Beyond the Identification of Transcribed Sequences: Functional and Expression Analysis

11th Annual Workshop

November 9-12, 2001
Washington D.C.

Co-sponsored by the U.S. Department of Energy

Abstracts * Speakers

Meeting Objective

Topics discussed included: mammalian gene and genome organization as determined from the construction of transcriptional maps and genomic sequence analysis; expression analysis of novel mammalian genes; analysis of genomic sequence, including gene and regulatory sequence prediction and verification, and annotation for public databases; expression and mutation analysis, and comparative mapping and genomic sequence analysis in model organisms (e.g. yeast, *C.elegans*, *drosophila*, zebrafish, pufferfish, chicken, mouse rat); construction and analysis of transgenic organisms; evolutionary comparisons; novel approaches for functional analysis of transcribed sequences; construction of full length and 5' specific cDNA libraries; cDNA array screening and analysis; RNA processing, including RNA editing, RNAi, antisense RNA and regulatory sequences; database construction, management and use in expression and functional analysis; and analysis of protein structure and function.

Organizers:

- Giorgio Bernardi, Stazione Zoologica, Anton Dohrn, Napoli, Italy
- Bernhard Korn, German Cancer Research Center, Heidelberg
- Katheleen Gardiner, Eleanor Roosevelt Institute, Denver, CO, USA
- Sherman Weissman, Yale University, New Haven, CT, USA
- Thomas Werner, Institute for Saeugertiergenetik, Oberschleissheim, Germany

For Questions or Additional Information, contact:

Katheleen Gardiner, gardiner@eri.uchsc.edu
Tel: 303-336-5652

Nan Matthews, themattgrp@qwest.net
Tel: 303-843-6414

Highlights from the 2000 Workshop

- See "Report of the 10th International Workshop on the Identification of Transcribed Sequences, 2000," *Cytogenetics and Cell Genetics*, Volume 92, pages 49-58, 2001

Beyond the Identification of Transcribed Sequences:
Functional and Expression Analysis
11th Annual Workshop
November 9-12, 2001
Washington D.C.

Abstracts

Bruce Aronow: Universal Reference Approach to the Creation and Mining of a General Purpose Mouse Gene Expression Database

Arman A. Bashirova: Comparative analysis of the CD209 (DC-SIGN)-related genes in primates

Giorgio Bernardi: Misunderstandings about isochores: A review

Juergen Bode: Fragile Sites, Bordering Elements and Chromatin Domains: Lessons from the Type 1 Human Interferon Gene Cluster on 9p22

Alvis Brazma: From Microarray Data to Gene Expression Networks

John G. Conboy: Tissue-specific regulation of alternative splicing in protein 4.1R pre-mRNA

Marc Damelin: Genome-wide localization of a chromatin-remodeling factor demonstrates its role in orchestrating the transcriptional program of the TOR signaling pathway

Robert B. Darnell: Identifying RNA-protein interactions important in neuronal function

Olof Emanuelsson: Elucidating a "theoretical" proteome of the *Arabidopsis thaliana* thylakoid

Tom Freeman: HGMP-RC Microarray Programme - Facing the Challenge of High Throughput Microarray Fabrication and Analysis

Elena I. Frolova: Functional analysis of the promoter for the long isoform of collagen a1(IX) in ovo

Katheleen Gardiner: Never-ending annotation of human chromosome 21: alternative splicing, antisense transcripts and mouse conservation

Mark Gerstein: Integrative Genomics

Yoram Groner: Spatial and temporal expression pattern of Runx3 (aml2) and Runx1 (aml1) indicates non-redundant functions during mouse embryogenesis

S. Blair Hedges: Human and ape molecular clocks and the origin of bipedalism

John Hesketh: mRNA targeting by sequences in the 3'untranslated region promotes nuclear protein import

Geoff Hicks: Functional Analysis of Genetic Determinants of Human Disease

Winston Hide: Diversity in gene expression: Assessment of exon skipping and expression states

Bernhard Korn: Characterization of novel Proteins by Analysis of Human Protein Networks: Part1 - cloning of human full ORFs of novel proteins

Bernhard Korn: Experimental verification of predicted splice variants of human genes

Anuj Kumar, Mike Snyder: Functional Genomics by Transposon-Tagging and ChIP-Chip Analysis in Yeast

Urban I. Liebel: Development of a Microscope-based High Content Screening Platform

Shuo Lin: PRODUCTION OF FERTILE TRANSGENIC ZEBRAFISH FROM LONG-TERM CULTURED CELLS BY NUCLEAR TRANSFER

C Louvet: Real-time quantitative PCR : A fast and efficient method to screen clones derived from Subtractive Suppression Hybridization

Christian Maercker: Gene expression profiling analysis shows differential expression of extracellular matrix genes in left ventricular hypertrophy of renal failure

Stephen Munroe: Alternative processing of thyroid hormone receptor mRNA: Role of cis-acting sequences and antisense RNA

Richard Mural: WHOLE GENOME SHOTGUN ASSEMBLY OF THE MOUSE GENOME

Axel Nagel: GABI Primary Database: Bioinformatics Resource for Plant Genome Data

Laszlo Nagy: Using quantitative and global expression profiling to explore and quantitate nuclear hormone receptor function in myeloid cell differentiation and metabolism

Helen Parkinson: ArrayExpress - a public repository for Gene Expression Data at the European Bioinformatics Institute

Rainer Pepperkok: Systematic subcellular localisation and functional characterization of novel proteins identified by large-scale cDNA sequencing

Graziano Pesole: STRUCTURAL AND COMPOSITIONAL ANALYSIS OF TRANSCRIPTION START SITES AND 5'-UNTRANSLATED REGIONS IN HUMAN MRNAS

K.A. Resing: Proteomics Analysis of Melanoma Cell Lines

Martin Ringwald: The Mouse Gene Expression Database (GXD)

Yikang S. Rong: A generally applicable gene targeting method in *Drosophila melanogaster*

Serge Saxonov: Annotating Proteomes using Highly Specific Protein Scoring Matrices

Sabrina Semprini: An improved method for cloning EST-specific full-length cDNAs

Jeremy C. Simpson: ‘Green Genes’ - Systematic subcellular localisation of novel proteins identified by large-scale sequencing identifies new regulators of membrane transport

Dobromir Slavov: Phylogenetic comparison of the RNA editase ADAR2 genes reveals conservation and diversity in editing site sequence and alternative splicing patterns

Richard D. Smith: New Methods for the Characterization of Proteomes: Initial Application to *D. radiodurans*

Marcelo Bento Soares: Strategies for construction of subtracted libraries enriched for full-length cDNAs and for preferential cloning of rare mRNAs

Stefan Stamm: The alternative exon database (AEDB)

Eugene Sverdlov: A technique for genome-wide identification of differences in the interspersed repeats integrations between closely related genomes

Attila Szanto: Interaction between PPAR-gamma and Retinoic Acid Receptor (RAR) pathways during the differentiation of monocytic leukemia cells

Annabel E. Todd: No title

P. Van Sloun: Functional Analysis of the Mammalian Genome by Large Scale Gene Trap Mutagenesis

Andreas Wagner: Robustness and Redundancy in Large Genetic Networks

Sherman M. Weissman: Analysis of myeloid cell development at the RNA and protein level

Ruth Wellenreuther: Generating and Sequencing Full-length cDNAs of Novel Human Genes within the German cDNA Consortium

Thomas Werner: Elucidation of Gene Function Must Include the Genomic Context

Stefan Wieman: IDENTIFICATION OF TRANSCRIBED SEQUENCES: MORE CHALLENGES?

Miles Wilkinson: An RNA Surveillance Mechanism that Detects Aberrant Transcripts

Andreas Winterpacht: Dissection of a complex phenotype by gene identification, gene function analysis and comparative sequencing between man and mouse

Michael Q. Zhang: CART CLASSIFICATION OF HUMAN 5'UTR SEQUENCES

Beyond the Identification of Transcribed Sequences:
Functional and Expression Analysis

11th Annual Workshop
November 9-12, 2001
Washington D.C.

2001 Confirmed Speakers

Giorgio Bernardi
Stazione Zoologica Anton Dohrn
Napoli, Italy

Michael Bittner
NIH/NHGRI
Bethesda, Maryland, USA

Jeurgen Bode
GBF, German Research Center for Biotechnology
Braunschweig, Germany

Alvis Brazma
European Bioinformatics Institute
Hinxton, Cambridge, UK

J. G. Conboy
Lawrence Berkeley National Laboratory
Berkeley, California, USA

Marc Damelin
Dana Farber Cancer Institute
Boston, Massachusetts, USA

R. B. Darnell
Rockefeller University
New York, New York, USA

Olof Emanuelsson
Stockholm University
Stockholm, Sweden

Thomas Freeman
The Sanger Centre
Hinxton, Cambridge, UK

Rainer Pepperkok
EMBL
Heidelberg, Germany

Graziano Pesole
Dipartimento di Fisiologia e Biochimica Generali
Milan, Italy

Katheryn Resing
University of Colorado
Boulder, Colorado, USA

Martin Ringwald
The Jackson Laboratory
Bar Harbor, Maine, USA

Yikang Rong
The Stowers Institute for Medical Research
Kansas City, Missouri, USA

Serge Saxonov
Brutlag Laboratory, Stanford University
Stanford, California, USA

Richard Smith
Pacific Northwest National Laboratory
Richland, Washington, USA

Marcelo Bento Soares
University of Iowa
Iowa City, Iowa, USA

Stefan Stamm
University of Erlangen-Nuremberg
Erlangen, Germany

Katheleen Gardiner
Eleanor Roosevelt Institute
Denver, Colorado, USA

Mark Gerstein
Yale University
New Haven, Connecticut, USA

S. Blair Hedges
Pennsylvania State University
University Park, Pennsylvania, USA

John Hesketh
University of Newcastle
Newcastle-upon-Tyne, UK

Geoffrey Hicks
Manitoba Institute of Cell Biology
Manitoba, Winnipeg, Canada

Winston Hide
South African National Bioinformatics Institute
Bellville, South Africa

Bernhard Korn
RZPD -- Ressourcenzentrum fur Genomforschung
Heidelberg, Germany

Anuj Kumar
Yale University
New Haven, Connecticut, USA

Shuo Lin
University of California - Los Angeles
Los Angeles, California, USA

Richard Mural
Celera Genomics
Rockville, Maryland, USA

Annabel Todd
University College London
London, UK

Andreas Wagner
University of New Mexico
Albuquerque, New Mexico, USA

Sherman Weissman
Yale University School of Medicine
New Haven, Connecticut, USA

Thomas Werner
Genomatix Software
Munchen, Germany

Stefan Wiemann
German Cancer Research Center
Heidelberg, Germany

M. F. Wilkinson
M.D. Anderson Cancer Center
Houston, Texas, USA

Edgar Wingender
AG Bioinformtik
Braunschweig, Germany

Michael Q. Zhang
Cold Spring Harbor Laboratory
Cold Spring Harbor, New York, USA

Universal Reference Approach to the Creation and Mining of a General Purpose Mouse Gene Expression Database

Bruce Aronow
CHRF 2048
3333 Burnet Ave
Cincinnati, OH 45229
telephone: 513 636-4865
fax:
email: bruce.aronow@chmcc.org
prestype: Platform
presenter: Bruce Aronow

Bruce J Aronow, Sarah Williams, Cathy Ebert, and a consortium of UC/CHMC investigators.
Divisions of Molecular Developmental Biology, Pediatric Informatics, and University of Cincinnati Genome Informatics Core. University of Cincinnati and Children's Hospital Medical Center, Cincinnati, OH, USA

We have analyzed mRNA expression profiles of 81 normal, developing and disease mouse tissues using Incyte MouseGEM1 microarrays and a single common reference mRNA with a strong emphasis on adult and developing lung, cardiac, CNS, GI, urogenital, immunologic, and endocrine tissues. Duplicate Cy3/Cy5 hybridizations with Agilent Bioanalyzer-graded mRNAs
and day 1 whole mouse mRNA reference demonstrated excellent reproducibility (even with respect to genes expressed at very low level in the reference mRNA), equivalency to dye reversal, and agreement with direct sample comparisons. Use of multiple independent normalization strategies greatly improved quality assurance, optimal replicate correlations, as well as extraction of tissue, organ, and gene ontology-specific expression pattern relationships. Excluding the most over-expressed genes reduced ability to classify tissue specificity, but less so organ origin. Tissues from CNS, immunologic, and GI systems exhibited impressive expression diversity and repertoire specificity suggesting both subtle and intense commitment to tissue-specific gene expression programming. Probing for correlated genes with known biologic relationships within multiple gene ontologies and other known biologic relationships demonstrated great potential of the database to implicate functional associations and potential pathway relationships for unknown genes. These results support the hypothesis that systematic database mining by cross-comparative analysis of diverse biologic systems will greatly augment gene discovery, annotation, and pathway knowledge.
(supported by multiple NIH grants of consortium members and the Howard Hughes Medical Institute)

Comparative analysis of the CD209 (DC-SIGN)-related genes in primates

Arman A. Bashirova
Laboratory of Genomic Diversity
NCI-Frederick
Bldg 560
Rm 21-70
Frederick MD, 21702
telephone: 301-846-5323
fax: 301-846-1909
email: bashirovaa@mail.ncifcrf.gov
prestype: Poster
presenter: Arman A. Bashirova

Arman A. Bashirova, Jie Cheng, Maureen P. Martin, and Mary Carrington

CD209(DC-SIGN) and CD209L(DC-SIGNR/L-SIGN) were recently identified on human chromosome 19 as highly similar genes encoding C-type lectins that serve as adhesion receptors for ICAM-2/3 and HIV-1 gp120. The two genes are expressed differentially in human tissues, with CD209 expressed mainly on dendritic cells and CD209L expressed on sinusoidal cells of liver and lymph node. As the two genes are obviously products of recent duplication, we were interested in their evolutionary history. Using PCR/sequencing and Southern blot techniques, we confirmed the presence of CD209 in old world primates. Also, we identified a third gene of this family that we named CD209L2 in old world monkeys and apes, including baboons, macaques, gibbons, orangutans, and chimpanzees. Interestingly, there was no evidence for an orthologue of CD209L2 in humans. On the other hand, CD209L (which we now propose to call CD209L1) appears to be absent in old world monkeys, and present only in apes and humans. Our data suggest that the CD209 gene family has undergone recent evolutionary processes involving duplication/deletion

Misunderstandings about isochores: A review

Giorgio Bernardi
Stazione Zoologica Anton Dohrn
Villa Comunale 80121 Napoli
Italy
telephone: 0039 081 5833300
fax: 0039 081 2455807
email: bernardi@alpha.szn.it
prestype: Platform
presenter: Giorgio Bernardi

Giorgio Bernardi

Laboratorio di Evoluzione Molecolare
Stazione Zoologica Anton Dohrn
Napoli 80121
Italy

The draft sequence paper published by the International Human Genome Sequencing Consortium (IHGSC) is different from previous sequence reports (including the paper published simultaneously by Venter et al., 2001), which presented data and addressed issues of sequence analysis and gene prediction, in that the IHGSC attempted to also present a general picture of a very broad and complex research area, that of the organization and evolution of the human genome. This attempt was apparently too ambitious, also in view of the time and space limitations imposed on the authors. As a consequence, some erroneous and controversial conclusions found their way into the paper. Since, in all likelihood, the IHGSC article will have a very wide circulation, it is important that such conclusions be corrected or critically discussed before they spread into the literature and become (at least temporarily) established truths. This discussion of the IHGSC paper is, in fact, already under way on some of the topics addressed. For instance, the proposed horizontal transfer of bacterial genes to vertebrates was shown to be explained, in most cases, by descent through common ancestry (Stanhope et al. 2001; Roelofs and Van Haastert, 2001). Here, the discussion will be focused on some subjects of the IHGSC paper previously dealt with in our laboratory, such as the broad genomic landscape, namely the isochore pattern of the human genome, the distribution of repeats and genes in the isochores, and the mutational bias, i.e. the non-randomness of the mutational input.

Fragile Sites, Bordering Elements and Chromatin Domains: Lessons from the Type 1 Human Interferon Gene Cluster on 9p22

Juergen Bode
Gesellschaft für Biotechnologische Forschung
mbH / EPI
Mascheroder Weg 1
38124 Braunschweig
telephone: +49 531 6181 251
fax: +49 531 6181 262
email: jbo@gbf.de
prestype: Platform
presenter: Juergen Bode

J. Bode 1), A. Baer 1), E. Ernst 1), S. Goetze 1), A. Knopp 1), K. Nehlsen 1), C. Benham2)
1)German Research Center for Biotechnology (GBF)/EPI, D-38124 Braunschweig, jbo@gbf.de;
2)UC Davis,Div. Biol. Saciences, Genome Center benham@blue.ucdavis.edu

S/MARs have been discovered more than a decade ago and have been defined as DNA-elements staying attached to or associating with the nuclear skeleton after the extraction of the histones and soluble factors from eukaryotic nuclei (1). While S/MARs do not conform to any obvious sequence consensus, their recognition is governed by structural features, most significantly a propensity to expose single strands under negative superhelical tension. This property has been used to localize S/MARs in SIDD (stress-induced duplex destabilization) profiles, to explain their illegitimate recombination potential (4) which becomes evident by

- their preference as integration for retroelements
- their coincidence with fragile genomic sites.

We are in the process of predicting their participation in transcription and replication processes and to use biomathematics to guide their design for various biological and biotechnological applications.

The first widely accepted activity of S/MARs was the augmentation of transcription initiation rates which is distinct from enhancement (1). Since then a variety of additional functions have been delineated acting on transcriptional competence, and providing enhancer or origin of replication support. This overlap of functions has made difficult the unambiguous demonstration of any of these components. It was originally for this reason that we have refined techniques based on site-specific recombination systems like Flp/FRT (3,7). With these techniques, complete chromatin domains cannot only be decomposed but also be elaborated at a predefined chromosomal locus. Our most advanced system, the recombinase-mediated cassette exchange (Flp-RMCE) permits the mutual exchange of cassettes which are flanked by an FRT-site and an FRT-mutant, resp. (3,7).

So far the stable modification of target cells is mostly achieved by integrating vectors(2). For gene therapy purposes, derivatives of retroviral vectors have found extensive use although their expression may be rapidly silenced. While methylation-dependent silencing can be controlled by the presence of S/MARs(1), major recent efforts have been devoted to the use of S/MAR-ori sequences from the human genome to obtain vectors which replicate autonomously providing a stable and high-level expression. 5,6).

Recent reviews:

1. J. Bode, C. Benham, A. Knopp and C. Mielke (2000). Transcriptional Augmentation: Modulation of Gene

Expression by Scaffold/Matrix Attached Regions (S/MAR Elements).Crit. Rev. Eukaryot. Gene Expr. 10, 73-90.
2. A. Baer, D. Schübeler and J. Bode (2000). Transcriptional Properties of genomic transgene integration sites marked by electroporation or retroviral infection. Biochemistry 39, 7041-7049.
3. J. Bode, T. Schlake, M. Iber, D. Schübeler, J. Seibler, E. Snezhkov and L. Nikolaev (2000). The transgeneticist´s toolbox - Novel methods for the targeted modification of eukaryotic genomes. Biol. Chem. 381, 801-813.
4. J. Bode, C. Benham, E. Ernst, A. Knopp, R. Marschalek, R. Strick, and P. Strissel (2000). Fatal connections: When DNA ends meet on the nuclear matrix. J. Cell. Biochem. Suppl. 35, 3-22, http://www3.interscience.wiley.com/cgi-bin/issuetoc?ID=82002725.
5. J. Bode, C. P. Fetzer, K. Nehlsen, M. Scinteie, Bok-Hee Hinrich, A.Baiker, C. Piechazcek, C. Benham and H. J. Lipps (2001). The Hitchhiking Principle: Optimizing episomal vectors for the use in gene therapy and biotechnology. Int. J. Gene Ther. Mol. Biol 6, 33-46. www.gtmb.org
6. H. J. Lipps and J. Bode (2001). Exploiting chromosomal and viral strategies: The design of safe and efficient non-viral gene transfer systems. Curr. Opin. Mol. Therapeut. 3, 133-141.
7. A. Baer and J. Bode (2001). Coping with kinetic and thermodynamic barriers: RMCE, an efficient strategy for the targeted integration of transgenes. Curr. Opin. Biotech. Curr. Opin. Biotech. 12, 473-480.

From Microarray Data to Gene Expression Networks

Alvis Brazma
European Bioinformatics Institute
Wellcome Trust Genome Campus, Hinxton, Cambridge, UK
Phone: 1223-494658; Fax: 1223-494468
Email: brazma@ebi.ac.uk

The handling and analysis of the huge amounts of microarray data are becoming the major bottlenecks in the utilization of the microarray technology. Storing and annotating these data is not a trivial problem due to many reasons. The raw microarray data are images, which have to be transformed into gene expression matrices -- tables where rows represent genes, columns represent various samples such as different tissues, and values at each position characterize the expression level of the particular gene in the particular sample. These matrices have to be integrated with other genomic data and analyzed further, if any knowledge about the underlying biological processes is to be extracted (see [1]). European Bioinformatics Institute initiated an international effort to establish standards for microarray data representation, annotation and exchange [2]. An XML based data exchange format - MicroArray Gene Expression Markup Language (MAGE-ML) it being developed in collaboration with Microarray Gene Expression Database (MGED) Group (see www.mged.org). EBI is establishing a public repository for microarray data ArrayExpress, which will accept data in MAGE-0ML format. Online tools for gene expression data analysis are available (www.ebi.ac.uk/microarray).

We study dependencies between the gene expression profiles in a dataset from genome wide yeast mutation studies, regarding the profiles as random variables. We build gene expression dependency networks from these data and study their properties. We look for 'important' genes, i.e., genes with high out-degree in the dependency graph, and genes with complex regulation, i.e., genes with high in-degree in the graph.

References

1. A. Brazma, and L. Vilo, Gene Expression Data Analysis. FEBS Letters 480 (2000) 17-24.
2. A. Brazma, A. Robinson, G. Cameron, M. Ashburner. One-stop shop for microarray data. Nature, Vol 403, 17 February, 2000, p. 699-700

Tissue-specific regulation of alternative splicing in protein 4.1R pre-mRNA

John G. Conboy
Lawrence Berkeley National Laboratory
Mailstop 74-157
1 Cyclotron Road
Berkeley, CA 94720
telephone: 510-486-6973
fax: 510-486-6746
email: jgconboy@lbl.gov
prestype: Platform
presenter: John G, Conboy

Victor C. Hou, Robert Lersch, Sherry L. Gee, Marilyn Parra, Michael Wu, Chris W. Turck, Mark Koury, Adrian R. Krainer, Akila Mayeda, and John G. Conboy

Lawrence Berkeley National Laboratory, Life Sciences Division/ Mailstop 74-157, 1 Cyclotron Road, Berkeley, CA 94720; University of California, Berkeley, Dept. of Molecular and Cellular Biology, Berkeley, CA 94720; University of California, San Francisco, Howard Hughes Medical Institute, Dept. of Medicine and Cardiovascular Research Institute, San Francisco, CA 94143; Dept. of Medicine, Vanderbilt University, Veterans Affairs Medical Centers, Nashville, TN 37232; Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724; University of Miami School of Medicine, Dept. of Biochemistry and Molecular Biology, Miami, FL 33136.

The single gene for erythroid skeletal protein 4.1R is an extraordinarily complex locus on human chromosome 1 that encodes a large family of isoforms varying in size, subcellular localization, and functional interactions with other proteins. Much of this diversity is generated through the mechanism of tissue-specific alternative pre-mRNA splicing and, preliminary evidence suggests, alternative transcriptional events at multiple alternative first exons. The magnitude of the complexity is such that 15 of the 26 known exons comprising this large gene (>220kb) are alternatively expressed via these processes, and several are regulated in tissue- and developmental-specific patterns. Similarly complex regulation of RNA processing also characterizes expression of three paralogous genes, 4.1G, 4.1N, and 4.1B. In the 4.1 gene family alone, there are multiple alternative exons with tissue-specific or tissue-restricted patterns of expression in erythroid cells, epithelial cells, muscle, and brain. Given recent estimates that a majority of human genes exhibit some form of alternative splicing, it seems clear that regulation of RNA processing is a major mechanism of cellular differentiation, and that identification of genomic regulatory sequences is of considerable importance.

We are studying the molecular mechanisms regulating three regulated splicing events in the 4.1R gene: activation of exon 16 in differentiating erythroid cells; activation of exon 17B in mammary epithelial cells; and regulation of exon 2' splicing that controls expression of an alternative translation initiation site. Here we will focus on recent experiments that provide new insights into the regulation of exon 16 (E16). E16 expression is controlled by a physiologically important alternative splicing "switch" that operates at a specific stage of erythroid differentiation. E16, 63nt in length, encodes a critical region of the spectrin-actin binding domain and is essential for normal membrane mechanical properties. E16 is skipped during 4.1R pre-mRNA splicing in early erythroid progenitors but efficiently included in 4.1R mRNA in later progenitors. A major goal of these studies is to identify both the regulatory sequences in 4.1 pre-mRNA and the splicing factors that bind to these sequences to mediate this splicing switch.

For these studies we constructed model 4.1R pre-mRNAs containing wild type or mutated E16 sequences, assayed the splicing of these pre-mRNAs in vitro (using HeLa cell nuclear extracts) or in vivo (in transfected

cells), and characterized the spliced products using RT/PCT techniques. Recent experiments have identified two elements within exon 16- a 5' purine-rich element (PRE16), and 3' evolutionarily conserved element (CE16)- that are critical to the repression of E16 splicing in early erythroid cells. A combination of experimental approaches including site specific mutagenesis of exon 16, RNA-protein binding assays, splicing factor depletion and add-back experiments with recombinant proteins, and Western blot analysis of splicing factors in differentiating erythroid cells, all support the following model. In early erythroid progenitors, high nuclear concentrations of hnRNP A1 protein, a known splicing inhibitor protein, represses exon 16 splicing by binding to two "silencer" elements within exon 16. One site is located in PRE16 and a second site is in CE16; both silencers are required for optimal repression. As erythroblasts mature, there is a several-fold decrease in cellular A1 levels that temporally correlates with the activation of exon 16 splicing. Control experiments suggest that these effects are specific, i.e., that other known splicing factors do not correlate in activity or expression patterns with the E16 splicing switch. Remarkably, the correlated changes in A1 expression and E16 splicing can be observed in culture using a mouse erythropoiesis model system. These findings demonstrate that natural developmental changes in hnRNP A protein expression can effect physiologically important switches in pre-mRNA splicing through interactions with silencer elements within E16. We speculate further that detailed biochemical studies of this type, in combination with computational analysis of nucleotide sequences near other regulated alternative exons, will eventually reveal the rules governing tissue-specific regulation of splicing in higher eukaryotes.

Genome-wide localization of a chromatin-remodeling factor demonstrates its role in orchestrating the transcriptional program of the TOR signaling pathway

Marc Damelin
Dana-Farber Cancer Institute
Smith Building 922
44 Binney Street
Boston, MA 02115
telephone: (617) 632-5104
fax: (617) 632-5103
email: damelin@fas.harvard.edu
prestype: Platform
presenter: Marc Damelin

Marc Damelin (1,2), Itamar Simon (3), Frederick P. Roth (1), Richard A. Young (3), Bradley R. Cairns (4), and Pamela A. Silver (1,2)

(1) Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115.
(2) The Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115.
(3) Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142.
(4) Howard Hughes Medical Institute and Department of Oncological Science, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT 84112.

We have found that the RSC chromatin-remodeling complex in Saccharomyces cerevisiae coordinates programs of gene expression regulated by the TOR pathway, which regulates cell growth in response to nutrient availability and stress. The connection became evident from the genome-wide localization of Rsc9, a novel RSC component. The experiment explores protein-DNA interactions at the genomic level by combining chromatin immunoprecipitation and microarray analysis (Ren et al., 2000), and has enabled us to identify the sites of RSC occupancy throughout the genome.

Our analysis of Rsc9-occupied sites focused on detecting over-representations of genes in categories defined by a given attribute, such as a biochemical function or subcellular localization. We noticed that several categories that are significantly enriched among Rsc9-occupied sites are comprised of genes transcriptionally regulated by TOR signaling. Indeed, treatment of cells with rapamycin, which specifically affects the TOR pathway, resulted in dramatic genome-wide changes in Rsc9 localization. Northern analysis demonstrated that Rsc9 both represses and activates transcription regulated by TOR, consistent with its rapamycin-induced redistribution. We conclude that RSC is directly involved in the transcriptional program directed by the TOR pathway. Our results illustrate the strength of functional genomics in revealing relationships between distinct cellular processes.

Identifying RNA-protein interactions important in neuronal function

Jennifer Darnell, Kate Dredge, Kirk Jensen and **Robert B. Darnell**
Laboratory of Moloecular Neuro-Oncology, The Rockefeller University, New York, NY

The brain appear to harbor a unique set of proteins that regulate neuron-specific RNA metabolism. Several sets of these proteins have been identified as target antigens in autoimmune neurologic diseases triggered by systemic malignancies. We have taken a multistep approach to identify the function of these neuronal RNA binding proteins. Several complimentary approaches have been undertaken to identify the array of RNAs to which these proteins bind. For example, RNA selection has been used to identify complex RNA targets for the Nova RNA binding protein, and for the related fragile-X mental retardation protein, FMRP. These targets have successfully been used to screen databases of RNAs and identify candidate in vivo RNA targets. We have combined biochemical assays with functional assays, relying on genetically null animals to confirm the relevance of these sets of RNAs.

Elucidating a "theoretical" proteome of the *Arabidopsis thaliana* thylakoid

Olof Emanuelsson
Stockholm Center for Physics, Astronomy and Biotechnology,
Stockholm University,
Stockholm Bioinformatics Center,
S-106 91 Stockholm
Sweden
telephone: +46-8-5537 8574
fax: +46-8-5537 8214
email: olof@sbc.su.se
prestype: Platform
presenter: Olof Emanuelsson

Olof Emanuelsson (1), Jean-Benoit Peltier (2), Gunnar von Heijne (1), Klaas J. van Wijk (2);
(1) Stockholm Bioinformatics Center, Stockholm University, Stockholm, Sweden.
(2) Department of Plant Biology, Cornell University, Ithaca, NY, USA.

The *Arabidopsis thaliana* genome offers nice opportunities to develop and test whole-genome based approaches to theoretical proteomics. Using subcellular localization predictions (TargetP followed by SignalP-HMM) and subsequent transmembrane predictions (TMHMM 2.0), we have predicted the total proteome size of the lumen of the chloroplast sub-compartment thylakoid to be somewhere between 200 and 500 different proteins, whereof a substantial part lacks any functional annotation and approximately 50% contain a TAT-pathway signal. We have also evaluated the combined predictor approach in several ways, specifically addressing the SignalP performance on the signal peptide-like thylakoidal transfering domain adjacent to the chloroplast transit peptide, and it has been clear that a thylakoid-dedicated signal peptide predictor would be useful.

HGMP-RC Microarray Programme - Facing the Challenge of High Throughput Microarray Fabrication and Analysis

Tom Freeman
Microarray Programme Leader
MRC Human Genome Mapping Programme-RC
Wellcome Trust Genome Campus
Hinxton
CB10 1SB
telephone: +44 1223 494552
fax: +44 1223 494512
email: tfreeman@hgmp.mrc.ac.uk
prestype: Platform
presenter: Tom Freeman

Tom Freeman, Richard Stephens, Debbie Williams, Jayn Wright, Rebecca Mayes, Clare Potter, Laurie Scott, Yagnish Umrania*, Naran Hirani*, Tony Attwood*, Jayne Vallance*, Alan Flook and Phil Gardner*
Microarray and *Informatics Development Groups, Human Genome Mapping Programme-Resource Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SB.

The application of microarrays for gene expression profiling has been demonstrated to be one of the most powerful and direct ways of using the sequence data for functional studies. However, the true potential of this approach has yet to be realised widely, because previously access to the technology has been limited. The Microarray Programme at the HGMP-RC has a remit to develop and distribute the technology, specifically microarrays of human and mouse gene probes, to a UK academic user base.

We currently distribute microarrays of human and mouse genes, the probes having been generated by insert PCR from clone libraries. However, we are now increasingly turning to oligonucleotides as a probe resource for microarray fabrication. I will discuss our experience with this approach and other technological issues with respect to the manufacture and distribution of microarrays. This will be contrasted to our experience with the Affymetrix GeneChip system which we have been running for over a year. In addition, a considerable effort is currently being put in to developing the bioinformatics support for this programme and this will be discussed also. Finally, the future plans for the Microarray Programme at the HGMP-RC will be outlined.

http://www.hgmp.mrc.ac.uk/Research/Microarray/index.jsp

This work is funded by the MRC and BBSRC.

Functional analysis of the promoter for the long isoform of collagen a1(IX) in ovo

Elena I. Frolova
UTMB
Dept. Microbiology and Immunology
301 University Blvd
Galveston, TX 77555-1019
telephone: (409)772-2373
fax: (409)772-5065
email: elfrolov@utmb.edu
prestype: Platform
presenter: Elena I. Frolova

Elena I. Frolova
Department of Microbiology and Immunology, The University of Texas Medical Branch at Galveston, Galveston, Texas, USA

A great deal is known about the early stages of eye development, the differentiation of the lens and the formation of the neural retina. However, the differentiation of the anterior part of the optic cup into the ciliary body and the iris structures remains obscure. One of the few proteins known to be expressed specifically and in abundance in the prospective ciliary epithelium is the long-isoform of collagen a1(IX). Its transcript is detected in the nonpigmented layer of the ciliary epithelium even before any morphological changes can be observed. Therefore, an understanding of the transcriptional regulation of the long-isoform of the collagen a1(IX) gene will give insight into the transcriptional mechanisms regulating the early steps of ciliary epithelium differentiation. It was shown that correct differentiation of the ciliary epithelium requires the presence of the lens and probably the neural retina. This makes difficult to create cell cultures that precisely imitate the differentiating ciliary epithelium in vivo to use for transient expression assays in vitro. To overcome this problem we employed a new method, in ovo microelectroporation, for functional study of the long-isoform of the collagen a1(IX) promoter.

We used in ovo microelectroporation to introduce reporter plasmids containing different parts of the collagen a1(IX) promoter into cells of the optic cup in living chicken embryos. Promoter constructs were designed in which the promoter and part of the coding sequence for collagen a1(IX) gene were fused in frame with the coding sequence of green fluorescent protein (GFP). The GFP fusion constructs were then injected into the optic vesicle of chicken embryos. After electroporation, embryos were allowed to develop for 2-3 days. GFP expression was observed by confocal microscopy in whole mounts and in sections.

We demonstrated that the presence of intron 1 of the collagen a1(IX) gene is essential to direct its expression to the optic cup. Despite the difference in the size of intron 1 in chickens, mice and humans, this part of the collagen 1 (IX) gene contains several highly conserved sequences.

The advantages and disadvantages of using in ovo microelectroporation for functional study of the promoters will be discussed.

Never-ending annotation of human chromosome 21: alternative splicing, antisense transcripts and mouse conservation

Katheleen Gardiner, Ph.D.
Eleanor Roosevelt Institute
1899 Gaylord Street
Denver, CO 80206-1210
USA
telephone: 303-336-5652
fax: 303-333-8423
email: gardiner@eri.uchsc.edu
prestype: Platform
presenter: Katheleen Gardiner, Ph.D.

Katheleen Gardiner and Muriel Davisson
Eleanor Roosevelt Institute, Denver Colorado and The Jackson Laboratory, Bar Harbor Maine

The 34 Mb finished sequence of human chromosome 21 was published in May 2000 with 225 genes/models reported. With hand curation, the gene number is now 250, however, much annotation of the genomic sequence remains to be added. We will present information on alternative splicing within coding regions, putative antisense transcripts, and conservation with homologous mouse sequences.

i) Alternative splicing: To determine the number of protein isoforms encoded by chromosome 21, information from dbEST entries was used to detect potential splice variants within coding regions. Of the 200 genes for which sufficient EST data were available, >40% showed two or more splice variants. These included novel splice variants of well studied genes (e.g. APP and HMG14), and prediction of possibly >30 forms for the Intersectin gene.

ii) Antisense transcripts: Information from spliced ESTs was also used to identify putative antisense transcripts to known protein coding genes. Eight examples were found where spliced opposite-strand transcripts contained exons complementary to one or more coding exons of the sense strand gene. Biological functions of the sense genes are diverse; preliminary experimental analysis of the antisense genes suggests that transcription is very low in the tissues tested.

iii) Comparison with mouse genomic sequence: Draft genomic sequences from the public sector and the Celera databases were searched for all segments with homology to human chromosome 21. Comparison of >30 Mb shows that most, but not all, “known” genes and complete cDNAs are present in both species; gene order and orientation is generally, but not always, conserved; a 7 Mb “gene desert” is conserved; intergenic and intronic regions, largely only within non-GC-rich regions, show numerous segments of significant conservation, but unknown function; and a large number of spliced EST sequences appear to be species specific.

Integrative Genomics

Mark Gerstein

P Harrison, J Qian, V Alexandrov, P Bertone, R Das, D Greenbaum, R Jansen, W Krebs, N Echols, J Lin, C Wilson, A Drawid, Z Zhang, Y Kluger, N Lan, N Luscombe

Molecular Biophysics & Biochemistry Department
Yale University
New Haven, CT 06520

http://bioinfo.mbb.yale.edu
http://bioinfo.mbb.yale.edu/lectures
http://bioinfo.mbb.yale.edu/genome/

My talk will focus on analyzing genomes and functional genomics data in terms of the finite list of protein "parts". I use the term "part" rather broadly, and depending on context, it can either be a protein family, fold, or motif. I will touch on SOME of the following topics: (i) How one can compare different genomes in terms of the occurrence of parts. (ii) How one can do the exact same operation on the pseudogenome -- the total complement on pseudogenes in an organism. (iii) How this idea can be further extended to compare the representation of parts in the genome versus the transcriptome.

References

J Qian, B Stenger, CA Wilson, J Lin, R Jansen, SA Teichmann, J Park, WG Krebs, H Yu, V Alexandrov, N Echols, M Gerstein (2001). "PartsList: a web-based system for dynamically ranking protein folds based on disparate attributes, including whole-genome expression and interaction information.," Nucleic Acids Res 29: 1750-64

PM Harrison, N Echols, M Gerstein (2001). "Digging for dead genes: an analysis of the characteristics of the pseudogene population in the Caenorhabditis elegans genome.," Nucleic Acids Res 29: 818-30

A Drawid , M Gerstein (2000). "A Bayesian system integrating expression data with sequence patterns for localizing proteins: comprehensive application to the yeast genome." J Mol Biol 301 : 1059-75 .

R Jansen , M Gerstein (2000). "Analysis of the yeast transcriptome with structural and functional categories: characterizing highly expressed proteins." Nucleic Acids Res 28 : 1481-8 .

R Jansen, D Greenbaum, M Gerstein. (in press) "Relating whole-genome expression data with protein-protein interactions." Genome Research.

J Qian, M Dolled-Filhart, J Lin, H Yu, M Gerstein. (2001) "Beyond synexpression relationships: Local Clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions."
J Mol Biol 314: 1053-1066.

Spatial and temporal expression pattern of Runx3 (aml2) and Runx1 (aml1) indicates non-redundant functions during mouse embryogenesis

Yoram Groner
The Weizmann Institute Rehovot
Israel
telephone: 972-8-9343972
fax: 972-8-9344108
email: yoram.groner@weizmann.ac.il
prestype: Platform
presenter: Yoram Groner

Ditsa Levanon[1], Ori Brenner[2], Varda Negreanu[1], David Bettoun[1], Eilon Woolf[1], Raya Eilam[2], Joseph Lotem[1], Yoram Groner[1]
Departments of [1]Molecular Genetics and [2]Experimental Animals, The Weizmann Institute of Science Rehovot 76100 ISRAEL

The human RUNX3/AML2 gene belongs to the “runt domain” family of transcription factors that act as gene expression regulators in major developmental pathways. We will describe the expression pattern of Runx3 during mouse embryogenesis compared to the expression pattern of Runx1. E10.5 and E14.5 to E16.5 embryos were analyzed using both immunohistochemistry and b-galactosidase activity of targeted Runx3 and Runx1 loci. We found that Runx3 expression overlapped with that of Runx1 in the hematopoietic system, whereas in sensory ganglia, epidermal appendages, and developing skeletal elements, their expression was confined to different compartments. These data provide new insights into the function of Runx3 and Runx1 in organogenesis and support the possibility that cross regulation between them plays a role in embryogenesis.

Human and ape molecular clocks and the origin of bipedalism

S. Blair Hedges
Pennsylvania State University

[abstract not available]

mRNA targeting by sequences in the 3'untranslated region promotes nuclear protein import

John Hesketh
Department of Biological and Nutritional Sciences,
University of Newcastle, Kings Rd.,
Newcastle-upon-Tyne, UK NE1 7RU
telephone: 44-191-222-8744
fax: 44-191-222-8684
email: j.e.hesketh@ncl.ac.uk
prestype: Platform
presenter: John Hesketh

John E. HESKETH and Marilyne LEVADOUX,
Department of Biological and Nutritional Sciences, University of Newcastle, Newcastle-upon-Tyne, NE1 7RU, UK

Both c-myc and c-fos mRNAs are localised to the perinuclear cytoplasm and associated with the cytoskeleton and these mRNAs encode important nuclear transcription factors. The association of these mRNAs with the cytoskeleton and their perinuclear localisation is due to sequences within their 3'untranslated regions (e.g. Dalgleish et al., 2001). In addition the mRNA encoding metallothionein-1 (MT) is also associated with the cytoskeleton and localised around the nucleus. The MT mRNA is also targeted to the perinuclear cytoplasm and the cytoskeleton by signals within its 3'untranslated region (3'UTR). The localisation signal has not been exactly defined in any of these three mRNAs but it is relatively small (less than 86, 145 and 41nt in myc, fos and MT mRNAs respectively).

MT, which has a role in metal metabolism and may also protect DNA against oxidants, has been found localised in the nucleus during S-phase and has been suggested to play a role in apoptosis. Using cells transfected with gene constructs differing in their 3'UTRs we have investigated the role of perinuclear mRNA targeting in the facilitation of the subsequent import of MT into the nucleus and its function. Perinuclear localization of MT-1 mRNA and its association with the cytoskeleton was necessary for the shuttling of MT protein into the nucleus during the early S-phase (Levadoux et al., 1999). Following exposure of the cells to a variety of agents the extent of protection from oxidative stress and DNA damage provided by MT was lower if the mRNA was not correctly targeted. The data also suggest that MT plays a role in protecting cells against apoptosis (Levadoux-Martin et al., 2001).

Our hypothesis is that targeting of mRNA to the cytoskeleton around the nucleus is a mechanism which promotes the subsequent nuclear import of a range of proteins; in this way it contributes to local synthesis of the proteins close to their site of function.

- Levadoux,M., Mahon,C., Beattie,J., Wallace,H.M. and Hesketh,J.E. (1999) J.Biol.Chem. 274, 34961-6
- Dalgleish,G.D, Veyrune,J-L., Blanchard,J-M., and Hesketh,J.E. (2001) J.Biol.Chem. 276, 13593-9
- LevadouxMartin,M., Hesketh,J.E., Beattie,J. and Wallace,H.M. (2001) Biochem.J. 355, 473-9

Functional Analysis of Genetic Determinants of Human Disease

Geoff Hicks
Manitoba Inst. Cell Biology/U Manitoba
Rm ON5029
675 McDermot Ave
Winnipeg, MB R3E 0V9
Canada
telephone: 204-787-2133
fax: 204-787-2190
email: hicksgg@cc.umanitoba.ca
prestype: Platform
presenter: Geoff Hicks

The Manitoba Institute of Cell Biology Centre for Mammalian Functional Genomics (MFG) uses a novel form of gene trapping called Tagged-Sequence Mutagenesis (TSM). Through random insertion of a gene-trap vector into the genome, genes are both tagged and invariably mutated. Each mutation is studied first by direct DNA sequencing. Comparison of the sequence tags (PSTs) with the existing databases identifies disruptions of known genes or genes which may be related by homology or functional domains. As a result, ES cell lines bearing disruptions in genes of particular interest can be studied directly or used for making mutant mice. ES cells lines will be made available to the academic community at large. The MFG database currently being developed, <www.EScells.ca>, will ensure widespread public access to this resource. The Centre is also integrated with the International Gene-Trap Consortium. Future developments will focus on high throughput sequence-based screens to generate ES cell libraries for cell-type specific gene expression, and for generation of hyper- and hypo-morphic mutations. It is hoped that this resource will significantly impact on the accessibility of genetic modeling in mice to understand gene function within the context of the whole organism – a context that is critical for assessing the role of genetic determinants of disease.

Diversity in gene expression: Assessment of exon skipping and expression states

Winston Hide
SA national Bioinformatics Institute
University Western cape
Private bag x17
Bellville
South Africa
telephone: +27 21 959 3645
fax: +27 21 959 2512
email: winhide@sanbia.c.za
prestype: Platform
presenter: Winston Hide

Winston Hide, Tzu-Ming Chern, Janet Kelso and Vladimir Babenko

Completion of the human genome sequence provides evidence for a gene count with lower bound 30 000 – 40 000. Significant protein complexity may derive in part from multiple transcript isoforms. Recent EST based studies have revealed that alternate transcription, including alternative splicing, polyadenylation and transcription start sites, occurs within at least 30-40 % of human genes. Transcript form surveys have yet to integrate the genomic context, expression, frequency, and contribution to protein diversity of isoform variation. We describe the degree to which protein coding diversity may be influenced by alternate expression of transcripts. 545 genes have been studied in this first intensive hand-curated assessment of exon skipping on chromosome 22. Combining manual assessment with software screening of exon boundaries provides a highly accurate and internally consistent indication of skipping frequency. 57 of 62 exon skipping events occur in the protein coding regions of 52 genes. A single gene, (FBXO7) expresses an exon repetition. 59% of highly represented multi-exon genes are likely to express exon-skipped isoforms in ratios that vary from 1:1 to 1:>100. The proportion of all transcripts corresponding to multi-exon genes that exhibit an exon skip is estimated to be 5%. A comparison with mouse orthologous genes reveals that common skipping events are not frequently detected, but that the frequency of skipping is similar between mouse and man. Comparitive ssessment of expression state and skip occurrence is discussed.

Characterization of novel Proteins by Analysis of Human Protein Networks: Part1 - cloning of human full ORFs of novel proteins

Bernhard Korn
RZPD - Ressourcenzentrum für Genomforschung
INF 506
69120 Heidelberg
Germany
telephone: +49 (6221) 42 4700
fax: +49 (6221) 42 4704
email: korn@rzpd.de
prestype: Platform
presenter: Korn

R. Schatten[1], S. Henze[1], M. Schick[1], S. Wiemann[2], A. Gödde[1], H. Göhler[3], E. Wanker[3], B. Korn[1]

1 RZPD Ressourcenzentrum für Genomforschung, INF 506, 69120 Heidelberg, Germany
2 Deutsches Krebsforschungszentrum, INF 506, 69120 Heidelberg, Germany
3 Max-Planck Institut für Molekulare Genetik, Ihnestr. 73, 14195 Berlin, Germany

The sequencing of the human genome has resulted in the identification of a large number of novel proteins, whose function and interactions with other proteins have to be determined. We aim at the characterization of unknown proteins and the assignment of these proteins to particular pathways or multimeric structures. We apply a systematic, high throughput approach for functional analysis. We have developed an automated, medium throughput cloning and verification strategy that allows us to clone human full ORF cDNAs at a rate of 100 per week. Up to now, we have cloned and completely sequence verified more than 2.400 ORFs. We plan to clone 6.000 ORFs within the next year.

These ORFs are so far not characterised on the functional level, except for computational predictions, many have been identified by the German cDNA Consortium. It is our intention to channel these clones to a medium size yeast two-hybrid system that allows the screening for protein-protein interactions. Within our current project, up to 5,000 defined human cDNAs will be inserted into a DNA binding domain vector for screens against different human cDNA libraries, cloned into activation domain vectors.

Special emphasis will be put on protein interactions with medically relevant proteins such as oncology related proteins, proteins involved in cardiovascular defects, and in neurodegenerative diseases. Protein-protein interactions will be confirmed by in vitro binding experiments, co-immunoprecipitations and co-localisation studies. Finally, the results of the two-hybrid screens and of the functional characterisations will be stored in the Primary Database at RZPD. With this novel systematic approach we will combine for the first time high quality sequence information with large-scale functional data in a common database. We suggest that this information will add value to newly identified genes and our overall understanding of protein networks. Finally, clones will be made available for public use via the RZPD.

Experimental verification of predicted splice variants of human genes

Bernhard Korn
RZPD - Ressourcenzentrum für Genomforschung
INF 506
69120 Heidelberg
Germany
telephone: +49 (6221) 42 4700
fax: +49 (6221) 42 4704
email: korn@rzpd.de
prestype: Platform
presenter: Zink

D. Zink1,2, S. Haas3, E. Coward3, M. Vingron3, B. Korn1

1 RZPD Ressourcenzentrum für Genomforschung, INF 506, 69120 Heidelberg, Germany
2 Deutsches Krebsforschungszentrum, INF 506, 69120 Heidelberg, Germany
3 Max-Planck Institut für Molekulare Genetik, Ihnestr. 73, 14195 Berlin, Germany

Public EST databases currently contain more than 3 million human EST sequences, representing probably 40-50.000 human genes/transcripts. Within these data exists a large redundancy. We take advantage of this redundancy by analysing the differences of sequences belonging to the same gene. The EST sequences are clustered and assembled to a consensus sequence. However, many clusters cannot be assembled into a single consensus sequence. The EST sequences then fall into multiple consensus sequences (contigs) within one clusters. The differences might be due to imperfect sequence data (e.g. partially unspliced sequence templates, sequencing errors) or due to alternative splicing. Instead of one gene coding for one mRNA leading to one protein, alternative splicing of transcripts may lead to different mRNA species and therefore to potentially different proteins. Splice variants are often due to alternative exon usage, which we verify by RT-PCR. We have set up a medium throughput strategy that does allow us to screen expression of genes in 25 different human tissues of multiple stages. We initiated this project by analysing genes that predominantly reside within the Down-Syndrome critical region and on Chromosome 22. Our results indicate, that the theoretical data represented in EST databases can be verified in many cases by our experimental design. Moreover, we do find additional splice products, that are not defined by any EST sequence. In order to gain more insight, we re-sequence PCR products in question, to confirm their origin and nature. Nevertheless, in more than 18% of the cases, we cannot experimentally support EST data by RT-PCR. In future we want to extend splice variant analysis to other chromosomes and gene families. We intend to automate RT-PCR and ultimately design a chip to discriminate a large number of different splice forms of medically relevant genes.

Functional Genomics by Transposon-Tagging and ChIP-Chip Analysis in Yeast

Anuj Kumar, Mike Snyder
Dept. of MCD Biology
Yale University
P.O. Box 208103
New Haven, CT 06520-8103
telephone: 203-432-9949
fax: 203-432-6161
email: anuj.kumar@yale.edu
prestype: Platform
presenter: Anuj Kumar

Anuj Kumar & Mike Snyder

Here we present two novel approaches by which gene function may be investigated on a genome-wide scale. The first approach utilizes transposon-based insertional mutagenesis as a means of conveniently generating informative mutant alleles for subsequent functional analysis. Transposons employed for this purpose have been modified such that a single insertion is sufficient to derive a reporter gene fusion, gene disruption, and epitope-tagged allele, while also serving as a simple gene trap. As applied to the genome of Saccharomyces cerevisiae, these multi-functional transposons have been used in conjunction with high-throughput screening to generate a collection of yeast strains mutated for nearly 4,000 different annotated genes (roughly two-thirds of the yeast genome). The cumulatively unique data set resulting from analysis of these insertions is cataloged in the Web-accessible database TRIPLES (http://ygac.med.yale.edu). While transposable elements are powerful tools for functional genomics, microarray technologies provide an equally exciting avenue for large-scale biological discovery. Utilizing microarray-based methods, we have recently developed a genomic approach (in collaboration with Pat Brown's group) to globally map binding sites of yeast transcription factors. In this approach, protein-DNA interactions are "captured" in vivo by crosslinking proteins to their genomic binding sites; crosslinked DNA is subsequently extracted, sheared, and purified by immunoprecipitation with antibodies directed against the chosen transcription factor. Bound DNA (enriched for transcription factor binding sites) is subsequently prepared as probes against a microarray of yeast intergenic regions. By this ChIP-Chip strategy (Chromatin ImmunoPrecipitation and microarray analysis), we have defined over 200 previously unidentified targets of the transcription factors SBF and MBF, two sequence-specific activators regulating gene expression during the G1/S transition in yeast. As a means of globally identifying protein-DNA interactions, our ChIP-Chip approach can be used to map a variety of functional sites in the yeast genome as well as in DNA from other organisms.

Development of a Microscope-based High Content Screening Platform

Urban I. Liebel
Cell Biology and Biophysics
EMBL-Heidelberg
Meyerhofstrasse 1
69117 Heidelberg
Germany
telephone: 49 6221 387232
fax: 49 6221 387306
email: liebel@embl-heidelberg.de
prestype: Poster
presenter: Urban I. Liebel

Urban I. Liebel, Jeremy C. Simpson and Rainer Pepperkok

The identification of large numbers of novel cDNAs about which no functional information exists, represents an enormous resource which should be exploited. Tagging each of these coding sequences with the green fluorescent protein (GFP), then expressing these fusion proteins in living cells allows a number of microscopic assays to be performed, which in turn generate functional data about these molecules. To this end, we have developed a fully automatic microscope which is integrated with a robotic liquid handling station.

We have written software which allows full control over objectives (10x, 20x, 40x, 63x), excitation/emission and neutral density filters. An image acquisition module for the cooled 12-bit CCD camera and a module to control the precision xy-stage are also integrated into the software. Assays have been developed in a 96-well format using thin glass-bottomed plates. A software autofocus has been developed to compensate for the well bottom roughness, and a fast hardware autofocus is under construction. This will decrease the screening time by a factor of ten. We are currently integrating this platform into a standard LIMS (Laboratory Information Management System) software environment in order to screen multiple plates without user intervention.

Different datamining modules, also developed in our lab, analyse the images taken by the microscope. The software can determine which cells are transfected with the GFP-tagged cDNAs, and quantify the fluorescence from multiple channels. The integration of these components allow us to rapidly carry out hundreds of individual experiments on these novel cDNAs which contribute the first information about these novel proteins.

PRODUCTION OF FERTILE TRANSGENIC ZEBRAFISH FROM LONG-TERM CULTURED CELLS BY NUCLEAR TRANSFER

Shuo Lin
Department of Molecular,
Cell & Developmental Biology,
University of California, Los Angeles,
621 Charles E. Young Drive South, LS4325,
Los Angeles, CA 90095-1606
telephone: 310-267-4970
fax: 310-267-4971
email: shuolin@ucla.edu
prestype: Platform
presenter: Shuo Lin

Ki-Young Lee #*, Haigen Huang #, Jae Yong Han*, and Shuo Lin #

Department of Molecular, Cell & Developmental Biology, University of California, Los Angeles, 621 Charles E. Young Drive South, LS4325, Los Angeles, CA 90095-1606
* School of Agriculture Biotechnology, Seoul National University, Korea

ABSTRACT

The zebrafish has become an important model for genetic and developmental biology studies. However, due to the lack of embryonic stem cells, one major limitation of this model is the inability to perform site-specific genetic modifications. In some other animals, generation of gene knockout by somatic cell cloning has been used to overcome this limitation. As the first step towards this goal, we established conditions to produce cloned transgenic zebrafish using a long-term cultured embryonic fibroblast cells. First, these cells were infected with a retrovial vector expressing the GFP marker gene. The nuclei of these cells were then transplanted into enucleated, unfertilized eggs and the resultant cloned zebrafish were shown to be fertile and continue to express the GFP reporter gene. From 10 experimental groups, a total of 34 embryos (36.2 %) reached the blastula stage while 16 embryos (17.0%) from 6 groups hatched, and 11 individuals reached adulthood. All of these individuals produced offspring after mating with wild type, and their F1 and F2 progeny expressed GFP with similar expression patterns as the F0 generation. This study sets up a foundation for further genetic manipulations in zebrafish.

Real-time quantitative PCR : A fast and efficient method to screen clones derived from Subtractive Suppression Hybridization

C Louvet
INSERM U437 - ITERT
30, bd Jean Monnet
CHU-HOTEL DIEU
44093 NANTES
telephone: (33)2 40 08 74 10
fax: (33)2 40 08 74 11
email: clouvet@nantes.inserm.fr
prestype: Platform
presenter: Louvet C

Louvet C, Heslan JM, Chiffoleau E, Cuturi MC.

Introduction. Subtractive Suppression Hybridization (SSH) has proven successful identification of differentially expressed genes between two types of cells or between two conditions. SSH technique uses selective exponential amplification by PCR to isolate target cDNA fragments. As a result, the subtracted library can contain rare differentially expressed transcripts enriched by ~1000-fold. However, SSH from high-complexity cDNA samples may generate numbers of background clones representing non-differentially expressed species, even after primary screening by differential hybridization. Northern blotting is generally used to confirm differential expression but this method remains semi-quantitative and often not sensitive-enough to detect low-abundant mRNAs. To overcome these problems, we introduced real-time quantitative PCR to precisely assess the expression levels of clones derived from SSH.

Materials and Methods. Acute rejection of heart allograft in an adult rat model can be prevented by donor-specific transfusion (DST) of the recipient before the engraftment. To identify novel genes involved in tolerance induction, we performed SSH on mRNAs from whole-heart grafts of untreated and DST-treated animals, followed by screening by differential hybridization with subtracted probes, sequencing, and SYBR® Green real-time quantitative PCR on cDNA pools of syngenic, allogenic-rejecting, and non-rejecting-allogenic grafts.

Results. Among the 288 randomly selected clones from SSH, 50 putative clones showed differential hybridization with subtracted probes. Based on their sequences, primers were designed for 25 candidates. Real-time quantitative PCRs were carried out and allowed to eliminate 23 false-positive clones and retain only two true clones corresponding to transcripts specifically accumulated in the non-rejecting allografts. Expression levels of these transcripts were then assessed in grafts from day 1 to 7 after transplantation and revealed over-expression for both clones from day 3 to 7 in the non-rejected allografts.

Conclusions. SSH can be performed on high-complexity cDNA samples derived from whole tissues like heart allografts when it is coupled with real-time quantitative PCR as assesment of the true levels of mRNA expression. In addition, kinetics analysis can be concurrently done on numerous samples to rapidly and precisely determine the modulation of expression for each transcript of interest.

Gene expression profiling analysis shows differential expression of extracellular matrix genes in left ventricular hypertrophy of renal failure

Christian Maercker
Resource Center for Genome Research
German Cancer Research Center H0600
Im Neuenheimer Feld 280
D-69120 Heidelberg
Germany
telephone: +49 6221 424741
fax: +49 6221 423454
email: c.maercker@dkfz.de
prestype: Platform
presenter: Christian Maercker

Christian Maercker[1], Diana Lutz[1], Eberhard Ritz[2], Gerhard Mall[3], Kerstin Amann[4]
Resource Center for Genome Research (RZPD)[1], Heidelberg, Dept. of Internal Medicine[2], University of Heidelberg, Heidelberg, Dept. of Pathology[3], Darmstadt, Department of Pathology[4], University of Erlangen, Erlangen,

Aims: Death from cardiac causes is the leading fatality in patients with chronic renal failure. In particular, left ventricular hypertrophy (LVH) is a frequent alteration which develops very early on in pa-tients and rats with renal failure. The pathomechanisms involved, however, are currently not fully understood. Thus, additional information from gene expression profiling could be extremely helpful.

Methods: Male Sprague-Dawley rats, which were subjected to sub-total nephrectomy (SNX), served as a model system for a gene expression profiling analysis. Poly(A)+ RNA from the hearts of SNX animals and from sham-operated rats (SHAM) as a control, isolated 2 weeks and 12 weeks after operation, was labeled and hybridized with Rat UniGene filters containing about 27.000 gene and EST sequences (Bento Soares, Univ. of Iowa). Phosphoimaging and software analysis revealed substantial changes in gene expression in SNX animals compared to SHAM.

Results: Interestingly, some genes were upregulated in SNX rats, which may be involved in activation and expansion of the non-vascular interstitial tissue in uraemic animals. These include genes like timp3, tgf-b1, osteonectin, paxillin, laminin a1, integrin b. Altogether 80 genes were found at least 3fold under or overexpressed. By hybridization of the RNA to human and mouse arrays (Human UniGene RZPD-1, 33.000 clones; Mouse UniGene RZPD-1, 25.000 clones) we could confirm the results with genes which are conserved between human, rat and mouse.

Conclusions: The findings showed (i) that extracellular matrix genes are differentially expressed in LVH of renal failure and that (ii) the complex hybridisation technique produced reliable results. This is an important prerequisite for ongoing experiments with candidate genes involved in cardiac structural changes in chronic renal failure.

Alternative processing of thyroid hormone receptor mRNA: Role of cis-acting sequences and antisense RNA

Presentation

Stephen Munroe
Department of Biology
Wehr Life Sciences Building
Marquette University
PO Box 1881
Milwaukee, WI 53201
telephone: 414-288-1485
fax: 414-288-7357
email: stephen.munroe@marquette.edu
prestype: Platform
presenter: Munroe
http://mendel.biol.mu.edu/faculty/munroe.html

Stephen H. Munroe, Michelle L. Hastings and Valerie K. Salato
Department of Biology, Marquette University

In mammals mRNA for the a-type thyroid hormone receptor (TRa1) is transcribed from a complex locus that encodes several nuclear receptors. Alternative processing give rise to two mRNAs encoding two functionally antagonistic proteins, TRa1 and TRa2. The latter is a variant receptor that lacks the C-terminal domain required for hormone binding and transcriptional activation. A third nuclear receptor, Rev-ErbAa (RevErb), is encoded on the opposite strand and overlaps TRa2 but not TRa1 mRNA. Cells expressing high levels of RevErb generally have higher amounts of TRa1 relative to TRa2. In vivo overexpression of RevErb leads to a shift in the TRa1/TRa2 ratio, and in vitro RevErb RNA efficiently blocks TRa2 splicing. These results support an antisense mechanism in which RevErb selectively blocks TRa2 mRNA splicing. TRa2 mRNA is formed by splicing out a portion of the 3' terminal exon of TRa1 mRNA and utilizing a downstream 3' exon specific to TRa2. In common with other regulated alternative processing sites, those specific to TRa1 or TRa2 mRNA are suboptimal. Point mutations within the TRa2-specific splice sites that increase the match to the consensus splice site sequences greatly increase the level of TRa2 mRNA. Similarly, replacing the TRa1-specific polyA site with a strong viral polyA site increases the level of TRa1 mRNA to the exclusion of TRa2 splicing. Recent work has focused on a highly conserved 80 splicing enhancer located near the 5' end of the TRa2-specific intron, immediately downstream of the TRa1 stop codon. This element includes a conserved pseudo-5'ss sequence that is critical for TRa2 mRNA splicing. A number of single nucleotide replacements within the pseudo-5'ss reduce TRa2 splicing. However, point mutation that improve the match to the 5'ss consensus sequence activate cryptic splicing. Several proteins, including hnRNP H and SF2/ASF, crosslink specifically to the enhancer element. Mutations within the pseudo-5'ss that inactivate TRa2 splicing also disrupt the crosslinking of these proteins, strongly indicating that these proteins play a role in splicing. Further experiments examine the role of a 3' intronic splicing enhancer, the TRa1 polyadenylation site and the role of the complementary RevErb mRNA in modulating the balance between TRa1 and TRa2 expression.

REFERENCES:
Hastings, M.L., H.A Ingle, M.A. Lazar and S.H. Munroe. (2000) *J. Biol. Chem.* **275**, 11507-11513.
Hastings, M.L., C. Milcarek, K.Martincic, M.L.Peterson, and S.H.Munroe. (1997) *Nucl. Acids Res.*, **25**, 4296-4300.

Hastings, M.L., C.M. Wilson and S.H. Munroe. (2001) *RNA* **7**, 859-874.
Kumar, M and G.G. Carmichael. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 3542-3547.
Munroe, S.H., and M.A Lazar. (1991) *J. Biol. Chem.*, **266**, 22083-22086.

CONTACT:
stephen.munroe@marquette.edu
http://mendel.biol.mu.edu/faculty/Munroe.html

WHOLE GENOME SHOTGUN ASSEMBLY OF THE MOUSE GENOME

Richard Mural
45 West Gude Drive
Rockville, MD 20850
telephone: 240-453-3665
email: richard.mural@celera.com
prestype: Platform
presenter: Richard Mural

Richard J. Mural, Celera Genomics. Celera Genomics, Rockville, MD. 20850. www.celera.com

We have generated whole genome shotgun sequence from 3 strains of the laboratory mouse (129X1/SvJ, A/J, DBA/2) to a total coverage of 5.5X coverage. Paired-end sequences from libraries with insert sizes of 2, 10 and 50 kbp (about 23 million reads) were assembled using the whole genome assembly protocol that we have previously described for the assembly of the human genome (Venter, et al. Science, vol. 291, 2001). Nearly 50% of this sequence assembles into scaffold >5 megabases in length. More recently we have completed a new assembly which incorporates data generated by the public mouse genome project from the C57/BL6 strain (this assembly has 7 to 8X coverage of the mouse genome).

We have also analyzed the assembled mouse genome with a modification of the computational analysis pipeline and auto-annotation system, Otto, that we used for the computational annotation of the human genome. This analysis pipeline included mouse full-length cDNAs as well as human genes. The results of this analysis and a preliminary comparison of the human and mouse gene complement will be presented.

GABI Primary Database: Bioinformatics Resource for Plant Genome Data

Axel Nagel
RZPD Deutsches Ressourcenzentrum
fuer Genomforschung GmbH
Heubnerweg 6
D-14059 Berlin
Germany
telephone: (+49 30) 32639-264
fax: (+49 30) 32639-111
email: nagel@rzpd.de
prestype: Poster
presenter: Axel Nagel

Axel Nagel, Svenja Meyer & Steffen Schulze-Kremer

Gabi, acronym for "Genomanalyse im biologischen System Pflanze", is the name of a large collaborative network of different plant genomic research projects. The project is funded by the German Ministry for Education and Research (BMBF) with the purpose, to support plant genome research in Germany, to yield information about commercial important plant genomes, and to establish a scientific network within plant genomic research. GabiPD (Primary database) is part of the large Gabi network. It is localized at the Resource Center (RZPD), which has been established in 1995 as the central infrastructure of the German Human Genome Project.

The main objectives of the GabiPD are to collect and integrate biological primary data, like sequence data, results of microarray studies and proteome analysis (e.g. data of 2D gelelectrophoresis) as well as mapping data etc. Interfaces are developed to allow database searches and to visualize data via internet. GabiPD facilitates the interchange of data within Gabi and combines data of different experiments and groups. This way a synergistic effect is created for all Gabi participants. Particular emphasis is put on high troughput expression analysis. To store this data a MIAME-compliant database schema was developed.

The poster will give you an impression of GabiPD contents, developed interfaces and underlying database structures.

Using quantitative and global expression profiling to explore and quantitate nuclear hormone receptor function in myeloid cell differentiation and metabolism

Laszlo Nagy
Department of Biochemistry and Molecular Biology
University of Debrecen, Medical and Health Science Center
98 Nagyerdei krt.
DEBRECE, H-4012
telephone: +36-52-416-432
fax: +36-52-416-432
email: lnagy@indi.biochem.dote.hu
prestype: Platform
presenter: Laszlo Nagy

Laszlo Nagy, Attila Szanto and Istvan Szatmari
Department of Biochemistry and Molecular Biology, University of Debrecen, Nagyerdei krt. 98, DEBRECEN, Hungary H-4012

Nuclear hormone receptor are ligand activated transcription factors. We are trying to understand how these receptors modulate cellular differentiation and function and what is their contribution to diseases such as atherosclerosis and immune response. Since the receptors' transcriptional activity can be regulated (switched on) by small lipid soluble molecules they are excellent targets of global expression profiling.

We have been using real time quantitative PCR for absolute transcript number determination and DNA microarrays for global transcription profiling to determine receptor levels, target genes profiles and to identify components of the genetic programs of receptor activation

A group of receptors such as PPARs (Peroxisome Proliferator Activated Receptor) and LXR (Liver X Receptor) can be activated by modified fatty acids and cholesterol molecules, respectively. Using natural and synthetic ligands we have characterized the role of these receptors in myeloid cell differentiation, lipid metabolism and dendritic cell differentiation.

We have found that oxidatively modified LDL (Low Density Lipoprotein) induces and activates PPARg in monocytes leading to macrophage differentiation, expression of the scavenger receptor CD36 and increased lipid uptake. This process operates as a positive feed back loop and contributes to foam cell formation. PPARg also increases lipid efflux from macrophages through induction of LXRa, a receptor activated by oxysterols. This leads to the increased expression of the transporter ABC1 (ATP Binding Cassette) and cholesterol efflux from macrophages. Our results suggest the existence of a transcriptional cascade and a complex and complementary role for PPARg and LXRa as key regulators of a coordinated cellular responses to oxidatively modified LDL in macrophages. Recently, we have also found that key components of this cascade also exist in maturing/differentiating antigen presenting cells (dendritic cells) of myeloid origin. The role of these transcriptional networks in normal cell function and diseases and their potential for therapeutic utilization will be discussed.

ArrayExpress - a public repository for Gene Expression Data at the European Bioinformatics Institute

Helen Parkinson
Microarray Informatics
European Bioinformations Institute
Wellcome Trust Genome Campus
Hinxton
CB10 1SD
telephone: +44 (0)1223 494672
fax: +44 (0) 1223494468
email: parkinson@ebi.ac.uk
prestype: Poster
presenter: Helen Parkinson

Helen Parkinson, Alvis Brazma, Alan Robinson, Ugis Sarkans, Mohammedezra Shojatalab and Jaak Vilo

ArrayExpress (www.ebi.ac.uk/microarray) is a public repository for microarray based gene expression data, which covers the requirements of Minimum Information About a Microarray Experiment (MIAME)developed by the MGED consortium.

It supports data in MAGE-OM format which is an XML based data exchange format (see www.mged.org)

Systematic subcellular localisation and functional characterization of novel proteins identified by large-scale cDNA sequencing

Dr. Rainer Pepperkok
Meyerhofstr.1
69117 Heidelberg,
Germany
telephone: 0049 6221 387332
fax: 0049 6221 387332
email: pepperko@embl-heidelberg.de
prestype: Platform
presenter: Rainer Pepperkok

Rainer Pepperkok, Urban Liebel, Veronika E. Neubrand, Jeremy C. Simpson, Ruth Wellenreuther (2), and Stefan Wiemann(2)
Department of Cell Biology and Biophysics, EMBL Heidelberg, Meyerhofstrasse 1, 69117 Heidelberg, Germany, email : pepperko@embl-heidelberg.de
(2) Molecular Genome Analysis, German Cancer Research Center, Im Neuenheimer Feld 280/506, 69120 Heidelberg, Germany

High-throughput cDNA sequencing projects worldwide have resulted in the identification of a large number of novel genes. The characterisation of the encoded proteins currently has remained limited to bioinformatic analyses, which could make functional predictions for only about half of the cDNAs sequenced. As a first step towards a more comprehensive functional characterisation of these cDNAs we have developed and tested a strategy, which allows their systematic and fast subcellular localization [1]. Further we have automated a number of simple microscopy based assays which aim at the better understanding of the function of these cDNAs in basic cellular processes like protein secretion, cell proliferation, apoptosis and signal transduction.

1. Simpson JC, Wellenreuther R, Poustka A, Pepperkok R, Wiemann S: Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. EMBO Reports 2000, 1:287-292.

STRUCTURAL AND COMPOSITIONAL ANALYSIS OF TRANSCRIPTION START SITES AND 5'-UNTRANSLATED REGIONS IN HUMAN MRNAS

Graziano Pesole
Dipartimento di Fisiologia e Biochimica
Generali
Università di Milano
via Celoria, 26
20133 Milano
Italy
telephone: +39-02-58354915
fax: +39-02-58354912
email: graziano.pesole@unimi.it
prestype: Platform
presenter: Graziano Pesole

Pesole G.*, Gissi C.*, Grillo G.§, Licciulli F.§, Mignone F.*, Iacono M.*, Liuni S.§
* Università di Milano, Milan, Italy (graziano.pesole@unimi.it)
§ CSMME-C.N.R., Bari, Italy

The 5' and 3' untranslated regions of eukaryotic mRNAs may play a crucial role in the regulation of gene expression controlling mRNA nucleo-cytoplasmic transport, subcellular localization, stability and translation efficiency. In order to study the general structural and compositional features of these sequences we have previously developed UTRdb, a specialized database of 5' and 3' UTR sequences of eukaryotic mRNAs cleaned from redundancy (Pesole et al. 2000a).

Utrdb (release 15.0) contains about 250,000 entries (>65,000,000 nucleotides) which are also annotated for the presence of functional sequence patterns whose biological activity has been experimentally demonstrated. All these patterns have been collected in the UTRsite database where for each functional pattern, corresponding to a specific entry, the consensus structure is reported with a short description of its biological activity and the relevant bibliography.

All Web resources we implemented for the retrieval and the analysis of UTR sequences are available at the UTR home page (http://bighost.area.ba.cnr.it/BIG/UTRHome/) we recently implemented. UTRdb entries can be retrieved through the SRS system where crosslinks to UTRsite as well as to the nucleotide or aminoacid primary database are also established. Through the Web facility UTRscan any input sequence can be searched for the presence of a functional pattern annotated in UTRsite and UTRblast allows to assess sequence similarity between a query sequence and UTRdb entries.

The analysis of complete UTR sequences contained in this database allowed us to define specific structural and compositional features of UTRs from mRNAs belonging to various eukaryotic taxa (Pesole et al. 2000b).

We now report some recent results obtained by analyzing the oligonucleotide context and the compositional features of the transcription start site for those human genes where the full length 5'UTR as well as the corresponding genomic region was available. Furthermore, we investigated the presence and features of upstream ORFs, of sequence tracts similar or complementary to 18S rRNA and the context of initiator ATG (Pesole et al. 2000c), which are known to affect the mRNA translation efficiency.

Pesole, G., S. Liuni, G. Grillo, F. Licciulli, A. Larizza et al., 2000a UTRdb and UTRsite: specialized databases of

sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs. Nucleic Acids Res 28: 193-196.

Pesole, G., C. Gissi, G. Grillo, F. Licciulli, S. Liuni et al., 2000b Analysis of oligonucleotide AUG start codon context in eukariotic mRNAs. Gene 261: 85-91.

Pesole, G., G. Grillo, A. Larizza and S. Liuni, 2000c The untranslated regions of eukaryotic mRNAs: structure, function and bioinformatic tools for their analysis. Briefings in Bioinformatics 1: 236-249.

Proteomics Analysis of Melanoma Cell Lines

K.A.Resing, K.Bernard, L.Aveline, K.Jonscher, N.G.Ahn

Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO, USA

Biochemical, molecular biological, and genetic studies have led to a model for carcinogenesis wherein tumor progression involves increasing mutational burden resulting in progressive release from growth regulation and apoptosis surveillance, and increasing likelihood of metastasis. In melanoma, this progression is reflected in morphological "stages": atypical naevus, radial growth phase (RGP), vertical growth phase (VGP), and metastasis. When melanoma cells are cultured from tumors, these stages are correlated with ability to produce tumors or metastases in nude mice, or with differences in motility, adhesion, or sensitivity to apoptosis signals. Although changes in these properties are driven by changes in signaling pathways, the complexity of these pathways makes it impossible to directly or quantitatively connect these functional assays to specific signaling events.

We are using immobiline dry strip method for isoelectric focusing and two-dimensional gel electrophoresis to provide a more precise method of delineating the phenotype of the cell in a quantifiable way that can be directly related to the underlying signaling status. Using pI range 4.5 to 6.5 gels, approximately 3500 protein spots can be identified. Initial analyses of 15 cell lines (5 RGP, 4 VGP, 6 Metastatic) in the size range 10-55 kDa, shows most of the variability in this pI and size range is related to the growth rate and cell morphology in culture. Eight proteins can be related to the functional stages; but in each case, there are some cell lines that do not fit the pattern. Identification of these proteins reveals that they include known carcinogenesis markers, as well as potential new markers. Using pharmacological agents and transient transfections we are testing the role of specific signaling pathways in the expression of the variable proteins that vary in comparisons between these cell lines and with primary melanocytes. For example, we have identified a group of the variable proteins as those that are turned on by TPA, but not growth factors, suggesting that they are regulated by a PKC dependent mechanism.

The Mouse Gene Expression Database (GXD)

Martin Ringwald
The Jackson Laboratory
600 Main Street
Bar Harbor, ME 04609
USA
telephone: +1 (207) 288-6436
fax: +1 (207) 288-6132
email: ringwald@informatics.jax.org
prestype: Platform
presenter: Martin Ringwald

D.A. Begley, I.J. McCright, T.F. Hayamizu, D.P. Hill, C.M. Smith, J.A. Blake, J.T. Eppig, J.A. Kadin, J.E. Richardson, and M. Ringwald.
The Jackson Laboratory, Bar Harbor, ME 04069, USA

The Gene Expression Database (GXD) is a community resource of gene expression information for the laboratory mouse. By integrating many different types of expression data, GXD provides increasingly complete information about what transcripts and proteins are produced by what genes; where, when and in what amounts these gene products are expressed; and how their expression varies in different mouse strains and mutants. Data are acquired from the literature by curation staff and via electronic submission from laboratories. New expression data are made available on a daily basis. GXD is integrated with the Mouse Genome Database (MGD) to enable a combined analysis of genotype, expression, and phenotype data. In conjunction with the Gene Ontology project we build shared controlled vocabularies for biological processes, molecular functions and cellular components and assign those terms to mouse genes and their products. These classification schemes provide important new search parameters for expression data. Extensive and continuously refined interconnections with sequence databases and with databases from other species place the gene expression information in the larger biological and analytical context. GXD is accessible through the Mouse Genome Informatics web site at http://www.informatics.jax.org/ or directly at http://www.informatics.jax.org/menus/expression_menu.shtml.

GXD is supported by NIH grant HD33745. The Gene Ontology project is supported by NIH

A generally applicable gene targeting method in *Drosophila melanogaster*

Yikang S. Rong
The Stowers Institute for Medical Research,
1000 East, 50th Street,
Kansas City, MO 64111
telephone: (816)926-4103
fax: (816)926-2058
email: yir@stowers-institute.org
prestype: Platform
presenter: Yikang S. Rong

YS Rong*, SWA Titen*, M Golic*, H Xie*, KG Golic*, M Bastiani**, PK Bandyopadhyay**, BM Olivera**, M. Brodsky***, GM Rubin***
*: Stowers Institute, **: Department of Biology, University of Utah, ***: Department of MCB, University of Califonia at Berkeley

Since the development of a new gene targeting method in *Drosophila* (Rong and Golic 2000), we and others have successfully targeted at least nine different loci throughout the genome. This proves that our method should be generally applicable for introducing specific modification to any endogenous locus in *Drosophila*. It will allow *Drosophila* researchers to take full advantage of the complete sequences of the genome in functional studies of *Drosophila* genes.

We showed that the mechanism of gene targeting by homologous recombination (HR) in *Drosophila* was essentially identical to that of traditional yeast and mouse targeting. The novelty of the *Drosophila* method lies in the way in which the donor DNA molecule for targeting was introduced into the cells. Instead of direct delivery by electroporation or chemical treatment of cells, the *Drosophila* donor was generated within the germ cells of intact animals. This was done by first introducing the donor randomly into the genome as a specially constructed P transposable element. Then, a free moving donor molecule was generated from the chromosomal P element by the concert actions of the yeast FLP site-specific recombinase and the yeast I-SceI site-specific DNA endonuclease which were synthesized in *Drosophila* cells. To accomplish this, flies were produced by crossing that contained a donor P element in the genome, as well as two transgenes expressing FLP and I-SceI respectively. In the germ cells of these flies, FLP excised the donor as an extrachromosomal circle, and I-SceI made a double strand break (DSB) at its recognition site imbedded in the donor circle. The presence of this DSB stimulated HR between the donor and the target leading to donor integration at the endogenous locus. The end result was a tandem duplication of the target gene. Targeting events were recovered as progeny of these flies after they were mated to a tester strain.

We showed that, similar to results from yeast and mouse targeting experiments, targeting efficiency in *Drosophila* (1) was locus dependent; (2) could be improved by increasing donor:target homology; and (3) would be affected by sequence heterology between the donor and the target. We also showed some special features of *Drosophila* targeting: (1) targeting was much more efficient in the female germline than in the male germline, which might be related to the fact that *Drosophila* males do not have meiotic recombination; and (2) targeting frequencies were highly variable among different insertions of the donor P element, which might be caused by chromosomal position effects on donor generation. Using two different modifications of the original targeting scheme, we and others have succeeded in generating mutant alleles of different genes. We have also developed a two-step allelic replacement method that allowed us to recover a single target gene harboring the desired sequence modification starting with a tandem duplication of the target locus. We propose that the *Drosophila* targeting scheme should be adaptable to other organisms that have been successfully transformed. This is based

on the facts that both the site-specific recombination system and the site-specific DSB system have been proven effective in many different heterologous systems.

Annotating Proteomes using Highly Specific Protein Scoring Matrices

Serge Saxonov
Beckman Center, Room B 403
Stanford University School of Medicine
Stanford, CA 94305
telephone: 650-723-5976
fax: 650-725-6044
email: saxonov@stanford.edu
prestype: Platform
presenter: Serge Saxonov

Serge Saxonov, Qiaojuan Su and Douglas L. Brutlag
Department of Biochemistry, Stanford University, Stanford California, 94305-5307

In order to identify new functions in protein families and superfamilies, we have developed eBLOCKS, a database of ungapped alignments (blocks) of highly conserved protein regions. eBLOCKS automatically builds blocks directly from protein sequences such as the SWISS-PROT database. Each unclassified protein sequence is used as a PSI-BLAST query and compared against the entire sequence database. The resulting PSI-BLAST alignments are analyzed by a modified K-means clustering algorithm to generate protein groups with different levels of similarity, representing protein families and super-families. Each group of conserved sequences are aligned heuristically and trimmed into ungapped regions. The current eBLOCKS database contains 81,413 blocks. The completely automated eBLOCKS database has several advantages over BLOCKS+ database, which is built from protein groups pre-defined in a number of protein family databases. The eBLOCKS database is more comprehensive: it covers the majority of BLOCKS+ database and yet 65% of its blocks are novel to eBLOCKS.

Instead of representing a region by only one block as in BLOCKS+, eBLOCKS database enumerates blocks representing different family levels for each conserved region and thus maximizes sensitivity and specificity when used to search an unknown sequence. Unlike BLOCKS+, eBLOCKS does not require three conserved positions in the blocks, and can thus incorporate blocks with variability allowed at all positions.

Evaluation tests show that eBLOCKS has greater sensitivity than BLOCKS+. In particular, we have used eBLOCKS and BLOCKS+ as scoring matrices to search the human proteome for significant hits at the specificity of 10-3. We were thus able to annotate 67% of the human proteins in RefSeq with eBLOCKS vs. 54% when using BLOCKS+. This rate approaches the level of what one can achieve with homology searches. Annotating proteins with blocks gives more accurate annotation information than homology-based searches since the use of blocks focuses precisely on the functionally important regions. The eBLOCKS database is available on the World Wide Web at http://fold.stanford.edu/eblocks/

An improved method for cloning EST-specific full-length cDNAs

Sabrina Semprini
Department of Medicine
Human Molecular Genetics Unit
University of Vermont
Health Science Research Facility 227
149, Beaumount Ave
Burlington, Vermont 05405
U.S.
telephone: 001-802-656-8921
fax: 001-802-656-8903
email: ssemprin@zoo.uvm.edu
prestype: Poster
presenter: Sabrina Semprini

S. Semprini[1], C. Collins[1], A. Colosimo[2], M. Pavoni[3], B. Dallapiccola[2], G Novelli[3], D.C. Gruenert[1]
1.Human Molecular Genetic Unit, University of Vermont, VT USA
2.CSS Mendel Institute, Rome, Italy
3.Dept. of Biopathology and Diagnostic Imaging, Tor Vergata University of Rome, Rome Italy

Primary approaches for cloning full-length cDNAs involve cDNA library screening and rapid amplification of cDNA ends (RACE). These techniques are laborious and time consuming. Recently, a technique developed to clone expression sequence tag (EST)-specific full-length cDNA has streamlined this process considerably. This strategy is based on EST-specific mRNA enrichment with EST-derived biotin labeled "capture" primers and streptavidin-coated magnetic beads. Enriched mRNA is then used to generate cDNA for cloning into plasmid vectors. This strategy has now been further modified to improve its efficiency and specificiy for cloning full-length cDNA. This modification involves the ligation of a double-stranded DNA (dsDNA) adapter to a RNA/cDNA hybrid molecule. As a proof-of-principle, first strand cDNA synthesis was performed on kanamycin mRNA using a Not I oligo(dT)15 primer. The second adapter, containing an MluI restriction site, was ligated to the mRNA-cDNA hybrid. Following adapter addition, a second strand of cDNA was synthesized. Resultant (ds)cDNAs were size fractionated prior to ligation into the pSPORT I vector. Bacterial cells were transformed with the vector with the cDNA insert. Resultant colonies were screened by PCR. Sequence analysis showed full-length kanamycin sequence in positive clones. These results indicate the potential of this technique in the rapid cloning of full-length cDNAs. This improved technique has now been applied to isolating full-length sequences corresponding to breast cancer candidate ESTs, using a novel cDNA cloning vector that optimizes the cloning of large cDNA inserts.

This work was founded by NIH Grant #CA80200-03

‘Green Genes’ - Systematic subcellular localisation of novel proteins identified by large-scale sequencing identifies new regulators of membrane transport

Jeremy C. Simpson
Cell Biology and Biophysics
EMBL-Heidelberg
Meyerhofstrasse 1
69117 Heidelberg
Germany
telephone: 49 6221 387232
fax: 49 6221 387306
email: simpson@embl-heidelberg.de
prestype: Platform
presenter: Jeremy C. Simpson

Jeremy C. Simpson, Veronika E. Neubrand, Stefan Wiemann and Rainer Pepperkok

Large-scale cDNA sequencing projects are providing us with vast numbers of open reading frames (ORFs) encoding novel proteins of completely unknown function. As a first step towards their characterisation we have tagged the first 200 of these with the green fluorescent protein (GFP), and examined the subcellular localisations of these fusion proteins in living cells. The cloning of ORFs on this scale has been made possible using a novel recombination-based cloning technology which can generate both the N-terminal and C-terminal fusions in a single reaction. Subcellular localisations to discrete compartments were determined for over 75% of the fusions, providing the first functional information about these proteins, which is of particular importance for those molecules where bioinformatic tools were unable to provide any information. The tools generated by such an approach combined with the ease of cloning, mean that automation of more functional assays for such a large set of molecules is now a reality.

Of the many molecules which we have identified which localise to organelles of the secretory pathway (endoplasmic reticulum, Golgi, endosomes, plasma membrane), we have identified a novel regulator of sorting at the level of the trans-Golgi network (TGN). This protein appears to interact with adaptor molecules, specifically regulating the export of passenger molecules en route to the plasma membrane. The overexpression of this protein acts as a potent inhibitor of membrane transport from this organelle. This novel approach to finding new molecules involved in membrane transport already appears to be fruitful.

Phylogenetic comparison of the RNA editase ADAR2 genes reveals conservation and diversity in editing site sequence and alternative splicing patterns

Dobromir Slavov
Eleanor Roosevelt Institute
1899 Gaylord Street
Denver, CO 80206-1210
USA
telephone: 303-336-5684
fax: 303-333-8423
email: slavov@eri.uchsc.edu
prestype: Poster
presenter: Dobromir Slavov

Dobromir Slavov and Katheleen Gardiner

ADAR2 (Adenosine Deaminase that Acts on RNA 2) is one of three known mammalian genes that encode A-to-I RNA editases, enzymes that deaminate specific adenosine (A) residues in specific pre-mRNAs to produce inosines (I). Known substrates of ADAR2 include sites within pre-mRNAs of the ionotropic glutamate receptors, GluR2-GluR7, and the serotonin receptor, 5HT2C. Because the ribosome reads I residues as G residues, and because the edited sites largely occur within coding regions, most of these editing events result in amino acid changes. Editing has been demonstrated to affect protein function, reducing calcium permeability and desensitization recovery times in the glutamate receptors and reducing receptor-G protein coupling in the serotonin receptor. Editing activity is highly regulated; levels are rarely 100% for any of the known pre-mRNA substrates and vary with developmental time and among brain regions. Lack of editing, at least of GluRB, is neonatal lethal. A-to-I RNA editing and the enzymes controlling it thus provide an important mechanism for regulation of neurological development and function by means of regulation of protein function and diversity.

Because of the important biological role of A-to-I editing, it is of interest to examine the evolutionary conservation of ADAR2 regulatory features. We therefore undertook a phylogenetic comparison of the genomic structure, editing and alternative splicing of the ADAR2 genes from human, mouse, chicken, fugu and zebrafish. Here we show that the genomic sequences and RNA secondary structures required for the ADAR2 self-editing within intron 2 are highly conserved among all organisms, more highly conserved than coding exons. Also conserved are general patterns of alternative splicing within the 5’ UTRs. There is, however, diversity in other respects. In chicken and mammalian ADAR2, but not in fish, alternative splicing that likely affects the catalytic domain is observed, although locations and mechanisms differ. Complex 3’ end alternative splicing appears to be mammalian-specific, as does conservation within the 3’UTR.

New Methods for the Characterization of Proteomes: Initial Application to *D. radiodurans*

Richard D. Smith
Pacific Northwest National Laboratory
P.O. Box 999 (K8-98)
Richland, WA 99352
telephone: 509-376-0723
fax: 509-376-7722
email: rd_smith@pnl.gov
prestype: Platform
presenter: Richard D. Smith

Richard D. Smith, Gordon A. Anderson, Mary S. Lipton, Ljiljana Pasa-Tolic, Margaret F. Romine, Jim Fredrickson, Yufeng Shen, Timothy D. Veenstra, Thomas P. Conrads, and Harold R. Udseth

The patterns of gene expression, protein post-translational modifications, covalent and non-covalent associations, and how these may be affected by changes in the environment, cannot be accurately predicted from DNA sequences. Approaches for proteome characterization are presently based upon mass spectrometric analysis of 2-D gel separated proteins. However, this approach remains constrained by the speed of the 2-D gel separations, the sensitivity needed for protein visualization, the speed and sensitivity of subsequent mass spectrometric analyses for identification, and the limitations of spot visualization for quantitation. Our objective is to circumvent the limitations of this approach by directly characterizing the cell's polypeptide constituents using a combination of high resolution separations and the mass accuracy and sensitivity obtainable with Fourier transform ion cyclotron resonance (FTICR) mass spectrometry. Protein identification is based upon global approaches for protein digestion and accurate peptide mass analysis; i.e. "Accurate Mass Tags" (AMTs). Our two-stage strategy exploits FTICR to based upon validation and subsequent routine measurement of peptide AMTs from "potential mass tags" initially identified using tandem mass spectrometry methods, and thus providing the basis for both extremely high confidence in identifications and subsequent high throughput proteome-wide measurements. A single high resolution capillary liquid chromatography separation combined with high sensitivity, high resolution and accurate FTICR measurements has been shown to be capable of characterizing peptide mixtures of more than 100,000 poly peptide components. Attractions of the approach include the capability for automated high-confidence protein identification, broader proteome coverage, and the capability for exploiting stable-isotope labeling methods to realize high precision for relative protein abundance measurements. Using this approach we have been able to identify >60% of the predicted ORFs for Deinococcus radiodurans, including ~50% of proteins previously designated only as "hypothetical". The status of our efforts will be presented in the context of results for D. radiodurans.

This work was supported by the Office of Biological and Environmental Research of the U.S. Department of Energy. Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute through Contract No. DE-AC06-76RLO 1830.

Strategies for construction of subtracted libraries enriched for full-length cDNAs and for preferential cloning of rare mRNAs

Marcelo Bento Soares
The University of Iowa
451 Eckstein Medical Research Building
Iowa City, IA 52242
telephone: 319-335-8250
fax: 319-335-9565
email: bento-soares@uiowa.edu
prestype: Platform
presenter: M. Bento Soares

Brian Berger, Sergey Malchenko, Irina Koroleva, Einat Snir, Tammy Kucaba, Maria de Fatima Bonaldo & Marcelo Bento Soares.

Subtracted libraries enriched for full-length cDNAs.
A major challenge of the ongoing NIH Mammalian Gene Collection Program is the identification of sufficient novel full-length cDNAs to enable achieving the yearly full-length sequencing goals of the project. In an effort to assist in the identification of novel full-length cDNAs we have constructed full-length-enriched libraries and we have developed a novel method for generation of subtracted libraries enriched for full-length cDNAs. Conventional subtractive hybridization procedures cannot be applied for full-length-enriched libraries because a truncated clone in the driver population has the potential to subtract its full-length counterpart from the library. Briefly, 100-150 bp single-stranded overhangs are generated at the 5' end of all clones in the library (tracer), for hybridization with a biotinylated driver population comprising representative clones of every sequence contig identified in the starting full-length-enriched library. The subtracted population is purified from the hybrids using streptavidin-coated magnetic beads, repaired and electroporated into bacteria for propagation of a subtracted full-length-enriched library. We have used this method successfully to generate a subtracted full-length-enriched library derived from germinal center B cells.

Preferential cloning of rare mRNAs.
Discovery of rare mRNAs in large-scale EST projects remain difficult and inefficient because of poor representation of such transcripts in cDNA libraries. In an attempt to expedite the identification of rare mRNAs, we developed a novel method for prerential cloning of rare mRNAs. Briefly, mRNA is hybridized with a driver comprising most/all already identified cDNAs and subsequently destroyed with RNAse H. THe remainder intact mRNA is linearly amplified and cloned for production of a library enriched for rare mRNAs. We have used this method to construct a mouse cDNA library enriched for rare mRNAs from hippocampus. The efficacy of our method was demonstrated by sequencing and by microarray hybridization analises.

The alternative exon database (AEDB)

Presentation

telephone: +49 9131 8524622
fax: +49-9131-85-24605
email: stefan@stamms-lab.net
prestype: Platform
presenter: Stefan Stamm

Stefan Stamm, Peter Stoilov, and Michael Q. Zhang

We compiled a comprehensive database (AEDB) of alternative exons from the literature and analyzed them statistically. Most alternative exons are cassette exons and are expressed in more than two tissues. Of all exons whose expression was reported to be specific for a certain tissue, the majority was expressed in the brain. Whereas the length of constitutive exons follows a normal distribution, the distribution of alternative exons is skewed towards smaller ones. Furthermore, alternative splice sites deviate more from the consensus: their 3' splice sites are characterized by a higher purine content in the polypyrimidine stretch and their 5' splice sites deviate mostly at the +4 and +5 positions from the consensus sequence. Furthermore, for exons expressed in a single tissue, adenosine is more frequently used at the -3 position of the 3' splice site. These results confim our previous findings form neuron-specific exons. The usage of exons with suboptimal splice sites is stimulated by regulatory sequences known as exonic enhancers. SELEX eperiments and analysis of model systems shown that two major groups of enhancers exist: purine rich and AC-rich. In addition to those sequence elements, analysis using a Gibbs algorithm identifies several motifs in exons surrounded by weak splice sites and in tissue-specific exons. Those naturally exonic enhancer motifs are more degenerate than sequences obtained by SELEX. Together, these data indicate a combinatorial effect of weak splice sites, atypical nucleotide usage at certain positions and functional enhancers as an important contribution to alternative exon regulation.

A technique for genome-wide identification of differences in the interspersed repeats integrations between closely related genomes.

Anton Buzdin[a*], Ilgar Mamedov[a], Konstantin Khodosevich[a], Tatyana Vinogradovaa, Yuri Lebedev[a], Gerhard Hunsmann[b], **Eugene Sverdlov**[a]

[a]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, 16/10 Miklukho-Maklaya St, Moscow 117871, Russia

[b]German Primate Centre, Department for Virology and Immunology, Kellnerweg 4, D-37077 Goettingen, Germany

ABSTRACT

A method of Targeted Genomic Difference Analysis (TGDA) was developed allowing genome-wide detection of differences in integration sites of interspersed repeats between related genomes. The method includes two principal steps: (i) a whole genome selective amplification of the flanks adjacent to target interspersed repetitive elements in both genomic DNAs under comparison and (ii) subtractive hybridization of the selected amplicons. Differences between human and chimpanzee's genomes in the integration sites of HERV-K(HML-2) human endogenous retroviruses and related solitary long terminal repeats (LTRs) were analyzed. Of 55 randomly chosen sequenced clones from a library enriched with human specific integration (HSI) sites, 33 (60%) represented HSIs, 14 (25%) were found also in the chimpanzee's genome and 8 (15%) could not be unambiguously assigned. The number of the HSI sites in the library was estimated to be 67 that corresponds to 2-3% of all HERV-K(HML-2) LTRs integrations in the human genome. Together with the HSIs described by other authors, the number of the characterized LTR HSIs is now 39. All the revealed HSI LTRs belong to two closely related evolutionary young groups, thus suggesting parallel activities of two master genes in the hominid lineage. No deletion/insertion polymorphism was detected for any of the LTR HSIs by genomic PCR of DNA samples from 25 unrelated Caucasoid individuals. The HSI LTRs were assigned to the human genome sequenced loci and the neighboring genes were identified. Many of the LTRs were found to be located close to genes or within gene introns. The technique developed can be applied to any type of moderate repeats in mammalian genomes, including young subfamilies of Alu and L1 repeats.

Interaction between PPAR-gamma and Retinoic Acid Receptor (RAR) pathways during the differentiation of monocytic leukemia cells

Attila Szanto
Department of Biochemistry and Molecular Biology
University of Debrecen, Medical and Health Science
Center
98 Nagyerdei krt.
DEBRECE, H-4012
telephone: +36-52-416-432
fax: +36-52-416-432
email: szantoa@indi.biochem.dote.hu
prestype: Poster
presenter: Attila Szanto

Attila Szanto and Laszlo Nagy
Department of Biochemistry and Molecular Biology, University of Debrecen, Nagyerdei krt. 98, DEBRECEN, Hungary H-4012

Both PPAR-gamma and Retinoic Acid Receptor (RAR) belong to the nuclear hormone receptor superfamily and they play a role in the differentiation of myelomonocytic cells. Activation of the RAR pathway leads to a granulocytic direction and activation of the PPAR-gamma pathway leads to monocytic differentiation. We looked for factors that determine the fate of the differentiating cells and characterized the interactions between the two pathways using methods of global transcription profiling and quantitative transcript determinations.

We used monocytic leukaemia cell lines and treated the cells with nuclear hormone receptor agonists and antagonist. We found that PPAR-gamma mRNA levels increased parallel with the degree of monocytic differentiation. Activation of either PPAR-gamma or its heterodimeric partner RXR led to inhibition of proliferation and subsequent differentiation. We monitored this process by determining the surface expression of the myeloid markers, CD14 and CD18 and several receptor target genes such as CD36, LXR-alpha (liver X receptor) for PPAR:RXR activation and CD38 and RAOH (retinoic acid hydroxilase) for RAR:RXR heterodimer activation.

We analyzed the pharmacological profile of differentiation and the mechanism of interactions between the receptor pathways and found that activation of the RAR leads to an increased expression of the PPAR-gamma target genes. This phenomenon is accompanied with an elevation of the PPAR-gamma receptor molecule number. On the other side PPAR-gamma activation leads to a decreased response of the RAR target genes. We characterized these processes and propose potential mechanism for them.

Annabel E. Todd (1), Christine A. Orengo (1) and Janet M. Thornton (1,2)

1. Department of Biochemistry and Molecular Biology, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK
2. Department of Crystallography, Birkbeck College, Malet Street, London, WC1E 7HX

The recent growth in protein sequence and structural databases has revealed the functional diversity of many protein superfamilies. An understanding of how such diversity has evolved through sequence and structural changes is essential for the accurate functional annotation of the large number of uncharacterised gene products identified in genome sequencing projects. Given the large number of genes in the human genome, but a comparatively small number of folds, extensive combination, mixing and modulation of existing folds has occurred during evolution to generate the multitude of functions necessary to sustain life. With the first working draft of the human genome complete, and the sequencing of other multi-cellular organisms underway, a grasp of these evolutionary processes is required if we are to benefit from this wealth of data.

We have analysed how functional changes are implemented by modulation of sequence and structure with reference to 31 diverse enzyme superfamilies, and thus provide an overview of the mechanisms by which functional diversity has evolved. This has involved extensive reading of the literature combined with analyses of our own. Functional variation occurs mostly in more distantly related proteins (<40%) and the structural data have been essential for understanding the molecular basis of observed functional differences. A large number of variations and peculiarities are observed, at the atomic level through to gross structural rearrangements. Using selected examples, we present the structural and functional attributes which are conserved within some superfamilies and those that differ, and what bearing, if any, these similarities and changes have on protein function. The implications these observations have on structural genomics projects will be discussed.

Functional Analysis of the Mammalian Genome by Large Scale Gene Trap Mutagenesis

P. Van Sloun
Laboratory for Molecular Hematology, University of Frankfurt Medical School
Frankfurt am Main, Germany.
Telephone: 49-69-6301-6390; Fax: 49-69-6301-7877
Email: vansloun@em.uni-frankfurt.de

Gene trap mutagenesis in mouse embryonic stem (ES) cells is a complementary approach to the functional annotation of the mouse genome. In concert with chemical (ENU) mutagenesis, the approach enables the analysis of gene function in the context of an entire organism and thus furthers our understanding of human disease. Identified disease genes will then greatly assist the process of drug discovery.

We have established a Gene Trap Consortium in the context of the German Human Genome Project (HUGO) with the goal to saturate the mouse genome with insertional mutations. By disrupting genes repeatedly, we anticipate to create an allelic series for each gene expressed in the ES cells.

Within the last four years, the GGTC has generated 14 000 mutant ES cell lines and has identified the gene trap integration sites in 6829 clones. Of the generated gene trap sequence tags (GTSTs), 5269 high quality sequences were used in NCBI-GenBank database searches. The sequences produced 1099 (21%) matches with non-redundant genes, 1552 (29%) matches with expressed sequence tags (EST's), and 2618 (50%) no matches (e£1.0e-30) in part reflecting novel genes. Of all integrations into previously characterized genes, more than 40 occurred in genes involved in human disease. Forty-six ES cell lines were passaged through the germ line and the phenotypes induced by transgenes are being presently assessed.

Data obtained from individual clones, such as GTSTs, blast results and OMIM references are deposited in the GGTC's database which is publicly accessible via http://genetrap.gsf.de. Clones of interest can be identified using blast or keyword search and corresponding cell lines, stored frozen at the GSF in Munich, are freely available upon request.

Since none of the gene trap vectors currently in use incorporates all features one would expect from an ideal vector, the Consortium is actively pursuing the development of novel gene trap vectors. One of these is the U3Ceo vector, a retroviral gene trap designed to specifically trap secreted and transmembrane proteins. Compared to previously reported secretion traps, U3Ceo is at least 4 times as efficient and thus ideally suited for large scale mutagenesis of signal sequence encoding genes.

Robustness and Redundancy in Large Genetic Networks

Andreas Wagner
Department of Biology, University of New Mexico
Albuquerque, NM, USA
Phone: 505-277-2021; Fax: 505-277-0304
Email: wagnera@unm.edu

Is redundancy caused by gene duplications responsible for the weak phenotype of many gene knock-out mutations? I will analyze the structure and evolution of a large protein interaction network in yeast to address this question. Structurally, the network belongs to the growing class of scale-free small-world networks found in a variety of unrelated areas of science, such as the internet, sociological networks, and metabolism. Evolutionarily, the network shows a high turnover of protein interactions. Abundant shared protein interactions among duplicate gene products in this network might indicate pervasive genetic redundancy. However, such common interactions are rare.

Analysis of myeloid cell development at the RNA and protein level

Presentation

Sherman M. Weissman
Sterling Professor of Genetics and Medicine
Yale University School of Medicine
295 Congress Avenue, 336 BCMM
New Haven, CT 06510
telephone: (203) 737-2282
fax: (203) 737-2286
email: barbara.gramenos@yale.edu
prestype: Platform
presenter: Sherman M . Weissman

Zheng Lian[1], Shigeru Yamaga[1], Yasuhiro Tsukahara[1], Yuval Kluger[2], Dov Greenbaum[1], Mark Gerstein[2], Nancy Berliner[3], Peter Newburger[4], Sherman M. Weissman[1]
[1]Department of Genetics, Yale University School of Medicine, New Haven, CT 06510, [2]Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, [3]Department of Genetics and Medicine, New Haven, CT 06510, [4]Department of Pediatrics, University of Massachusetts Medical School, Worcester, MA 01605

Polymorphonuclear cells are a highly differentiated post-mitotic cell population that differentiates through a series of steps from a common hematopoietic precursor. Mature PMN can be readily isolated from the blood of normal individuals in amounts sufficient for biochemical analysis. Model systems for inducing myeloid differentiation or polymorphonuclear cell development in vitro also exist. One favorable system is the MPRO murine promyelocytic cell line. In this line development is blocked by a dominant negative retinoic acid receptor. Upon addition of all trans retinoic acid, the cells differentiate over about four days into essentially mature neutrophils.

To analyze the processes of neutrophil differentiation and activation we have studied RNA expression in normal resting and activated neutrophils and in developing MPRO cells, using both gel display methods and oligonucleotide chip analysis. In addition, we have compared RNA patterns in less differentiated hematopoietic cells including hematopoietic stem cells and cells differentiated into other hematopoietic lineages. To complement this approach we have analyzed protein expression in the myeloid cells, using two-dimensional gel electrophoresis with a variety of pH range isoelectric focusing dimensions, and protein identification by MALDI-TOF mass spectroscopy. Technically, there are well known limitations in both the detection and quantitation of proteins by the 2D methods, and there is only a moderate correlation between mRNA levels or the changes in mRNA levels and the corresponding protein levels.

Analysis of mRNA levels show interesting changes in the expression of genes for certain transcription factors, some but not all of which had been previously implicated in myeloid development. To further investigate these we are attempting 2D gel electrophoresis of transcription factor enriched protein preparations from myeloid cells and also exploring introduction of factors into developing cells by the use of “Trojan Horse” peptides.

Generating and Sequencing Full-length cDNAs of Novel Human Genes within the German cDNA Consortium

Ruth Wellenreuther
German Cancer Research Center (DKFZ)
Dept. Molecular genome Analysis (H0600)
Im Neuenheimer Feld 506
69120 Heidelberg
telephone: ++49-6221-42 4711
fax: ++49-6221-42 52 4711
email: r.wellenreuther@dkfz.de
prestype: Poster
presenter: Ruth Wellenreuther

Ruth Wellenreuther, Stefan Wiemann, Daniela Heiss, Annemarie Poustka

We generate human cDNA libraries that are enriched in full length clones i.e. from the translation start to the poly A tail. These libraries are used for a) systematic sequencing within the cDNA consortium of the Genome Project aiming at the identification and analysis of as many new genes as possible and b) for screening to isolate full length clones of partial genes.

Libraries are created by directional cloning of cDNA's into plasmid vectors. Full-length enrichment is achieved via Clontech's SMART technology. In this strategy, which is PCR-based, we amplify and clone selective size windows of the cDNA fraction above 3 kb.

Clones from the libraries generated within this project are the major source for the cDNA sequencing effort of the German Human Genome Project that is carried out by a consortium of nine laboratories. Over 800 full-length cDNAs of novel human genes could be identified by the consortium within the last 3 years.

The clones generated in this project are available through the Resource Center of the German Human Genome Project (www.rzpd.de).

Elucidation of Gene Function Must Include the Genomic Context

Thomas Werner
Institute of Experimental Genetics, GSF-National Research Institute for Environment and Health, Neuherberg, Germany and Genomatix Software GmbH, München, Germany
Phone: 49-89-3187-4050; Fax: 49-89-3187-4400
E-mail: Werner@gsf.de

The annotation of human genomic sequences is far from complete and "anonymous" still describes a major part of the genomic sequences because experimental evaluation is unable to keep pace with sequencing. Therefore, rapid and precise functional annotation of the collected large-scale sequences by in silico methods will be mandatory. In order to understand biological functions of genes it will be especially important to elucidate also features that depend on the genomic context rather than being intrinsic to individual gene. For example, analysis of the amino acid sequences derived from a gene allows elucidation of intrinsic features of the protein. However, this should not be confused with the elucidation of the function(s) of the gene because a significant portion of the gene functions cannot be determined from the amino acid sequence. The trivial case is that the protein only has a partial function (e.g. as a part of a functional heterodimer). Unless the partner(s) are known there is no way to determine the function. Another less obvious reason is that the amino acid sequence only provides the information which protein can be encoded, but no clue to where and when this might be happening.

The central processor of gene regulation is the gene promoter, which by definition comprises the 5'-end of the transcribed sequence. There is a variety of other elements like matrix attachment (S/MAR) or locus control regions (LCR), enhancers, silencers influencing transcription. However, none of these other control regions can have any effect on the actual transcription of the gene unless its signals are integrated and executed within the promoter of the gene, as only there the RNA polymerase binds. Therefore, promoters can be seen as the central processing unit of gene transcription. Nevertheless, promoters tell relatively little about gene regulation if looked at in isolation. The actual behavior of a promoter is determined by the nature and quantities of transcription factors (specialized proteins) within the cell and cannot be determined solely from the promoter sequence. As in case of the amino acid sequence, the promoter sequence just determines, what kind of regulation is possible, not what is really happening in a given cell and situation.

A short survey will be given about how to find promoters in large genomic sequences followed by the reasoning of how to compare promoter sequences in order to elucidate the functional context of genes. Promoters can no longer be treated as functional units as the same promoter may have quite different functions in a different cellular context. This illustrates that there is no way to determine the promoter function in a single analysis as often promoters have not one but several different functions.

IDENTIFICATION OF TRANSCRIBED SEQUENCES: MORE CHALLENGES?

Stefan Wiemann
Dept. Molecular Genome Analysis
Genomics&Proteomics
German Cancer Research Center
Im Neuenheimer Feld 506
69120 Heidelberg
Germany
telephone: +49 6221 42 4702
fax: +49 6221 4252 4702
email: s.wiemann@dkfz.de
prestype: Platform
presenter: Stefan Wiemann

Stefan Wiemann, Ruth Wellenreuther, Petra Heidrich, Regina Albert, Ingo Schupp, Vladimir Kuryshev, Jiancong Xu, Jeremy Simpson[2], Rainer Pepperkok[2], Annemarie Poustka and the German cDNA Consortium
[1]Molecular Genome Analysis, German Cancer Research Center, Im Neuenheimer Feld 506, D-69120 Heidelberg, Germany
[2]Cell Biology Program, European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany

We have formed a network in the frame of the German Genome Project aiming at the generation and sequencing of novel full-length cDNAs, and the comprehensive functional analysis the deduced proteins. The project started in September 1997. Over 4,500 cDNAs (> 11.6 Mb) have been sequenced since. The set of fully sequenced clones in combination with EST-sequenced clones is used to generate a minimal set of full-length clones for employment in subsequent functional analysis. A progress report of the network activities and achievements will be presented.

Using the full-length cDNAs in a systematic analysis of gene predictions, which are based on genomic sequence, we observe that most (>90%) gene predictions do not correctly annotate the true 5'-end of coding regions. As a consequence the amplification of ORFs that are based on gene predictions is not reliable enough to obtain the majority of protein coding regions. The analysis of full-length cDNAs, therefore, still remains the method of choice for correctly identifying complete protein coding sequences.

An RNA Surveillance Mechanism that Detects Aberrant Transcripts

Miles Wilkinson, Ph.D.
Professor
Immunology Department, Box 180
U. T. M. D. Anderson Cancer Center
1515 Holcombe Blvd.
Houston, TX 77030
telephone: (713) 794-5526
fax: (713) 745-0846
email: mwilkins@mdanderson.org
prestype: Platform
presenter:= Miles Wilkinson

Jun Wang, John I. Hamilton, Mark Carter, Shulin Li, and Miles F. Wilkinson
Department of Immunology, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030, U.S.A.

Nonsense codons terminating translation prematurely generate potentially deleterious truncated proteins. The hazards of such dominant-negative proteins are diminished by nonsense-mediated decay (NMD), an RNA surveillance pathway that recognizes and degrades transcripts harboring premature termination codons (PTCs). Here we report that PTCs in a T-cell receptor (TCR) gene not only decrease the levels of the normally spliced (norm) mRNA but also surprisingly increase the levels of an alternatively spliced (alt) transcript that has skipped the offending PTC. RNA half-life analysis demonstrated that the increase in the amount of alt mRNA did not result from an increase in its stability. Instead, it resulted from increased splicing from the alternative splice sites used to generate the alt mRNA, as demonstrated by analyzing the levels of alt and norm spliced introns in vivo. One model that has been proposed to explain how nonsense mutations could influence RNA splicing is that such mutations disrupt splicing enhancers. We demonstrated that this splicing-enhancer disruption model does not apply in our case, as the alt TCR transcript was specifically induced by PTCs generated by point mutations and frameshifts at various positions in the rearranged VDJ exon but was not induced by missense or silent mutations. Thus, our results indicate that bona fide nonsense codons stimulate nuclear RNA splicing, which is paradoxical, as the only known machinery that recognizes nonsense codons is the translation machinery operating in the cytoplasm. We therefore examined whether alt mRNA induction in response to PTCs occurs in the nucleus or cytoplasm and whether it depends on a translation-like mechanism. Our results showed that alt mRNA induction occurred in the nuclear fraction of mammalian cells and that it had the following features of translation: (1) it depended on an inititator AUG, (2) it depended on a Kozak consensus sequence surrounding the AUG codon, (3) it was inhibited by a stem loop and suppressor tRNAs. We therefore propose that the induction of alt mRNA in response to PTCs results from a tRNA- and AUG-dependent scanning mechanism that acts in the nuclear fraction of cells.

Dissection of a complex phenotype by gene identification, gene function analysis and comparative sequencing between man and mouse

Andreas Winterpacht, Ph.D
Institute of Human Genetics
University of Hamburg
Butenfeld 42
D-22529 Hamburg
Germany
telephone: ++49-40-42803-5535
fax: ++49-40-42803-5128
email: winterpacht@uke.uni-hamburg.de
prestype: Platform
presenter: Andreas Winterpacht, Ph.D.

Winterpacht, Andreas (1); Schlickum S. (1); Pfarr N. (1,2); Steglich C. (1); Heller R. (1); Endele, S.(1)
(1) Institute of Human Genetics, University of Hamburg, D-22529 Hamburg, Germany; (2) Children´s Hospital, University of Mainz, D-55101 Mainz, Germany

The Wolf-Hirschhorn syndrome (WHS) is a complex and variable malformation syndrome associated with the deletion of the terminal short arm of one chromosome 4. The WHS critical region (WHSCR) has been confined to a 165 kb gene rich region on chromosomal subband 4p16.3. Clinical and cytogenetic data indicate that WHS is a contiguous gene syndrome, which means that haploinsufficiency of more than one gene in the affected chromosome region contribute to the phenotype. In order to elucidate the etiology and pathogenesis of this syndrome our studies aim at identifying all genes and regulatory regions contributing to the phenotype.

We sequenced the corresponding WHSCR in the mouse genome and applied interspecies sequence comparison, computer assisted analysis of the present human sequence, as well as functional studies. Two known and three novel genes were identified. In addition, we detected several EST matches as well as predicted exon sequences that did not correspond to any of these transcription units. By extensive RT-PCR analysis and rapid amplification of cDNA ends (RACE) we could demonstrate that most of these sequences correspond to a single novel transcript which is expressed at a low level in several tissues. The transcript contains repetitive sequences, does not display any significant open reading frame and undergoes complex alternative splicing. The 5´-end of the transcript overlaps with exon 1 of a gene on the opposite strand indicating that it might function as an antisense transcript.

By detailed expression and functional studies we are currently investigating which of the genes contribute to the WHS phenotype. Moreover we elucidate the possibility of coregulation of these genes.

CART CLASSIFICATION OF HUMAN 5'UTR SEQUENCES

Prof. Michael Q. Zhang
Watson School of Biological Sciences
Cold Spring Harbor Laboratory
1 Bungtown Road
Cold Spring Harbor, NY 11724
telephone: 516-367-8393
fax: 516-367-8461
email: mzhang@cshl.org
prestype: Platform
presenter: Michael Q Zhang

Ramana V. Davuluri,[1] Yutaka Suzuki,[2] Sumio Sugano[2] and Michael Q. Zhang[1]

[1]Cold Spring Harbor Laboratory; 1 Bungtown Road; PO Box: 100; Cold Spring Harbor, New York 11724; Tel: 516-367-6956; Fax: 516-367-8461; Email: {ramana, mzhang}@cshl.org

[2]Department of Virology; Institute of Medical Sciences; The University of Tokyo; 4-6-1, Shirokanedai, Minato-ku; Tokyo 108-8639, Japan

A non-redundant database of 2312 full-length human 5'UTRs was carefully prepared using state of art experimental and computational technologies. A comprehensive computational analysis of this data was conducted for characterizing the 5'UTR features. Classification and Regression Tree analysis was used to classify the data into three distinct classes. Class I consists of mRNAs that are believed to be poorly translated with long 5'UTRs filled with potential inhibitory features. Class II consists of TOP mRNAs that are regulated in growth dependent manner and Class III consists of mRNAs with favorable 5'UTR features that may help efficient translation. The most accurate tree we found has 92.5% classification accuracy as estimated by cross validation. The classification model included presence of TOP (terminal oligopyrimidine tract), secondary structure, 5'UTR length and presence of upstream AUGs (uAUGs) as the most relevant variables. The present classification and characterization of the 5'UTRs provide precious information for better understanding the translational regulation of human mRNAs. Furthermore, this database and classification can help people to build better computational models for predicting the 5'terminal-exon and separating the 5'UTR from the coding region.