripts. The clones are localized to all chromosomes with the exception of 22 and include one clone which is X-Y homologous. Cross hybridization with mouse and hamster DNA was evident for 70% of the clones demonstrating sequence conservation of these transcripts. 10% of the clones hybridized with numerous (six or more) fragments which hybridized specifically with more than one hybrid. These clones may represent conserved gene families. The expression of 12 of 60 transcripts is testis specific as determined by Northern analysis. In addition many of the clones detected a testis specific band which was usually smaller than a more widely expressed band.

Characterizing human brain ESTs of unknown function: expression patterns, mapping, and additional sequencing

D. Maglott¹, W. Nierman¹, T. Feldblyum¹, M. Adams², and S. Durkin¹. ¹American Type Culture Collection, Rockville, MD 20852-1776; ²The Institute for Genomic Research, Gaithersburg, MD 20878.

A large number of expressed sequence tags (ESTs) are being characterized by single-pass sequencing in laboratories worldwide. The majority of these sequences do not exhibit sequence homology with any known gene or gene product. If the function of these unknown genes is established, a significant amount of new basic biological information may be revealed.

Some of these "unknown" ESTs do match other single-pass sequences, and match more than one. These constitute an interesting class of "unknown" sequences, namely those that appear to be moderately expressed because of their frequent isolation.

Multiple approaches were applied to characterize a set of "abundant unknown" ESTs isolated at least once from the human brain. Somatic cell hybrid mapping panels and PCR were used to establish chromosome and in many cases subregional assignments. Expression patterns were revealed by Northern blots from multiple adult human tissues, and by two types of screening of cDNA libraries: PCR and sequence matches from single-pass sequencing projects using libraries of diverse origins. Also, assembling overlapping sequences from matching ESTs and additional sequencing of clones were used to expand the available sequence information. These sequence data were then analyzed for sequence motifs and other function-related characteristics of the putative gene products.

Human cDNA Mapping by Hybridization to High-density megaYAC Dot Blots.

Lucy Ling, Betty Borsody, JoAnn Dubois, Jonathan Norcross, Kathy Falls, Romina Bashirzadeh, Liam Haveran, Teresa Kanjuparamban, Ron Lundstrom, and Donald Moir,
Collaborative Research, Inc., Waltham, MA 02154

The emerging YAC-based physical map of the human genome provides a unique substrate for assignment of genes to chromosomal locations. The resulting gene map will be useful for positional cloning strategies and for improving the resolution of comparative mapping between organisms such as mouse and human. We are developing technology for high throughput, large scale mapping of human cDNAs by hybridization to high density filter grids of DNA from megaYAC clones. Our approach involves redundant pooling of megaYAC DNAs and of cDNA probes to reduce errors and save labor. For example, two 96-well source plates of megaYAC DNA are pooled into one 96-well target plate in a manner which represents each clone in three different ($r=3$) pools of six clones each ($p=6$). This reduces the number of required dots by a factor of two (p/r) but at the same time provides multiple signals for each authentic hybridization to a megaYAC. Probes are also pooled in a similar manner to reduce the number of hybridizations required. We have completed isolation of about 100 ug of total yeast DNA from each of 9,000 megaYAC clones (CEPH plates 887-980) to provide 3-fold coverage of the human genome. Despite the chimeric nature of many megaYACs, calculations indicate that 62% of cDNAs will be mapped unambiguously to the correct chromosomal location by hybridization to megaYACs representing 3 human genome equivalents. Multiple possible chromosomal locations, including the correct location, will be obtained for most of the remaining probes. Results from probing of filters containing DNA from 3,000 megaYAC clones (one genome equivalent) in a 3x3 array will be described. Random cDNA probes (provided by G. Lennon) were from a normalized human infant brain cDNA library prepared by Bento Soares (Columbia U.), gridded by Greg Lennon (Livermore), and sequenced by Charles Auffray (Gnthon). Probes for known members of multi-gene families included the heterotrimeric G-protein α -subunits and β -actin. Positive dot blot signals are translated into unique megaYAC addresses by computer, and genetic/physical map positions of resulting megaYACs are determined from the Gnthon QUICKMAP database and the Whitehead Institute/MIT Center for Genome Research data releases. Experiments were designed to address the following questions: (1) What is the rate and accuracy of cDNA mapping achievable with this approach? (2) How interpretable are results obtained with probes representing members of homologous multi-gene families? For example, can the multitude of hybridization signals obtained from such probes be disambiguated into unique megaYAC addresses? And, are signals resulting from hybridization to the authentic gene distinguishable from signals resulting from hybridization to related gene family members? (3) Is it possible to map cDNA probes containing repetitive elements by pre-reassociation of the probe with Cot-1 DNA prior to hybridization? (4) Can low-stringency inter-*Alu* element PCR products from megaYAC DNA be substituted for total yeast DNA from megaYAC clones as a substrate for cDNA mapping by hybridization? What fraction of cDNA probes find hybridization targets among the inter-*Alu* PCR products of megaYACs? Answers to these questions will determine the feasibility of this approach for rapid construction of a total human gene map.

**The IMAGE Consortium : A collaborative effort to
characterize arrayed cDNA libraries**

G. Lennon¹, C. Auffray², M. Boguski³, and C. Tolstochev³

¹Human Genome Center, L-452, Lawrence Livermore National Laboratories, Livermore CA 94550; ²GenExpress, Genethon, 1, rue de l'Internationale, Evry, France ; ³National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda MD 20894

We will describe the Integrated Molecular Analysis of Gene Expression (IMAGE) Consortium. IMAGE is a collaborative group of over 28 laboratories working on the characterization of clones from shared arrayed cDNA libraries. Information derived from the study of specific clones is shared through the use of a collaborative database maintained at the NCBI. This database periodically transfers data directly to the public domain database dbEST. Most work to date has focused on the normalized infant brain cDNA library (from Dr. M. Soares, Columbia Univ.) as arrayed at LLNL, and then replicated and distributed worldwide. We are currently arraying other high-quality normalized cDNA libraries, and invite the participation and support of others willing to abide by the Consortium guidelines.

Work of G.L. was performed under the auspices of the U.S. DOE by LLNL under contract no. W-7405-Eng-48.

Informatics

GENE EXPRESSION INFORMATION RESOURCE: PHASE 1

M. Ringwald, R. Baldock*, J. Bard#, D. Davidson*, J. Eppig, M. Kaufman#, J. Nadeau, J. Richardson, A. Smith, and L. Taylor. Jackson Laboratory, Bar Harbor; *MRC Human Genetics Unit, Western General Hospital, Edinburgh; #Dept. of Anatomy, Edinburgh University, Edinburgh.

We are developing a comprehensive information resource of gene expression information for embryonic development of the laboratory mouse. Ultimately, the resource will combine 3D graphical images of mouse embryos with detailed text descriptions to enable the storage and analysis of expression patterns. Integration with other relevant databases of DNA and protein sequence information, genetic and physical maps, and descriptions of deviant phenotypes and disease states will place the resource into the proper biological context. Query mechanisms will be provided for asking complex questions requiring information from each of these resources. Current work focuses on database design and data acquisition. We are taking advantage of the literature triage system developed by the Mouse Genome Informatics Project to identify published articles dealing with gene expression information from two gene families as a way of exploring database requirements. A key requirement for storage and analysis of expression data is use of a dictionary of standardized nomenclature for embryonic structures, a thesaurus of alternative terminology, and a fate table documenting the names for derived anatomical structures. User friendly interfaces for annotating gene expression data and prototypes of electronic data submission forms are under development. The ability to describe molecular and cellular interactions in time and space will have profound effects on storage and analysis of biological data. Long-range applications will be discussed with respect to the design of the gene expression information resource.

Supported by the Keck Foundation, the MRC, and the Deutsche Forschungsgemeinschaft.

Entering the vast territory of uncharted ESTs

Jean-Michel Claverie & David Lipman

National Center for Biotechnology Information, National Library of Medicine,
NIH, Bethesda MD 20894, USA

Following the concept of Expressed Sequence Tags (ESTs) originated by C. Venter et al. (1) an increasing number of publicly funded laboratories and private companies are now engaged in large scale efforts for the partial sequencing of cDNAs from a variety of organisms (vertebrates: *H.sapiens*, *M.musculus*; invertebrates: *C.elegans*; plants: *A.thaliana*, *O.sativa*, *Z.mays*; Yeast; and *P.falciparum*). EST sequencing is usually presented as a mean to speed up gene discovery, and provide new targets for drug development or the functional approach to human disease genes. In the present study, we first have determined carefully the proportion of "unknown" ESTs (with no similarity within the protein databases) and find this proportion to be very high, on the order of 69% (23115/33566 for dbEST v2.9). This result is compatible with our previous observation (2) (and theoretical prediction) that less than 50% of newly determined sequences have a relative in the current databases.

Up to now, the application of EST data has been limited to the 31% of cases where the encoded putative peptides exhibit significant similarities with already characterized proteins, thus providing a hint about their functions. Characterizing the remaining 69% of "unknown" cDNAs constitutes a major challenge in the Genome project. A first step in that direction, is to try to prioritize the "unknown" ESTs and define a subset of potentially interesting ones. This subset includes ESTs exhibiting strong conservation across species and/or constituting families of related genes (paralogs or orthologs). We will present a robust cross-organism comparison strategy that, despite the inherent redundancy and high error rate of the EST data, allowed us to discover 180 additional protein domains, previously uncharacterized. Only 52 (an upper bound estimate) of them fall into our previously defined category of ancestral proteins (predating the coelomate radiation, 600 Myear ago, (2)), thus in agreement with the notion that most of ancestral proteins are known (3). The 52 new ancestral protein domains are expected to be involved in basic, yet uncharacterized, cellular functions.

- (1) Adams M.D., Dubnick M., Kerlavage A.R., Moreno R.F., Kelley J.M., Utterback, T.R., Nagle, J.W., Fields C., & Venter J. (1992) Sequence Identification of 2,375 human brain genes. *Nature* 355, 632-634.
- (2) Claverie J.-M. (1993). Database of ancient sequences. *Nature* 364: 19-20
- (3) Green, P., Lipman, D., Hillier, L., Waterston, R., States, D. & Claverie, J.M. (1993). Ancient conserved regions in new sequences and the protein databases. *Science* 259: 1711-1716 .

New Feature of dbEST - NCBI's Public Repository of "Single Pass" Sequences and Mapping Data

Mark S. Boguski¹, Douglas E. Bassett Jr.^{1,2}, Carolyn Tolstoshev¹.

¹National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda MD 20894; ²Department of Molecular Biology and Genetics, Center for Medical Genetics, The Johns Hopkins University School of Medicine, Baltimore MD 21205.

dbEST has been operating for 2.5 years and usage has increased approximately ten fold over the past twelve months. It is now the second most frequently searched database at NCBI. dbEST release 2.22 (July 22, 1994) contains 40,218 partial DNA sequences, half of which are derived from human cDNA libraries with 20 other species making up the remainder; nearly 13,000 cDNAs are of plant origin. The dbEST schema was extended to include products of exon trapping or amplification as well as genetic and physical mapping data. A new email service has been implemented for retrieval of dbEST records based on chromosome map location.

World Wide Web/Mosaic access to dbEST is now available from the NCBI Home Page at <http://www.ncbi.nlm.nih.gov/>. The dbEST page provides information on the current release, a summary of other ways to access dbEST data, and instructions on submission of sequence data, map information and updates. In addition, one can now do complex (Boolean) text queries on dbEST annotations including, library and contact information, mapping and homology data. dbEST records are directly linked to their corresponding GenBank entries.

dbEST infrastructure is being used to support the IMAGE Consortium, a collaborative effort to characterize arrayed cDNA libraries (G. Lennon, C. Auffray et al.). dbEST will also be used to provide public access to mapping data produced under a grant from the National Center for Human Genome Research to identify human EST homologues of yeast genes and map these in the mouse and human genomes (P. Hieter, M. Boguski, F. Spencer, M. Goebel, R. Reeves).

INTEGRATED GENE RECOGNITION, MODELING AND DATABASE SEARCHING IN THE GRAIL AND GENQUEST ENVIRONMENTS. Richard J. Mural¹, Donna M. Buley¹, J. Ralph Einstein², Xiaojun Guan², Sheri Matis², Ying Xu², and Edward C. Uberbacher². ¹Biology and ²Engineering Physics and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831. E-mail: GRAILMAIL@ornl.gov

GenQuest is an integrated sequence comparison server which facilitates rapid and sensitive analysis of DNA and protein sequences by comparing them to existing sequence and motif databases. The system allows characterization of newly obtained sequences by homology based methods through accessing a number of databases using a number of comparison methods including: FASTA, BLAST and a parallel implementation of the Smith-Waterman algorithm which utilizes a specialized computing environment at the Oak Ridge National Laboratory.

GRAIL is a modular system which supports the recognition of gene features and gene modeling for the analysis and characterization of DNA sequences. The core of the GRAIL analysis engine continues to be based on multiple sensor-neural networks whose output is interpreted by a rule based system. In its original implementation, GRAIL analyzed sequence by determining the protein coding potential of a fixed sequence window (100 bases). As a result, short coding regions (<100 bases) were not reliably located. In the current version of the system a "windowless", length independent, version of the coding recognition module has been developed which recognizes 80% of the coding regions which are less than 100 bases in length. Overall about 92% of exons are found by this system. Using these accurately predicted exons as a starting point, a gene assembly program (GAP) uses heuristic methods and dynamic programming to construct gene models. The gene assembly process, which is constrained by maintaining the predicted reading frame for each newly added exon, can delete or insert exons if error conditions are encountered. This leads to an overall improvement in gene characterization, particularly in the recognition of very short exons. Genes modeled by this system contain about 94% of the all true exons, have only 7.4% false positive information, and 79% of predicted exons have both edges correctly defined with an additional 19% having at least one correct edge after gene modeling.

All of this information is presented to the user in graphic form in the X-window based client-server system XGRAIL. In addition the user can view the positions of a number of other features including poly-A addition sites, potential promoters, CpG islands and repetitive DNA elements using this system. XGRAIL also has a direct link to the genQuest server allowing the user to search various databases with the translations of predicted exons or gene models. Following an analysis session the user can use an annotation tool which is part of the XGRAIL system to generate a "feature table" describing the current sequence and its properties.

The e-mail version of GRAIL can be accessed at grail@ornl.gov and the e-mail version of genQuest can be accessed at Q@ornl.gov. Instructions can be obtained by sending the word "help" to either address. The XGRAIL and XgenQuest client software is available by anonymous ftp from [arthur.epm.ornl.gov](ftp://arthur.epm.ornl.gov) (128.219.9.76). Communications with the GRAIL staff should be addressed to GRAILMAIL@ornl.gov. (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.)

Alphabetical List of Abstract Authors

(Page numbers in bold indicate the abstract that author is presenting; page numbers in normal type indicate abstracts to which the author contributed.)

	<u>Page</u>		<u>Page</u>
Ackerman, Susan	17	Gilad, Shlomit	38
Bar-Shira, Anat	38	Gong, Weilong	48, 11
Bench, Anthony	26	Gruber, Christian	21
Biunno, Ida	46	Hadano, Shinji	20
Boguski, Mark	61, 58	Jackson, Cynthia	27
Bonaldo, M.	24	Jones, Michael	55
Brady, Kevin	54	Justice, Monica	7
Brent, Roger	8	Kohrman, David	53
Brookes, Anthony	29	Korn, Bernard	47, 36
Budarf, Marcia	11, 48	Krishnan, B. Raj	12
Burn, Timothy	32	Kurnit, David	15
Camper, Sally	51	Lambert, Catherine	50
Claverie, J-M	60	Landes, Greg	32
Connors, Timothy	32	Lau, Chris	33
den Dunnen, Johan	31	Lebedev, Yuri	14, 39
Drew, Paul	22	Lennon, Greg	41, 58, 14, 39
Fan, Wufang	35, 23	Liang, Teresa	30
Fontes, Michel	44	Maglott, Donna	56
Fujita, Ricardo	28	Meisler, Miriam	53
Gardiner, K.	43, 42		

	<u>Page</u>
Moir, Donald	57
Mural, Richard	62
Neri, Christian	13
Nguyen, Catherine	18
Nisson, Paul	34
Patanjali, S	25, 23, 38, 49
Peterfy, Miklos	52
Pietu, Genevieve	19
Rao, Srinivasa	9
Ringwald, Martin	59
Rommens, Johanna	40
Rotman, Galit	38
Savitski, Kinneret	38
Sanseau, Phillipe	36
Shiloh, Yossi	25, 38

	<u>Page</u>
Sirotkin, Howard	49
Smith, Carissa	37
Soares, Bento	24
Sundaresan, Vasi	10
Swaroop, Anand	28
Tagle, Danilo	38
Takacs, Laszlo	52
Tassone, Flora	43
Toniolo, Daniela	16, 45
Udar, Nitin	30
Volik, Stanislav	39, 14
Weissman, S.	23, 25, 35, 38, 42, 43, 49
Xu, Hongxia	42, 23, 25, 43

1994 Rooster

Susan Ackerman, Ph.D.
The Jackson Laboratory
600 Main Street
Bar Harbor, ME 04609
Tel: 207-288-3371 ext. 1615
Fax: 207-288-0370
sal@aretha.jax.org

Megan Adamson, M.D.
NIAAA, NIH
Lab of Neurogenetics
12501 Washington Avenue
Rockville, MD 20852
Tel: 301-496-2727
Fax: 301-443-8579

Anat Bar-Shira, Ph.D.
Department of Human Genetics
Sackler School of Medicine
Tel Aviv University
Ramat Aviv 69978
Israel
Tel: 972-3-640-8584
Fax: 972-3-640-7471

Anthony Bench, Ph.D.
Clinical Oncology and Radiotherapeutics Unit
MRC Centre
Hills Road
Cambridge CB2 2QH
United Kingdom
Tel: 44-223-402-362
Fax: 44-223-402-180
ajb@mrc-lmb.cam.ac.uk

Ida Biunno, Ph.D.
Consiglio Nazionale delle Ricerche
via Ampere 56
Milano
Italy
Tel: 39-2-7064-3391
Fax: 39-2-7063-6469

Mark S. Boguski, M.D., Ph.D.
National Center for Biotech. Info.
National Library of Medicine
Bldg 38A, Rm 85-810
8600 Rockville Pike
Bethesda, MD 20894
Tel: 301-496-2475
Fax: 301-480-9241
boguski@ncbi.nlm.nih.gov

Maria de Fatima Bonaldo, Ph.D.
Columbia University
722 West 168th Street, Unit #41
New York, NY 10032
Tel: 212-960-2315
Fax: 212-781-3577

Giuseppe Borsani, Ph.D.
Telethon Institute for Genetics and
Medicine
Via Olgettina 58
20132 Milano
Italy
Tel: 02-26434893
Fax: 02-26434771
gborsani@bcm.tmc.edu

Kevin Brady, Ph.D.
Division of Genetics
Brigham & Women's Hospital
Harvard Medical School
75 Francis Street
Boston, MA 02115
Tel: 617-732-7798
Fax: 617-232-4623
brady@rascal.med.harvard.edu

Miles B. Brennan, Ph.D.
Unit on Genomics
NIMH, IRP, CNG
9000 Rockville Pike
Bldg. 10, Rm 4N320
Bethesda MD 20892
Tel: 301-402-1769/496-0373 x250
Fax: 301-402-2140/402-6438
genunit@helix.nih.gov

MSC 4405
49 Convent Drive
Bldg 49, Rm B1EE16

Roger Brent, Ph.D.
Department of Molecular Biology
Wellman Bldg, 8th Floor
Massachusetts General Hospital
50 Blossom Street
Boston, MA 02114
Tel: 617-726-5925 6 5956
Fax: 617-726-6893
brent@opal.mgh.harvard.edu

Anthony J. Brookes, Ph.D.
MRC, Human Genetics Unit
Western General Hospital
Crewe Road
Edinburgh EH4 2XU
United Kingdom
Tel: 44-31-332-2471
Fax: 44-31-343-2620
tony@uk.ac.mrc.hgu.

Marcia L. Budarf, Ph.D.
Children's Hospital of Philadelphia
Division of Genetics
34th Street and Civic Center Boulevard
Philadelphia PA 19104
Tel: 215-590-3856
Fax: 215-590-3764
marcia@cit.med.upenn.edu

Timothy Burn, Ph.D.
Integrated Genetics
One Mountain Road
Framingham, MA 01701-9322
Tel: 508-872-8400
Fax: 508-620-1203
gmilleriglab@delphi.com

Sally A. Camper, Ph.D.
Department of Human Genetics
University of Michigan Medical School
Medical Science II M4708
Ann Arbor, MI 48109-0618
Tel: 313-763-0682
Fax: 313-763-3784
sally.camper@med.umich.edu

Settara Chandrasekharappa, Ph.D.
Laboratory of Gene Transfer
NCHGR, NIH
Bldg 49, Room 3A76
49 Convent Drive, MSC 4470
Bethesda, MD 20892-4470
Tel: 301-402-2344
Fax: 301-402-4929
chandra@helix.nih.gov

Jean-Michel Claverie, Ph.D.
NCBI, NLM, NIH
Building 38A, Room 8S804
8600 Rockville Pike
Bethesda MD 20894
Tel: 301-496-2475 33-91-71-78-96
Fax: 301-486-9241
jmc@ncbi.nlm.nih.gov

Timothy Connors
Integrated Genetics
One Mountain Road
Framingham, MA 01701-9322
Tel: 508-872-8400
Fax: 508-620-1203
gmilleriglab@delphi.com

Sevilla Detera-Wadleigh, Ph.D.
Clinical Neurogenetics Branch
NIMH/NIH
Bldg 10, Rm 3N218
Bethesda, MD 20892
Tel: 301-496-1641
Fax: 301-402-0859
sevilla@helix.nih.gov

Johan T. den Dunnen, Ph.D.
Department of Human Genetics
Leiden University
Wassenaarseweg 72
2333 Al Leiden
The Netherlands
Tel: 31-71-276105
Fax: 31-71-276075
ddunnen@ruly46.LeidenUniv.nl

Paul Drew, Ph.D.
Neuroimmunology Branch
NIH
Bldg. 10, Rm. 5B16
Bethesda, MD 20892
Tel: 301-496-0520
Fax: 301-402-0373
pddrew@nih.helix.gov

Brigitte Ecarot, Ph.D.
Genetics Unit
Shriners Hospital
1529 Cedar Avenue
Montreal, Quebec
Canada H3G 1A6
Tel: 514-842-5964
Fax: 514-842-5581
czeb@musica.mcgill.ca

Wufang Fan, Ph.D.
Human Genome Center
University of California
P.O. Box 808
LLNL, L952
Livermore, CA 94551-9900
Tel: 510-422-5712
Fax: 510-422-2282

Michel Fontes, Ph.D.
INSERM U406
BP 24
27, Bd Jean Moulin
13385 Marseille Cedex 5
France
Tel: 33-91-78-44-77
Fax: 33-91-80-43-19

Ricardo Fujita, Ph.D.
W.K. Kellogg Eye Center, #540
University of Michigan
1000 Wall Street
Ann Arbor, MI 48105
Tel: 313-936-8370
Fax: 313-936-2340
ricardo.fujita@um.cc.umich.edu

Katheleen Gardiner, Ph.D.
Eleanor Roosevelt Institute
1899 Gaylord Street
Denver CO 80206
Tel: 303-333-4515
Fax: 303-333-8423

Shlomit Gilad, M.Sc.
Department of Human Genetics
Sackler School of Medicine
Tel Aviv University
Ramat Aviv 69978
Israel
Tel: 972-3-640-8584
Fax: 972-3-640-7471

Weilong Gong, M.D.
Children's Hospital of Philadelphia
Division of Genetics
34th Street and Civic Center
Boulevard
Philadelphia, PA 19104
Tel: 215-590-3867
Fax: 215-590-3764

Antonio Gonzalez, Ph.D.
Instituto de Parasitologia y
Biomedicina
Calle Ventanilla, 11
18001 Granada
Spain
Tel: 34-58-203802
Fax: 32-58-203323
aaguilar@samba.cnb.uam.es

Christian E. Gruber, Ph.D.
Life Technologies, Inc.
8717 Grovemont Circle
Gaithersburg, MD 20877
Tel: 301-670-7714
Fax: 301-921-2116

Shinji Hadano, Ph.D.
IGSP/ERATO
Tokai University School of Medicine
Bohseidai, Isehara
Kanagawa 259-11
Japan
Tel: 81-463-91-4056
Fax: 81-463-91-4110
shinji@igsp1.med.u-tokai.ac.jp

Barbara Incerti, M.D.
Telethon Institute for Genetics and
Medicine
Via Olgettina 58
20132 Milano
Italy
Tel: 02-26434893
Fax: 02-26434771
borsang@dibit.hsr.it

Cynthia L. Jackson, Ph.D.
Department of Pathology
Rhode Island Hospital
593 Eddy Street
Providence, RI 02903
Tel: 401-444-4370
Fax: 401-444-8514

Michael H. Jones, Ph.D.
Cambridge University
Department of Pathology
Human Molecular Genetics Group
Tennis Court Road
Cambridge CB2 1QP
England
Tel: 44-223-333-709
Fax: 44-223-333-346
mj@mbuc.bio.cam.ac.uk

Monica J. Justice, Ph.D.
Division of Biology
312 Ackert Hall
Kansas State University
Manhattan, KS 66506-4901
Tel: 913-532-6164
Fax: 913-532-6653
mjj@ksu.edu

David C. Kohrman, Ph.D.
Department of Human Genetics
University of Michigan
4708 Medical Sciences II
Ann Arbor, MI 48109-0618
Tel: 313-763-5546
Fax: 313-763-9691
dave.kohrman@med.umich.edu

Bernard Korn, Ph.D.
German Cancer Research Center
Institute for Virus Research
D-6000 Heidelberg
Germany
Tel: 49-6221-423-409
Fax: 49-6221-423-454

B. Rajendra Krishnan, Ph.D.
Washington University School of Medicine
Div of Allergy & Immunology
Department of Internal Medicine
Box 8122
660 South Euclid
St Louis, MO 63110
Tel: 314-362-9038
Fax: 314-454-0486
krishnan@borcim.wustl.edu

David M. Kurnit, M.D., Ph.D.
Howard Hughes Medical Institute
Univ of Michigan Medical School
1150 W. Medical Center Drive
3520 MSRB I, Box 0650
Ann Arbor MI 48109-0650
Tel: 313-747-4747
Fax: 313-936-9353
david.kurnit@med.umich.edu

Catherine Lambert, Ph.D.
Department of Physiology
FUNDP School of Medicine
61, rue de Bruxelles
B-5000 Namur
Belgium
Tel: 32-81-72-42-74
Fax: 32-81-72-42-80
clambert@quick.cc.fundp.ac.be

Greg Landes, Ph.D.
Integrated Genetics
One Mountain Road
Framingham, MA 01701-9322
Tel: 508-872-8400
Fax: 508-872-9080

Yun-Fai Chris Lau, Ph.D.
Division of Cell Development Genetics
Department of Medicine
VA Medical Center, 111C5
Univ of California, San Francisco
4150 Clement Street
San Francisco, CA 94121
Tel: 415-476-8839
Fax: 415-502-1613
clau@itsa.ucsf.edu

Yuri B. Lebedev, Ph.D.
Shemyakin-Ovchinnikov Institute of
Bioorganic Chemistry
16/10 Miklukho-Maklaya
Moscow, 117871
Russia
Tel: 7-095-330-6992
Fax: 7-095-330-6538
yuri@humgen.siobc.msk.su

Greg Lennon, Ph.D.
Human Genome Center, L-452
Lawrence Livermore National Labs
Livermore CA 94551-9900
Tel: 510-422-5711
Fax: 510-422-2282
greg@mendel.llnl.gov

Teresa Liang
Department of Pathology
Center for the Health Sciences
10833 LeConte Ave.
Los Angeles, CA 90024
Tel: 310-825-7200
Fax: 310-825-7618

Donna R. Maglott, Ph.D.
American Type Culture Collection
12301 Parklawn Drive
Rockville MD 20852-1776
Tel: 301-231-5586
Fax: 301-770-1541
dmaglott@atcc.org

Miriam H. Meisler, Ph.D.
University of Michigan
Department of Human Genetics
4708 Medical Sciences II
Ann Arbor, MI 48109-0618
Tel: 313-763-5546
Fax: 313-763-9691
miriam.meisler@med.umich.edu

Donald T. Moir, Ph.D.
Collaborative Research, Inc.
1365 Main Street
Waltham MA 02154
Tel: 617-487-7979, ext 114
Fax: 617-891-5062
moir@cric.com

Richard J. Mural, Ph.D.
Biology Division
Oak Ridge National Laboratory
P. O. Box 2009
Oak Ridge TN 37831-8077
Tel: 615-576-2938
Fax: 615-574-1274
M9L@stc10.ctd.ornl.gov

Christian Neri, Ph.D.
Centre D'Etude du Polymorphisme
Humain (CEPH)
27, rue Juliette Dodu
75010 Paris
France
Tel: 33-1-42-49-98-59
Fax: 33-1-40-18-01-55
neri@cephb.fr

Catherine Nguyen, Ph.D.
Centre d'immunologie de Marseille-
Luminy
INSERM/CNRS
Parc Scientifique de Luminy-Case 906
13288 Marseille Cedex 9
France
Tel: 33-91-26-94-96
Fax: 33-91-26-94-30

Paul E. Nisson, Ph.D.
Life Technologies, Inc.
8717 Grovemont Circle
P.O. Box 6009
Gaithersburg MD 20877
Tel: 301-670-7774
Fax: 301-921-2116
nisson@biotechnet.com

B.A. Oostra, Ph.D.
Department of Clinical Genetics
Erasmus University Rotterdam
P.O. Box 1738
3000 DR Rotterdam
The Netherlands
Tel: 31-10-4088136
Fax: 31-10-4087200
oostra@kgen.fgg.eur.nl

Sankhavaram R. Patanjali, Ph.D.
Yale University School of Medicine
Boyer Ctr for Mol Medicine, Rm 333
295 Congress Avenue
New Haven, CT 06510
Tel: 203-737-2277
Fax: 203-737-2286
sankhavaram@biomed.med.yale.edu

Miklos Peterfy, Ph.D.
AMGEN, Inc.
MS 8-1-A236
1840 Dehavilland Drive
Thousand Oaks, CA 91320-1789
Tel: 805-447-6596
Fax: 805-498-8674
mpeterfy@amgen.com

Hemant S. Pawar, Ph.D.
W.K. Kellogg Eye Center
University of Michigan
1000 Wall Street
Ann Arbor, MI 48105
Tel: 313-764-0059
Fax: 313-936-2340
hemant.pawar@um.cc.umich.edu

Genevieve Pietu, Ph.D.
Genexpress
Genethon, BP60
91002, Evry
France
Tel: 33-1-69-47-28-30
Fax: 33-1-60-77-86-98

Srinivasa K. Rao, Ph.D.
Department of Surgery
Long Island Jewish Medical Center
New Hyde Park, NY 11042
Tel: 718-470-7060
Fax: 718-343-3429
kalasapu@aecom.yu.edu

Vincent Raymond, M.D., Ph.D.
Laboratory of Molecular Endocrinology
Centre de Recherche du Chul
2705, Laurier Blvd
Quebec, G1V 4G2
Canada
Tel: 418-654-2296
Fax: 418-654-2761
raymond@adn.chul.quebec.ca

Martin Ringwald, Ph.D.
The Jackson Laboratory
600 Main Street
Bar Harbor, ME 04609
Tel: 207-288-3371
Fax: 207-288-2516
ringwald@informatics.jax.org

Johanna Rommens, Ph.D.
Department of Genetics
The Hospital for Sick Children
555 University Avenue
Toronto, Ontario M5G 1X8
Canada
Tel: 416-813-7095
Fax: 416-813-4931

Galit Rotman, Ph.D.
Department of Human Genetics
Sackler School of Medicine
Tel Aviv University
Ramat Aviv 69978
Israel
Tel: 972-3-640-8584
Fax: 972-3-640-7471

Phillipe Sanseau, Ph.D.
GLAXO Research and Development
Ltd.
Molecular Pathology Dept
Building 40, Room 118
Greenford Road
Greenford, Middlesex UB6 OHE
United Kingdom
Tel: 44-81-966-2451
Fax: 44-81-422-2054
PS14446@ggr.co.uk@relay

Kinneret Savitski, M.Sc.
Department of Human Genetics
Sackler School of Medicine
Tel Aviv University
Ramat Aviv 69978
Israel
Tel: 972-3-640-8584
Fax: 972-3-640-7471

Yossi Shiloh, Ph.D.
Department of Human Genetics
Sackler School of Medicine
Tel Aviv University
Ramat Aviv 69978
Israel
Tel: 972-3-640-9760
Fax: 972-3-640-7471
yossih@ccsg.tau.ac.il

Pierre M. Sinet, M.D., Ph.D.
Laboratoire de Biochimie Genetique
CNRS1335,
Hopital Necker-Enfants Malades
149, rue de Sevres
75743 Paris Cedex 15
France
Tel: 33-1-43-06-92-64
Fax: 33-1-42-73-06-59
zrahm@citi2.fr

Guru Siradanahalli, Ph.D.
NCHGR/NIH
Room 3C28, Bldg 49
Bethesda, MD 20814
Tel: 301-402-2345
Fax: 301-402-4929
guru@helix.nih.gov

Howard Sirotkin, Ph.D.
Albert Einstein College of Medicine
Department of Molecular Genetics
1300 Morris Park Ave
Bronx, NY 10461
Tel: 718-430-2848
Fax: 718-430-8778
hsirotki@aecom.yu.edu

Carissa M. Smith
Washington Univ School of Medicine
Department of Genetics
Box 8232
4566 Scott Avenue
St. Louis, MO 63110
Tel: 314-362-1714
Fax: 314-362-7855
carissa@sequencer.wustl.edu

Marcelo Bento Soares, Ph.D.
Department of Psychiatry
Columbia University, Unit #41
722 West 168th Street
New York NY 10032
Tel: 212-960-2313
Fax: 212-781-3577
cuc@cuccfa.ccc.columbia.edu

Vasi Sundaresan, Ph.D.
MRC Clinical Oncology &
Radiotherapeutics Unit
MRC Centre
Hills Road
Cambridge, CB2 2QH
England
Tel: 44-223-245133
Fax: 44-223-402180
vsl@mrc-lmb.cam.ac.uk

Anand Swaroop, Ph.D.
University of Michigan 540 Kellogg Eye Center
W.K. Kellogg Eye Center Univ of Michigan
~~3481 Yellowstone~~ 1000 Wall Street
Ann Arbor MI 48105
Tel: 313-936-9547
Fax: 313-936-2340
anand.swaroop@med.umich.edu

Danilo A. Tagle, Ph.D.
Laboratory of Gene Transfer
Ntl Ctr for Human Genome Research
National Institutes of Health
Bldg 49, Room 3A18
49 Convent Drive MSC 4470
Bethesda, MD 20892-4470
Tel: 301-402-2530
Fax: 301-402-4929
datagle@helix.nih.gov

Laszlo Takacs, M.D., Ph.D.
AMGEN, Inc.
M/S 8-1-A-236
1840 Dehavilland Drive
Thousand Oaks, CA 91320-1789
Tel: 805-447-2214
Fax: 805-498-8674
ltakacs@amgen.com

Flora Tassone, Ph.D.
Eleanor Roosevelt Institute
1899 Gaylord Street
Denver, CO 80206
Tel: 303-333-4515
Fax: 303-333-8423

Daniela Toniolo, Ph.D.
I.G.B.E., CNR
Via Abbiategrasso 207,
27100 Pavia
Italy
Tel: 39-382-546340
Fax: 39-382-422286
toniolo@ipvgbe.unipv.it

Nitin S. Udar, Ph.D.
Department of Pathology
School of Medicine
UCLA
10833 LeConte Ave
Los Angeles, CA 90024
Tel: 410-825-7200
Fax: 410-206-5178
nudar@pathology.schmed.ucla.edu

Jean Vacher, Ph.D.
Clinical Research Institute of Montreal
110 Pins Avenue West
Montreal, Quebec H2W 1R7
Tel: 514-987-5734
Fax: 514-987-5675

Stanislav V. Volik, Ph.D.
Shemyakin-Ovchinnikov Institute of
Bioorganic Chemistry
16/10 Miklukho-Maklaya
Moscow, 117871
Russia
Tel: 7-095-330-6992
Fax: 7-095-330-6538
stas@humgen.siohc.msk.su

Sherman M. Weissman, M.D.
Yale University School of Medicine
336 BCMM
295 Congress Avenue
New Haven CT 06510
Tel: 203-737-2282
Fax: 203-737-2286
weissman@biomed.med.yale.edu

Guoqiang Xing, Ph.D.
Clinical Neurogenetics Branch
NIMH/NIH
Building 10, Room 3N218
Bethesda, MD 20892-1274
Tel: 301-496-1641
Fax: 301-402-0859

Hongxia Xu, M.D., Ph.D.
Yale University School of Medicine
Department of Genetics
P.O. Box 208005
New Haven, CT 06520
Tel: 203-737-2277
Fax: 203-737-2286
weissman@biomed.med.yale.edu

Huanming Yang, Ph.D.
Department of Medical Genetics
Peking Union Medical College
5 Dong Dan San Tiao
Beijing, 100005
China
Tel: 86-1-522-9323
Fax: 86-1-512-4876
nudar@pathology.schmed.ucla.edu

Takeo Yoshikawa, M.D.
Clinical Neurogenetics Branch
Bldg 10, Rm 3N218
NIMH/NIH
Bethesda, MD 20892
Tel: 301-496-1641
Fax: 301-402-0859