CONF-9110446--Abst.

DOE/ER/61245--1

DE93 005046

Identification of Transcribed Sequences in the Human Genome
Conference Grant

Final Report

for Period September 15, 1991 - September 14, 1992

Katheleen Gardiner

Eleanor Roosevelt Institute
1899 Gaylord Street
Denver, CO 80206

December 1992

Prepared for

THE U. S. DEPARTMENT OF ENERGY
AGREEMENT NO. DE-FG02-91ER61245

DOE/ER/61245-1
Katheleen Gardiner, Ph.D.

Grant Title: **First International Workshop on the Identification of Transcribed Sequences in the Human Genome**

Abstract

Because of rapid progress being made in the construction of physical maps of large chromosomal regions, many investigators are finding it timely to begin comprehensive gene searches from YAC and cosmid contigs. Thus, it was deemed important to bring together interested workers in this field for discussions of old and new techniques in gene identification, to determine what progress is being made and what additional types of innovations may be required for the construction of transcriptional maps of the human genome.

Final Report

The workshop was held at the National Institutes of Mental Health, Bethesda, Maryland, on October 4 and 5, 1991. Twenty-four investigators attended from England, Germany and the United States. The topics discussed included: i) genome sequence analysis using computer assisted detection of open reading frames, splice sites and hexamer patterns, ii) direct exon identification using trapping of internal and 3' exons, and a recombination based system, iii) cDNA library construction and screening, including the use of normalization and subtraction procedures, Alu and splice donor site PCR from hybrid cell lines, and microdissection clones as probes, iv) use of labeled CDNAS as probes to screen lambda and cosmid libraries, and v) sequencing of random cDNAs. Significant time was devoted to discussion of the strengths and weaknesses of the various approaches, and to complementarities among them. In general, it was agreed that further experience in comprehensive gene searches and in the development of transcriptional maps is necessary before the most efficient approaches can be defined.

All participants agreed that a second workshop, again to be appended to the annual American Society of Human Genetics meeting, would be crucial. Many techniques are too new to be properly evaluated, and additional approaches are sure to appear. It was suggested that the second workshop be increased in size to 40 participants, to represent more laboratories and projects, but that the informal workshop style of all oral presentations and much discussion be maintained.

A report of the workshop was published in Human Genome News, January 1992 (copy enclosed). Also enclosed are copies of the workshop agenda, abstracts presented and list of attendees.

#1203

Meeting Reports

First International Workshop on the Identification of Transcribed Sequences

Twenty-four investigators met on October 4–5, 1991, at the NIH National Institute of Mental Health (NIMH) in Bethesda, Maryland, for the First International Workshop on the Identification of Transcribed Sequences, sponsored by the DOE Human Genome Program. The purpose of the workshop was to exchange information on the systematic identification of transcribed sequences and the construction of transcriptional maps for large chromosomal regions.

Investigators discussed the broad areas of (1) identification of expressed sequences from genomic clones and (2) cDNA library analysis. The group also considered strategies for the most reliable and exhaustive search for gene sequences from any chromosomal region.

J. Gregor Sutcliffe (Scripps Research Institute) defined three data requirements for including a gene on a transcriptional map: sequence, physical and genetic map location, and pattern of expression. Katheleen Gardiner (Eleanor Roosevelt Institute) described the striking variation in gene and potential CpG island density found on human chromosome 21 and suggested that different approaches may be required for gene-rich and gene-poor regions.

Sequence Analysis

Several presentations were devoted to the analysis of genomic sequence information. Andrzej Konopka (National Cancer Institute) discussed basing the detection of coding sequences on statistical characteristics (such as complexity) of textual elements, using as an example protein-coding sequences.

Steen Knudson (Boston University) is investigating a neural network approach to predicting splice sites and open reading frames. Richard Mural (Oak Ridge National Laboratory) reported considerable success with a neural network/rule-based inference system. This program identifies about 90% of protein-coding exons 100 or more bases long and has predicted 14 exons in a region near the Huntington's disease locus that have been experimentally confirmed.

(see cDNA Sequences, p. 8)

Sequence Workshop Considers Strategies for Gene Searches in any Chromosomal Region

Meeting Reports

cDNA Sequences *(from p. 7)*

Direct Exon Identification

Four speakers addressed three approaches to direct isolation of coding sequences from genomic clones. Geoffrey Duyk (Harvard Medical School) discussed current and proposed modifications of the "exon trapping" procedure; Alan Buckler (Massachusetts Institute of Technology) described experience with the related "exon amplification" system. Nine of ten putative exons obtained with the latter system subsequently identified clones in a cDNA library.

Susan Berget (Baylor College of Medicine) presented a scheme for trapping 3' exons; such an approach would have the advantage of isolating larger exons (typically, around 600 nucleotides, compared with 100 to 200 nucleotides for internal exons).

These three approaches still require library screening to obtain a complete cDNA. Alternatively, David Kurnit (Howard Hughes Medical Institute and University of Michigan Medical School) described his recombination-based assay system in the isolation of cDNAs encoded in yeast artificial chromosome (YAC) clones. A lambda cDNA library is replicated in the presence of a plasmid library made from YAC DNA. The progeny phage are then grown in an *Escherichia coli* host that requires plasmid sequences; only those phage that have integrated a plasmid are viable. This system has been used to obtain many cDNAs from several human chromosome 21 YACs.

Approaches to Direct Isolation of Coding Sequences:

- Exon Trapping
- Exon Amplification
- 3' Exon Trapping
- Isolation of cDNAs Encoded in YACs

cDNA Library Construction and Screening

Two approaches for identifying human cDNAs from somatic cell hybrids were discussed. David Nelson (Baylor College of Medicine) has used *Alu*-specific polymerase chain reaction (PCR) primers to amplify human hnRNA from the Xq28 region. Michael Siciliano (M. D. Anderson Cancer Center) used splice donor site-specific PCR primers to construct libraries of amplified material and screened the libraries with human repeat sequences. False positives continue to present some difficulties, although both methods are rapid and successful.

Bento Soares (Columbia University) discussed the use of lafmid vectors in cDNA library construction. Such libraries can be normalized efficiently and used in subtractive hybridizations.

Fa-Ten Kao (Eleanor Roosevelt Institute) described the identification of 7 chromosome 21 cDNAs, obtained by screening cDNA libraries with 200 microdissection genomic clones. Large pools of clones facilitated the screening, as did the use of normalized cDNA libraries [Sherman Weissman (Yale University School of Medicine)].

cDNAs as Probes

Direct cDNA screening of large (> 10,000) arrayed genomic libraries has been used to identify genomic clones containing transcribed sequences. Ute Hochgeschwender (NIMH) presented results for mouse chromosome 16, reporting improved sensitivity by using cDNA probes with (1) decreased complexity or (2) enrichment for low-abundance transcripts. A second application, presented by Anne Marie Poutska (German Cancer Research Fund, Heidelberg), used pig cDNA probes to screen a human Xq28–specific genomic library. This is an alternative solution to the problem of repetitive sequences and identifies conserved, transcribed sequences.

Hans Lehrach (Imperial Cancer Research Fund, London) discussed using arrayed cDNA and region-specific genomic libraries to map particular cDNAs to genomic clones. Possibilities include the use of clone pools and cDNAs from various tissues and oligonucleotide fingerprinting of both cDNA and genomic libraries.

Hybrid-selection schemes to isolate cDNA clones from YACs and from pools of cosmids were described by Mike Lovett (Genelabs, Inc.) and Weissman. In these methods, cDNAs are annealed to immobilized clones of genomic DNA, and the annealed fraction is recovered, amplified, and cloned. Impressive enrichments of > 1000-fold were reported for specific cDNAs.

MaryKay McCormick (Los Alamos National Laboratory) outlined an alternative strategy that would use homologous recombination and fragmentation to locate the gene position within a YAC. Then cDNAs of interest are cloned into an appropriate vector and transformed into a yeast clone containing a YAC. Truncation of the original YAC will occur where it contains sequences homologous to the cDNA.

(see cDNA Sequences, p. 9)

This newsletter is prepared, at the request of the DOE Office of Health and Environmental Research and/or the NIH National Center for Human Genome Research by the Biomedical and Environmental Information Analysis Section of the Health and Safety Research Division at Oak Ridge National Laboratory, which is managed by Martin Marietta Energy Systems, Inc., for the U.S. Department of Energy, under Contracts DE-AC05-84OR21400.

cDNA Sequences *(from p. 8)*

Sequencing cDNAs

James Sikela (University of Colorado Health Sciences Center) and Mark Adams [National Institute of Neurological Disorders and Stroke (NINDS)] reported on projects to sequence 100 to 200 nucleotides from the 3' (and possibly 5') ends of a large number of random human brain cDNAs. The usefulness of this approach will depend partly on the generation of sufficient sequence to permit protein motif identification and also on the ability to map accurately the genomic sequences. Some regional clone localization by fluorescence in situ hybridization has been proposed (Adams).

Chris Fields (NINDS) discussed database formats for the storage of cDNA sequence information.

Summary

Participants agreed that conventional techniques, including cDNA library screening with YAC clones and searches of CpG islands and conserved sequences, can be informative but are not likely to be comprehensive. However, at this time no one technique is completely satisfactory; an exhaustive gene search will require several complementary methodologies. Many techniques discussed are still very new and have not been applied extensively.

The group recommended that another workshop be held in 1992, when experience in different laboratories will allow more critical technique evaluation. Future considerations also will include how best to approach the thorny problem of determining expression patterns, both in developmental timing and tissue specificity. ◊

Reported by Katheleen Gardiner
Eleanor Roosevelt Institute
and
Miles Brennan
National Institute of Mental Health

Abstracts of papers presented
at the 1991 International Workshop on

THE IDENTIFICATION OF TRANSCRIBED SEQUENCES

October 4 - October 5, 1991

Arranged by
Katheleen Gardiner, *Eleanor Roosevelt Institute*
Miles Brennan, *National Institute of Mental Health*

Sponsored by
The Department of Energy

INTERNATIONAL WORKSHOP ON THE IDENTIFICATION OF TRANSCRIBED SEQUENCES

FRIDAY, Oct. 4 - 8:30 AM

SESSION 1 cDNA LIBRARY SCREENING AND SEQUENCING

Chairman: Greg Sutcliffe

FRIDAY, Oct. 4 - 1:00 PM

SESSION 2 EXON IDENTIFICATION - COMPUTATIONAL METHODS

Chairman: Richard Mural

SESSION 3 EXON IDENTIFICATION - EXPERIMENTAL APPROACHES

Chairman: Glen Evans

SATURDAY, Oct. 5 - 9:00 AM

SESSION 4 HYBRIDIZATION BASED APPROACHES - DIRECT SCREENING AND SELECTION

Chairman: Mike Lovett

SATURDAY, Oct. 5 - 1:00 PM

SESSION 5 HYBRIDIZATION BASED APPROACHES - DIRECT SCREENING AND SELECTION - CONTD.

cDNA LIBRARY SCREENING AND SEQUENCING

J. Gregor Sutcliffe, The Scripps Research Institute

Complete understanding of human physiology will require substantially complete identification of the protein molecules expressed in the body and their sites of synthesis as well as knowledge of their biochmeical activities and contributions to organismic function. Exhaustive, systematic cDNA cloning and nucleotide sequence analysis allow expressed genes to be identified and their encoded proteins recognized. How many genes are there to be identified, and what obstacles will one encounter in traversing from cDNA cloning to physiological understanding?

Systematic studies on cDNA cloning brain mRNAs indicate that there are about 30,000 distinct mRNAs expressed in the brain. Extrapolation to the whole oranism suggests that between 50,000 and 100,000 might be a reasonable estimate of total gene number. Technical obstacles to hurdle include the isolation of complete sets of cDNA clones, obtaining full-length clones and accurate sequences. Intellectual obstacles will be encountered in bridging between a protein's sequence and pattern of expression and its actual biochemical and physiological function. The deliberate production of mutant animals will be one step in the analysis.

ISOLATION AND MAPPING OF EXPRESSED SEQUENCES OF HUMAN CHROMOSOME 21 USING MICRODISSECTION CLONES

Fa-Ten Kao

Human chromosome 21 is important because it carries genes responsible for Down syndrome. We constructed a large library of chromosome 21 comprising 700,000 recombinant plasmid clones using microdissection and PCR microcloning procedures (ref. PNAS 88, 1844-1848, 1991). These microclones contain an average insert size of 400 bp and about 60% of the microclones are devoid of repetitive sequences. Thus, large numbers of unique sequence microclones can be readily isolated for use in direct screening of cDNA libraries.

In our screening procedure, the unique sequence inserts in the microclones were amplified by PCR and pooled, 10 or 20 in a group, to prepare labeled probes. After screening 100,000 phage from a human liver λgt11 cDNA library using 200 unique sequence microclones, 5 different cDNA clones were identified and plaque purified. One of the cDNA clones, 21E-C1, was shown to be derived from microclone 21E-181. This cDNA clone contains an insert of 1.5 kb and was regionally localized to 21q21.3.

In addition, we used the same 200 unique sequence microclones to screen a normalized human thymus cDNA library, kindly provided by Drs. Patanjali and Weissman (PNAS 88, 1943-1947, 1991). In these screens, we identified the same 5 cDNA clones as found in the previous library, but occurring at higher frequencies. We also identified 2 new cDNA clones not found in the previous library.

The direct screening of cDNA libraries using unique sequence microclones appears to be simple and efficient. We estimate that for a microdissected region of 10 mb, it is entirely practical to use this procedure to isolate most of the cDNA clones from the dissected region that are present in the cDNA library. It also appears feasible to use this approach to isolate a majority of the cDNA clones even for the entire chromosome 21.

cDNA CHARACTERIZATION BY ARRAYED LIBRARY HYBRIDIZATION

Greg Lennon

We (G. Lennon and H. Lehrach) have developed the first sets of distributed, arrayed, cDNA library filters. Now, work is focussing on the use of these filters for genome scale characterization. These projects include: 1) the isolation of specific genes using "motif" oligonucleotides, 2) the isolation of cDNAs hybridizing to whole cosmid probes, 3) the use of these filters to increase the efficiency of mass cDNA sequencing efforts, and 4) the use of these filters to study gene expression, including the isolation of differentially expressed genes.

INTEGRATION OF cDNA AND GENOMIC LIBRARIES

Hans Lehrach

Identification of coding regions in long stretches of DNA will be an essential step in understanding the function of the mammalian genome. This will require integration of information from genomic, cDNA and exon libraries. As a general approach to this problem, we have constructed a number of different types of libraries in high density filter grids. Libraries can be fingerprinted, partially sequenced by oligonucleotide hybridization, or cross-screened by hybridizing clones or pools of clones to the same or different libraries. Specific experiments include the hybridization of pools of cDNA clones to cDNA, exon, YAC and cosmid filters, hybridization of radiolabelled cDNAs to cDNA or exon clones, and the partial sequence determination of exon and cDNA clone sequences by oligonucleotide hybridization.

HIGH THROUGHPUT AUTOMATED SEQUENCING OF HUMAN BRAIN cDNAs

James M. Sikela

Rapid characterization of large numbers of cDNAs by automated single pass sequencing provides a major resource to the development of a master set of cDNAs. Such a collection of cDNAs would represent a natural starting point for a large scale effort to determine the complete coding region sequences of full-length cDNAs. In addition, the partially sequenced cDNAs would provide tools that could be used for chromosomal localization studies that, if done on a large scale, would generate an important resource for the identification of human genetic disease genes.

In order to avoid repeated sequencing of highly represented cDNAs, we prescreen human brain cDNA libraries with cDNA derived from human brain mRNA in the presence of an excess of cold human genomic DNA. Thus, highly represented clones can be identified and only non-hybridizing clones, likely representing rare or moderately expressed mRNAs, selected. In this way, selection of only the non-hybridizing clones results in an enrichment for cDNAs that are unique from one another. Data will be presented illustrating the efficiency of this strategy.

Once plaques are identified they are individually picked and used for PCR amplification with vector-specific primers that flank the cloning site. After rapid removal of primers, dNTPS, etc., the PCR products are directly sequenced using an ABI 373 automated sequencer and a cycle sequencing protocol.

Sequenced clones can be used to generate STS primers, which can then be used to assign the cDNA to a chromosome by PCR analysis of somatic cell hybrid DNAs. Coordination of cDNA sequencing among the laboratories involved has the potential to significantly improve the efficiency of the process by increasing the overall rate of sequencing while minimizing duplication of effort.

EXPRESSED SEQUENCE TAGS FROM HUMAN *DROSOPHILA* AND *C. elegans*

M.D. Adams, W.R. McCombie, R.F. Moreno, A. Martin-Gallardo and J.C. Venter

One goal of our laboratory is to identify as completely as possible the expressed gene complement of the human brain. Toward this end, we have collected partial sequence data (expressed sequence tags, ESTs) from over 2000 cDNA clones. Over half of these sequences are unrelated to any previously known sequences. The remaining ESTs are split between "junk" sequences (mitochondrial, ribosomal, etc.), and those with significant database similarities. We have found that various brain libraries contain an excellent diversity of clones and that with judicious screening of a relatively small number of highly and moderately represented sequences, large numbers of different genes can be tagged. In order to examine the difficulties to be encountered in achieving closure (finding the last several percent of the expressed genes) and to provide additional model systems in which to study novel genes identified by the EST program, we have begun EST sequencing using *C. elegans* and *Drosophila* cDNA libraries. Over five hundred *C. elegans* clones have been sequenced, revealing a somewhat lower level of diversity than the human libraries. Nonetheless, a relatively small pool of clones could be used as a probe to eliminate a majority of the multiple represented clones, thus enhancing the efficiency of new clone selection. Redundancy probes of human and *C. elegans* clones are currently being constructed. We have also begun construction of a Sybase database application for storage and integration of cDNA sequence analysis data. The Sybase database currently consists of tables related to the sequencing reaction, database searches, mapping information, and tentative formal identification of the gene and gene family. It is anticipated that the extensive genetic manipulations possible in *C. elegans* and *Drosophila* will prove to be useful systems in which to study well conserved, but unknown genes that are identified in all three species.

INTEGRATION OF RESULTS FROM LARGE-SCALE cDNA AND GENOMIC SEQUENCING

Chris Fields

Sequence data are currently being obtained using a variety of strategies, including traditional directed genomic and cDNA sequencing, random or selected genomic STS sequencing, cosmid sequencing, and large-scale partial cDNA (EST) sequencing. Results obtained with one strategy can often aid in analyzing or interpreting those obtained with a different strategy; e.g. a single-pass EST sequence may identify a transcribed region of an otherwise anonymous cosmid sequence. Efficient integration of data obtained using different strategies can, therefore, significantly increase the overall efficiency with which sequence data can be analyzed and genes identified.

Large-scale EST sequencing has proven to be a very efficient means of identifying clones of protein-coding genes. Over 2000 human ESTs and 500 *C. elegans* ESTs have been sequenced at NINDS thus far; of these, roughly 75% correspond to protein-coding genes, of which two-thirds are previously unidentified. This rate can be compared with 1% to 2% of random genomic STS sequences obtained at LANL that can be identified as coding sequences similar to known genes. The accumulation of new coding sequences by EST sequencing will rapidly increase the sequence diversity in the protein sequence databases, which will further increase the efficiency of new gene identification.

Efficient use of this data requires mechanisms for effective data sharing, and analysis tools that allow the incorporation of data from different sources. A cDNA database that will serve the EST sequencing and mapping community is being developed at NINDS, together with tools for automatically searching new cDNA sequences and their translations against both the existing collection and the public databases. These tools are also appropriate for use with genomic STS sequences. The gm genomic sequence analysis system has been enhanced to allow use of partial cDNA sequence data to initiate construction of gene models, and to provide map integration tools in the user interface. These systems are being employed to search for gene in both humans and worms.

PREDICTION OF HUMAN mRNA DONOR AND ACCEPTOR SITES FROM THE DNA SEQUENCE

Soren Brunak, Jacob Engelbrecht and Steen Knudsen

NETGENE is a combination of artificial neural networks that identify splice sites and coding regions. Jointly these networks can predict exons, or parts of exons.

The performance of NETGENE allows it to predict 95% of the donor sites in test sequences, with, on average, only one and a half false predictions per true site.

Another of the capabilities of NETGENE is to predict donor sites that are certain (around 1/5 of the total donor sites can be predicted with certainty).

A RULE-BASED APPROACH TO THE PREDICTION OF GENE STRUCTURE

Roderic Guigo, Steen Knudsen, Neil Drake, and Temple Smith

GENEID is a rule-based system that incorporates a large number of algorithms to predict the exon structure of genes given the pre-mRNA sequence. The current performance of GENEID allows it to predict the exon structure of 4 out of 28 test genes entirely correct, whereas the remaining genes are predicted with varying accuracy. The accuracy of the prediction can be enhanced by giving some experimental information, like the knowledge of a last exon or part of a coding sequence (expressed tagged site). We are currently directing our efforts to localizing transcription signals as well.

BIOLOGICAL CODING PROBLEM AND PRACTICAL ASPECTS OF DISCRIMINANT ANALYSIS

Andrzej K. Konopka

It is well known that statistical characteristics of textual elements (such as frequency distribution of letters) are well preserved in any intelligible text written in a given (sequential) language. By analogy, it is believed that statistical analyses may reveal the structure of "textual elements" associated with a given function of nucleic acid or protein fragments.

The main goal of discriminant analysis is to identify putative functional domains in unannotated nucleic acid sequences. The research leading to this goal has two aspects:

(1) Identify (by computational experiments) function-associated sequence patterns in large collections of functionally equivalent sequences (i.e., determine a classification "code").

(2) Design protocols that will help "translate" observed sequence patterns into discriminant functions usable for mapping.

The presentation is devoted to the first of the above aspects. Four systematic methods of segmenting a text (i.e., string of symbols that belong to a finite alphabet) into morphologically distinct domains will be introduced. Then, it will be shown that each of these methods leads to a good correspondence between known functional domains and morphological domains determined in known nucleic acid sequences.

ARTIFICIAL INTELLIGENCE BASED TOOLS FOR GENE RECOGNITION AND ASSEMBLY FROM DNA SEQUENCE DATA

R. Mural, R. Einstein, X. Guan, R. Mann and E. Uberbacher

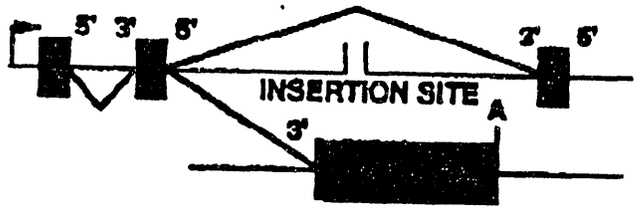
We are building a hybrid neural network and rule-based inference system to examine and characterize regions of anonymous DNA sequence which is based on a new approach to feature identification that uses a multiple sensor-neural network formalism. For example, we have developed a module which correctly recognizes 90% of the coding exons of 100 or more bases long and which has a very low noise level (Uberbacher and Mural, in press). We are constructing additional modules to recognize splice junctions and other features which represent powerful tools that can be used in a stand-alone manner or in an integrated system. These tools have recently been combined with a rule-based interpreter and user interface to allow analysis of DNA sequence over the Internet (Gene Recognition and Analysis Internet Link;GRAIL). User E-mail DNA sequences to the system and have the analysis returned automatically by E-mail. The current analysis includes potential exon positions, with strand assignment and preferred reading frame determination. the GRAIL system has successfully located a number of genes, including 11 verified exons in a 60 kb fragment near the Huntington's locus. The integrated system we are constructing uses the blackboard CLIPS rule-based expert system shell to automatically integrate recognized features into hypothetical gene models. This system consists of a series of hierarchically arranged "blackboard panels" on which connection of potential features and construction of gene fragments is accomplished by a set of logically independent knowledge modules. A tentative coding message will be automatically compared to sequence databases using highly parallel methods. (This research was sponsored by the Office of Health and Environmental Research, U.S. Department of Energy, under contract DE-AC05-840R21400 with the Martin Marietta Energy Systems, Inc.)

EXON TRAPPING SPECIFIC FOR 3' TERMINAL EXONS

David Krizman, Christine Chang and Susan Berget

We have designed an exon trapping scheme directed at isolation of 3' terminal exons. Vertebrate exons are typically quite short. Internal exons have an average length less than 150 bases. Some exons are extremely short; 6 and 7 nucleotide exons are constitutively expressed in certain muscle genes. 3' terminal exons, in contrast, are larger than their internal counterparts. A 1988 study including over 1000 vertebrate genes indicated an average length of terminal exons of 643 nucleotides. Thus, 3' terminal exons on average contain enough contiguous information for applications that directly map genes to chromosomes. The technique we are developing selects terminal exons from a source of genomic DNA and creates a terminal exon library for use in gene mapping.

3' terminal exons are selected by the processing machinery by virtue of the processing sequences that demarcate them. They are preceded by a 3' splice site and they terminate with a poly(A) site. The bulk of the work in our laboratory is on understanding the orchestration of vertebrate splicing. In the last several years we have accumulated a body of evidence that indicates that splice sites are selected a pairs ACROSS EXONS during assembly of the spliceosome. For internal exons, the pair is the 3' splice site preceding the exon and the 5' splice site following the exon. We find that these sites must be within 300 bases of each other, both in vitro and in vivo; this requirement perhaps in part explains the present day small size of vertebrate internal exons. 3' terminal exons have no 5' splice site; instead they terminate with consensus polyadenylation sequences. We have published several reports indicating that the splicing and polyadenylation factors interact during the recognition of terminal exons. Briefly, mutation of splicing signals depresses polyadenylation in vitro and in vivo and mutation of poly (A) signals depresses splicing of terminal introns (but not introns further upstream). These results suggest that poly(A) sites must reside within terminal exons to be maximally recognized in vertebrate precursor RNAs containing multiple introns. In support of this view, isolated poly(A) sites are ignored when placed within vertebrate introns and the gene is spliced normally.



Our exon trapping scheme takes advantage of this property of last exons. Randomly-sheared genomic DNA is inserted within the middle of an intron of a multi-intron gene lacking a terminal exon. In the absence of an insert no stable cytoplasmic mRNA should be produced from this gene. If a terminal exon, however, is inserted, it should be recognized leading to the production of stable poly(A)+ mRNA in the cytoplasm. Isolated 3' splice sites, or isolated poly(A) sites should not score in this assay, reducing non-specific background of the selected sequence population. In fact, we find that a wild type 3' splice site and wild type poly(A) site are required for recognition of a poly(A) site placed within an intron. mRNA populations are expanded and cloned using standard RT-PCR techniques to provide a last exon library. This approach should isolate exons of a reasonable size, and offers the advantage of a POSITIVE selection for clones containing exons.

THE IDENTIFICATION OF CODING SEQUENCES IN MAMMALIAN GENOMIC DNA USING EXON AMPLIFICATION

A.J. Buckler, D.M. Church, A.C. Rogers, J.D. Brook, M.E. MacDonald, J.F. Gusella, and D.E. Housman

Advances in molecular biology have enabled researchers to rapidly clone and sequence extensive portions of several viral and prokaryotic genomes as well as a fraction of the much larger human genome. Physical and genetic organization of the genome are useful for mapping the location of interesting phenotypes with some precision, but do not address the critical problem of identifying structural genes which may comprise less than 1% of the total sequence. To address this problem we have developed a method, exon amplification, to rapidly and efficiently isolate exon sequences from cloned genomic DNA by virtue of selection for functional 5' and 3' splice sites (Buckler et.al., PNAS 88:4005-4009, 1991). Random segments of chromosomal DNA are inserted into an intron present within a mammalian expression vector and, following transfection into a host cell, are amplified and transcribed at high levels. Cytoplasmic mRNA derived from tansfectants is screened by RNA-based PCR amplification for the acquisition of an exon from the genomic fragment. All indications are that this method will be extremely useful for coding sequence identification.

Our preliminary results indicate that exon amplification will facilitate rapid isolation of coding sequences from genomic DNA sources as complex as lambda phage (15-20 Kbp) or cosmid (35-40 Kbp) clones. We have begun to test the effectiveness of this approach by extending this analysis to a large number of genomic clones derived from a variety of human chromosomes, including chromosomes 4, 9, and 19. Approximately 92% of cosmid genomic clones and 65% of lambda phage genomic clones have yielded RNA/PCR products containing potential exon sequences. Although extensive analysis has, at present, been perfromed on only a small number of the cloned products, the majority of these have demonstrated cross-species sequence conservation or have identified cDNAs. Sequence analysis of cloned RNA/PCR, products also identified a class of false positives which arise due to the presence of a cryptic 5' splice site in pSPL1 and a cryptic 3' junction. These products are easily identifiable and can be removed from subsequent analyses. Even in the presence of these false positives, however, this procedure represents a high level of purification of coding sequences from genomic DNA sources.

These preliminary findings have led us to our current research plan, which entails improving the method by devising means to eliminate false positives, and increase its sensitivity and capacity. The results of these studies will enable us to generate libraries of exons from complex sources of mammalian genomic DNA.

EXON TRAPPING

Geoffrey Duyk

We and others have proposed utilizing splicing signals as primary identifiers for genes within cloned genomic DNA. Using genetic screens based on this idea, we have analyzed ten cosmids from the HD region in 4p16.3 and have identified at least three transcribed sequences. This experience has lead to the development of strategies, based on the primary sequence information obtained from the "trapped exons", to identify transcripts and recover cDNA clones. In addition we have developed second and third generation vectors which allow more rapid screening of larger genomic regions and may allow the integration of physical mapping and gene hunting strategies.

PHYSICAL MAP OF HUMAN CHROMOSOME 11

Glen Evans

We are constructing a large scale map of human chromosome 11 using a combination of mapping techniques. These include 1)high resolution *in situ* hybridization using cosmids from an arrayed cosmid library to establish reference landmarks, 2) establishing STS identifiers from each reference landmark, 3) isolating yeast artificial chromosome clones from each reference landmark and 4) filling in gaps in the map using hybridization-based YAC fingerprinting with high density cosmid arrays and recombination-based YAC "walking". This approach has resulted in a reference landmark map of chromosome 11q as well as the construction of several cosmid/YAC contigs of > 2 mb. Inherent in this project is an interest in isolating and characterizing genes within these large contigs by identification of HTF islands, hybridization of YAC clones to cDNA libraries, and expression of genes by introduction of YACs into mammalian cells.

A RECOMBINATION-BASED ASSAY TO ISOLATE GENES

A.J. Hanzlik, M.A. Hauser, M.M. Osemlak, A. Hajra, M. Ray, G.D. Stewart and D.M. Kurnit.

We utilize a rapid recombination-based assay to replace the considerably more time-consuming technique of nucleic acid hybridization. As a result, we can assay efficiently whether few copy sequences are transcribed in a given tissue at a given time of development.

To select for *supF* carriage by λ, we constructed and *E. coli* strain, DM43, that carried *dnaBam* and *lacZam* mutations. This strain selects for phages that have Incorporated a *supF* plasmid by homologous recombination. We also constructed a plasmid based on RGK (not homologous to pBR322) that allows us to screen cDNA libraries via recombination that have sequences homologous to pBR322, including common cDNA libraries constructed in λgt10 and λgt11.

(1) Isolation of few copy sequences.

Copy number is determined by the frequency with which a given insert in a *supF* plasmid retrieves sequences from a genomic library. This rapid methodology enables us to separate few copy, moderately repetitive and highly repetitive sequences.

(2) Determination of the transcriptional activity of few copy fragments in different tissues at different times coupled with concomitant isolation of the gene.

A bank of human fetal (brain, spinal cord, eye, kidney and muscle) and HeLa cDNA libraries was screened via recombination with few copy probes (see (1)) subcloned from 2 YACs localized to the region of 21q22 responsible for Down syndrome. We determined thereby whether transcription occurred in particular tissues at particular times of development. Two thirds of 210 few copy sequences were transcribed by this assay in at least one of the cDNA libraries. Many of the clones were represented in multiple (but not all) cDNA libraries. DNA from flow sorted material enriched for chromosome 21 will be compared to determine whether the distal R-band on chromosome 21 is particularly enriched in expressed sequences. The high degree of transcription that we observe is in accord with our previous hybridization results and indicates that a large proportion of single copy sequences are transcribed. The ease, generality and rapidity of application combine to make the recombination-based assay a method of choice for coupling a genic initiative to the genomic initiative: this allows us to screen for time and tissue of transcription as well as to isolate transcribed sequences.

Our finding that a significant plurality of few copy sequences are transcribed, which was not readily feasible by other technologies, has several corollaries:

(1) A considerable proportion of the few copy segment of the genome is transcribed; thus, the search for genes must be efficient.

(2) Much of the transcription that occurs is nonabundant.

(3) It is necessary to decipher the timing and tissue of transcription of a given sequence.

(4) We do not know as yet if the apparent tissue specificity for nonabundant sequences represents true tissue specificity, "leaky" transcription (whereby a gene is expressed at a low level in a tissue in which the gene is supposed to be inactive), or a combination of the two effects.

IDENTIFICATION OF TRANSCRIBED SEQUENCES BY DIRECT cDNA SCREENING OF GENOMIC LIBRARIES

Ute Hochgeschwender and Miles B. Brennan

Our approach uses direct cDNA screening of arrayed genomic libraries as a method for systematically identifying transcribed sequences in genomic DNA. The cDNA probes can be derived from mRNA of any tissue at any time. The genomic libraries are collections of phage clones (lambda DASH) carrying 10 - 20 kb inserts and are permanently arrayed. The material for the libraries can be total genomic DNA, DNA from specific chromosomes or from parts of chromosomes (YACs).

This approach poses two problems: The background hybridization of highly repeated sequences in the cDNA probe and the limited sensitivity of detection in hybridizations using complex cDNA probes.

To remove highly repeated sequences from the cDNA probes, we prehybridize the probe with a 1000-fold excess of genomic DNA coupled to finely divided cellulose. This method quantitatively removes repeated sequences from the cDNA probes.

To increase the sensitivity of the cDNA probes, we want to reduce the complexity and/or "normalize" the unequal representation of sequences in the cDNA probes. To reduce the complexity of the cDNA probes we size fractionate the template mRNA and then use the separate fractions for making cDNA. While increasing the number of cDNA probes, this method decreases the complexity, and thereby increases the sensitivity, of each probe. Probes are normalized and thereby enriched for low abundance sequences by annealing with an excess of biotinylated "driver" mRNA, and subsequent purification of the single strand cDNA.

The application of these techniques to the screening of arrayed genomic libraries allows the rapid and systematic identification of expressed sequences and the determination of their patterns of expression. This approach is amenable to large-scale applications in identifying transcribed sequences in the human genome.

Hn-cDNA LIBARIES FROM HYBRID CELLS - ENRICHED FOR HUMAN CHROMOSOME REGION SPECIFIC EXONS

Pu Liu and Michael J. Siciliano

Methods to isolate transcribed sequences representing specific regions of the human genome assist the cloning of mapped genes of biomedical interest and the identification of chromosome specific cDNAs. A procedure developed in our laboratory to accomplish that involves the construction of an hn-cDNA library enriched for exon sequences from an interspecific hybrid cell containing the human genomic region of interest as its only human DNA content (Liu et al., 1989, Science 246:813). This procedure enables one to take advantage of the numerous monochromosomal and radiation hybrids available to isolate chromosome specific and region specific genes. Since the procedure requires the use of consensus 5' splice sequences to prime reverse transcriptase cDNA synthesis from poly-A tailed RNA extracted from the hybrid cell, conditions for such hexamer priming are analyzed in a series of primer extension experiments. A sample of sequenced inserts from clones of such a library made from a hybrid cell with portions of human chromosome 17 and 19 indicate that their sequence anatomy is consistent with the philosophy of library construction and screening. The ability of the system to isolate human sequences expressed at low levels is examined as well as methods to evaluate the transcription status of isolated inserts.

IDENTIFICATION OF GENES IN Xq28

Annemarie Poustka

We are currently in the process of establishing ordered cosmid and YAC libraries of the Xq28 region of the human X chromosome (Kloschis and Poustka, unpublished) and will use this region, which has been characterised exceptionally well by physical (Poustka et al., in press) and genetic analysis to test a number of gene identification strategies. This will mainly involve the screening of cosmids with radioactive cDNAs, the hybridisation of cosmid and YAC clones to cDNA libraries, as well as the isolation of conserved and transcribed sequences by a coupled hybridisation/PCR technique. In this protocol biotinylated DNA from pools of cosmid clones is hybridised with either pig DNA fragments, which had been ligated to an adaptor sequence, cDNA clones, or double stranded cDNAs after ligation to vector or adaptor. After isolation of the hybrid, we will use PCR using the flanking oligonucleotide sequences to selectively amplify the pig respectively cDNA sequences. The recovered material can either be labelled radioactively and hybridised to colony filters or Southern blots of restriction digests of the cosmids used in the hybridisation experiment, or can be cloned to give a library enriched in genomic or cDNA sequences with homology to the clones from the region.

IDENTIFICATION OF EXPRESSED SEQUENCES IN DISTAL Xq: EXPERIENCE FROM FRAGILE X AND LOWE'S SYNDROMES AS WELL AS *ALU*-PRIMED hncDNA.

D.L. Nelson, R.L. Nussbaum, S.T. Warren, B.A. Oostra and C.T. Caskey.

Identification of expressed sequences derived from specific genomic regions has been an area of interest in our group for several years. The Alu-primed hncDNA method we developed has been used to identify expressed regions in Xq24-qter present in the X3000-11 cell line. Extension of this method to transfected DNA from yeast artificial chromosomes is under development.

Our successful identification of the *FMR-1* gene at the Xq27.3 fragile site suggests that simple hybridization-based approaches using YAC and cosmid DNAs can identify genes once a locus is represented in cloned genomic fragments. This approach has also been applied with success to genomic DNA crossing a translocation breakpoint associated with Lowe's syndrome, a rare disease involving cataracts, mental retardation and kidney dysfunction. In each of these cases, more exotic methods have not proven of utility, however this does not necessarily rule out their applicability. These methods and the likely reasons for their failure will be presented along with the strategies that have led to identification of transcripts in the fragile X and Lowe's syndromes.

GENE AND CpG ISLAND SEARCH ON HUMAN CHROMOSOME 21

Katheleen Gardiner

The distribution of known genes on human chromosome 21 is non random. Of 31 expressed sequences, 24 map to the distal 1/3 of the long arm, the region which also is highest in G+C content. One way to examine the significance of this distribution is to determine the distribution of CpG islands. This has been approached in two ways. Four megbases of YAC clones have been mapped for clusters of rare restriction sites. In YACs mapping to R bands in the distal 1/3 of the long are, potential clusters occur every 70 Kb (27 clusters in 1700 Kb). In YACs from the proximal 2/3 and G band 21q22.2, only 5 clusters were observed in 2200 Kb (1/450 Kb). Because in these latter regions, CpG islands may be more subtle, eleven NotI and EagI boundary clones were examined for the presence of SmaI, HpaII, HhaI and HaeIII sites. No clusters were observed; in all cases the frequency of such sites was what is expected for a 40% G+C region. Unique sequences from these clones were examined in zoo blots, and exon amplification tests are in progress. These results suggest thtat efficient gen identification strategies may need to be chosen based on characteristics of the region of interest.

APPROACHES FOR IDENTIFYING EXPRESSED SEQUENCES ON YEAST ARTIFICIAL CHROMOSOMES

Mary Kay McCormick

A human chromosome 21 specific YAC library has been constructed from flow sorted chromosomes and is being utilized to generate a physical map of a region which may be important in Down syndrome. Several approaches will be used to identify genes which may be present on YACs mapping to this region, which has been shown to be conserved in several other species. To determine if there are conserved sequences present on the YACs, they will be purified from the yeast background and hybridized to "zoo" blots. YACs containing conserved sequences will be hybridized to cDNA libraries constructed from tissues that are affected in Down syndrome. The possibility of identifying genes present on YACs by homologous recombination with cDNAs or exons introduced into the strain will also be explored. Fragmentation of YACs at repetitive sequences has been documented, and it may be possible to adapt this procedure so it can be used to screen, for example, a YAC contig with a collection of CDNAs or exons. Currently, a contig containing four YACs, representing >500 kb, has been localized by in situ to an area of chromosome 21 involved in Down syndrome. Expansion of this contig and screening for expressed sequences is ongoing.

IDENTIFICATION OF TRANSCRIBED SEQUENCES

S. Parimoo, S.R. Patanjali, N. Baskaran, V. Goei, J.R. Gruen, D.D. Chaplin, H. Shukla and S.M. Weissman

We are in the process of preparing normalized libraries (Patanjali et al., Proc. Natl. Acad. Sci. USA 88:1943-47) of random-primed cDNA fragments from human thymus, spleen, fetal brain, whole fetus, thyroid, duodenum, liver, and retina. Each library will be derived from 3-10 million inserts and will be cloned in lambda gt10 and/or gt11.

We have developed a convenient polymerase chain reaction (PCR) based procedure for hybrid selection to enrich cDNA fragments encoded by large fragments of genomic DNA. We are currently applying the method to a series of yeast artificial chromosomes (YACs) covering about four megabases of human chromosome 6, and including most of the human MHC and the probable locus of the hemochromatosis gene, so as to identify most of the coding sequences of this region. This procedure gave ten-thousand-fold or more enrichment for low abundance cDNAs encoded by a human DNA fragment cloned in a YAC. Only a very low background of repetitive or ribosomal sequences was observed in the selected material. The selection procedure is potentially applicable to parallel analysis of groups of YAC clones by using pulsed-field gel electrophoretically purified and extracted YAC DNA, or directly to DNA blots of fractionated YACs.

The cDNA selection procedure requires relatively small amounts of target DNA fragment, encouraging us to explore further applications such as assigning coding sequences to arrays of clones. In addition, we are examining the overall error rate due to PCR and cloning artifacts, and the possibility of any bias in selecting coding sequences.

DIRECT SELECTION: A METHOD FOR THE SELECTIVE ISOLATION OF cDNAS ENCODED BY LARGE GENOMIC CLONES.

Michael Lovett, John Morgan and Linda M. Hinton.

We have developed a strategy that facilitates the rapid enrichment and identification of cDNAs that are encoded by large genomic regions. The basis of this scheme is the hybridization of an entire library of cDNAs to an immobilized cloned genomic region. Non-specific hybrids are eliminated and selected cDNAs are eluted. These molecules are then amplified and are either cloned or subjected to further selection/amplification cycles. We have used a 550kb yeast artificial chromosome (YAC) clone that contains the erythropoietin (EPO) gene, to select cDNAs from human fetal kidney. We have achieved a 1000-fold enrichment of EPO cDNAs in these experiments and have investigated several of the non-EPO cDNAs that were selectively enriched. The first "anonymous" cDNA that we evaluated is indeed encoded by the EPO YAC. The cDNA was present at a frequency of 1 in 10^6 cDNAs in the starting library but comprised 2% of the selected library of cDNAs, again reflecting an approximately 1000-fold enrichment. DNA sequence analysis of this cDNA and limited regions of the YAC clone revealed that this gene encodes the beta-2 subunit of the human guanine nucleotide-binding proteins (G proteins). This gene GNB2, was previously localized to the long arm of chromosome 7, but its close proximity to the EPO gene was not known. The selective isolation and mapping of GNB2 confirms the feasibility of the direct selection strategy and suggests that it should be useful for the rapid isolation of coding sequences from extensive portions of the human genome.

CHROMOSOME-SPECIFIC cDNAs/ESTs

Marcelo Benro Soares, Pierre Jelec, Maria de Batina Bonaldo, Stephen Brown and Argiris Efstratiadis

We propose to participate in the overall effort directed towards the mapping and sequencing of the human genome by constructing and normalizing cDNA libraries from human fetal and adult tissues, developing methods for assignment of cDNAs to particular chromosomes and chromosome regions, and using the assigned cDNAs to generate ESTs. Our strategy involves the generation of cDNA libraries using noval vectors (lafmids) that we have constructed and tested. Following cDNA cloning in lafmids, single-stranded versions of the libraries can be obtained and used for library normalization by a kinetic approach. Hybridization between normalized libraries and chromosome-specific libraries can then be employed for assignment of cDNAs to chromosomes. Variations of this approach also allow assignment of cDNAs to chromosomal regions. The assigned cDNAs and the corresponding exon-containing genomic clones will be used as sources of ESTs and STSs, respectively.

ALPHABETICAL LIST OF PARTICIPANTS

Dr. Susan Berget
Dept. of Biochemistry
Baylor College of Medicine
1 Baylor Plaza
Houston, TX 77030
Tel 713-798-5758
Fax 713-796-9438

Dr. Miles B. Brennan
Unit on Genomics
NIMH
Bldg. 10, Rm 4N320
Bethesda, MD 20892
Tel 301-402-1769
Fax 301-402-2140

Dr. Alan Buckler
Massachusetts Institute of Technology
Center for Cancer Research
77 Massachusetts Avenue, E17-543
Cambridge, MA 02139
Tel 617-253-3015
Fax 617-253-5202

Dr. Carol Dahl
HGR, OSR
Bldg. 38A, Room 606
NIH
Bethesda, MD 20892
Tel 301-402-0838
Fax 301-480-2770

Dr. Geoffrey Duyk
Dept. of Genetics
Harvard Medical School, HHMI
25 Shattuk St.
Boston, MA 02115
Tel 617-232-7667
Fax 617-732-7663

Dr. Linda Engel
HGR, OSR
Bldg. 38A, Room 606
NIH
Bethesda, MD 20892
Tel 301-402-0838
Fax 301-480-2770

Dr. Glen Evans
The Salk Institute
P.O. Box 85800
San Diego, CA 92138-9216
Tel 619-453-4100 x279
Fax 619-558-9513

Dr. Chris Fields
RBMB/LMCN/NINDS/NIH
Park Bldg Room 405
Bethesda, MD 20892
Tel 301-496-8800
Fax 301-480-8588

Dr. Kathleen Gardiner
Eleanor Roosevelt Institute
1899 Gaylord St.
Denver, CO 80206
Tel 303-333-4515
Fax 303-333-8423

Dr. Betty Graham
HGR, OSR
Bldg. 38A, Room 606
NIH
Bethesda, MD 20892
Tel 301-402-0838
Fax 301-480-2770

Dr. Ute Hochgeschwender
Unit on Genomics
NIMH
Bldg. 10, Rm 4N320
Bethesda, MD 20892
Tel 301-402-1769
Fax 301-402-2140

Dr. David Housman
Professor of Biology
Massachusetts Institute of Technology
Center for Cancer Research
77 Massachusetts Avenue, E17-543
Cambridge, MA 02139
Tel 617-253-3013
Fax 617-253-5202

Dr. Fa-Ten Kao
Senior Fellow
Eleanor Roosevelt Institute for Cancer Research
University of Colorado
1899 Gaylord Street
Denver, CO 80206
Tel 303-333-4515
Fax 303-333-8423

Dr. Steen Knudsen
Engineering Research Ctr.
Boston University
36 Cummington St.
Boston, MA 02215
Tel 617-353-7123
Fax 617-353-7020

Dr. Andrzej Konopka
NCI/DCBDC Cancer Research
Bldg 469/Rm 151
Fredrick, MD 21702
Tel 301-846-5793
Fax 301-846-5598

Dr. David Kurnit
Howard Hughes Medical Institute
University of Michigan Medical School
Ann Arbor, MI
Tel 313-747-4747
Fax 313-936-9353

Dr. Hans Lehrach
ICRF
Lincoln's Inn Fields
London, WC2A 3PX
U.K.
Tel 071-269-3308
Fax 071-405-4303

Dr. Greg Lennon
ICRF
Lincoln's Inn Fields
London, WC2A 3PX
U.K.
Tel 071-269-3308
Fax 071-405-4303

Dr. Mike Lovett
Genelabs, Inc.
505 Penobscot Dr.
Redwood City, CA 94063
Tel 415-369-9500
Fax 415-368-0709

Dr. Mary Kay McCormick
M-886
Los Alamos National Laboratory
Los Alamos, NM 87545
Tel 505-665-4438
Fax 505-665-3024

Dr. Richard Mural
Biology Division
Oak Ridge National Lab
P.O. Box 2009
Oak Ridge, TN 37831
Tel 615-576-2938
Fax 615-574-1274

Dr. David Nelson
Baylor College of Medicine
Rm T8-17
One Baylor Plaza
Houston, TX 77030
Tel 713-798-6552/4787
Fax 713-798-6521

Dr. Nancy Pearson
HGR, OSR
Bldg. 38A, Room 606
NIH
Bethesda, MD 20892
Tel 301-402-0838
Fax 301-480-2770

Dr. Anne Marie Poutska
In Neuenheimer Feld 506
German Cancer Research Center
Institute for Virus Research
9600-Heidelberg
Germany
Tel 49-6221-42-3409
Fax 49-6221-470-333

Dr. Michael Siciliano
Dept. of Molecular Genetics, Box 45
M.D. Anderson Cancer Center
Univ. of Texas
Houston, TX 77030
Tel 713-792-2910
Fax 713-794-4295

Dr. James Sikela
Assistant Professor
University of Colorado Health Sciences Center
4200 East 9th Avenue, Box C236
Denver, CO 80262
TEL 303-270-8637
Fax 303-270-7097

Dr. Karen J. Skinner
Developmental Neurology Branch
Division of Convulsive, Developmental and Neuromuscualar Disorders
Federal Building, 8C04
7550 Wisonsin Ave
Bethesda, MD 20892
Tel 301-496-5821
Fax 301-402-0302

Dr. Bento Soares
Dept. of Psychiatry
Columbia University
722 W. 168th St.
Box 41
New York, NY 10032
Tel 212-960-2313
Fax 212-795-5886

Dr. Marvin Stoldolsky
Dept. of Energy
Health Effects and Life Sciences
Research Division, ER-72
Office of Health & Environmental Research
Office of Energy Research
Washington, DC 20585

Dr. J. Gregor Sutcliffe
RISC, MB10
10666 N. Torrey Pines Rd.
La Jolla, CA 92037
Tel 619-554-8064
Fax 619-554-6112

Dr. J. Craig Venter
Dr. Antonio Martin
Dr. Mark Adams
Section Receptor Biochemistry and Cellular Neurobiology
NIH/NINDS
Bethesda, MD 20814
Tel 301-496-8800
Fax 301-480-8588

Dr. Sherman M. Weissman
Dept of Human Genetics
Yale University School of Medicine
333 Cedar St.
P.O. Box 3333
New Haven, CT 06510-8005
Tel 203-785-2667
Fax 203-785-3033

END

DATE FILMED

2 / 10 / 93