Send to OSTI

CONF-9310297--Absts.

DOE/ER/61670-1

# Workshop on Identification of Transcribed Sequences
New Orleans, LA, October 2-4, 1993

Final Report

for Period August 15, 1993 - August 14, 1994

Katheleen Gardiner

Eleanor Roosevelt Institute
1899 Gaylord Street
Denver, CO 80206

November 1993

Prepared for
THE U.S. DEPARTMENT OF ENERGY
AGREEMENT NO. DE-FG02-92ER61490

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or 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.

DOE/ER/61670--1

GRANT #: DE-FG03-93ER61670
TITLE: Third International Workshop on the Identification of Transcribed Sequences
P.I.: Katheleen Gardiner, Ph.D.
Eleanor Roosevelt Institute for Cancer Research, Inc.
1899 Gaylord Street
Denver, CO 80206
GRANT DATES: 08/15/93 - 08/14/94

Abstract

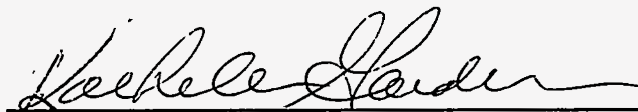
Tremendous progress has been made in the construction of physical and genetic maps of the human chromosomes. The next step in the solving of disease related problems, and in understanding the human genome as a whole, is the systematic isolation of transcribed sequences. Many investigators have already 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 Third Annual Workshop on the Identification of Transcribed Sequences was held in New Orleans, Louisiana, October 2-4, 1993, and had 58 attendees. An important focus of this meeting was discussion of the relative strengths and weaknesses of the more popular techniques and potential complementaries among them. This has permitted some direct determination of the robustness and portability of particular methods, as well as some evaluation of their relative efficiencies for mapping projects of different scales. Variations and improvements on these techniques, as well as the development of alternative approaches also provided important topics for discussion.

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:



11/23/94
Date

Katheleen Gardiner, Ph.D.
Institute Fellow
Principal Investigator

Approved by:

11/23/94
Date

David Patterson, Ph.D.
President

## DISCLAIMER

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

# 3rd International Workshop on the Identification of Transcribed Sequences

New Orleans, LA, USA
October 2-4, 1993

Organizers
Katheleen Gardiner, Ph.D.
Eleanor Roosevelt Institute, Denver

Ute Hochgeschwender, M.D.
NIMH, Bethesda

Gail Bruns, M.D., Ph.D.
Children's Hospital, Boston

Sponsored by the
Department of Energy
and the
National Institutes of Health

DISCLAIMER

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

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

**INTERNATIONAL WORKSHOP ON THE IDENTIFICATION OF TRANSCRIBED SEQUENCES**
**October 2-4, 1993 New Orleans, LA**

## AGENDA

**Saturday, October 2, 1993**

| | |
|---|---|
| **8:30 am** | Arrive |
| **8:45 am** | Introduction (Katheleen Gardiner) |
| **9:00-11:00 am** | **METHODS FOR GENE IDENTIFICATION**<br>Chair: Ute Hochgeschwender and Paul Nisson |
| Sherman Weissman, M.D. | Motif libraries and hybridization selection techniques. |
| Danilo Tagle, Ph.D. | Magnetic bead capture and CpG island rescue PCR: Two strategies for isolating transcribed sequences from large genomic regions. |
| Ruchira Das Gupta, Ph.D. | An integrated method for isolating and mapping human transcribed sequences. |
| Michel Fontes, Ph.D. | Isolation of expressed sequences localized in a YAC by a "physical trapping" strategy (direct cDNA selection). |
| Bento Soares, Ph.D. | Subtractive hybridizations of tagged normalized cDNA libraries and selection of chromosome-specific cDNAs. |
| Anthony Brookes, Ph.D. | Coincident sequence cloning; Gene identification and genome analysis. |
| **11:00-11:30** | **BREAK** |
| **11:30-1:00 pm** | **METHODS FOR GENE IDENTIFICATION (continued)** |
| P. Jagadeeswaran, Ph.D. | Novel strategy for isolating unknown coding sequences from genomic DNA by generating genomic-cDNA chimeras. |
| John C. Hozier, Ph.D. | Preparative in situ hybridization: Isolation of cDNA clones according to chromosome location. |
| Anand Swaroop, Ph.D. | Two methods for identifying cDNA clones from a specific genomic region. |

| | |
|---|---|
| William Nierman, Ph.D. | PCR mapping of human cDNA clones. |
| Donald Moir, Ph.D. | Mapping cDNAs by hybridization to gridded arrays of DNA from YAC clones. |
| David Beier, M.D., Ph.D. | Mapping cloned sequences in the mouse by SSCP analysis of RI strains and interspecific crosses. |

## Sunday, October 3, 1993

| | |
|---|---|
| **8:00-10:00 am** | **TRANSCRIPTIONAL MAPS [large chromosomal regions]**<br>Chair: Gail Bruns and Johanna Rommens |
| Annemarie Poustka, Ph.D. | Construction of integrated physical and transcription maps. |
| Daniela Toniolo, Ph.D. | Identification of genes and construction of a transcriptional map in XQ28. |
| Katheleen Gardiner, Ph.D. | Towards a transcriptional map of human chromosome 21. |
| Marie-Laure Yaspo, Ph.D. | Towards a transcriptional map of human chromosome 21. |
| Andrew Peterson, Ph.D. | Approaches towards comparative transcript maps: Mouse chromosome 16 vs human chromosome 21. |
| Johanna Rommens, Ph.D. | Towards a transcriptional map of the q22 region of chromosome 7. |
| **10:00-10:30 am** | **BREAK** |
| **10:30-12:00 PM** | **TRANSCRIPTIONAL MAPS [confined regions]** |
| Gregory Dolganov, Ph.D. | Organization of the genes in human 5q23-31. |
| T. Conrad Gilliam, Ph.D. | Gene identification in a region on chromosome 5q13 containing low-copy repeat sequences. |
| Hubert Smeets, Ph.D. | Identification of disease genes in Xq28. |
| Jerome Gorski, Ph.D. | Use of evolutionarily conserved sequences and cDNA selection to isolate regional cDNA clones from Xp11.21. |
| Jozef Gecz, Ph.D. | YAC-mediated direct cDNA selection transcriptional mapping of 1 Mb DXS56-PGK1 region (Xq13.3) using human and mouse cDNAs. |

| | |
|---|---|
| Roderic Guigo, Ph.D. | An integrated system to identify genes in DNA sequences. |
| Richard Mural, Ph.D. | Enhancement to Grail and the introduction of a client-server version for gene assembly and annotation. |
| J-M. Claverie, Ph.D. | Similarity search methods for the analysis of real and putative transcribed sequences. |
| C. Tolstoshev, Ph.D. | dbEST- Database for expressed sequence tags. |
| **11:00-11:30 am** | **BREAK** |
| **11:30-1:00 PM** | **DISCUSSION** |
| **2:00 pm** | **ADJOURNMENT** |

**Isolation of cDNA's from YAC's mapping to the critical region of the 5q-chromosome in pre-leukaemia.**

**Boultwood J, Fidler C, Cotter F, Wainscoat JS.**

LRF Molecular Haematology Unit, Department of Haematology, John Radcliffe Hospital, Headington, Oxford, U.K. LRF Centre for Childhood Leukaemia at the Institute of Child Health, London, U.K.

We have used molecular mapping techniques to identify the critical region of gene loss of the 5q-chromosome in pre-leukaemia. The critical region is localised to an approximately 5.6 mbp region between FGFA and the NKSF1 genes. We are constructing a YAC contig encompassing the entire critical region and are using these YAC clones to isolate coding sequences. We are using the direct selection method of Lovett et al (PNAS, 88, 9628, 1991) which involves hybridizing cDNA to genomic DNA and eluting the specifically bound selected cDNA. A number of sequences have been isolated from our YACs and we are currently involved in the characterization of these products. The molecular analysis of both candidate and newly identified genes mapping within the critical region should identify the putative myeloid tumour suppressor gene on 5q.

**IDENTIFICATION AND ISOLATION OF GENES INVOLVED IN VCFS: A CHROMOSOME 22 MICRODELETION SYNDROME** Marcia.L. Budarf[1,2], Scott Baldwin[1], Deborah Driscoll[1], Bruce Roe[3] and Beverly S. Emanuel[1,2]. [1]Children's Hospital of Philadelphia, the [2]Human Genome Center for Chromosome 22, Philadelphia, Pennsylvania, [3]University 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 dysmorphia, conotruncal cardiac malformations and learning disabilities. It is possible to mimic many features of VCFS in animal embryos by ablation of cephalic neural crest cells or treatment with teratogens early in development. In humans, however, the majority of VCFS patients (>80%) studied have been shown to have micro-deletions of 22q11.2. This suggests the presence of a gene(s) in the deleted region responsible for craniofacial and cardiac morphogenesis. The minimal region of overlap between VCFS patient deletions is large, suggesting that the loss of function of more than a single gene is required for the pathogenesis of this disorder.

A pulsed-field gel map of the 1.6 Mb critical region has been completed. Several of the probes mapping to the minimal deleted region contain CpG rich islands. We have successfully used two such probes to screen human cDNA libraries. Probe N41 detected a 1.5 kb cDNA clone in a human fetal liver library and three additional overlapping clones were identified in a fetal brain library. They have been completely sequenced and a search using BLASTX showed significant similarity to the amino acid sequence of a yeast cell adhesion glycoprotein. Northern blot analysis detects a 4.5 kb transcript in several human tissues. The human N41 cDNA cross hybridizes strongly with mouse genomic DNA and detects a similar sized transcript on Northern blots of RNA from 11.5 day mouse embryos. The N41 cDNA was used to screen a mouse cDNA library and a 2.1 kb mouse homologue has been isolated. The mouse homologue maps back to the VCFS critical region by Southern hybridization to a human chromosome 22 somatic cell hybrid mapping panel. Sequence analysis demonstrates an 82% homology between the human and mouse cDNA. *In situ* hybridization to whole mount 8-9 day mouse embryos using the mouse N41 cDNA shows enriched expression in the neural crest cell population.

Using cosmids containing the N25 NotI linking clone to directly probe the same cDNA libraries, a second gene has been isolated. A single 2.2 kb clone was determined to be specific after secondary screens, has been completely sequenced and does not have any obvious homology to other sequences in the data bases. Northern blot analysis of several adult tissues reveals a 5.5 kb transcript only in skeletal muscle. Four additional N25 cDNA clones have been isolated from an adult skeletal muscle cDNA library using the original cDNA clone as a probe including one with a poly A tail.

Over 30 YACs and 200 cosmids have been isolated using probes from the 22q11 commonly deleted region. These reagents are being used to isolate cDNAs by direct selection with biotinylated genomic DNA. We have demonstrated an enrichment of several thousand fold for the N41 cDNA. In addition, we are performing large-scale sequencing and computer analysis to identify potential coding regions. Using this approach, we have completed the sequence of a 45 kb N25 cosmid which contains the N25 NotI site. Additional sequence analysis has also been performed on the anonymous DNA marker pH160b. pH160b maps to the central portion of the VCFS critical region, contains a CpG island and detects cross-hybridizing sequences in several different species. The region of this clone which cross-hybridizes most strongly was subcloned and completely sequenced. The sequence was analyzed for open reading frames using GenLang (David Searls, personal comm.). Two potential open reading frames were identified and a BLASTX search revealed that one of the putative exons had sequence similarity to the Brachyury gene. We are now developing PCR primers for RT-PCR to determine tissue specific expression for pH160b.

## An Integrated Method for Isolating and Mapping Human Transcribed Sequences

Ruchira Das Gupta[1], Howard Sirotkin[1], Satish Parimoo[2], Sherman Weissman[2], Arthur Skoultchi[3], Raju Kucherlapati[1].
[1]Dept. of Molecular Genetics, [3] Dept. of Cell Biology, Albert Einstein College of Medicine, Bronx, NY; [2] Boyer Center for Molecular Medicine, Yale University School of Medicine, New Haven, CT.

We have developed an integrated method for identifying and mapping transcribed sequences in particular regions of chromosomes. In this method, random hexanucleotide primed tissue specific or comprehensive cDNA libraries are hybridized with isolated YAC DNA or with total genomic DNA from yeast containing a YAC to obtain a YAC-specific cDNA minilibrary. The members of this library are cloned and sequenced to generate ESTs. The affinity purified cDNAs are cloned into a yeast chromosome fragmentation vector and used to transform yeast containing the YAC. Exons within the targeting cDNA are expected to recognize their corresponding exons on the YAC and undergo homologous recombination, generating fragmented YACs. Each fragmented site represents the location of an exon. The common vector sequence present in truncations allow recovery of DNA sequence flanking the recombination junction, permitting access to intron-exon junctions and 5' or 3' regulatory regions.

The feasibility of targeting individual exons was demonstrated with a human X-chromosome YAC containing the clotting factor IX (F9) gene. We have used this method to identify several cDNA fragments corresponding to YACs from human chromosomes 6 and 22. Sequencing individual cDNAs permitted assembly of the cDNA fragments into overlapping contigs. Chromosome fragmentation with individual cDNA clones has allowed us to make gene maps. Comparison of one of these gene maps with that generated by conventional restriction enzyme digestion methods allowed us to conclude that the maps generated by this method are accurate We are currently planning to extend this method to rapidly isolate genes for larger parts of human chromosomes.

**Organization of the genes in human 5q23-31.**

Burr John[1], Mike Lovett[2], Dennis LePaslier[3], and Gregory Dolganov[1].

(1) Genelabs Technologies, Inc.; (2) The Eugene McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center at Dallas; (3) Centre d'Etude du polymorphisme Humain, Paris.

The 5q23-31 region of human chromosome 5 encodes an unusually large number of growth factors, growth factor receptors, transcription regulatory factors and hormone/neurotransmitter receptors, including IL3, GM-CSF, IL4, IL5, IL9, IL13, IRF1, TCF7, CDC25c, EGR1, CD14, FGFA, GRL1, etc. Several disease genes have also been mapped to this region, including 5q-syndrome, an autosomal form of hereditary deafness, Treacher-Collins syndrome, Diastrophic dysplasia, etc. This locus is expected to contain roughly 1,000 genes, although less than 2% of the expected number of genes have been identified. Current work has been focused on the development of framework physical and transcriptional maps of 5q23-31 (about 25 Mb), by making contigs of overlapping YAC and chromosome 5-specific cosmid clones, and identifying genes within the contigs. Primary cDNA pools, including human fetal brain, adult bone marrow and a mixture of several activated T and B-cell lines, clones, and lymphomas with specific tags to track their origin, have been constructed and used in direct selection with expressed sequences within 5q23-31. The selected cDNAs from the IL13-IL4-IL3 gene area (1.1 Mb) have been analyzed by hybridization to arrayed genomic clones and known cDNAs. Now they are being sequenced and compared with databases, and assessed for temporal and spatial expression patterns. These data will be incorporated into the existing framework physical map of the region.

**Using Defined Anatomical Regions and Single Cells to Characterize Expressed Sequences**

Jim Eberwine
University of Pennsylvania

Cellular identify is dictated by the cadre of genes that are expressed. Understanding the mechanisms by which regulation of the coordinate expression of multiple genes occurs is required to be able to meaningfully manipulate physiological processes. Important to this task is identification of the expressed sequences which are regulated by any given physiological manipulation. In neurons of the central nervous system this identification not only requires information about levels of mRNA in the soma of a neuron but also that in the processes. Using *in situ* transcription, aRNA amplification and differential display we have characterized classes of mRNAs which are present in the neuronal cell body as well as those present in its cognate cell processes. The relationship of mRNA level changes and sub-cellular transport to synaptic plasticity will be discussed during the presentation.

Identification of Expressed Sequences on Human Chromosome 9q32-34

Jeffrey D. Falk, Hiroshi Usui and J. Gregor Sutcliffe, Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037.

To identify the gene for the neuromuscular disease idiopathic torsion dystonia, DYT1, as well as other neurologically important genes localized in the human 9q32-34 region, we are identifying expressed sequences within chromosome-specific libraries and elucidating their patterns of tissue expression. Candidates for the dystonia gene are being assessed based on their expression in dystonic target tissues and their localization with respect to the DYT1 locus. We have assembled an arrayed collection of 3,000 9q32-34-specific clones which have been screened with rat cDNA probes from various brain regions and peripheral tissues to elucidate the tissue expression patterns of the genes corresponding to each clone. This screening has identified 141 genes within our collection, 33 of which are expressed in a brain-specific manner. Screening with a caudate cDNA probe from which cerebellar cDNA sequences have been subtracted has identified several additional clones containing genes expressed specifically within the brain. These expressed sequences are currently being mapped with respect to YAC and cosmid clones that are proximal to the DYT1 locus. A PCR-based screening procedure is also being used to enable us to identify 9q32-34-specific genes that are transcribed at levels too low for our current screening methods to detect. These techniques should allow us to determine the tissue expression patterns of nearly all the genes in this region, ultimately enabling us to identify the dystonia and other brain-specific genes.

**Towards a transcriptional map of human chromosome 21.**

K. Gardiner[1], H. Xu[3], W. Bonds[3], F. Tassone[2], S. Parimoo[3], R. Sivakamasundari[3], F. Hisama[3], A. Rynditch[4] and S. Weissman[3]. [1]Eleanor Roosevelt Institute, Denver CO, USA; [2]Institute of Human Genetics, "A Gemelli" School of Medicine, Rome, Italy; [3]Yale University School of Medicine, New Haven CT, USA; [4]Institute of Molecular Biology and Genetics, Kiev, Ukraine.

A transcriptional map of human chromosome 21 is of interest both for studies on human genome organization and for the identification of genes relevant to Down Syndrome. Towards these ends, a gene isolation effort has been implemented based on the technique of cDNA hybrid selection from YAC clones.

For cDNA selection, mixtures of cDNAs from fetal brain, thymus, spleen, liver, testes and whole fetus, as well as neuroblastoma and neuroglioma cell lines, are being used to construct individual selected cDNA libraries from each of approximately 70 YAC clones, comprising an essentially complete contig of the long arm. Initial detailed characterization of the selected libraries has focused on 4 YACs comprising approximately 2 Mb: a >800 kb YAC (containing the APP gene) from 21q21, 600 kb (D21S54) from q22.1, 300 kb (containing the ERG gene) in q22.2, and 200 kb (D21S171) near the telomere. YACs are non-chimeric based on FISH and/or end cloning. Selected cDNAs so far have included controls (APP and ERG), as well as novel zinc finger genes, a cDNA corresponding to D21S60, and several previously unknown genes. Characterization includes sequence analysis with GenBank and Prosite, mapping within the YAC clones to estimate gene size and CpG island association and screening of long insert cDNA libraries.

**USE OF EVOLUTIONARILY CONSERVED SEQUENCES AND cDNA SELECTION TO ISOLATE REGIONAL cDNA CLONES FROM Xp11.21**

JL Gorski[1,2], EN Burright[1], NG Pasteris[1], and N Karuppiah[2]. [1]Departments of Human Genetics and [2]Pediatrics, University of Michigan, Ann Arbor, 48109 USA

Several disease genes implicated in the differentiation of the neural crest have been mapped to the proximal short arm of the human X chromosome. Among these loci, incontinentia pigmenti type 1 (IP1) and Aarskog syndrome (facio-genital dysplasia; FGDY) have been localized to region Xp11.21. As part of a systematic effort to clone the IP1 and FGDY loci, we examined regional YAC inserts to identify transcribed sequences. A 1.5 Mb YAC contig spanning the IP1 and FGDY disease-specific X-chromosomal translocation breakpoints has been constructed. A restriction map of YACs spanning the disease-specific breakpoints identified at least six potential HTF islands. Bacteriophage subclones containing two of the HTF islands have been used to identify evolutionarily conserved sequences. Each conserved sequence has been successfully used to isolate cDNA clones from a human fetal brain cDNA library; a 2.4 kb cDNA clone was found to span a genomic distance in excess of 200 kb, a 1.1 kb cDNA clone spanned at least 80 kb of genomic DNA. To isolate additional regional transcripts, a 700 kb YAC encoding the 2.4 kb cDNA clone was used for cDNA selection. By immobilizing YAC DNA to a membrane, a single round of selection yielded an estimated $10^3$ fold enrichment. Approximately 3000 independent selected cDNA clones were isolated and arrayed for analysis. Of these, at least four unique regional cDNA clones have been identified.

PREPARATIVE IN SITU HYBRIDIZATION: ISOLATION OF cDNA CLONES ACCORDING TO CHROMOSOME LOCATION.

John C. Hozier, Regina Graham, Theresa Westfall, and Lisa M. Davis. Applied Genetics Laboratories, Melbourne, FL.

We have developed a new procedure, which combines *in situ* hybridization and chromosome microdissection in a unique way to isolate cDNA clones according to chromosome location. This procedure combines the best of related strategies for isolating cDNAs: selection of coding sequences directly by genomic DNA, without the need for cloning the chromosomal region first, as well as isolating cDNAs directly for characterization, without the need for subsequent cytogenetic mapping. cDNA populations with PCR primer binding sites ligated to their termini are hybridized to metaphase chromosome spreads under conditions typically used to hybridize complex probes, such as "painting" probes, except that the cDNAs are not labeled with biotin or any other "reporter" molecule. After hybridization, the chromosomes are washed free of non-specifically hybridized cDNAs, and the chromosomes are banded for identification. The chromosome band of interest is then dissected and the hybridized cDNAs are recovered by amplification with the PCR. Using this procedure, we have prepared a mouse liver cDNA library specific to chromosome band 4C3-7, biotin labeled the library, rehybridized to mouse chromosomes, and detected signal at 4C3-7. In addition, we have recovered the cDNA for human beta-2-microglobulin from a single dissection of chromosome 15 hybridized with a placenta cDNA library, and we have prepared a fetal brain cDNA library specific to 11p13. We will present data demonstrating successful uses of this gene finding strategy, and discuss additional advantages and applications of the procedure.

**Novel Strategy for Isolating Unknown Coding Sequences From Genomic DNA By Generating Genomic-cDNA Chimeras.**

Pudur Jagadeeswaran, Michael J. Odom and Edward J. Boland
Department of Cellular and Structural Biology, The University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, Texas 78284

The essential information contained in the human genome is encoded in short segments of DNA (genes), which are interspersed within much more abundant non-coding regions. The task of identifying the unknown coding sequences within the morass of the genome using conventional screening methods is difficult for two major reasons. First, only 3% of genomic DNA constitutes coding sequences that are homologous to cDNAs. Second, a significant proportion of cDNAs contain repetitive DNA elements, complicating their use as probes in the screening of genomic libraries. Recently, two important approaches have been described to isolate coding sequences from large genomic regions. They are exon trapping (1-3), and selection by hybridization (4-8). Although these techniques have certain advantages, they still have some drawbacks when used to identify several coding regions from large areas of the chromosome. In this abstract, we describe a novel strategy for the specific retrieval of coding sequences by the Polymerase Chain Reaction (PCR), using selective in vitro recombination, to form a genomic-cDNA chimeric molecule. The fundamental principle is to use the terminal free 3'OH-groups of restriction-enzyme digested cDNA fragments as primers for chain elongation. The elongation generates genomic-cDNA chimera molecules. These are then amplified by PCR to identify the coding sequences from the genomic DNA.

The details of the principle are as follows. The large genomic fragment is first converted into a short fragment genomic library by shotgun subcloning into a vector. The genomic inserts are subsequently amplified by PCR using vector-specific primers P1 and P2. At the same time, a cDNA library is generated in a different vector and is amplified using different vector-specific primers X1 and X2. The amplified cDNA is then digested by a restriction enzyme that cuts infrequently in the repetitive DNA, but does not cut between the primers and the insert. When the amplified genomic DNA and the restriction enzyme-digested cDNAs are mixed under conditions to allow hybridization, genomic-cDNA hybrids are formed. The hybrid molecule is then amplified by primers, one flanking the cDNA and the other flanking the genomic DNA. Thus, coding sequences can be retrieved as genomic-cDNA chimeras.

The feasibility of this strategy was tested by isolating a control chimera using cDNA and genomic sources of varying complexities. A number of inherent advantages of this novel PCR strategy are discussed. In summary, the method described in this abstract is a powerful tool for retrieving the coding sequences from unknown areas in genomic DNA and should be useful in isolating genes from YAC or cosmid clones and from chromosomal specific libraries.

1. D. Auch and M. Reth, *Nucleic Acids Res.*, 18,6743 (1990).
2. A.F. Buckler *et al., Proc. Natl. Acad. Sci. U S A.*, 88,4005 (1991).
3. M. Hamaguchi *et al., Proc. Natl. Acad. Sci. U S A.*, 89,9779 (1992).
4. S. Parimoo *et al., Proc. Natl. Acad. Sci. U S A.*, 88,9623 (1991).
5. M. Lovett *et al., Proc. Natl. Acad. Sci. U S A.*, 88,9628 (1991).
6. J.G. Morgan *et al., Nucleic Acids Res.*, 20,5173 (1992).
7. B. Korn *et al., Hum. Mol. Genet.*, 1,235 (1992).
8. D.A. Tagle *et al., Nature*, 361,751 (1993).

**Down Syndrome: Cloning the Genes for Congenital Heart Disease**

Korenberg, J.R., Ahmanson Department of Pediatrics, Division of Genetics, Cedars-Sinai Medical Center, UCLA

Down Syndrome (DS) is a major cause of mental retardation and congenital heart disease (CHD) affecting the welfare of more than 300,000 individuals and their families in the United States. Although CHD is clinically present in 46% of DS, it is seen at autopsy in 66% of which about 60% are defects of the endocardial cushions (ECD). The ultimate goal of this research is to identify the genes responsible for DS-CHD. Important steps in doing so are to identify small chromosomal regions that are responsible for this specific phenotype, to clone and characterize large DNA fragments that cover this region, and to identify the genes in these that are expressed in the developing heart. cDNA libraries of these embryonic tissues are under construction. We have defined four individuals with DS and CHD and small duplications that overlap in a small region of band 21q22 below the limit of cytogenetic detection. Yeast artificial chromosomes (YACs) and bacteria artificial chromosomes (BACs) defining these markers have been identified and a contiguous map is being constructed. Genes located in this contig and are expressed in the human fetal heart are identified using hybrid selection. Expression will be evaluated using tissue in situ hybridization.

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

Greg Lennon
Human Genome Center, L-452, Lawrence Livermore National Laboratory, Livermore, CA 94550

The goal of this effort is to isolate, sequence, and map coding regions in the form of cDNAs or exons from human chromosome 19. Three aspects of this work will be discussed. First, selected cosmids are being used directly as probes against arrayed cDNA libraries with the aim of isolating associated cDNAs. Using 23 cosmids as probes, 130 cDNAs have been identified. Of the cDNAs that have been fully characterized, 55% (11/20) are from chromosome 19 and represent novel genes. Second, as a pilot study in new methods of chromosomal assignment, we have produced and used stamp-sized filters containing flow-sorted chromosome dots to chromosomally map 20 new cDNAs. Third, we will discuss a strategy for sequencing genomic clones that is based on the increasingly complete physical map of chromosome 19. THis strategy is intended to both identify coding regions and form a sequence framework for the entire chromosome.

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

IDENTIFICATION OF TRANSCRIBED SEQUENCES IN THE TWIRLER REGION OF MOUSE CHR 18. Miriam H. Meisler, Daniel L. Burgess, David Kohrman, Karin Wiebauer and Wonhee Jang. Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109-0618.

The mouse mutant Twirler is characterized by abnormalities of the vestibular apparatus and by cleft palate ± cleft lip (Lyon, 1958). We have identified a transgenic insertional mutation which appears to be a re-mutation of the Twirler locus (Ting et al, 1993). In order to clone the mutated gene, a cosmid library was prepared using genomic DNA from the transgenic mutant and screened with transgene sequences. A clone containing one of the transgene flanks was isolated and partially sequenced. A 100 kb P1 contig was then generated by screening a mouse P1 library (Genome Systems, Inc.,St. Louis) using PCR primers derived from the original cosmid clone. Transcribed sequences within the P1 contig are being identified by exon trapping and sequence analysis. The corresponding human genomic region is also being cloned and we are developing improved methods for isolation of sequences which are conserved in the mouse and human P1 clones. With better methods, evolutionary conservation may provide a practical approach to identification of functionally important sequences in the human genome.

ENHANCEMENTS TO GRAIL AND THE INTRODUCTION OF A CLIENT-SEVER VERSION FOR GENE ASSEMBLY AND ANNOTATION. R. J. Mural[1], J. R. Einstein[2], X. Guan[2], S. Matis[2], M. Shah[2], Y. Xu[2], and E. C. Uberbacher[2]. [1]Biology Division and [2]Engineering Physics and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN. 37831.

GRAIL continues to be a widely used tool for the recognition of protein coding regions in anonymous DNA sequences. The e-mail version of GRAIL now allows the user to automatically search the translation of each predicted coding region against the PROSITE dictionary of sites and patterns in proteins (Bairoch, A. 1993, Nucl. Acids Res. **21**, 3097-3103) as well as the previously supported search, using a parallel implementation of the Smith-Waterman method, against the SWISPROT protein sequence database.
We have recently developed a client-server version of GRAIL which allows the user to interactively explore many features of genomic DNA sequence. This tool assembles predicted coding regions, within a user specified interval, into gene models, allows for database searches of the translation of gene models and locates a variety of sequence features such as potential poly-A addition sites and various classes of repetitive DNA elements, providing an environment to facilitate the annotation of new genome sequence. This system will be discussed the greater detail and instructions will be given for accessing and using the system.

PCR MAPPING OF HUMAN CDNA CLONES

W. C. Nierman, A. S. Durkin and D. R. Maglott,. American *Type Culture* Collection, Rockville, MD 20852-1776.

The laboratory of J. Craig Venter is identifying genes expressed in the human brain by sequencing portions of cDNA clones (Adams *et al.*, Science 252:1651-1656, 1991; Adams *et al.*, Nature 355:632-634,1992). In collaboration with the Venter group, we are developing methods for rapid localization of newly identified cDNA sequences to human chromosomes. We are using the ABI automated DNA sequencer to analyze fluorescently-tagged PCR products. Primer pairs are designed from the partial cDNA sequence data and tested for specific amplification from human genomic DNA. Primers permitting resolution of human products from rodent products are tested with DNA from somatic cell hybrid cell mapping panels. The presence or absence of specific amplification products in each cell line DNA is determined electrophoretically using the ABI sequencer, and chromosomal assignments are made by discordancy analysis. By using chromosome-specific panels we are able to make subregional assignments for cDNAs from several chromosomes. We are multiplexing the amplification reactions and analysis of the reaction products to achieve high productivity with a minimum allocation of resources. This project will determine chromosomal assignments for "Expressed Sequence Tags" (ESTs), provide primer sequence data for subsequent subchromosomal localizations, and generate a broad data set from which to evaluate strategies to identify functional primer sequences from cDNA sequence data.

# ISOLATION OF CODING SEQUENCE FROM YACS BY EXON AMPLIFICATION

Michael North*, Fernando Gibson^, Marie-Laure Yaspo*, Stephen Brown^ and Hans Lehrach*.

* Department of Genome Analysis, Imperial Cancer Research Fund, London WC2A 3PX, UK.

^ Department of Biochemstry and Molecular Genetics, St. Mary's Hospital Medical School, Imperial College of Science, Technology and Medicine, London, W2 1PG, UK.

The pSPL1 exon amplification system (Buckler *et al.*, P.N.A.S 88, 4005-4009) is now well established as a rapid and reliable method to isolate coding sequences from individual cosmids and complements other methods of gene isolation (North *et al.*, Mammalian Genome, *in press*). We have tested the applicability of the system to the rapid isolation of coding sequence from mouse and human YACs. We have successfully isolated exons from gel purified YACs (including YACs containing known genes) and have tested the feasibility of exon amplification directly from yeast genomic DNA containing a YAC of interest. Isolated exons can be used in direct library screens (Yaspo *et al.*, Nucleic Acids Res. 21, 2271-2272) or can be rapidly subcloned and sequenced. As an example, a 175kb mouse YAC (in pYAC4) was gel purified, digested with *BamHI/BglII* and a library constructed in pSPL1 corresponding to a 5-fold coverage of the YAC, with a 3% non-recombinant background. Exon amplification was carried out essentially as described by Buckler *et al.*, P.N.A.S 88, 4005-4009 and a minimal set of 9 unrelated products were distinguished. Clones from each of these groups were sequenced. Two displayed a high degree of homology to repetitive elements HSAG-1 and B1 SINE. One product, though not repetitive, did not contain an open reading frame in the correct orientation. The remaining six products were of average size 90bp, on average 55% GC rich and 4/6 were chimaeric with a portion of the HIV tat intron present in pSPL1 (a event which we have found not to significantly compromise the specificity of the system for amplifying exonic sequences; North *et al.*, Mammalian Genome, *in press*). The two largest products (one chimaeric with tat, one not) detected sequences conserved across a wide variety of species and corresponding cDNAs were characterised (the other potential coding sequences identified were not successfully crosshybridised to other species, possibly due to their small size-a problem also encountered by Hamaguchi *et al.*, P.N.A.S 89, 9779-9783.).

Repetitive elements containing sequences with good homology to acceptor and donor splice junctions (eg. including some subclasses of the *Alu* family) are are most important non-genic contaminant and human YACs typically give higher levels of repetitive clones than mouse YACs. Additionally, direct exon amplification from total yeast DNA leads to a significant background of yeast ribosomal sequences. Both types of artefact can be quickly eliminated by hybridisation screening subcloned exontrap products.

MAPPING OF 500 cDNAs AND THEIR DISTRIBUTION IN THE HUMAN GENOME.

Mihael H. Polymeropoulos, Laboratory of Biochemical Genetics, NIMH, Washington, D.C.

We have determined the chromosomal assignment of 500 brain expressed genes by studying the segregation of PCR products in human rodent somatic cell hybrids and by genetically mapping polymorphic cDNAs using the CEPH reference pedigrees and database. These mapped genes can function as markers on the physical map of the human genome, as well as serve as candidate disease gene loci. Distribution of these genes to the human chromosomes correlates well with the degree of G+C richness of the human chromosomes but does not correlate well with the cytogenetic length of each chromosome. Our current data of the chromosomal distribution of genes together with statistics on the localization of known genes and recent evidence on the non-uniform concentration of genes may suggest a new direction in the development of genetic polymorphisms. The characterization of polymorphic cDNAs can have a dual function of expressed sequence tags and at the same time serve as genetic markers in the construction of genetic maps of the human genome and the isolation of disease genes by means of genetic linkage analysis.

**Analysis of the X chromosome breakpoint region at Xq13 in patients with acquired isodicentric X chromosome associated with leukaemia.**

S. H. Rider[1], K.A. Rack[2], J. Chelly[1], R.G.Lafrenière[3] H.F.Willard[3], V.J. Buckle[2] and A.P. Monaco[1]. [1] ICRF, IMM, Oxford OX3 9DU U.K., [2]. MRC, MHU, IMM, Oxford OX3 9DU U.K., [3] Department of Genetics, Case Western Reserve University, Cleveland, Ohio 44106 USA.

Elderly female patients presenting with acquired sideoblastic anaemia have been shown to have a common cytogenetic abnormality of the acquisition of an isodicentric X chromosome, idic(X q13). These patients also show an increased incidence of leukemia. We have characterised the breakpoints of two patients with acquired idic(X) chromosomes associated with leukemia to a 350kb region of DNA within Xq13. The X-inactivation center(XIC) has also been localised to Xq13. Ubiquitous expression of XIST occurs only from the inactive X chromosome and has been postulated to play a major role in the process of X-inactivation. We have demonstrated that the idic(X) chromosome are subject to X-inactivation but are deleted for the XIST gene; suggestive that XIST expression is not required for the maintenance of the inactive state. These idic(X) chromosomes arise somatically, X-inactivation may be maintained automatically in such circumstances, or factors either primed by or independent of XIST and proximal to Xq13 breakpoints may have a role to play in maintenance of X-inactivation. The consistency of the leukemic features associated with this rearrangement however do suggest that the disruption of critical sequences in the vicinity of the breakpoint may lead to leukemogenesis. In order to isolate a candidate gene for leukemogenesis, we are in the process of constructing a cosmid contig of the region. For this aim we have made a cosmid library from a YAC clone which was shown by in situ hybridization to cross the breakpoint in a patient with an idic(Xq13) chromosome. In addition we have also obtained some additional cosmids by screening the ICRF gridded X chromosome cosmid library with YAC DNA. The cosmid contig is being constructed by 'walking' using end specific RNA probes. As the cosmids are localized in the contig they are examined for the presence of HTF islands. Our primary approach to isolate a candidate gene for leukamogenesis, will be to utilize those cosmids known to contain an HTF island, to directly hybridize against filters of cDNA libraries under competitive conditions. We have demonstrated that this approach can successfully identify coding sequences.

**A catalogue of the genes of man**

Joachim Rothe, Sebastian Meier-Ewert, Richard Mott and Hans Lehrach
Imperial Cancer Research Fund

The understanding of the information contained in the human genome requires the handling and analysis of hundred of thousands of genomic or cDNA clones, and will involve acquisition of terabytes of information. To be able to generate such large amounts of information at acceptable cost, it is necessary to develop new approaches, based on close integration between experimental strategies, automation and informatics components. We have concentrated in our work on the use of high throughput hybridization techniques for both genomic mapping and gene analysis, based on the hybridization of different types of probes to high density filter grids of colonies or DNA samples to establish genomic maps (Meier et al, *Nature Genetics* **1**: 273-277 1992; Hoheisel et al, *Cell* **73**: 1-20 1993), as well as partial sequence information of cDNA clones (Meier-Ewert et al, *Nature* **361**: 375-376 1993). To establish a catalogue of most or all genes expressed in a given cDNA library, we have developed a partial sequencing approach based on the use of short oligonucleotides (hepa- or octanucleotides) as probes in hybridizations carried out under conditions allowing the identification of perfectly matched hybrids. This strategy is currently being applied to the analysis of a human embryo cDNA library of 100,000 clones, and will be expanded to analyze cDNA libraries from a number of different tissues, giving access to a significant fraction of the full complement of genes of the human genome.

**Rapid Identification of Gene Sequences by Direct cDNA Screening of Genomic Reference Libraries**

Wolfgang Schwabe, Brenda Lawrence, Rene Hopfinger, Miles B. Brennan and Ute Hochgeschwender

Unit on Genomics, NIMH, 9000 Rockville Pike, Bethesda, MD 20892, USA

We developed direct cDNA screening of large, arrayed genomic libraries (reference libraries) as a method for identifying cloned transcribed sequences (Hochgeschwender et al., PNAS 86:8482-8486, 1989).
Unlabeled first strand cDNA is used as template in a random hexamer labeling reaction. This probe is then quantitatively depleted of repeated sequences by annealing with a large excess of high molecular weight genomic DNA immobilized on finely divided cellulose. Replica filters of a reference library are then hybridized with the depleted cDNA probe.
We have used this direct cDNA screening protocol to identify genomic sequences transcribed in developing and adult brain in inter- and intraspecific hybridizations to arrayed genomic cosmid and phage clones. We derive coding sequences from these genomic clones by first identifying fragments containing transcribed sequences and subjecting these to exon trapping or to partial sequencing and analysis by Grail.
The approach is amenable to large scale applications and should be useful in isolating candidates for disease genes as well as in assembling integrated transcriptional maps from large genomic regions.

# Identification of disease genes in Xq28

*Hubert J. M. Smeets*

*Department of Human Genetics*

*University Hospital Nijmegen*

*Nijmegen, The Netherlands*

The chromosomal region Xq28-Xqter is rich in disease genes. Recently, most of this region has been delineated by genetic and physical mapping strategies and the larger part of the region has been cloned in cosmid vectors and/or yeast artificial chromosomes. Our group has been working on this region for several years, in particular on the elucidation of the genetic defect in Nephrogenic Diabetes Insipidus and on linkage analysis in a variety of other disorders in Xqter. Our next goal is the identification of genes in a region of about 2 Mb, where the gene for the rare disorder myotubular myopathy, a muscle developmental disorder, has been localized. To isolate the transcribed regions, a variety of different techniques will be used, including the screening of cDNA libraries with cosmids from the region, cDNA selection and hybridization of cDNA probes to cosmids. Candidate genes will be selected on the basis of their tissue specific expression and, where possible, their sequence features. Subsequent screening for structural gene rearrangements, altered expression and deleterious point mutations will lead to the definite identification of the gene involved. Furthermore, we have a general interest in developing methods to isolate genes from this area based on their functional characteristics. The main problem at the moment seems not to be the generation of new cDNAs and genes anymore, but to get to know the functions of the proteins these genes code for.

**Subtractive Hybridizations of Tagged Normalized cDNA Libraries and Selection of Chromosome-Specific cDNAs**

Su[1], L. Lawton[1], D. Warburton[2] and M. Bento Soares[1]

Department of Psychiatry[1] and Genetics and Development[2]
Columbia University

We are constructing normalized human cDNA libraries with sequence tags (tissue-specific IDs). Library-specific-tags are provided by the sequence of the oligonucleotide [Pac I - $(N)_6$ - $(dT)_{18}$] used to prime first-strand cDNA synthesis. Double-stranded cDNAs are ligated to Eco RI adaptors, digested with Pac I and directionally cloned into a phagemid vector. Normalized cDNA libraries with tags will facilitate the assessment of tissue and/or temporal specificity of a subtracted library. We are using these libraries for selection of chromosome-specific cDNA clones. DNa from an entire chromosome-specific genomic library is filter immobilized and hybridized to a normalized library in the form of single-stranded circles. After washing, the cDNA circles are eluted and electroporated into bacteria, to generate a mini-library highly enriched for chromosome-specific cDNAs.

**Magnetic bead capture and CpG island rescue PCR: Two strategies for isolating transcribed sequences from large genomic regions.**

Danilo A. Tagle[1], John Valdes[1], Fergus Couch[1], Yossi Shiloh[2] and Francis S. Collins[1,3]
[1]Department of Human Genetics and Internal Medicine, University of Michigan Medical Center, Ann Arbor, MI 48109; [2] Department of Human Genetics, Tel Aviv University, Ramat Aviv, israel 69978; [3] National Center for Human Genome Research, National Institutes of Health, Bethesda, MD 20892.

One of the key steps in positional cloning of any human inherited disorder involves the rapid identification and isolation of transcribed sequences from the genomic region that define the candidate interval. Conventional methods of finding transcripts tend to be laborious and cumbersome when the candidate interval spans several hundred thousand to a few million base pairs. We describe here two strategies that will greatly facilitate the identification of cDNAs encoded within a large genomic segment cloned into a contig of YACs or cosmids. In the first strategy, called magnetic bead capture, YAC or cosmid DNA is digested and ligated onto linkers. the linkered genomic segments are amplified by PCR using a biotinylated linker-specific primer and they hybridized at high stringency to PCR-amplified inserts from cDNA libraries. The biotinylated genomic-cDNA complexes are captured using streptavidin-coated paramagnetic beads and the unbound, nonspecific cDNAs are washed off. The captured cDNAs are eluted, PCR-amplified and subcloned as a region-specific sublibrary. We have applied this method to isolate cDNAs from the Huntington's Disease (HD) locus at 4p16.3 using either a pool of 8 YACS that form a contig of 2 Mb or each YAC separately, and also from a pool of 17 cosmids that form a contig of 550 kb; from the early onset breast cancer (BRCA1) locus at 17q21.2 using 5 pools of 8 cosmids each; and from the ataxia-telangiectasia (A-T) locus at 11q23 using a pool of 10 cosmids. Subchromosomal-specific transcripts found in low abundance can be enriched several thousand fold with an associated decrease in non-specific transcripts. this technique allows parallel analysis of several large genomic segments of varying complexities and can be applied to the isolation of expressed sequences from various tissue sources.

The second strategy takes advantage of the observed high density of rare-cutter sites clustered into CpG islands which usually define the 5' end of genes and gene clusters. This strategy, called Island Rescue PCR, entails digestion of total YAC DNA with rare-cutting restriction enzymes (*BssH II, EagI, NaeI, SacII,* and *SmaI*) whose restriction sites commonly occur in CpG islands. Uni-directional or vectorette linkers are ligated to the restricted YAC DNA. Sequences adjacent to the CpG islands are amplified by PCR using a linker-specific primer and Alu repeat primers, thus generating a human specific probe which extends from the CpG island to the nearest Alu repetitive element. the island rescued product, which is gel purified away from the Alu-Alu PCR products, is likely to include the 5' end of a gene associated with the island and can be used to directly screen cDNA libraries. The method is capable of rapidly screening several YACs within a given candidate interval yielding cDNA clones containing the 5'-end of most genes associated with CpG islands. We have also applied this method of YACs in the 4p16.3 and 17q21.2 regions and have successfully identified known genes as well as novel genes in these regions.

## dbEST - DATABASE FOR EXPRESSED SEQUENCE TAGS

*M. S. Boguski, T. Lowe, B.F.F. Ouellette, C. M. Tolstoshev.*
*National Center for Biotechnology Information, National Library of Medicine,*
*National Institutes of Health, Bethesda, Maryland 20894.*
*(for more information send Email to info@ncbi.nlm.nih.gov)*

The accumulation and analysis of incomplete, "single-pass" cDNA sequences (ESTs or "Expressed Sequence Tags," also known as: "Transcribed Sequence Fragments", "Putatively Transcribed Partial Sequences" or "Sequence-Tags Sites") has become an important component of genome research. Uses of EST data include the identification of previously unknown gene products for genetic mapping purposes, the study of tissue differentiation and ontogeny and molecular evolution, and the accelerated cloning of human genes for which homologues in other organisms have already been functionally characterized.

EST data differ from traditional sequence database entries in terms of accuracy, completeness, rate of accumulation and intended uses. For the past year, NCBI has been operating a special resource for these data. dbEST now contains more than 24,000 sequences from the major model organisms as well as other species. NCBI accepts direct bulk submissions of data and issues GenBank accession numbers but also obtains input via daily updates from Los Alamos, EMBL and DDBJ for data submitted to these locations. Because up to 70% of new ESTs cannot be characterized by homology with known sequences, we have developed a special system based on a BLAST function library (W. Gish, pers. comm) for automatic and continual re-annotation by periodic re-screening of all ESTs against the general-purpose nucleotide and protein sequence databases. This system includes prior filtering to minimize the number of non-specific, spurious and uninformative matches caused by sequence contaminants, the presence of repetitive elements, and that of the occurrence of "low complexity" subsequences in conceptually translated ORFs.

Information submitted by contributors, as well as the results of our analyses, are stored in a relational database and made available to the public as follows: 1) the data may be searched using the BLAST Internet e-mail servers (send "help" to blast@ncbi.nlm.nih.gov for more information);
2) "full reports" on ESTs may be obtained by sending e-mail to the EST server (send "help" to est_report@ncbi.nlm.nih.gov for more information);
3) sequences from dbEST are now included in a new EST Division of GenBank;
4) a FASTA-formatted version of all sequences in dbEST (along with descriptive header lines) are available by anonymous ftp from NCBI (ncbi.nlm.nih.gov) in the /repository/dbEST directory as dbBEST.vX.X, where X.X is the version number. Full reports on ESTs also contain information on the availability of physical DNA clones (e.g. ATCC numbers and ordering information) and any additional mapping data contained within the Genome Data Base.

**Abstract**

Dominique Weil, Ph.D.
INSTITUT PASTEUR

28, Rue du Dr. Roux, 76724 Paris, Cedex 15

The research projects in the lab are focused on different aspects of the hearing defects and imply isolation of specific expressed sequences:

1. Chromosomal localisation of some of the genes responsible for autosomal recessive deafness (we have already mapped one of them), and subsequent cloning of these genes by exon trapping.

2. Isolation and characterization of genes expressed in the development of the inner ear.

We have already constructed a cDNA library using RNA extracted from 16 day old mice cochlea and are on the way to construct libraries using RNA extracted from 2 day old mice as well as fetuses (embryonic day 17th) cochlea.

**Towards a Transcriptional Map of Human Chromosome 21**

Yaspo Marie-Laure, Nizetic Dean, Mott Richard and Lehrach Hans.

Genome Analysis Laboratory, Imperial Cancer Research Fund,
44, Lincoln's Inn Field, POBox 123, WC 2A 3 PX London, UK

Chromosome 21 serves as a prototype for extensive genetical and physical mapping but only 25 genes have been cloned so far among at least 500 expected coding sequences. Moreover, this chromosome is implicated in several severe diseases for which candidate genes are still unknown, such as the main morphological features of Down syndrome, leukaemias associated with constitutional or acquired trisomy 21 and a specific form of progressive myoclonic epilepsy. Definition of a large scale transcriptional map requires the construction of expressed sequence libraries allowing efficient and rapid screening in order to integrate physical and transcriptional information.

Isolation of genes from genomic clones is a challenging task but hybridization-based procedure involving cDNAs are unreliable and represent a tremendous work. To circumvent this problem, we are using alternative strategies, as the exon-amplification technique. We aim to construct regional exon-trapped libraries derived from groups of cosmids already mapped on chromosome 21 (ICRF reference library). Eventually, cloned exons will be picked and arrayed onto high density filters, providing tools for extensive identification and analysis of chromosome 21 coding sequences.

Ruchira Das Gupta, M.S.
Department of Molecular Genetics
Albert Einstein College of Medicine
1925 East Chester Road
Bronx NY 10461
TEL 718-430-2848
FAX 718-822-0845
dasgupta@accom.yu.edu

Nicole A. Datson, M.Sc.
Department of Human Genetics
Leiden University
Wassenaarseweg 72, 2300 RA Leiden
PO Box 9503
The Netherlands
TEL 31-71-276046
FAX 31-71-276075
wagdunne@rulgl.LeidenUniv.nl

Lisa Davis, Ph.D.
Applied Genetics Laboratories, Inc.
1335 Gateway Drive, Suite 2001
Melbourne FL 32901
TEL 407-768-2048
FAX 407-727-2643

Dr. Johan T. den Dunnen
Department of Human Genetics
Leiden University
Wassenaarseweg 72, 2333 Al Leiden
PO Box 9503, 2300 R.A. LEIDEN
The Netherlands
TEL 31-71-276065
FAX 31-71-276075
wagdunne@rulgl.LeidenUniv.nl

Gregory Dolganov, Ph.D.
GENELABS, Inc.
505 Penobscot Drive
Redwood City, CA 94063
TEL 415-369-9500
FAX 415-366-8748
dolgreg@netcom.com

Radoje Drmanac, Ph.D.
Integral Genetics Group
Biological and Medical Research Division
Argonne National Laboratory
9700 South Cass Avenue
Argonne IL 60439-4833
TEL 708-252-3175
FAX 708-252-3387
rade@everest.bim.anl.gov

James Eberwine, Ph.D.
Assistant Professor
Department of Pharmacology
University of Pennsylvania
Medical Center
36th Street and Hamilton Walk
Philadelphia PA 19104-6084
TEL 215-898-0420
FAX 215-898-9927

Mark Erlander, Ph.D.
The Scripps Research Institute
RISC, MB10
10666 N. Torrey Pines Road
La Jolla CA 92037
TEL 619-554-8064
FAX 619-554-8841

Jeffrey Falk, Ph.D.
Department of Molecular Biology
Scripps Research Institute, MB10
10666 N. Torrey Pines Road
La Jolla CA 92037
TEL 619-554-4314
FAX 619-554-6112

Michel Fontes, Ph.D.
INSERM U262
Institut De Pediatrie
BP4 13351
Marseille Cedex 5
FRANCE
TEL 33-91-78-44-77
FAX 33-91-80-43-19

Pudur Jagadeeswaran, Ph.D.
Associate Professor
Division of Human Genetics
Department of Cellular
& Structural Biology
The University of Texas Health Science
Center at San Antonio
7703 Floyd Curl Drive
San Antonio, TX 78284-7762
TEL 210-567-3848
FAX 210-567-3803

Julie R. Korenberg, M.D., Ph.D.
Director, Molecular Diagnostic Lab
Department of Pediatrics
Medical Genetic Birth Defects Center
Cedars-Sinai Hospital
110 George Burns Road
Davis Building, Suite 2069
Los Angeles, CA 90048-1869
TEL 310-855-7627
FAX 310-652-8010

David M. Kurnit, M.D., Ph.D.
Investigator
Howard Hughes Medical Institute
Professor, Pediatrics and Human Genetics
University 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@um.cc.umich.edu

Greg Lennon, Ph.D.
Human Genome Center, L-452
Lawrence Livermore National Labs
Livermore CA 94550
TEL 510-422-5711
FAX 510-423-3608
greg@mendel. llnl.gov

Wai-Choi Leung, Ph.D.
Professor of Pathology
Director, Division of Molecular Pathology
Tulane University School of Medicine
Department of Pathology
1430 Tulane Avenue
New Orleans LA 70112
TEL 504-584-2436
FAX 504-587-7389

Miriam H. Meisler, Ph.D.
Professor of Human Genetics
University of Michigan
Department of Human Genetics
2806 Medical Science Building II
Ann Arbor MI 48109-0618
TEL 313-763-5546
FAX 313-764-3784
Miriam.Meisler@med.umich.edu

Donald T. Moir, Ph.D.
Research Director
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

Sue Rider, Ph.D.
Imperial Cancer Research Fund
University of Oxford
Institute of Molecular Medicine
John Radcliffe Hospital
Headington Oxford OX3 9DU
TEL 44-865-222432
FAX 44-865-222431
s.rider@alpha.lif.icnet.uk

Johanna Rommens, Ph.D.
Scientist/Assistant Professor
Genetics Department
The Hosptial for Sick Children
555 University Avenue
Toronto, Ontario M5G 1X8
Canada
TEL 416-813-7095
FAX 416-813-4931

Joachim Rothe, Ph.D.
Genome Analysis
Imperial Cancer Research Fund
PO Box 123
44 Lincoln's Inn Fields
London, WC2A 3PX
United Kingdom
TEL 44-71-269-3080
FAX 44-71-269-3068
J.ROTHE@uk.icnet.icrf

Claire Rougeulle, Ph.D.
Unite de Genetique Moleculaire Murine
Institut Pasteur
25 rue du D'Roux
75726 Paris Cedex 15
France
TEL 33-1-45-68-86-25
FAX 33-1-45-68-86-56

Wolfgang Schwabe, Ph.D.
Unit on Genomics
NIH/NIMH
Bldg. 10, Rm 4N320
9000 Rockville Pike
Bethesda MD 20892
TEL 301-402-1769
FAX 301-402-2140
Genunit@helix.nih.gov.

Yosef Shiloh, Ph.D.
Department of Human Genetics
Sackler School of Medicine
Tel Aviv University
Ramat Aviv 69978
ISRAEL
TEL 972-3-6409760
FAX 972-3-6409900

James Sikela, Ph.D.
Assistant Professor
Department of Pharmacology, C-236
University of Colorado
Health Sciences Center
4200 East 9th Avenue
Denver CO 80262
TEL 303-270-8637
FAX 303-270-7097
js@tally.hsc.colorado.edu

Hubert Smeets, Ph.D.
Department of Human Genetics
University Hospital Nijmegen
P.O. Box 9101
NL-6500 HB
Nijmegen, The Netherlands
TEL 31-80-613799
FAX 31-80-616658

copyrighted material removed
JR

Carla Tribioli, Ph.D.
I.G.B.E., CNR
UIA Abbiategrasso 207,
27100 Pavia, Italy
TEL 39-382-546367
FAX 39-382-422286
IN%"IPVG03OT IPVGBE.UNIPV.IT"

Dominique Weil
Unite de Genetique Moleculaire Humaine
Institute Pasteur
25, rue du Docteur Roux
75724 PARIS-Cedex 15
TEL 33-1-45 68 88 93
FAX 33-1-45 67 69 78

Sherman M. Weissman, M.D.
Sterling Professor of Genetics
and Medicine
Yale University School of Medicine
336A BCMM
295 Congress Avenue
New Haven CT 06510
TEL 203-737-2285
FAX 203-737-2286
WeissmaN@biomed.med.yale.edu

Marie-Laure Yaspo, Ph.D.
Genome Analysis Lab
ICRF
44 Lincoln's Inn Field
London WC2A 3PX
ENGLAND
TEL 44-71-269-3080
FAX 44-71-269-3068