Beyond the Identification of Transcribed Sequences: Functional and Expression Analysis

10th Annual Workshop

October 28-31, 2000
Heidelberg, Germany

Co-sponsored by the US Department of Energy

Speakers * Abstracts * Home

Meeting Objective

The 2000 Workshop is the tenth in this series. Originally designed to discuss methods for isolation and verification of transcribed sequences from the mammalian genome, progress in this and other areas has prompted significant broadening of this focus.

Topics to be discussed include but are not limited to: 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 in model organisms (e.g. yeast, C. elegans, drosophila, zebrafish, pufferfish, chicken, mouse, rat); construction and analysis of transgenic organisms; novel approaches for functional analysis of transcribed sequences; construction of full length and 5' specific cDNA libraries; database construction, management and use in expression and functional analysis; and analysis of protein structure and function.

Functional determinations promise to be more challenging than transcript identification and even genomic sequencing. Methods for and results of mutational analysis in model organisms, including mouse, rat, zebrafish, drosophila, C.elegans and yeast, are informative and are now often being scaled up for genome wide applications. Evolutionary approaches, including cross species sequence comparisons and studies on features of genome organization, evolution and conserved synteny are critical. Non-coding sequences, in DNA and RNA, are fraught with possibilities for regulation of expression and structural features.

The aim of this workshop is to bring together investigators interested in these, and related topics, for informal presentations and discussions. The format is designed to facilitate the exchange of ideas across disciplines, and to aid in the forming of new interactions and new collaborations.

Organizers

- Katheleen Gardiner, Eleanor Roosevelt Institute, Denver, CO USA
- Bernard Korn, German Cancer Research Center, Heidelberg, Germany
- Richard Mural, Celera Genomics, Rockville, MD, USA
- Sherman Weissman, Yale University, New Haven, CT, USA
- Thomas Werner, Institute for Saeugertiergenetik, Oberschleissheim, Germany

Highlights from the 1999 Workshop

(See "Report of the 9th International Workshop on the Identification of Transcribed Sequences, 1999", Cytogenetics and Cell Genetics, Volume 88, Pages 1-10, 2000)

For Questions or Additional Information, contact:

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

Bernhard Korn, b.korn@dkfz-heidelberg.de
Tel. 49-6221-42-4700

Peggy Woodson, Peggy@eri.uchsc.edu
Tel: 303-336-5612

SPEAKERS
(Confirmed)

Catherine André
UPR41, CNRS
Rennes, France

David Baillie
Simon Fraser University
British Columbia, Canada

Thorsten Bangsow
Technische Universitat
Darmstadt, Germany

Douglas E. Bassett, Jr.
Rosetta Inpharmatics, Inc.
Kirland, Washington USA

Michael Bittner
National Human Genome Research Inst
Bethesda, Maryland USA

Juergen Bode
GBF, National Inst for Biotechnological Research
Braunschweig, Germany

Alwis Brazma
Wellcome Trust Genome Campus
Cambridge, UK

Anthony J. Brookes
Medical Genetics Unit
Uppsala, Sweden

Jüergen Brosius
University of Munster
Munster, Germany

Marcia L. Budarf
The Children's Hospital of Pennsylvania
Phildelphia, Pennsylvania USA

Jean Marie Burstedde

Daniel C. Christophe
Institute de Biologie et de Médecine Moléculaires (IBMM)
Gosselies, Belgium

Daniel Morse
University of Utah
Salt Lake City, Utah USA

Takahiro Nagase
Kazusa DNA Research Institute
Kisarazu, Japan

Christine Orengo
University College London
London, UK

Xing Hua Pan
Yale University School of Medicine
New Haven, Connecticut USA

Stuart Peltz
Robert Wood Johnson Medical School
Piscataway, New Jersey USA

Graziano Pesole
University of Milan
Milan, Italy

John H. Postlethwait
University of Oregon
Eugene, Oregon USA

Thomas Preiss
EMBL
Heidelberg, Germany

Martin Reese
ValiGen
Paris-La-Defense, France

Martin Ringwald
The Jackson Laboratory
Bar Harbor, Maine USA

Yikang Rong
University of Utah
Salt Lake City, Utah USA

Ulrich Schäfer
Max-Planck-Institut für biophysikal. Chemie
Göttingen, Germany

Sally H. Cross
MRC Human Genetics Unit
Edinburgh, UK

Marc Damelin
Dana Farber Cancer Institute
Boston, Massachusetts USA

Olof Emanuelsson
Stockholm University
Stockholm, Sweden

Sylvia E.J. Fisher
Hubrecht Laboratory
Utrecht, The Netherlands

Michel Fontes
INSERM U491
Marseille, France

Katheleen Gardiner
Eleanor Roosevelt Institute
Denver, Colorado USA

Elizabeth Gavis
Princeton University
Princeton, New Jersey USA

Mark Gerstein
Yale University
New Haven, Connecticut USA

Yoram Groner
Weizmann Institute
Israel

Blair Hedges
Pennsylvania State University
University Park, Pennsylvania USA

Bernhard Herrmann

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

Kazuho Ikeo
National Institute of Genetics
Mishima, Japan

Michel Simonneau
INSERM
Paris, France

Holger Sültmann
Deutsches Krebsforschungszentrum
Heidelberg, Germany

Eugene Sverdlov
Russian Academy of Science
Moscow, Russia

Hans-Juergen Thiesen
University of Rostock
Rostock, Germany

Marc Vidal

Jean Weissenbach
Genoscope
Evry Cedex, France

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

Ruth Wellenreuther
German Cancer Research Center
Heidelberg, Germany

Thomas Werner
GSF-National Research Centre for Environment and Health
Neuherberg, Germany

Stefan Wiemann
German Cancer Research Center
Heidelberg, Germany

Edgar Wingender
GBF
Braunschweig, Germany

S. Yokoyama
Syracuse University
Syracuse, New York, USA

Bernhard Korn
German Cancer Research Center
Heidelberg, Germany

Anuj Kumar
Yale University
New Haven, Connecticut USA

I.R. Lemischka
Princeton University
Princeton, New Jersey USA

Shuo Lin
Medical College of Georgia
Augusta, Georgia USA

Speakers * Abstracts * Home

**Beyond the Identification of Transcribed Sequences:
Functional and Expression Analysis**

**10th Annual Workshop
October 28 through October 31, 2000**

Speakers * Abstracts * Home

Submitted Abstracts

Catherine Andre: A 1500 Markers Radiation Hybrid Map of the Dog Genome

David L. Baillie: Comparative Genomic Analysis as a Tool in Revealing Gene Features

Giorgio Bernardi: Isochores and Chromosomal Bands

Douglas E. Bassett, Jr.: The Rosetta Resolver™ System: A Comprehensive Storage, Visualization, and Analysis Tool for High-Volume Gene Expression Data

Michael Bittner: Estimating the Reliability of Inferences Based on cDNA Microarray Ratio Data

Juergen Bode: Definition of Autonomously Regulated Chromatin Domains and Utilization of their Active Elements

Alvis Brazma: From Gene Expression Data to Gene Regulation

Anthony J Brookes: Assessing the Use of Genic SNP's for Complex Disease Analysis and Discovery of a Role for Apoptosis in Alzheimer's Disease

Jürgen Brosius: Retroposition - A Persistent and Pervasive Force in Genome Evolution

Jean-Marie Buerstedde: Identification and Characterization of Genes Involved in B-cell Specific Homologous Recombination
and
The DT40 Cell Line as a Genetic Model to Study Vertebrate Gene Function

Daniel C. Christophe: Cloning of DNA Sequences Encoding Proteins Targeted to the Nucleus in Transiently Transfected Cells

C. Christophe-Hobertus: Cloning and Characterization of the cDNA Encoding Brain Cell Membrane Protein 1 (BCMP1) Identifies a Novel Subclass of Four Transmembrane Proteins

Sally Cross: Screening for ENU-Induced Dominant Eye and Vision Mutants

Marc Damelin: Mapping Protein-Protein Interactions in Living Cells: Insights into the Mechanism of Nucleocytoplasmic Transport

Lars Ebert: High Throughput Subcloning of Complete Protein Coding Open Reading Frames (ORF)

Olof Emanuelsson: Predicting Subcellular Localization of Proteins Based on Their Amino Acid Sequence

Sylvia E.J. Fischer: Transposon Silencing, RNA Interference and Cosuppression in *Caenorhabditis elegans*

Thomas Floss: Functional Analysis of Mammalian Genes by a Large Scale Gene Trap Approach in Mouse Embryonic Stem Cells

Michel Fontes: Construction of Animal Models of Human Disorders by Injection of Large Genomic Fragment: From Function to Physiopathology and Therapy

Katheleen Gardiner: Human Chromosome 21 Gene Annotation I: Evolutionary Conservation and Organizational Features

Elizabeth R. Gavis: Dependence of Nanos mRNA Localization and Translational Control on Structure and Organization of 3'UTR Sequences

Mark Gerstein: Analyis of Genomes and Transcriptomes in terms of the Occurrence of Protein Parts and Features

Yoram Groner: Upstream and Downstream of the Human Leukemia Associated Transcription Factor RUNX1/AML1.

S. Blair Hedges: Genomic Clocks and Animal Evolution

Oliver Heil: Towards Optimized Arrays: Identification of Specific, Non-Conserved, Single Copy Sequence Fragments of all Genes and EST Clusters in Human

Anna Henger: The European Renal Biopsy cDNA Bank (ERCB): A Muticenter Study for Disease Specific mRNA Expression Analysis in Human Kidnet Biopsies

Bernhard Herrmann: Large Scale Screen for Genes Controlling Mammalian Embryogenesis

John Hesketh: The c-fos and c-myc 3'Untranslated Regions are Multi-Functional Regulatory Regions Influencing mRNA Localisation and Stability

Anja Kolb: Tissue-Specific Expression Analysis by RNA in situ Hybridisation of Mouse Genes in the Region Syntenic to Human Xq28

Bernhard Korn: Transcriptional Profiling for the Classification of Novel Natural Product Drugs

Anuj Kumar: Functional Genomics by Transposon Tagging in Yeast

Shuo Lin: Negative and Positive Regulation of Erythropoiesis

Harald von Melchner: Gene Trapping Identifies Transiently Induced Survival Genes During Programmed Cell Death

Daniel P. Morse: Adenosine Deaminases that Act on RNA Edit Non-Coding Regions of mRNAs in Both

Human and *C. elegans*

Takahiro Nagase: Analysis of Human cDNAs Encoding Large Proteins: Towards and Beyond the Identification of Protein-Coding Transcribed Sequences

Masumi Nakazawa: Planarian Gene Expression Profiles using cDNA Chip Technology

Christine Orengo: Insights into Protein Evolution and Function from the CATH Structural Database

Xinghua Pan and Sherman M. Weissman: A Novel and High Throughput Approach for SNP Screening, Scoring and Association Analysis Using Genomic or cDNA Pools

Graziano Pesole: Analysis of Oligonucleotide AUG Start Codon Context in Eukariotic mRNAs

John Postlethwait: Mapping of Zebrafish Expressed Sequences and the Evolution of the Vertebrate Genome

Thomas Preiss: Analysis of mRNA Recruitment to Ribosomes in Eukaryotic Cells

Martin G. Reese: Computational Gene Annotation of the Genome of *Drosophila melanogaster*

Martin Ringwald: GXD: Integrated Access to Gene Expression Data from the Laboratory Mouse

Yikang Rong: Gene Targeting By Homologous Recombination in *Drosophila*

Ulrich Schäfer: Gene Functions on the X-Chromosome of *Drosophila melanogaster*

Thomas Schlitt and Alvis Brazma: Reverse Engineering of Gene Regulatory Networks: a Finite State Linear Model

Mathias Schmid: KIAA0537, A New Tumor Suppressor Gene Candidate, Is Inactivated in Human Brain Tumors

Michel Simonneau: Identification of Novel Genes Involved in Physiological Networks Using Differential Display RT-PCR

Dobromir Slavov: Human Chromosome 21 Gene Annotation II: Refining Gene Structures and Expression Analysis

Holger Sültmann: Gene Expression Profiling in Renal Cell Carcinoma

Eugene D. Sverdlov: Towards Understanding Functional Significance of a Great Many of Solitary LTRs in the Human Genome

Hans-Juergen Thiesen: Differential Proteome Analysis of Germ-Line Maturaturation-Dependent Proteins of *C. elegans*

Marc Vidal: The *C. elegans* Protein Interaction Mapping Project

Jaak Vilo, Alvis Brazma: Expression Profiler: An Integrated Tool for Gene Expression and Sequence Analysis

Jean Weissenbach: Searching for the Protein Coding Genes in the Human Genome Sequence

Ruth Wellenreuther: Generation of Full Length cDNA Libraries in the Course of the German Human Genome Project

Thomas Werner: In Silico Prediction of Matrix Attachment Regions in Large Genomic Sequences

Stefan Wiemann: Sequencing and Analysis of Full Length cDNAs in the German cDNA Network

Shozo Yokoyama: Molecular Genetics and Evolution of Ultraviolet Vision in Vertebrates

---

Speakers * Abstracts * Home

A 1500 Markers Radiation Hybrid Map of the Dog Genome

Catherine Andre
CNRS UPR41 "Recombinaisons Génétiques"
Faculté de médecine
2, Avenue du Pr. Leon Bernard
35043 Rennes cedex FRANCE
telephone: 33 (0) 2 99 33 62 47
fax: 33 (0) 2 99 33 62 00
email: candre@univ-rennes1.fr
prestype: Platform
presenter: Catherine Andre

Catherine André, Angélique Chéron, Christophe Hitte, Zhihua Jiang, Sophie Jouquand, Catherine Priat, Corinne Renier, Françoise Vignaux & Francis Galibert
UPR41 CNRS, Recombinaisons Génétiques, Faculté de Médecine, 2 avenue du Pr. Léon Bernard, 35043 Rennes cedex, France.
Tel : 33 (0) 2 99 33 62 47 Fax : 33 (0) 2 99 33 62 00
E.mail : candre@univ-rennes1.fr.

Dog as a species stands out as a powerful genetic model, as the 300 well-established breeds harbour remarkable interbreed heterogeneity contrasting with strong intrabreed uniformity. This particular feature makes dog an invaluable resource for dissecting the molecular basis of genetic traits and diseases.

To this end we have constructed a whole genome dog /hamster radiation hybrid panel: RHDF5000 (Vignaux et al., 1999). In addition, 300 dog gene markers and 1000 microsatellite markers have been identified (Priat et al., 1999; Jouquand et al., 2000). With these tools we constructed a radiation hybrid map of the dog (i) by positioning 400 markers (Priat et al.,1998) and (ii) by integrating through a collaborative efforts (Mellersh et al., 2000), the Radiation hybrid map containing 600 markers into the Linkage map. At the present time, each of the 39 canine chromosomes (CFA1 to 38 and X and Y) has been assigned to the corresponding radiation hybrid group (Breen et al., in preparation) and a 1500-markers radiation hybrid map has recently been achieved (Jouquand et al., in preparation). This map, harbouring 1200 microsatellites and 300 genes covers 95% of the dog genome and allowed us to characterize conserved segments between the canine and human chromosomes. Using this map we plan to identify diseases or phenotypic /behavioural traits by linking the trait with polymorphic markers using the candidate gene approach.

Data available on: http://www-recomgen.univ-rennes1.fr/doggy.html.

Speakers * Abstracts * Home

Comparative Genomic Analysis as a Tool in Revealing Gene Features

David L. Baillie
Department of Mol. Biology and Biochem.
Simon Fraser University
8888 University Drive
Burnaby, B.C.
Canada V5A 1S6
telephone: 1-604-268-6590
fax: 1-604-291-5583
email: baillie@sfu.ca
prestype: Platform
presenter: David Baillie

David L. Baillie[1], Nigel O'Neil[2], Greg Vatcher[3], Steve Jones[3] and Marco Marra[3]
[1]Simon Fraser University, Canada
[2]Sanger Centre, UK
[3]B.C. Genome Research Centre, Canada

The *Caenorhabditis elegans* genome was completed almost two years ago. The sequence is 98 million bases and contains the about 19,000 protein encoding elements. A number of approaches are being taken to reveal the function of each of these coding elements. An international consortium has undertaken the selection of psoralen induced knockouts of all of the coding elements. This is being done in a community driven manner and requests for specific knockouts can be made at http://elegans.bcgsc.bc.ca/knockout.shtml. In addition chips containing arrays of all the putative genes are available for analysis of gene expression at http://cmgm.stanford.edu/~kimlab/wmdirectorybig.html. A large scale EST project with expression information for each cDNA is being done by Yugi Kohara and his data is accessable at http://www.ddbj.nig.ac.jp/htmls/c-elegans/html/CE_INDEX.html. We have been analyzing an area defined by the deficiency sDf125 (about 400,000 BPs) to determine the number of essential genes in this region. We have produced EMS induce lethal alleles of 15 of the 90 genes in the region. Over half of the genes have only one alleles at this time, thus it appears that the number of genes that have essential roles is greater than 30% in this region. We have used a library of overlapping cosmid transgenic containg animals to determine the genes associated with many of our alleles. We have used the sequence of the closely related nematode *C. briggsae* aid in the determination of the structure of these genes. Examples will be given of the types of information that can be derived from this approach.

Speakers * Abstracts * Home

Isochores and Chromosomal Bands

Giorgio Bernardi
Stazione Zoologica Anton Dohrn
Villa Comunale
80121 Napoli Italy
telephone: 0039 081 5833 300/402
fax: 0039 081 2455807
email: bernardi@alpha.szn.it
prestype: Platform
presenter: Giorgio Bernardi

S. Saccone[1,2], A. Pavlicek[3], C. Federico[1] and G. Bernardi[4]

[1]Dipartimento di Biologia Animale, University of Catania, via Androne 81, 95124 Catania, Italy;
[2]Dipartimento di Protezione e Valorizzazione Agroalimentare, University of Bologna, via F.lli Rosselli 107, 42100 Reggio Emilia, Italy;
[3]Institute of Molecular Genetics, Czech Academy of Sciences - Flemingovo nam. 2 - 166 37 Prague 6, Czech Republic
[4]Laboratorio di Evoluzione Molecolare, Stazione Zoologica, Villa Comunale, 80121 Napoli, Italy; Tel: (+39) 081 5833300; Fax: (+39) 081 2455807; E-mail: bernardi@alpha.szn.it

The human genome is formed by isochores belonging to five families, L1, L2, H1, H2, and H3, that are characterized by increasing GC levels and gene concentrations. In situ hybridization of DNA from different isochore families, provides, therefore, information not only on the correlation between isochores and chromosomal bands, but also on the distribution of genes in chromosomes. Three subsets of R(everse) bands, H3+, H3* and H3, that contain large, moderate, and no detectable amounts, respectively, of the gene- richest H2 and H3 isochores, and replicate very early and early, respectively, in S phase of the cell cycle, were identified.

Recent investigations on the GC levels, replication timings and DNA compaction of G (iemsa) bands, showed that G bands comprise two different subsets of bands, one of which is predominantly composed of L1 isochores, replicates at the end of the S phase, has a higher DNA compaction relative to H3+ bands and corresponds to the darkest G bands of Francke (Cytogenet. Cell Genet. 6: 206-219, 1994). In contrast, the other subset is composed of L2 and H1 isochores, has less extreme properties in replication and composition and corresponds to the less dark G bands of Francke.

Finally, the correlation between base composition of DNA and chromosomal bands as investigated using the DNA sequences of chromosomes 21 and 22 will be reported.

Speakers * Abstracts * Home

The Rosetta Resolver™ System: A Comprehensive Storage, Visualization, and Analysis Tool for High-Volume Gene Expression Data

Douglas E. Bassett, Jr.
Rosetta Inpharmatics, Inc.
12040 115th Avenue NE
Kirkland, WA 98034
telephone: 1-425-823-7324
fax: 1-425-821-5254
email: bassett@ncbi.nlm.nih.gov
prestype: Platform
presenter: Douglas E. Bassett, Jr.

The Rosetta Resolver™ system is the leading turnkey solution for storing, retrieving, and analyzing large quantities of gene expression data generated using cDNA microarrays, oligonucleotide arrays, and other technologies. Applied to large gene expression data sets, the Resolver serves as a powerful discovery tool for pathway interrogation, gene functional assignment, and compound analysis. Multi-experiment analyses are possible through correlation plots, cluster trees, and the Rosetta Array Search Tool (ROAST™), which enables the user to perform BLAST-like similarity searches across entire datasets of tens of thousands of gene expression profiles. Details of the Rosetta Resolver software and hardware will be discussed, and examples of the system's features will be shown.

Speakers * Abstracts * Home

Estimating the Reliability of Inferences Based on cDNA Microarray Ratio Data

Michael Bittner
NIH/NHGRI
Building 49, Room 4A52
9000 Rockville Pike
Bethesda, MD 20892 USA
telephone: 1-301-496-7980
fax: 1-301-402-3241
email: mbittner@nhgri.nih.gov
prestype: Platform
presenter: Michael Bittner

Y. Chen[1], Y. Jiang[1], E. Dougherty[2], S. Kim[2], Z. Yakhini[3], A. Ben-Dor[3], N. Sampas[3], M. Radmacher[4], R. Simon[4], M. D. Gubitoso[5], M. Brun[6], J. Barrera[6], C. Gooden, A. Glatfelter, P. Meltzer[1], J. Trent[1], and M. Bittner[1]

[1]Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892
[2]Department of Electrical Engineering, Texas A & M University, College Station, Texas 77843
[3]Chemical and Biological Systems Department, Agilent Laboratories
[4]National Cancer Institute, DCTDC, NIH, Bethesda, Maryland 20852
[5]Department of Computer Science, University of São Paulo
[6] Institute of Mathematics and Statistics, University of São Paulo

The ratio data from cDNA microarray experiments is used in a variety of analyses that compare the patterns of expression across a series of samples. These analytical tools are used to group either the samples, on the basis of overall similarities in expression pattern, or the genes themselves, on the basis of their individual similarities in expression pattern. Both the strength and breadth of the inferences that can be made in these analyses ultimately depend on the accuracy and reproducibility of the initial measurements. In this presentation we will show studies of the strength of inferences made on the basis of ratio data from a variety of perspectives. Using repeated measurement of the same samples and external validation it has been possible to directly examine accuracy and reproducibility. We have used these estimates to determine the magnitude of effect they would have in typical analyses. Using perturbation and randomization of actual data, it has been possible to demonstrate that observed consistencies of gene expression associated with sample type are extremely unlikely to arise from error in the measurements or from the inherent biological order existing in all individual samples. Simulation studies of the effects of sample variance on gene by gene clustering provides a way to find forms of class discovery tools most likely to correctly identify genes with similar expression patterns at the levels of variance typical of microarray results.

Speakers * Abstracts * Home

Definition of Autonomously Regulated Chromatin Domains and Utilization of Their Active Elements

Juergen Bode
GBF, German Research Center for
Biotechnology,
Mascheroder Weg 1
D-38124 Braunschweig
telephone: (0531)6181251
fax: (0531)6181262
email: jbogbf.de
prestype: Platform
presenter: J. Bode

J. Bode[1], A. Baer[1], E. Ernst[1], S. Goetze[1], A. Knopp[1], K. Nehlsen[1], C. Benham[2], H.-J. Lipps[3] and A. Baiker[4]
[1]GBF - German Research Center for Biotechnology/Epigenetic Regulation, D-38124 Braunschweig; jbo@gbf.de
[2]The Mount Sinai Med. Center, New York, NY 10029 , USA; benham@msvax.mssm.edu:
[3]Institut für Zellbiologie, Universität Witten/Herdecke, D-58448 Witten, lipps@uni-wh.de
[4]Stanford University/Department of Pediatrics, Stanford, CA 94305-5208; abaiker@stanford.edu

The human interferon gene cluster on the short arm of chromosome 9 comprises 26 genes the functional members of which are separated by highly efficient scaffold/matrix attached regions (S/MARs). We have refined in vitro and biomathematical methods to predict the activity of these elements which are characterized by their pronounced propensity to separate strands under superhelical tension. Other features of individual chromatin domains which become apparent in stress-induced duplex-destabilization (SIDD) profiles are the localization of regulatory elements (promoters and DNAse I hypersensitive sites) and of chromosomal breakpoints which explain the genomic instability of this genomic locus. LMPCR techniques have been used to verify the predicted state of DNA at base resolution under superhelical tension and in the living cell.

We have developed a number of in vivo assays to define the spectrum of biological activities for S/MAR elements, among these the augmentation of transcription initiation rates which is distinct from enhancement. S/MARs also interfere with the methylation of transgenes and can be applied to stabilize the long-term expression, particularly of retroviral vectors. Utilizing recombinase-mediated cassette exchange techniques (RMCE) we have initiated a series of experiments to unravel the potential of S/MARs to function as an insulator, a cis-acting or a targeting element. PCR techniques have been applied to identify their role as hotspots of recombination and their participation in deletion and translocation events.

Conventional integrating vectors for eukaryotic cells suffer from a number of limitations which include the possibilities of insertional mutagenesis or silencing of the transgene. So far, only viral vectors derived from SV40-, BPV- or EBV replicate episomally in some eukaryotic cells. The replication origins of these vectors require the support by virally encoded trans-acting factors which can lead to cellular transformation. We present an entirely new vector type in which the function of a viral origin depends of an active S/MAR element. This vector replicates episomally in a number of cell lines and primary cells and is maintained at a constant low copy number in the absence of selection pressure. The functional relevance of matrix association is demonstrated by in situ hybridization and nuclear matrix fractionation procedures.

Recent reviews:

- J. Bode, J. Bartsch, T. Boulikas, M. Iber, C. Mielke, D. Schübeler, J. Seibler, and C. Benham (1998). Transcription-Promoting Genomic Sites in Mammalia: Their Elucidation and Architectural Principles Gene Ther. Mol. Biol. 1, 551-880.
- 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.
- 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, Sept/Oct issue, in press

---

Speakers * Abstracts * Home

From Gene Expression Data to Gene Regulation

Alvis Brazma
EMBL Outstation -- Hinxton
European Bioinformatics Institute
Wellcome Trust Genome Campus
Hinxton, Cambridge CB10 1SD UK
telephone: +44-(0)1223 494658
fax: +44-(0)1223 494468
email: brazma@ebi.ac.uk
prestype: Platform
presenter: Alvis Brazma

Alvis Brazma, Jaak Vilo and Thomas Schlitt

DNA arrays are enabling researchers to make snapshots of gene expression levels for tens of thousands of genes at a given moment. This technology is already producing floods of valuable data, and the analysis and handling of these data is one of the most important and interesting problems of computational biology. We will discuss various informatics approaches to the analysis of DNA array data as well as our initiative to establish a public repository for these data. In particular we will discuss how these data can be used for finding putative regulatory signals in the genome and our approaches to the problem of reverse engineering of gene regulatory networks from gene expression data.

Speakers * Abstracts * Home

Assessing the Use of Genetic SNPs for Complex Disease Analysis and Discovery of a Role for Apoptosis in Alzheimer's Disease

Anthony J Brookes
Vice-Prefect, Clinical Genomics Coordinator
Center for Genomics Research
Karolinska Institute
Theorells väg 3
S-171 77 Stockholm Sweden
telephone: +46-(0)8-7286630
fax: +46-(0)8-331547
email: Anthony.Brookes@cgr.ki.se
prestype: Platform
presenter: Anthony J Brookes

A.J. BROOKES
Center for Genomics Research, Karolinska Institute, 171 77 Stockholm, Sweden.

Everyone hopes or assumes that SNP-based association analysis will help us to understand the genetic basis of complex disease. However, many problems could limit this possibility, e.g., phenotypic and genetic/allelic heterogeneity, population sub-structure, variance in Linkage Disequilibrium (LD), method inadequacies. Despite years of intense study in the field the number of success stories are few - yet it is difficult to draw strong conclusions due to intangible issues such as publication biases, variable experimental design quality, and etiologic differences across population differences.

We have been conducting association analysis in a large-scale manner using an optimal study designed focused strongly upon intragenic SNPs (primarily non-synonymous variants) - assuming that these will comprise and/or be in high LD with pathogenic variants. The study also i) followed a 'pathway-based hypothesis testing' strategy covering high candidature systems to look for clustering of weak-medium signals within pathways, ii) analyzed population-based sets of twins and phenotypic extreme clinical materials from Swedish and Scottish origins, since these will probably have more homogeneous and readily apparent genetic etiologies, and iii) targeted diseases that are well studied biochemically and known to possess high heritability - principally Alzheimer's Disease (AD) and obesity/diabetes.

To enable the above, we i) constructed the Human Genic Bi-Allelic SEquences (HGBASE) database to gather and extensively curate gene-linked SNPs from all public sources, ii) developed Dynamic Allele Specific Hybridization (DASH) for virtually error-free SNP genotyping (>99.9% accuracy) that is robust (less than 1% of assays fail), and cheap (<25c per genotype), and iii) screened 300 AD candidate genes for SNPs, yielding over 500 potential SNPs, many that change amino-acid residues.

Testing over 50 SNPs (60+ candidate AD genes) in the above experimental setting gives a good indication of the many problems and limitations of this approach. From all of these efforts, one very significant new AD association was uncovered (P=0.002), involving a promoter variation in a key apoptosis regulator gene and an interaction with APOE. This result provides strong genetic evidence in support of a role for apoptosis control differences in the causation of AD in many individuals, and may help explain the mechanism by which the E4 allele of APOE mediates an increased AD risk.

Speakers * Abstracts * Home

Retroposition - A Persistent and Pervasive Force in Genome Evolution

Jürgen Brosius
Institute of Experimental Pathology/Molecular Neurobiology
University of Münster
Von-Esmarch-Str. 56
D-48149 Münster, Germany
telephone: 0251 835 8511
fax: 0251 835 8512
email: RNA.world@uni-muenster.de
prestype: Platform
presenter: Jürgen Brosius

During transition from the RNP world (when RNA carried genetic information and RNA as well as protein had structural and catalytic assignments in the cell) to modern cells, the conversion of genetic information into DNA was probably achieved by retroposition i.e. reverse transcription of RNA into DNA and integration into the nascent DNA genome(1). One may assume that after all genomic RNA had been transformed into DNA, this process should have ceded. On the contrary, in many lineages this process is still rampant: all types of RNAs(2) can be reverse transcribed and their cDNA copies reintegrated into genomes as retronuons (a nuon is any stretch of a definable nucleic acid sequence(3). About 40% of the human genome consist of discernible retronuons. If one extrapolates this figure to "decayed", non- discernible retronuons, it is conceivable that a very large proportion of the 95-97% non- coding segments (including intergenic regions and introns) is of retronuon origin. In most cases this process leads to "junk DNA". However, there are many examples where mRNA-derived retronuons gave rise to active genes, often with different expression patterns, when compared to their respective founder genes(3,4). Retronuons derived from small non-messenger RNAs (snmRNAs) can give rise to novel snmRNA genes (such as neuronal dendritic BC1 and BC200 RNAs) and, quite often, are exapted (recruited) as regulatory elements that may alter expression or processing of targeted genes(3,5) (for compilations see http://www-ifi.uni-muenster.de/exapted-retrogenes/tables.html). Thus, retronuons are a major driving force of evolution. For example, comparison of the human genome with that of its closest relative, the chimpanzee, will reveal that neither species will contain many additional genes and if, they will not differ much from the genes they originated from via duplication. Instead, we will observe exaptation of novel exons (often involving alternative splicing) from previously nonaptive intronic and flanking sequences (naptonuons) originally generated by retroposition and different patterns of expression of otherwise shared genes with respect to development and cell-type specificity, often being influenced by de novo insertions of retronuons.

1. Brosius, J. (1999) Transmutation of tRNA over time. Nat. Genet. 22, 8-9.
2. Brosius, J. (1999) RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene 238, 115-134.
3. Brosius, J., Gould, S.J. (1992) On Genomenclature: A comprehensive (and respectful) taxonomy for pseudogenes and other 'junk DNA'. Proc. Natl. Acad. Sci. U.S.A. 89, 10706- 10710.
4. Brosius, J. (1991) Retroposons - seeds of evolution. Science 251, 753.
5. Brosius, J. (1999) Genomes were forged by massive bombardments with retroelements and retrosequences. Genetica 107, 209-238.

Speakers * Abstracts * Home

Identification and Characterization of Genes Involved in B-cell Specific Homologous Recombination

Jean-Marie Buerstedde
Department of Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52
D-20251 Hamburg, Germany
telephone: 0049 40 48051 290
fax: 0049 40 48051 296
email: buersted@genetics.hpi.uni-hamburg.de
prestype: Platform
presenter: Jean-Marie Buerstedde

Hiroshi Arakawa, Dima Lodygin, and Jean-Marie Buerstedde

Homologous recombination modifies the genome during meiotic and somatic cell development by enabling chomosomal crossovers and genetic exchanges between homologous sequences. It also plays a critical role for the repair of spontaneous and induced DNA damage. Gene conversion, as originally defined in yeast, is a special form of homologous recombination in which a donor sequence is duplicated and a target sequence is lost.

Homologous recombination plays a special role for the development of the immunoglobulin (Ig) gene repertoire, as B-cells in the chicken, rabbit and cow diversify their Ig genes at high frequency by gene conversion using pseudo V genes as donors. Ig gene conversion coexists with and resembles Ig somatic hypermutation in many vertebrate species.

The mechanism and enzymology of Ig gene conversion is poorly understood. To identify factors involved in this process we are establishing a comprehensive EST database from bursal B-cells which possess high Ig gene conversion activity. Blast searches of the already sequenced bursal ESTs against the public databases identified a number of promising candidates for DNA recombination factors. Among these are the genes encoding the DNA mismatch recognition factors MSH2, MSH3, MSH4 and MSH6, which have recently been implicated in Ig somatic hypermutation and switch recombination. In addition, we discovered a new type IV DNA polymerase (pol lambda) sharing homology to the patch repair DNA polymerase beta and the error prone terminal desoxytransferase. The pol lambda polymerase could either be involved in somatic hypermutation or in gene conversion.

Another interesting gene expressed in bursal B cells encodes a second structural homologue of the yeast RAD52 gene, which defines the double-strand break repair pathway by homologous recombination. We previously showed that the disruption of the first RAD52 gene in the chicken B cell line DT40 produces only a mild DNA repair defect, a finding which surprised many yeast geneticists. We now speculate that the second RAD52 homologue can compensate for the loss of the first RAD52 gene in vertebrate cells.

The exact functions of these genes are now determined by disruption in the DT40 cell line. DT40 is particularly suited, since it continues Ig gene conversion during cell culture and integrates transfected gene constructs at high ratios into the endogeneous loci. As we can recycle the our drug resistant marker by Cre-mediated excision, we shall also analyse the phenotype of multiple gene disruptions. We are for example planning to disrupt multiple MSH genes, DNA polymerases and RAD52 homologues within the same cell line to check for synergistic effects.

It is hoped that the isolation and characterization of B-cell specific recombination genes will eventually provide insight into the mechanism of immunoglobulin gene conversion. This research may also explain why targeted integration occurs at exceptionally high rates after transfection of DT40.

Speakers * Abstracts * Home

The DT40 Cell Line as a Genetic Model to Study Vertebrate Gene Function

Jean-Marie Buerstedde
Department of Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52
D-20251 Hamburg, Germany
telephone: 0049 40 48051 290
fax: 0049 40 48051 296
email: buersted@genetics.hpi.uni-hamburg.de
prestype: Platform
presenter: Jean-Marie Buerstedde

Jean-Marie Buerstedde, Igor Abdrahmanov and Bernhard Korn

Due to the high ratios of targeted integration the chicken B cell line DT40 cell line is a popular geneticsystem to study the functions of genes by disruption. However, relatively the sequences of few chicken genes are known and the isolation of the chicken cDNA by cross-hybridization or by reverse PCR is cumbersome.

To improve this situation we have started to build a comprehensive bursal B-cell EST database which currently holds over 7000 ESTs. Sequences corresponding to interesting candidate genes can be easily identified by online BLAST or keyword searches. Since the database reflects the gene expression profile of bursal lymphocytes, it provides valuable hints as to which genes might be involved in B-cell specific processes related to immunoglobulin repertoire formation, signal transduction, transcription and apoptosis.

This large collection of chicken ESTs will also be useful for gene expression studies and comparative gene mapping within the chicken genome project. In collaboration with the Department of Olli Lassila at Turku University in Finnland we are building filter and glas slide gene arrays which could be used to quantitate gene expression profiles in wild-type and mutant DT40 cells.

This bursal EST sequencing project will continue for the next three years. To avoid redundant sequencing of highly expressed genes we have successfully normalized the bursal cDNA library by hybridization with already sequenced ESTs. Since our current bursal library has only a low percentage of full-length cDNAs, we have also produced a second bursal cDNA library containing a high percentageof full length cDNA inserts with the help of the Riken Center in Japan. This library is currently orderedin the German Genome Resource Center. Using the normalized library and the library enriched for full-length clones we are planning to add new sequences at a rate of 1000 sequences per month for the next three years. All ESTs are freely available to all researchers.

We are producing other generic resources for genetic studies in DT40. In the past, six drug resistance markers were available, but for analysis of complex pathways a higher number of mutations are needed.

We therefore constructed new markers flanked by loxP sites which are recognized by the CRE-recombinase. These markers allow the generation of unlimited number of mutations in DT40 through the reuse of the markers after CRE-mediated excision.

Details of the bursal EST sequencing project and other resources can be found via the DT40 web site maintained by our Department.

Speakers * Abstracts * Home

Cloning of DNA Sequences Encoding Proteins Targeted to the Nucleus in Transiently Transfected Cells

Daniel C. Christophe
Institut de Biologie et de Médecine Moléculaires (IBMM)
IRIBHN-Faculté de Médecine, Université Libre de Bruxelles
Rue Profs. Jeener et Brachet, 12
B-6041 GOSSELIES, BELGIUM
telephone: +32-2-650 9828
fax: +32-2-650 9820
email: dchristo@ulb.ac.be
prestype: Platform
presenter: D. Christophe

D. Christophe[1], B. Pichon[1], D. Mercan[2], V. Pouillon[1], C. Christophe-Hobertus[1]
[1]Institut de Biologie et de Médecine Moléculaires (IBMM-IRIBHN), Université Libre de Bruxelles, B-6041 Gosselies and
[2]Service de Chimie, Hôpital Erasme, B-1070 Bruxelles, Belgium.

Determining the subcellular localization of a gene product is instrumental in the elucidation of its function in the cell. We have developped an experimental strategy allowing the direct identification of DNA sequences encoding nuclear proteins by transient transfection of cDNA libraries in cultured cells.

Briefly, our method involves the following steps: (i) the construction of a green fluorescent protein (GFP) fusion proteins library in an expression vector harboring the SV40 origin of replication and the kanamycin resistance gene; (ii) the dilution of the library DNA in a 100-fold excess of unrelated plasmid DNA prior to transfection into COS- 7 cells, which results in the expression of individual library constructs in most of transfected cells; (iii) the permeabilization of the transfected cells by saponin, which results in the release of most of the cytoplasmic GFP fusions proteins, and immediate sorting by flow cytometry of those cells that still contain large amounts of (essentially nuclear) GFP fusions proteins; and (iv) the recovery of the transfected library DNA by extraction of the DNA from the pool of sorted cells, transformation of bacteria and selection of kanamycin-resistant clones. The experimental conditions used in these steps were optimized using a mixture of DNA constructs encoding nuclear and cytoplasmic GFP fusion proteins.

The technique was then used to screen a dog thyroid cDNA library. Fluorescence microscopy observation of cells transfected with the library DNA revealed that fluorescence was essentially nuclear in only 1% of the GFP-expressing cells . Starting from 1,000,000 transfected cells (containing about 100,000 GFP-expressing cells), we isolated by FACSing a pool of 116 cells, and 560 bacterial clones were obtained after transformation with the extracted DNA. Individual transfection of 110 out of these clones and fluorescence microscopy observation showed that 73% of them encoded nuclear GFP-fusion proteins. Partial DNA sequencing of the clones revealed that 45 out of the 63 different sequences isolated corresponded to known cDNAs, more than half of these (25) encoding known nuclear proteins. Thirteen other clones harbored uncharacterized DNA sequences. Further characterization of 4 of the cDNAs led to the clear identification of 2 novel nuclear proteins.

The method described here is easy to apply and may be scaled up, which opens the way to a wide search for novel nuclear proteins.

Speakers * Abstracts * Home

Cloning and Characterization of the cDNA Encoding Brain Cell Membrane Protein 1 (BCMP1) Identifies a Novel Subclass of Four Transmembrane Proteins

Daniel C. Christophe
Institut de Biologie et de Médecine Moléculaires (IBMM)
IRIBHN-Faculté de Médecine, Université Libre de Bruxelles
Rue Profs. Jeener et Brachet, 12
B-6041 GOSSELIES, BELGIUM
telephone: +32-2-650 9828
fax: +32-2-650 9820
email: dchristo@ulb.ac.be
prestype: Poster
presenter: C. Christophe-Hobertus

C. Christophe-Hobertus and D. Christophe
Institut de Biologie et de Médecine Moléculaires (IBMM-IRIBHN), Université Libre de Bruxelles, B-6041 Gosselies, Belgium.

A partial cDNA sequence from dog thyroid presenting a similarity overlap with the mouse cDNA clone MNCb-0941 (GenBank acc. #: AU035837) was fortuitously isolated in the laboratory. Re-screening of a dog thyroid cDNA library resulted in the isolation of a 4kb- long cDNA exhibiting a putative open-reading frame of only 543bp followed by a 3.2kb-long 3' untranslated region containing several ATTTA instability motifs. Bioinformatic analysis of the encoded protein sequence identified the presence of four putative transmembrane domains and of a short intracellular N-terminal domain. Similarity searches in protein and protein domains databases identified partial sequence conservations with peripheral myelin protein 22 (PMP22) and epithelial membrane proteins (EMPs), both of which belonging to the large family of four-transmembrane proteins [1]. Northern-blot analysis of the expression of the corresponding mRNA in adult dog tissues revealed the presence of a huge amount of the 4kb transcript in the brain. Lower amounts were observed in a number of other tissues, including the thyroid. The subcellular localization of the encoded protein was determined by expressing a green fluorescent protein (GFP) fusion protein in transfected COS-7 cells. A patchy fluorescence pattern was observed all over the cell surface and a strong accumulation of fluorescence in the perinuclear region was observed in most cells, which was consistent with a membranous localization of the protein.

Very recently, the complete sequence of the mouse cDNA clone MNCb-0941 was made available (GenBank acc. #: AB041540). Comparison of the encoded protein sequences with the closest members of the four-transmembrane proteins family revealed that our dog cDNA sequence and the mouse MNCb-0941 clone defined the existence of a novel subclass in this family of proteins. In view of the high level of mRNA found in the brain, we termed the protein "brain cell membrane protein 1" (BCMP1).Bioinformatic search for the sequences corresponding to human BCMP1 led to the identification of the cDNA sequence DKFZp564E153 (GenBank acc. #: AL049257) which exhibited a very high degree of sequence conservation over 2.5kb with the 3' untranslated region of dog BCMP1 mRNA. The corresponding locus had been localized on chromosome X in man. A fragment of DNA sequence from human chromosome 8 (clone RP11-31H18 map 8, GenBank acc. #: AC041003) also presented a significant similarity with the coding region of BCMP1 mRNA. Whether this observation indicates the existence of two closely related genes in man or results from inaccurate sequence assignment in the database is currently being tested experimentally.

[1] Jetten A.M. & Suter U. (2000) Prog. Nucleic Acid Res. Mol. Biol. 64, 97-128.

Speakers * Abstracts * Home

Screening for ENU-Induced Dominant Eye and Vision Mutant

Sally Cross
MRC Human Genetics Unit
Western General Hospital
Crewe Road
Edinburgh EH4 2XU
Scotland, UK
telephone: 44 131 332 2471
fax: 44 131 343 2620
email: sally.cross@hgu.mrc.ac.uk
prestype: Platform
presenter: Sally Cross

Sally Cross[1], Caroline Thaung[1,2], Katrine West[1], Karen Arnold[1,2], Lisa McKie[1], Pat Nolan[2], Jo Peters[2], Sohaila Rastan[3], Jackie Hunter[3], Steve Brown[2], Ian Jackson[1]

[1]MRC Human Genetics Unit, Edinburgh
[2]MRC Mammalin Genetics Unit, Harwell
[3]SmithKline Beecham Pharmaceuticals, Harlow

There are about 150 genes in which mouse mutations affect vision or eye development, many of these are associated with other defects and ~25% result in cataracts. In order to identify additional genes important for eye development we have been screening for novel dominant eye and vision mutants in the mouse as part of the major ENU mutagenesis programme being carried out at Harwell. We chose to concentrate on the eye because it is a tractable developmental system and because of the value of creating new models of human eye disease. Using a phenotype-driven approach where no assumption is made about gene function we hope to identify novel genes and pathways. We have screened over 6000 potential mutant mice for vision defects using an optokinetic drum and have physically examined the eyes of about 6000 using both a slit-lamp biomicroscope to detect anterior segment defects and an indirect ophthalmoscope to detect retinal defects. The screen will also detect new mutant alleles of the recessive mutation Pdebrd1. To date we have found about 40 mutant phenotypes, affecting all parts of the eye, of which 15 are inherited, 14 were not inherited and the rest are still undergoing inheritence testing. Including those still undergoing testing we have at least 3 probable mutant alleles of Pax6, and up to 8 alleles of Pdeb. Amongst the others are novel mutations with specific effects on the cornea, the iris or the retina and others that affect multiple systems.

Speakers * Abstracts * Home

Mapping Protein-Protein Interactions in Living Cells: Insights into the Mechanism of Nucleocytoplasmic Transport

Marc Damelin
Department of Biological Chemistry and Molecular Pharmacology
Harvard Medical School and the Dana-Farber Cancer Institute
44 Binney Street
Boston, MA 02115 U.S.A.
telephone: 01-617-632-5104
fax: 01-617-632-5103
email: damelin@fas.harvard.edu
prestype: Platform
presenter: Marc Damelin

Marc Damelin and Pamela A. Silver
Department of Biological Chemistry and Molecular Pharmacology; Harvard Medical School and the Dana-Farber Cancer Institute; Boston, MA 02115

Rapid progress in genome sequencing has made feasible large-scale screening of protein- protein interactions. We have developed an assay based on fluorescence resonance energy transfer (FRET) to map dynamic protein interactions in living cells. DNA encoding either cyan or yellow fluorescent protein (CFP, YFP) is integrated into the genome at the locus of a target gene. The endogenous copy of the gene is thus replaced by the tagged version, and the resulting fusion is expressed off the genome. The expression and functionality of all tagged proteins are confirmed. Living cells co-expressing CFP and YFP fusions are examined under an epifluorescence microscope. Digitized images captured with a charge-coupled device (CCD) camera are analyzed quantitatively to determine significant FRET signals.

Using this approach, we have investigated the mechanism of nucleocytoplasmic transport in the model organism Saccharomyces cerevisiae. In all eukaryotic cells, cargo to be transported between the nucleus and cytoplasm is carried by a transport receptor through the nuclear pore complex (NPC). The yeast NPC is composed of 30 proteins (nups) in multiple copies and has a mass of ~60MDa. We have studied two aspects of the mechanism of NPC translocation with the FRET assay. First, we screened for interactions between two transport receptors and nups, and found that these transport receptors have overlapping pathways through the NPC. The distinct receptor-nup contacts in each pathway may serve as a receptor-specific regulatory mechanism. Second, we screened for nup-nup interactions using a panel of 15 nups, and found 11 interactions. These nup-nup interactions have led us to propose a more refined structural model of the 60MDa NPC.

Speakers * Abstracts * Home

High Throughput Subcloning of Complete Protein Coding Open Reading Frames (ORF)

Bernhard Korn
RZPD REsource Center for Genome Research
Im Neuenheimer Feld 506
69120 Heidelberg Germany
telephone: +49 (0)6221 42 4700
fax: +49 (0)6221 42 4704
email: b.korn@dkfz-heidelberg.de
prestype: Poster
presenter2 = Ebert, Lars

L. Ebert, R. Schatten, A. Poustka and B. Korn

In order to progress from genomic sequence information to mRNA expression analysis to proteomics, it is of increasing interest to have the functional gene products at hand. As only the protein determines the function of a given gene that harbors translation potential, the characterization of these final gene products will lead us to the experimental verification and determination of protein- (and therefore gene-) function. As cloning of native open reading frames of proteins is a labor, time and cost consuming process, we have initiated a project that aims to clone as many complete open reading frames of human genes as possible. We intend to provide the respective clones as part of the RZPD clone collection to all interested groups. In order to have the ORFs available for many different applications (e.g. in situ transcription (and translation), expression in eukaryots and prokaryotes, expression as fusion proteins, …), we decided to clone the complete ORF (excluding the stop codon) into a vector system which does allow base-specific shuttling of the insert into a wide range of other vector types without restriction and ligation, rather making use of in vitro recombination between vectors. In a first attempt, we have concentrated on human ORFs, which are potentially present in EST clones arising from the RZPD and IMAGE clone collection. However, most of these clones are only single pass sequenced, and only partial sequence of each clone is available (ESTs).

To solve this problem we developed a bioinformatic method to select EST clones, which contain the full ORF, although they are not yet completely sequenced.

1. We extract the relevant information - such as the sequence of the ORF region of a gene - from the complete human unigene dataset.
2. We use alignment algorithms to select clones, whose EST sequences cover the start and stop region of the ORF. The quality of the potential full ORF clone is checked via several bioinformatic selection procedures.
3. Primer pairs are designed that allow the amplification of the complete ORF while deleting the stop codon.
4. PCR is performed and the products are checked on agarose gels. PCR product size is verified by comparing the experimentally generated PCR product with the expected ORF size.
5. A secondary PCR is performed to add recombination sequences that allow the cloning of the ORF PCR product by homologous recombination.
6. Recombination clones are verified by: a) PCR of the inserts and b) expression of a GFP fusion protein.

Up to now, we have amplified more than 360 ORFs of known human genes, and aim to increase this number steadily while streamlining the whole process by partial automation. We expect to distribute the first clones from this set by beginning of year 2001 via the distribution service of RZPD.

Speakers * Abstracts * Home

Predicting Subcellular Localization of Proteins Based on their Amino Acid Sequence

Olof Emanuelsson
Stockholm Bioinformatics Center
Stockholm University
telephone: +46 (0)8 16 21 10
fax:
email: olof@sbc.su.se
prestype: Platform
presenter: Olof Emanuelsson

Olof Emanuelsson[1], Henrik Nielsen[2], Gunnar von Heijne[1]
[1]Stockholm Bioinformatics Center, Stockholm University, Stockholm, Sweden
[2]Center for Biological Sequence Analysis, The Technical University of Denmark, Lyngby, Denmark

The subcellular localization of a protein is an important characteristic with implication for protein function. We have developed TargetP, which is a neural network-based tool for large-scale subcellular location prediction of newly identified proteins. TargetP is able to discriminate between proteins destined for the mitochondrion, the chloroplast, the secretory pathway, and "other" localizations with a succes rate of 85% (plant) or 90% (non-plant) on redundancy-reduced test sets. Scanning the newly sequenced A. thaliana chromosomes 2 and 4 and the Ensembl human set with TargetP, we estimate that 10% of all plant proteins are mitochondrial and 14% chloroplastic, and that the abundance of secretory proteins, in both Arabidopsis and Homo, is around 10%. A recent extension of TargetP is a peroxisomal prediction module that at a well above-random rate is able to tell apart peroxisomal proteins with a C-terminal targeting tripeptide (PTS1, 22 different motifs found) from non-peroxisomal proteins whose three C-terminal residues coincide with any of the accepted PTS1 motifs. Current research also focus on increasing the prediction accuracy by incorporating multiple alignment and secondary structure information, and on further testing of prediction ability as more and more sequences become available. TargetP is found at http://www.cbs.dtu.dk/services/TargetP/.

Speakers * Abstracts * Home

Transposon Silencing, RNA Interference and Cosuppression in *Caenorhabditis elegans*

Sylvia E.J. Fischer
Hubrecht Laboratory
Uppsalalaan 8
3584 CT Utrecht, The Netherlands
telephone: +31 302121965
fax: +31 302516554
email: sfischer@niob.knaw.nl
prestype: Platform
presenter: Sylvia E.J. Fischer

Sylvia E.J. Fischer, Rene F. Ketting, Femke Simmer, Marcel Tijsterman, Marieke van der Horst, Karen Thijssen, Kristy Okihara and Ronald H.A. Plasterk

All known natural isolates of the nematode *C. elegans* have active transposition of the Tc1 transposon in the soma. In the germ line of many strains however, Tc1 transposition is silenced. 43 mutants in which Tc1 has been activated in the germ line ("mutators") were isolated after random mutagenesis of a silenced strain. These mutants have several phenotypes in addition to the alleviation of transposon silencing: temperature sensitive sterility and high incidence of males. Half of the mutants are also resistance to RNA interference, the silencing of gene expression in response to exogenous dsRNA. One of these RNAi-resistant mutators, mut-7, was cloned, and shown to encode a protein with homology to RNaseD. This mutant is also resistant to cosuppression, a phenomenon in which introduction of transgenic copies of certain genes results in reduced expression of both the transgene and the endogenous gene. Interestingly, several RNAi-sensitive mutators are also sensitive to cosuppression. On the other hand, several mutants that are resistant to RNAi like mut-7, do not show alleviation of transposon silencing nor resistance to cosuppression. We will discuss progress in identifying more genes of the three classes of mutants: RNAi-resistant mutators, RNAi-sensitive mutators and RNAi-deficient non-mutators.

Speakers * Abstracts * Home

Functional Analysis of Mammalian Genes by a Large Scale Gene Trap Approach in Mouse Embryonic Stem Cells

Thomas Floss
GSF Institute for Environment and Health
Institute of Mammalian Genetics
Ingolstaedter Landstr. 1
85764 Oberschleissheim
telephone: 089-3187-2214
fax: 089-3187-3099
email: tfloss@gsf.de
prestype: Poster
presenter: Thomas Floss

Thomas Floss[1], Franz Vauti[2], Michael Wiles[3], Ernst-Martin Füchtbauer[4], Hans-Henning Arnold[2], Harald von Melchner[5], Jürgen Otte[5], Thomas Metz[1] and Wolfgang Wurst[1]

[1]GSF - National Research Center for Environment and Health, Institute of Mammalian Genetics, Ingolstaedter Landstrasse 1, D-85764 Neuherberg;
[2]Department of Cell- and Molecular Biology, Institute of Biochemistry and Biotechnology, TU Braunschweig, Spielmannstr. 7, D-38106 Braunschweig;
[3]Max-Planck-Institute for Molecular Genetics, Ihnestr. 73, D-14195 Berlin;
[4]Max-Planck-Institute of Immunobiology, Department of Developmental Biology, Stübeweg 51, D-79108 Freiburg;
[5]Laboratory for Molecular Hematology, University of Frankfurt Medical School, Theodor-Stern Kai 7, D-60590 Frankfurt am Main.

We have established a research center within the German HUGO project in order to perform a large scale functional analysis of mammalian genes taking advantage of the gene trap technology in combination with mouse embryonic stem (ES) cells. The gene trap technology provides an important tool of the human genome project for the identification and functional characterization of mammalian genes.

We are currently performing a large scale insertional mutagenesis screen in mouse ES cells. The mutagenic events are based on gene trap vector integrations into genes expressed in ES cells, generating a large number of mutant ES clones. The mutated genes are identified using RACE-PCR strategies allowing us to establish an archive of mutated genes generated in ES cells. We will establish mutant mouse lines by germ line transmission in case the genes mutated relate to human diseases.

At present, we have established 14.000 individual, mutated ES cell clones from which 6800 trapped genes have been sequenced (49%). Of the 3848 good quality sequences obtained, 1818 (47%) show homology to relevant genes. Another 903 sequences (23%) have been identified as EST´s and the remaining 1127 sequences (29%) show no homologies to known sequences present in the NCBI-GenBank database. The individual data for each clone, i.e. sequence, expression pattern and eventually mutant phenotype is stored in a public database which is freely accessible to the scientific community (http://tikus.gsf.de). These data will extend the international EST program and will contribute to unravel gene function genome wide.

Speakers * Abstracts * Home

Construction of Animal Models of Human Disorders by Injection of Large Genomic Fragment: From Function to Physiopathology and Therapy

Michel Fontes
INSERM U491
Medical genetics and Development
Fac de Médecine de la Timone
Marseille, France
telephone: +33 4 91 25 71 59
fax: +33 4 91 80 43 19
email: fontes@medecine.univ-mrs.fr
prestype: Platform
presenter: Michel Fontes

E. Passage[1], Sanguedolce V.[1], J. Pizant[1], J.F. Pellissier[2] and M. Fontés[1]
[1]INSERM U491, Medical genetics and Development, Fac de Médecine de la Timone, Marseille, France
[2]Laboratoire de Biopathologie Nerveuse et Musculaire, Fac de Médecine de la Timone, Marseille, France

A major challenge of functional genomic is to understand the function of genes which have been characterized through the human genome sequencing effort. This goal is particularly important in medical genetics, as we did not know (or poorly) the function of genes, cloned through positional cloning strategies, and involved in inherited disorders. We have thus developed the construction of animal models in order to understand the function in normal and pathological conditions of inherited disorders. Moreover, these models can also be used as"preclinical models", in order to test new therapeutical approaches. We have used a new approach to create these models: a transgenic strategy using large cloned genomic fragments (YAC, PAC, BAC) which present the following advantages: One would expect there to be stable expression which. The use of standard methods results in a low level of transgene expression. In contrast, however, several experiments carried out using large fragments of DNA have shown expression levels similar to those found for endogenous copies. The control of transgenic expression should perfectly reproduce that of the endogene. Another advantage of this approach is that expression is not dependent on the site of integration.

Using this strategy, we have constructed two models: one for the Charcot-Marie-Tooth type 1A (CMT1A) disorder and one for the Creutzfeld-Jacob disorder. The construction of a model for PKD1 is in progress. From these models we have already gained important insights in the physiopathological processes involved in these disorders. Two examples will be given. CMT1A disorder is not a demyelinating disorder but a dysmyelinating disease, giving us some clue to the function of the gene PMP22 involved in the disorder. Overexpression of the prion protein, did not affect the development of the central nervous system. Finally, we will present our strategy to test molecules that seems to revert the phenotype.

Speakers * Abstracts * Home

Human Chromosome 21 Gene Annotation I: Evolutionary Conservation and Organizational Features

Katheleen Gardiner
Eleanor Roosevelt Institute
1899 Gaylord Street
Denver, CO 80206 USA
telephone: 1-303-336-5652
fax: 1-303-333-8423
email: gardiner@eri.uchsc.edu
prestype: Platform
presenter: Katheleen Gardiner

Katheleen Gardiner and Dobromir Slavov

Analysis of the recently completed genomic sequence of human chromosome 21 (Hattori et al 2000) identified ~225 genes and gene models. Analysis of the evolutionary conservation of these genes and investigation of gene-specific organizational features can be expected to aid in functional determinations and in understanding regulation of gene expression.With these goals in mind, we have carried out several analyses. i) Protein sequences of 122 of the 225 genes/models showing similarity to complete proteins or protein domains of experimentally verified function were divided among 20 broad functional catagories; such classifications are useful but can be limited, misleading and/or only marginally informative. ii) A search with chromosome 21 mRNA sequences of the complete sequences of Drosophila, C.elegans and S.cerevisiae identified >60, >45 and >30 homologous genes, respectively, in each organism; Drosophila, in particular, provides interesting opportunities for further mutational analyses. iii) A similar search for orthologous genes in mouse identified >125 candidates; interestingly, more than 25% of these show strong, gene-specific conservation in untranslated regions. iv) A search of the protein and draft sequence databases showed that a number of chromosome 21 genes have homologues elsewhere in the human genome; this suggests interesting possibilities for overlapping expression patterns and functions that may have broad implications for research in Down syndrome and other chromosome 21 associated diseases. Lastly, analysis of expression patterns and regulation have revealed uncommon features of gene structures, including antisense transcripts, overlapping genes and possible non-coding RNAs.

Speakers * Abstracts * Home

Dependence of Nanos mRNA Localization and Translational Control on Structure and Organization of 3'UTR Sequences

Elizabeth R. Gavis
Department of Molecular Biology
Princeton University
Washington Road
Princeton, NJ 08544
telephone: (609) 258-3857
fax: (609) 258-1343
email: lgavis@molbio.princeton.edu
prestype: Platform
presenter: Elizabeth Gavis

Sherri Evans Bergsten, Seema Chatterjee, Ira E. Clark, Susan Crucs, Tao Huang, Bingyi Yao, and Elizabeth R. Gavis
Department of Molecular Biology, Princeton University

In multi-cellular organisms, localization of proteins to particular domains within a cell is fundamental to the generation of asymmetry during development and to the polarization of differentiated cells. Apart from protein targeting, many cell types employ localized translation of mRNAs to ensure restricted accumulation of proteins. As a model system, my laboratory studies the Drosophila nanos (nos) mRNA, whose localized translation is essential for patterning of the anterior-posterior body axis during embryonic development. Remarkably, two post-transcriptional mechanisms, subcellular mRNA localization and translational control, are coupled to restrict Nos protein synthesis to the posterior of the embryo. Nos protein is required in the posterior of the embryo for abdominal development but must be excluded from the anterior to permit head and thorax development. Localization of nos RNA to the posterior pole of the embryo is required to activate nos translation; when localization is abolished, nos mRNA remains translationally repressed. Localization is not sufficient to restrict nos to the posterior, however; more than 95% of nos mRNA in the embryo is unlocalized. Translational repression of this unlocalized mRNA is essential for anterior development. We have shown that translational repression of unlocalized nos mRNA is mediated by a 90 nucleotide translational control element (TCE) within the nos 3' untranslated region (3'UTR).

Both the primary sequence and predicted secondary structure of the TCE is conserved between *D. melanogaster* and *D. virilis* and TCE function is conserved as well. Through a systematic analysis of TCE mutations, we have shown that TCE function in vivo requires formation of a bipartite structure consisting of two stem-loops. One RNA binding protein, Smg, has been identified that interacts with TCE stem-loop II; TCE function requires at least two additional, as yet unidentified factors, however.

Posterior localization is mediated by a 540 nucleotide cis-acting localization signal that can be subdivided into partially redundant localization elements, one of which is coincident with the TCE. Each element contains a domain whose sequence is conserved between the *D. melanogaster* and *D. virilis* nos 3'UTRs. Results from our analysis of localization element function indicate that wild-type nos RNA localization requires recognition of multiple sequence or structural motifs. Our demonstration that the D. virilis nos 3'UTR confers wild-type localization in a D. melanogaster embryo suggests that the conserved domains define functionally relevant motifs that are the targets for binding by putative localization factors. Through biochemical assays, we have identified several candidate localization factors that interact specifically with these sequences.

Our results indicate that the coincidence of cis-acting translational regulatory sequences and localization signal sequences is functionally significant. Our data suggest that interaction of the nos 3'UTR with translational repressors and localization factors is mutually exclusive; consequently interaction with localization factors is sufficient to activate nos translation. Recent results suggest that the nos 3'UTR exists in alternate structural conformations depending on whether nos RNA is localized or unlocalized. Formation of alternate 3'UTR structures may underlie the switch between translational repression and localization-dependent translational activation.

Speakers * Abstracts * Home

Analyis of Genomes and Transcriptomes in terms of the Occurrence of Protein Parts and Features

Mark Gerstein
Molecular Biophysics & Biochemistry
Yale University
PO Box 208114
New Haven, CT 06520-8114 USA
telephone: 203 432 6105
fax: 360 838 7861
email: Mark.Gerstein@yale.edu
prestype: Platform
presenter: Mark Gerstein

Mark Gerstein, Ronald Jansen, Amar Drawid, Dov Greenbaum
Molecular Biophysics & Biochemistry Department, Yale University, New Haven, CT 06520

SUMMARY

My talk will focus on analyzing genomes and gene-expression data in terms of the finite list of protein "parts". Depending on context, a part could be a structural fold or sequence superfamily. I will touch on the following topics:

* How one can compare different genomes in terms occurrence of various parts in them. And how this idea can be extended to compare the representation of parts in the genome versus the transcriptome. In particular, this allows one to see what protein features are enriched in highly expressed proteins.

* How one can analyze the relationship between where a part is located and its transcriptome occurrence -- i.e. between a protein's subcellular localization and its level of gene expression. We extend this work to develope a formal Bayesian system for predicting subcellular localization, partially based on gene expression data.

* To what degree is protein function and protein-protein interactions related to similarities in the level of gene expression. Based on developing a statistical significance formalism, I will argue that while there is a definite relationship for certain classes of protein functions and protein-protein interactions, the relationship is not general and global. The absence of correlation is principally due to the inconsistent way protein function is defined.

REFERENCES

- http://bioinfo.mbb.yale.edu
- M Gerstein & R Jansen (2000). "The current excitement in bioinformatics, analysis of whole-genome expression data: How does it relate to protein structure and function?" Curr. Opin. Struc. Biol. (in press).
- A Drawid, R Jansen & M Gerstein (2000). "Gene Expression Levels are Correlated with Protein Subcellular Localization," Trends in Genetics 16: 426-430.
- 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 Broad Structural and Functional Categories: Characterizing Highly Expressed Proteins," Nuc. Acids Res. 28:1481-1488
- M Gerstein (1998). "Patterns of Protein-Fold Usage in Eight Microbial Genomes: A Comprehensive Structural Census," Proteins 33: 518-534.

Speakers * Abstracts * Home

Upstream and Downstream of the Human Leukemia Associated Transcription Factor RUNX1/AML1

Yoram Groner
Dept. of Molecular Genetics
The Weizmann Institute of Science,
Rehovot, ISRAEL
telephone: 972-8-9343972
fax: 972-8-9344108
email: yoram.groner@weizmann.ac.il
prestype: Platform
presenter:

Amir Pozner, Talia Polak, Thorsten Bangsow, Yael Bernstein, Varda Negreanu, Dalia Goldenberg, Tali Ben Aziz-Aloya, Carmen Anghel-Bangsow, Nir Rubins, Joan D. Bettoun, Ditsa Levanon, and Yoram Groner
Dept. of Molecular Genetics, The Weizmann Inst. of Science, Rehovot, ISRAEL

The RUNX1 gene on human chromosome 21q22.1 belongs to the "runt domain" gene family of transcription factors (also known as AML/CBFA/PEBP2). In human three highly conserved runt domain genes have been identified: RUNX1 on chromosome 21q22.1, RUNX2 on chromosome 6p21 and RUNX3 on chromosome 1p36.1. RUNX1 is a key regulator of hematopoiesis and a frequent target of leukemia associated chromosomal translocations. RUNX2 on the other hand functions as a key regulator of osteogenesis. The three RUNX gene-products bind to the same DNA sequence; therefore, their pleotropic functions must result from a regulated, temporally specific and tissue-specific expression as well as interactions with other transcriptional regulators.

Complete DNA sequence of RUNX1 was recently obtained within the framework of chromosome 21 mapping and sequencing consortium and the structural/functional features of the gene were established in our lab.

We found that two distinct promoter regions, designated distal (D) and proximal (P), control the transcription of RUNX1. These promoter regions are separated by a particularly large intron of 160 Kb. RUNX1 D and P promoters mediate transcription of mRNAs with discrete 5' UTRs, the D-UTR (0.45 Kb) and P-UTR (1.6 Kb). These 5'UTRs function as translational regulators thereby creating an additional level of control. RUNX1 mRNAs bearing D-UTR are readily translated via cap-dependent mechanism; wherease translation of mRNAs with the uncommonly long P-UTR is mediated through usage of the internal ribosome entry site (IRES) of this UTR. The temporal and spatial expression of RUNX1 promoters during mouse embryogenesis is studied using promoter-lac-z constructs in transgenic mice. RUNX1 expression also involves complex patterns of alternative splicing generating a diverse collection of mRNAs with coding regions varying in size between 188-480 aa. The short proteins usually lack the so-called transactivation domain (TAD). We showed that one of the short isoform RUNX1/p26, but not the full length RUNX1/p49, suppressed in vivo tumor growth and differentiation of ES cells derived teratocarcinomas. Most TAD containing RUNX proteins terminate with an identical C-terminal VWRPY motif which serves to recruit the co- repressor Groucho/TLE. We demonstrated direct interaction between Gro/TLE and RUNX1/p49 and show that this interaction lead to transcriptional repression of RUNX1 regulated target genes. The protein isoform RUNX1/p51, which is almost identical to RUNX1/p49 but lacks VWRPY, did not interact with Gro/TLE. These data highlight the biological significance of the alternatively spliced RUNX1 isoforms.

Speakers * Abstracts * Home

Genomic Clocks and Animal Evolution

S. Blair Hedges
Department of Biology
208 Mueller Lab
Pennsylvania State University
University Park, PA, USA
telephone: 814-865-9991
fax: 814-865-9131
email: sbh1@psu.edu
prestype: Platform
presenter: S. Blair Hedges

Hedges, S. Blair
Department of Biology, Institute of Molecular Evolutionary Genetics, and Astrobiology Research Center, Pennsylvania State University, University Park, PA, USA

Four billion years of mutational change in the genomes of organisms has left a trail of evidence for the evolutionary detectives. A major goal is to reconstruct the branching order and times of divergence of populations, species, and genes. Until recently this goal has been difficult to achieve for many reasons, but primarily for lack of sufficient data. The genome projects are solving the data problem but analytical aspects have become more complex. Available computational tools are useful for grouping similar sequences, but true gene (orthology group) identification involves evolutionary considerations that have yet to be fully automated. New methods are needed for constructing trees and estimating divergence times from hundreds or thousands of genes. Evolutionary genomic studies have shown promise in yielding robust phylogenies and precise time estimates. The results of such studies have revealed that animals diverged from fungi and plants about 1.6 billion years ago, and many animal phyla arose before the Cambrian explosion. Recent evidence bearing on the phylogeny of vertebrates and their timescale of evolution will be discussed.

Speakers * Abstracts * Home

Towards Optimized Arrays: Identification of Specific, Non-Conserved, Single Copy Sequence Fragments of All Genes and EST Clusters in Human

Bernhard Korn
RZPD REsource Center for Genome Research
Im Neuenheimer Feld 506
69120 Heidelberg Germany
telephone: +49 (0)6221 42 4700
fax: +49 (0)6221 42 4704
email: b.korn@dkfz-heidelberg.de
prestype: Poster
presenter: Heil, Oliver

O. Heil, M. Roth, A. Poustka and B. Korn

The technology of complex hybridization for expression profiling is dependent on: a) high quality, well characterized RNA material and b) highly specific array elements. These array elements might be oligonucleotides of a given size or melting temperature (oligo arrays) or DNA fragments that are usually amplified by PCR (DNA array). Elements on DNA arrays carry in most cases DNA fragments that represent the complete inserts of given cDNA clones. However, the clone based strategy has many drawbacks, because amplifying inserts by vector primers do always generate products harboring all problematic sequence parts of a given cDNA clone, e.g. polyA tails (length of which is often not even known), repetitive sequences (e.g. Alu), low complex sequences, sequences conserved between different genes (domains, motifs, gene families), and parts of multicloning sites. Furthermore, vector PCR has the intrinsic problem of varying insert sizes, which leads to differences in efficiency in amplification, depending on insert size and GC content of the amplified sequence, resulting in amplification differences of more than 50-fold in molar terms.

In order to circumvent these problems, we have developed algorithms that detect single copy sequence fragments in all human Genes and EST clusters (also applicable for other species or whole genomes). In a second step, these sequence parts can be used for oligonucleotide design: a) for oligo arrays, or b) for PCR primer design, the approach we currently prefer.

The determination of PCR primers does allow us to choose the sequence part that suits best for PCR. The quality criteria applied are: homogeneous annealing temperature, homogeneous size of PCR product, relative position of the PCR product within each gene, GC-content of PCR product, and potential access for cloning of the individual PCR products.

We have started to amplify all human genes and EST clusters using primer pairs that are designed by this new software tools described above. In total we aim to generate about 70.000 different PCR products within the next 9 months. The algorithms, software tools and data handling will be discussed in detail. In addition, we will present amplification results and discuss automation of data and material handling as well as advantages and disadvantages of this approach towards optimized DNA arrays.

Speakers * Abstracts * Home

The European Renal Biopsy cDNA Bank (ERCB): A Multicenter Study for Disease Specific mRNA Expression Analysis in Human Kidney Biopsies

Dr. Anna Henger
Medizinische Poliklinik
Klinische Biochemie
Schillerstr. 42
80336 München
telephone: 089-5996841
fax: 089-5996860
email: Anna.Henger@medpoli.med.uni-muenchen.de
prestype: Poster
presenter: Anna Henger

A. Henger, C. D. Cohen, D. Schlöndorff, M. Kretzler
Medizinische Poliklinik der LMU, Pettenkoferstr. 8a, 80336 München, for the ERCB Consortium

As a consequence of the current molecular advances gene expression analysis could become an important diagnostic tool in nephrology. The generation of a comprehensive renal cDNA biopsy bank will be a critical pre-requisite to test the diagnostic and prognostic value of this approach. To this end, an interdisciplinary European collaboration of renal research centers was initiated and the following protocol for the collection of cDNAs from microdissected renal biopsies has been established: In participating centers informed consent is obtained from patients undergoing routine diagnostic renal biopsy, clinical parameters are collected in a standardized data sheet and 10% of a renal biopsy cylinder is obtained for molecular analysis. Glomeruli and tubulo-interstitial segments are manually microdissected and mRNA expression analysis is performed. Currently, the biopsy material is analyzed for expression changes of candidate marker genes with a semi-automated 'real-time' RT-PCR system. Glomerular specific cDNAs are employed together with housekeeping mRNAs for quality control of the material obtained. Using this system, cDNAs of interest can be reproducibly quantified from as little as 1% of a microdissected glomerulus, enabling parallel quantification of up to 100 different cDNAs from one biopsy segment. Microdissection effectively separated glomeruli and tubulointerstitial compartments with no significant cross contamination.

The ERCB will allow expression analysis in defined human renal diseases for specific cDNAs of interest. In addition, correlation of mRNA expression pattern could reveal diagnostic and prognostic molecular markers, giving nephrologists a fundamentally novel kind of information to guide their patient care. Finally, complex molecular networks activated in human renal disease processes can be identified as potential novel targets for disease management.

Members of the ERCB:
A. McKinnon, P. A. J. Brown, A. Rees, Aberdeen; P. Mertens, J. Floege, Aachen; L. Gesualdo, F. P. Schena, Bari; M. Saleem, Bristol; M. Clarkson, H. R. Brady, Dublin; F. Strutz, G. Müller, Göttingen; M. Kuhlmann, H. Köhler Homburg/S.; J. K. Gerth, R. Fünfstück, G. Stein, Jena; M. P. Rastaldi, G. D'Amico, Milan; F. Delarue, J. D. Sraer, Paris; M. Fischereder, B. Krämer, Regensburg; R. Wütrich, St. Gallen; N. Braun, T. Risler, Tübingen; R. Oberbauer, D. Kerjaschki, Vienna; D. Mönks, C. Wanner, Würzburg.

Speakers * Abstracts * Home

Large Scale Screen for Genes Controlling Mammalian Embryogenesis

Bernhard Herrmann
Max-Planck-Institut für Immunbiologie
Dept. Developmental Biology
Stuebeweg 51
79108 Freiburg
telephone: +49 761 5108 582
fax: +49 761 5108 569
email: Herrmann@Immunbio.mpg.de
prestype: Platform
presenter: Bernhard Herrmann

Lorenz Neidhardt, Stephan Gasca, Karin Wertz, Andrea Schepler, Franz Obermayr, Alexander Aulehla, Susanne Worpenberg and Bernhard G. Herrmann

The molecular analysis of cell differentiation, tissue patterning and organ development requires the isolation of genes involved in these processes. Large scale EST and genome sequencing have been employed by many laboratories in order to identify all genes encoded by the human genome. However, sequence information alone does not reveal the tissue type, process or organ in which a particular gene is acting.

We have developed a high-throughput strategy for determining the expression of genes in 9.5 day mouse embryos on a large scale. Genes are selected from cDNA libraries or cDNA arrays and assayed for their expression patterns by multiple whole mount in situ hybridisation analyses, thus providing expression data at high resolution. A genome set of 40,000 genes could be assayed in 20 person years (i.e. 5 persons, 4 years). Gene expression and sequence information together represent essential information about the cell type and process a gene is acting in, and often also about the biochemical function. This information can be utilized to develop strategies for further functional analyses, and to identify potential target genes for pharmaceutical drugs. In addition, chromosomal mapping of the genes identifies candidates for known human disorders and mouse mutations.

In a pilot screen, appr. 800 specifically expressed genes have been identified. Of 607 clones sequenced, about 43% represent known genes and about 44% novel genes or ESTs. The remaining cDNAs represent mouse orthologues of known genes or new members of known gene families.

Appr. 25% of the known genes isolated are involved in transcriptional regulation, such as Hox genes, Tbx genes, Msx, Pax9. Appr. 16% of the known genes isolated represent mouse orthologues of human disorders. Three mouse genes have been isolated for which a classical mutant locus is known (Brachyury, pudgy, vibrator). This suggests that a similar proportion of mouse and human mutations is represented among the novel genes and ESTs identified. A data base providing expression, sequence and homology search data is being built up at the Resource Center (RZPD). This project will have a strong impact on several fields of biomedical research.

Speakers * Abstracts * Home

The c-fos and c-myc 3' Untranslated Regions are Multi-Functional Regulatory Regions Influencing mRNA Localisation and Stability

John Hesketh
Department of Biological & Nutritional Sciences
University of Newcastle
Newcastle-upon-Tyne
UK NE1 7RU
telephone: 44-191-222-8744
fax:
email:
prestype: Platform
presenter: John HESKETH

Dalgleish, G.D.[1], Veyrune, J-L.[2], Blanchard, J-M.[2] and Hesketh, J.E.[1,3]

[1]Rowett Research Institute, Bucksburn, Aberdeen, AB21 9SB, UK

[2]IGMM, CNRS, 1818 Route de Mende, Montpellier, France

[3]Department of Biological and Nutritional Sciences, University of Newcastle, Kings Rd, Newcastle-upon-Tyne, NE1
7RU, UK

3'untranslated regions (3'UTR) are emerging as important in post-transcriptional control of gene expression, regulating mRNA stability, translation and localisation. The mRNAs encoding c-myc and c-fos are unstable and this instability is partly determined by sequences within the 3'UTR. C-myc mRNA is localised around the nucleus and both c-myc and c-fos mRNAs are found associated with the cytoskeleton. Studies with cells transfected with chimaeric reporter constructs in which 3'UTR sequences are linked to a ß-globin reporter show that both c-myc and c-fos 3'UTRs contain sequences capable of targeting globin to the cytoskeleton and the perinuclear cytoplasm. In both cases the part of the 3'UTR containing the localisation signal has been partly defined (to within 86nt for c- myc, to within 145 for c-fos) and the regions responsible are distinct from those that affect mRNA stability. Deletion analysis of the both 3'UTRs shows that removal of specific regions which affect mRNA localisation do not affect stability and vice versa. Thus, it appears that these two 3'UTRs contain multiple regulatory elements, one leading to mRNA instability and one to mRNA localisation. The nature of the RNA sequence/structure that forms the localisation signal is unknown. Furthermore, if translation is prevented localisation of the mRNA does not occur suggesting that some form of translation initiation complex is required for localisation of the mRNAs. Thus, the 3'UTRs of c-myc and c-fos play a key role in the post-transcriptional control of the expression of these genes and in determining the site of the synthesis of the encoded proteins.

Speakers * Abstracts * Home

Tissue-Specific Expression Analysis by RNA in situ Hybridisation of Mouse Genes in the Region Syntenic to Human Xq28

Bernhard Korn
RZPD REsource Center for Genome Research
Im Neuenheimer Feld 506
69120 Heidelberg Germany
telephone: +49 (0)6221 42 4700
fax: +49 (0)6221 42 4704
email: b.korn@dkfz-heidelberg.de
presenter3 = Kolb, Anja

Anja A. Kolb[1], Nina S. Heiss[1], Petra Kioschis[1,2], Annemarie Poustka[1]

[1]Deutsches Krebsforschungszentrum, Heidelberg, Germany
[2]Fachhochschule Mannheim, Hochschule für Technik und Gestaltung, Mannheim, Germany

The Xq28 region is a very well characterised part of the human genome mainly because it harbours a large number of disease genes. Although most of these genes have been identified and sequenced, the extent of characterisation at the level of expression and function varies. To continue our systematic analysis of genes in Xq28, we are investigating the developmental and tissue-specific expression of a number of orthologous mouse genes by RNA in situ hybridisation. One example is the DKC1 gene which is responsible for causing X-linked recessive dyskeratosis congenita (DKC) and the allelic variant of the disease, Hoyeraal-Hreidarsson syndrome (HHS). RNA in situ hybridisations revealed a ubiquitous expression pattern of the Dkc1 transcript, although considerably higher expression levels were detected in embryonic epithelial and neuroectodermal tissues, as well as in differentiated neurons of the adult brain. This dispels the notion that the major functions of the DKC1 gene are confined to rapidly dividing cells and is in agreement with some of the phenotypic features of DKC and HHS patients. This will aid in gaining a better understanding of the pathomechanism of the disease and will complement our studies on the respective mouse models.

Further, we have examined the tissue-specific and cell-type specific expression pattern of the chloride channel-like gene, CLIC2, and XAP139, a gene of unknown function. This is providing useful hints towards understanding the function of these genes. We hope to carry out such analyses for the majority of genes in Xq28. In the long run such data will be of value not only in identifying the remaining disease-associated genes in Xq28, but also in providing us with integrated knowledge about the functional and evolutionary aspects of this model region.

Speakers * Abstracts * Home

Transcriptional Profiling for the Classification of Novel Natural Product Drugs

Bernhard Korn
RZPD Resourcenzentrum für Genomforschung
INF 506
69120 Heidelberg, Germany
telephone: +49 (0)6221 42 4700
fax: +49 (0)6221 42 4704
email: b.korn@dkfz.de
prestype: Platform
presenter: Bernhard Korn

Korn, B.[1], Gernold, N.[1], Seranski, P.[2], Möckel, B.[2], Eck, J. [2], Zinke, H. [2], Poustka, A.[3]

[1]RZPD Resourcenzentrum für Genomforschung , INF 506, 69120 Heidelberg, Germany
[2]B·R·A·I·N GmbH, 64673 Zwingenberg, Germany
[3]DKFZ, Molecular Genome Analysis, 69120 Heidelberg, Germany

With the growing number of functionally characterised genes and proteins, the identification of novel pharmaceutical compounds can be supported by the use of transcriptional profiling of already known genes, gene clusters and complex pathways. The specific aim of our work was the identification and functional characterisation of novel compounds from natural product libraries. These compounds are analysed for their activity using cell assays and the functional classification is assigned by the application of array based transcriptional profiling with known cDNAs.

A focus in our drug discovery program lies in the field of oncology and immunomodulation. Therefore, we are using a target oriented cDNA array systems (RZPD onco array) that allows the identification of transcriptional changes in various intra- and intercellular signal transduction pathways leading e.g. to apoptosis, cell-cycle regulation or cytokine production. Unknown products can be classified by comparison of transcriptonal profiles of already known, cytotoxic, cytostatic or antibiotic drugs with arrays from surrogate hosts treated with compounds to be investigated.

We set up an array based classification system for the molecular characterization of recombinant mistletoe lectin, rViscumin, a plant derived ribosome inactivating protein [1,2]. rViscumin is a potent apoptosis inducing drug [3] with immunomodulatory activity in vitro. This drug was applied on the monocytic THP-1 cell line and compared to the expression profile changes induced by other well characterised drugs, e.g. taxol. In response to rViscumin treatment signal transduction pathways were induced, showing that the MAPKinases SAPK/JNK and p38 are activated and transcription factors are induced which regulate gene expression, e.g. cytokine genes. The accumulated data help to understand the molecular mechanism of rViscumin activity and furthermore lead to the identification of predictor gene classes that are subsequently used for screening and characterisation of new drugs.

References

[1] Eck, J. et al.(1999) Eur. J. Biochem. 264, 775-784
[2] Eck, J. et al. (1999) Eur. J. Biochem. 265, 788-797
[3] Möckel, B. et al. (1999) Canc.Res.Clin.Oncol. Suppl. 125, 55-4

Speakers * Abstracts * Home

Functional Genomics by Transposon Tagging in Yeast

Anuj Kumar
Project Manager, Yale Genome Analysis Center
Department of Molecular, Cellular and Developmental Biology
Yale University
P. O. Box 208103
New Haven, CT 06520-8103
telephone: 203-429-9949
fax: 203-432-6161
email: anuj.kumar@yale.edu
prestype: Platform
presenter: Anuj Kumar

Transposable elements are powerful tools for insertional mutagenesis, notable for the scale of throughput and ease of analysis they offer. As a result, transposon-based systems provide an excellent method of genome-wide mutagenesis. Illustrating this point, we have employed transposon mutagenesis as a means of elucidating gene function on a genome-wide scale in Saccharomyces cerevisiae. Our approach is unique in its versatility: by designing specially modified multi-functional mini-transposons, we can measure gene expression, generate gene disruptions, and localize gene products; all from a single transposon insertion event. Using these multi-functional transposons, we have generated a collection of roughly 17,000 yeast mutants, each carrying a single transposon insertion within a region of the genome expressed during vegetative growth and/or sporulation. This collection has been used to determine disruption phenotypes for nearly 8,000 strains under 20 different growth conditions. Additionally, we have identified over 300 previously non-annotated open reading frames and determined the subcellular localization of transposon-tagged proteins in approximately 9,000 yeast mutants. In total, our study encompasses over a quarter of a million data points, providing insight into the function of nearly 3,000 yeast genes. In addition to the wealth of data generated by this approach, our genome-wide collections of defined mutants constitute a valuable laboratory reagent, streamlining future genetic screens by eliminating the need to identify the affected gene within a given mutant of interest. As such, these collections of defined alleles promise to greatly expedite studies of gene function on a genomic scale.

Speakers * Abstracts * Home

Negative and Positive Regulation of Erythropoiesis

Shuo Lin
Institute of Molecular Medicine and Genetics
Medical College of Georgia
Augusta, GA 30912
telephone: 706-721-8762
fax: 706-721-8752
email: shuolin@immag.mcg.edu
prestype: Platform
presenter: Shuo Lin

The production of red blood cells is under both positive and negative control. Dysregulation of either of these may cause human diseases such as anemia or leukemia. Identification of lineage-specific regulators involved in erythroid cell proliferation should help elucidate the molecular basis underlying these diseases. We have performed a large-scale expression screen using a cDNA library constructed from the GATA-1 positive embryonic hematopoietic cells of transgenic zebrafish and have isolated more than 50 genes that are specifically expressed in erythroid cells. A number of negative regulators including receptors, transcription factors, and enzymes of the MAP kinase pathway have been identified. We report that ZH-DR, a novel hematopoietic death receptor gene isolated from zebrafish embryonic erythroid progenitor cells, plays a critical role in regulating erythropoiesis. While overexpression of ZE-DR induced rapid apoptosis, transgenic zebrafish expressing a dominant negative form of ZE-DR produced an excess of red blood cells. This study provides the first in vivo model of erythroid dysregulation caused by the down-regulation of a lineage-specific death receptor. In addition, we will discuss another zebrafish gene that can enhance GATA-1 expression in transgenic zebrafish embryos. Our studies in zebrafish have revealed novel information regarding negative and positive regulation of erythropoiesis and should be useful for understanding the similar process in higher vertebrates.

Speakers * Abstracts * Home

Gene Trapping Identified Transiently Induced Survival Genes During Programmed Cell Death

Harald von Melchner
Laboratory for Molecular Hematology
University of Frankfurt Medical School
60590 Frankfurt am Main, Germany
telephone: +49-69-63016696
fax: +49-69-63016390
email: melchner@em.uni-frankfurt.de
prestype: Platform
presenter: Frank Wempe

Frank Wempe, Ji-Yeon Yang, and Harald von Melchner
Laboratory for Molecular Hematology, University of Frankfurt Medical School, Frankfurt am Main, Germany

One of the most common forms of physiologic apoptosis occurs as aconsequence of "survival" factor deprivation. To identify genes that are transiently induced during this process, we have used a strategy based on gene trap mutagenesis and site specific recombination (Cre/ loxP). We show here that hematopoietic cells undergoing apoptosis by growth factor (IL3) deprivation upregulate survival genes prior to irreversible death commitment. The activation of protective mechanisms during the early stage of programmed cells death results in reduced apoptosis and improved survival of cells treated with a transient apoptotic stimulus. Thus, we conclude that apoptosis in hematopoietic cells is the end result of a conflict between death and survival signals, rather than a simple death by default.

Speakers * Abstracts * Home

Adenosine Deaminases that Act on RNA Edit Non-Coding Regions of mRNAs in Both Human and *C. elegans*

Daniel P. Morse
Department of Biochemistry/HHMI
50 North Medical Drive
Salt Lake City, Utah 84132
telephone: 801-585-3110
fax: 801-581-5379
email: dmorse@howard.genetics.utah.edu
prestype: Platform
presenter: Daniel P. Morse

Daniel P. Morse, P. Joe Aruscavage, and Brenda L. Bass

Adenosine deaminases that act on RNA (ADARs) are RNA editing enzymes that convert adenosines to inosines within double stranded regions of RNA. Using a method that detects inosine-containing messenger RNAs, we previously identified a number of ADAR substrates in *C. elegans* [Morse and Bass (1999) PNAS 96, 6048]. We chose to study *C. elegans* hoping that it would be a good model system for the function of ADARs in more complex organisms. At the time we began these studies, a few mammalian ADAR substrates had been discovered. In each of these, ADARs produced functionally significant codon changes. Thus, it was surprising that the *C. elegans* ADAR substrates were edited only in non-coding regions. To determine whether ADAR function in *C. elegans* is atypical, we initiated a search for human ADAR substrates. Using an improved version of the method we have identified a number of new ADAR substrates from human brain mRNA. As in *C. elegans*, these substrates are edited exclusively in non-coding regions suggesting that codon changes are the exception rather than the rule. The implications for the biological roles of ADARs will be discussed.

Speakers * Abstracts * Home

Analysis of Human cDNAs Encoding Large Proteins: Towards and Beyond the Identification of Protein-Coding Transcribed Sequences

Takahiro Nagase
Department of Human Gene Research
Kazusa DNA Research Institute
1532-3, Yana, Kisarazu
Chiba 292-0812, Japan
telephone: +81-438-52-3930
fax: +81-438-52-3931
email: nagase@kazusa.or.jp
prestype: Platform
presenter: Takahiro Nagase

Takahiro Nagase, Reiko Kikuno, and Osamu Ohara
Department of Human Gene Research, Kazusa DNA Research Institute

Over the past six years, we have been studying the protein-coding sequences of unidentified human genes. In particular, our recent analyses have been focused on previously unidentified genes that encode large proteins (>50 kDa) in the human brain because large proteins are known to be frequently involved in various important cellular processes. To achieve this, we have sequenced the full length of large cDNAs (>4 kb) selected according to novel sequences at their 5' and 3' ends and their protein-coding potentials. These newly identified genes were systematically designated KIAA plus a four-digit number. The number of cDNA sequences determined has reached 2000 and their average size is approximately 5 kb. Thus, nearly 10 Mb of human cloned cDNA sequence has been determined with high accuracy. Since the number of genes encoding large proteins is expected to be only about 10% of the total number of human genes, the number of KIAA genes in the public databases (1642 entries, August 2000) is quite significant when it is considered that this represents a set of genes expressed in the brain. We also have analyzed chromosomal loci and expression profiles of many of these KIAA genes, and have made all the data accessible at our Web site (http://www.kazusa.or.jp/huge). From our experience in human cDNA sequencing over the past six years, we have learnt how to deal with the many different problems that arise in cDNA analysis. As the human genome sequencing project enters the last phase in which the draft sequences are finalized, cDNA sequence data will serve as an important tool for the interpretation of the sequence of the human genome. Furthermore, the cDNA data will offer a variety of information regarding post-transcriptional events, such as alternative splicing and RNA editing, which cannot be predicted from the genome sequence at present. On the other hand, the genome sequence can help considerably with the resolution of problems in cDNA technology, most of which originate from the fact that cDNAs are nothing but artificial copies of mRNAs. Therefore, integration of the genomic and cDNA data should be an urgent and critical concern. The ultimate goal of our project goes beyond the identification of protein-coding sequences in the human genome, as we believe that cDNA analysis will play a key role in bridging gaps in understanding between the genome, the transcriptome, and the proteome.

Speakers * Abstracts * Home

Planarian Gene Expression Profiles Using cDNA Chip Technology

Masumi NAKAZAWA
National Institute of Genetics
Center for Information of Biology
Yata1111, Mishima-city, Shizuoka pref.
411-8540 Japan
telephone: +81-559-81-6847
fax: +81-559-81-6848
email: mnakazaw@lab.nig.ac.jp
prestype: Poster
presenter: Masumi NAKAZAWA

Masumi NAKAZAWA[1], Katsuhiko MINETA[1], Kazuho IKEO[1], Kiyokazu AGATA[2], Takashi GOJOBORI[1]
[1]National Institute of Genetics, [2]Okayama University, Japan

The primitive brain equipped animal, planarian (Platyhelminthes) gives us good evolutionary insights for brain study. We made a planarian cDNA microarray including 1570 genes, following our planarian EST project (*Dugesia japonica*) that has produced about 9000 sequences (Mineta et al, In preparation). One of neuron specific genes, Synaptotagmin (Djsyt) was found out from our ESTs and further study was reported (Tazaki et al. BBRC 260: 426-432, 1999). Our planarian cDNA chip is containing 17 neuron related genes, 13 transcription factors, 65 protein kinases, 270 hypothetical protein coding genes, 18 unknown genes and so on. This cDNA chip is used to detect brain specific genes doing competitive hybridization between mRNAs derived from head and body part of planarians. The current result shows planarian brain specific genes contain not only human neural gene homologues but also unknown genes, which seem planarian specific genes. We also attempt RNA interference technique to make brain-specific genes knocked out planarian. We will report the planarian brain specific genes and its expression profiles during regeneration and on knocked out planarian.

Speakers * Abstracts * Home

Insights into Protein Evolution and Function from the CATH Structural Database

Christine Orengo
Department of Biochemistry and Molecular Biology
University College
Gower Street
London WC1E 6BT, UK
telephone: 00 44 020 7419 3284
fax: 00 44 020 7380 7193
email:
prestype: Platform
presenter: Christine Orengo

Orengo, C.A, Pearl, F., Lee, D., Bray, J., Todd, A., Sillitoe, I. & Thornton, J.M.
Biomolecular Structure and Modelling Unit, University College, London

The rapid progress of the international genome projects has resulted in a wealth of sequence information for protein families. Although the structural data has lagged behind the sequences, analysis suggests that with the advent of structural genomics initiatives we may soon have structural representatives for many evolutionary protein families

At UCL, we have clustered all the well-resolved protein structures, in the PDB, into structural families. Proteins are first divided into separate domains and both sequence and structure alignment methods used to identify relationships. Data on families is stored within the CATH database (Class, Architecture, Topology or fold and Homologous superfamily). To date, there are ~26,000 domains within CATH, which cluster into ~1000 homologous superfamilies and ~600 fold groups. Recently, profile based methods (PSIBLAST, Altschul et al. 1997) have been used for identifying 160,000 sequence relatives in the genomes and integrating these into CATH families, using conservative thresholds.

Functional analysis has been undertaken within each superfamily in CATH, using protein ligand interaction plots (DOMPLOT, Todd et al. 1998). Correlations between sequence and structure motifs are captured using a new dictionary of functional information (DHS, Bray et al, 1999). Analysis of CATH enzyme superfamilies reveals that function is completely conserved in about one third whilst in a further quarter the catalytic mechanism is conserved though the substrate or ligand may vary. In some superfamilies function was observed to vary widely.

Speakers * Abstracts * Home

A Novel and High Throughput Approach for SNP Screening, Scoring and Association Analysis using Genomic or cDNA Pools

Xinghua Pan, Sherman M. Weissman
Department of Genetics
Boyer Center for Molecular Medicine
Yale University School of Medicine
295 Congress Ave., BCMM 331
and Molecular Staging, Inc.
New Haven, Connecticut 06510, USA
telephone: (203)737-2277
fax: (203)737-2286
email: xhpan@biomed.med.yale.edu
prestype: Platform
presenter: Xinghua Pan

Xinghua Pan[1,2] and Sherman M. Weissman[1]
[1]Department of Genetics, Boyer Center for Molecular Medicine, Yale University School of Medicine, 295 Congress Ave., BCMM 331
[2]Molecular Staging, Inc., New Haven, Connecticut 06510, USA.

SNP association studies may be the most powerful approach for identifying the genetic determinants of common disease. Developing a very high throughput and robust approach for screening/scoring the SNPs within one pool, and identifying the differences in SNP quality and quantity between two pools of human DNA samples is a critical enterprise in human genetics. Current techniques for SNPs analysis for the most part rely on discriminating the single base difference (allele) of each SNP with previously known sequence, and on typing all of a panel of DNAs with a limited number of candidate SNPs. These limitations seriously hinder complex trait mapping.

We have been developing an alternative approach that potentially allows one to map all SNPs in a pool of DNA samples or between 2 pools of DNA samples in one procedure. First we denature and reanneal DNA fragments and separate the SNP-containing DNA fragment pool from the non-SNP-containing DNA fragment pool. We then propose to apply these two pools in-parallel to oligonucleotide microarrays, gel display or other techniques to determine the relative allele frequencies within a given sample pool or between the tested sample and its control. Alternatively, another approach, SNPs - Representation Differential Analysis (SNPs-RDA) strategy may be used to distinguish the neutral polymorphism of SNPs from the SNP's in strong linkage disequilibrium with a monogenic trait. MFISH on metaphase chromosomes with the SNP fragment pool is proposed for the physical mapping of a complex trait, besides its genetic mapping. A highly specific as well as highly sensitive technique for separating SNP fragments from non-SNP fragments with immobilized DNA glycosylases has been established. With this tool, 250-500 folds of enrichment for fragments containing a single nucleotide mismatch, or for perfectly matched fragments, was obtained after one cycle of treatment. An efficient technique for reducing a complex DNA pool into full coverage and non-overlapping subsets with different sequences at the two ends of the subset fragments has been developed. A strategy for selective recovery of either one of the homphybrids or the heterohybrids from a mixture of denatured and reannealed DNA samples was also designed. As a first test, we are employing this approach in cDNA pool of a lymphoblast cell line. Initial data will be discussed.

The present approach has the potential to provide a simple, accurate, high throughput and efficient flexible

platform for new SNPs discovery and identification, for SNP screening/scoring, and most importantly for whole genome/ cDNA SNP association of a given trait.

Speakers * Abstracts * Home

Analysis of Oligonucleotide AUG Start Codon Context in Eukariotic mRNAs

Graziano Pesole
Dipartimento di Fisiologia e Biochimica Generali
via Celoria, 26
20133 Milan, ITALY
telephone: +39-02-70644803
fax: +39-02-70632811
email: graziano.pesole@unimi.it
prestype: Platform
presenter: Graziano Pesole

Graziano Pesole[1]*, Carmela Gissi[1], Giorgio Grillo[2], Flavio Licciulli[2], Sabino Liuni[3] and Cecilia Saccone[3,4]

[1]Dipartimento di Fisiologia e Biochimica Generali, Università di Milano, Milan, Italy
[2]Area di Ricerca del CNR, Bari, Italy
[3]Centro di Studio sui Mitocondri e Metabolismo Energetico del CNR, Bari, Italy
[4]Dipartimento di Biochimica e Biologia Molecolare, Università di Bari, Italy

The AUG start codon context features have been investigated by analyzing eukaryotic mRNAs belonging to various taxonomic groups. The functional relevance of each specific position surrounding the AUG start codon has been established as a function of the measured shift between base composition observed at that particular position, and base composition averaged over all the 5'untranslated regions. A more detailed analysis carried out on human genes belonging to different isochores showed significant isochore-specific features that cannot be explained only by a mutational bias effect.

The most represented heptamers spanning from position 3 to +4 with respect to the initiator AUG have been determined for mRNAs belonging to different taxonomic groups and a web page utility has been set up (http://bigarea.area.ba.cnr.it:8000/BioWWW/ATG.html) to determine the relative abundance of a user submitted oligonucleotide context in a given species or taxon.

Speakers * Abstracts * Home

Mapping of Zebrafish Expressed Sequences and the Evolution of the Vertebrate Genome

John Postlethwait
Institute of Neuroscience
University of Oregon
Eugene OR 97403 USA
telephone: 541-346-4538
fax: 541-346-4538 or 541-346-4548
email: jpostle@oregon.uoregon.edu
prestype: Platform
presenter: John H. Postlethwait

John H. Postlethwait
Institute of Neuroscience, University of Oregon, Eugene, OR 97403

Comparative mapping of zebrafish and mammalian genomes reveals large regions of conserved synteny, sometimes about as long as individual mammalian chromosomes. For example, for example results suggest that most of the loci on human chromosome 17 have remained syntenic for the last 450 million years, since the divergence of ray-fin and lobe-fin fish. Within conserved regions, however, frequent intrachromosomal rearrangements have altered gene order. Duplicated chromosome segments suggest that a genome duplication occurred in the lineage leading to zebrafish. Comparison with other teleosts suggest this event may have occurred deep in the ancestry of teleost fish. Genetic maps show that about 30% of zebrafish genes may be retained from this event. Analysis of expression patterns of duplicated genes derived from this event suggest that duplicated gene pairs have frequently partitioned ancestral functions between the two copies. Despite this genome duplication event, teleost fish and mammals have about the same number of chromosomes. Mapping data were analyzed to test the hypotheses that current chromosome numbers are due to either an excess of chromosome fissions in the human lineage or an excess of chromosome fusions in the fish lineage, and the latter seems more likely. These investigations will help improve connectivities between mammalian and fish genomes, and thus speed our understanding of gene function in both groups of animals.

Speakers * Abstracts * Home

Analysis of mRNA Recruitment to Ribosomes in Eukaryotic Cells

Thomas Preiss
EMBL
Gene Expression Programme
Meyerhofstrasse 1
D-69117 Heidelberg Germany
telephone: +6221 387 139
fax: +6221 387 518
email: preiss@embl-heidelberg.de
prestype: Platform
presenter: Thomas Preiss

Thomas Preiss, Julie Baron, Ennio De Gregorio, and Matthias W. Hentze
Gene Expression Programme, EMBL, Meyerhofstrasse 1, D-69117 Heidelberg, Germany
PREISS@EMBL-Heidelberg.de

Nearly all eukaryotic mRNAs carry both a 5’ cap structure and a 3’ poly(A) tail as posttranscriptional modifications. The two modified mRNA termini are known to jointly and synergistically activate translation mediated by bridging interactions of eIF4E and PABP, with the translational adapter eIF4G, enabling a "circular" conformation of the mRNA. eIF4G also binds to eIF4A, an ATP-dependent RNA helicase, and through contacting the small ribosomal subunit-bound factor eIF3, is thought to effect it’s recruitment to the mRNA, a central step in translation initiation. In addition to these ‘core’ interactions, eIF4G also displays a growing number of interactions with other factors, which can modulate translation initiation or have the potential to link it to other aspects of the mRNA’s life cycle. We are investigating the functional consequences of interactions involving eIF4G using in vitro translation systems as well as cultured cells.

A tethered-function approach has been developed to study initiation factor function in transfected HeLa cells. The factor under investigation is expressed as a fusion to an RNA- binding domain, in the presence of a bicistronic reporter mRNA bearing a specific binding site for the fusion protein in the intercistronic space. Specific ribosome recruitment activity is then scored as translation of the downstream cistron. This assay allows the dissection of central steps in the early initiation pathway in the context of unperturbed cellular translation. We demonstrated that the central part of eIF4G, devoid of eIF4E- and PABP-interaction domains, functions as a "ribosome recruitment core". eIF4E, but not eIF4A, is also active in this assay, in a manner that does not require an interaction with the cap structure, but depends on binding to eIF4G.

Genetic and biochemical analyses in yeast (in collaboration with the labs of Iain Mattaj and Alan Sachs) demonstrated an interaction between eIF4G and the nuclear cap-binding complex CBC and that this interaction is antagonised by eIF4E. Furthermore, we found that CBC can stimulate translation in extracts containing an eIF4G protein deficient for eIF4E binding. These data suggest that eIF4E binding to the eIF4G-CBC complex on newly exported mRNA displaces CBC, thus facilitating an exchange of nuclear for cytoplasmic factors, and that the first round of translation on mRNA may occur via a different mechanism than subsequent rounds.

Speakers * Abstracts * Home

Computational Gene Annotation of the Genome of *Drosophila melanogaster*

Martin G. Reese
Director of Informatics Discovery
ValiGen
Tour Neptune
98086 Paris-La-Defense
FRANCE
telephone: 33 (0)1 47 67 66 00
fax: 33 (0)1 47 67 66 00
email: reese@valigen.net
prestype: Platform
presenter: Martin G. Reese

The DNA sequence of the euchromatic portion of *Drosophila melanogaster* has been determined to 98.2% completion. An initial analysis and preliminary gene annotation and interpretation have been performed. The genome encodes a total of ~13,600 genes. I will present the annotation process focusing on the evaluation of genome annotation methods in the GASP (Genome Annotation Assessment Project; http://www.fruitfly.org/GASP) experiment. The results of this experiment are essential to fully understand the quality of the final gene predictions.

At least 30% of the initial predicted proteome in *Drosophila* is based solely on gene finding predictions. Many of the remaining genes were fully refined using the computer program Genie. An introduction into Genie, a generalized hidden Markov model, will be given and the strength and weaknesses will be described. Results of a comparison of the 13,187 Genie predictions against the final hand curated complete set of 13,601 annotated genes will be discussed.

Speakers * Abstracts * Home

GXD: Integrated Access to Gene Expression Data from the Laboratory Mouse

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

D.A. Begley, D.P. Hill, J.S. Beal, J.P. Corradi, G.L. Davis, J.T. Eppig, J.A. Kadin, I.J. McCright, 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. Designed as an open-ended system that can integrate many different types of expression data, such as RNA in situ hybridization, immunohistochemistry, Northern blot, Western blot, RT-PCR, and microarray data, GXD aims to provide 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. Expression patterns are described in standardized ways using an extensive dictionary of anatomical terms that has been established in collaboration with the Mouse 3D atlas project in Edinburgh, UK*, and database records are linked to digitized images of original expression data.

GXD is available at http://www.informatics.jax.org/. The database is updated and new expression data are made available on a daily basis. Data are acquired from the literature by curation staff and, increasingly, via electronic submission from laboratories. We have developed the Gene Expression Notebook, a user-friendly tool to manage expression data in the laboratory and to submit data for inclusion in GXD. GXD then places the expression data in the larger biological context, thereby providing a comprehensive framework for data storage and analysis.

GXD is integrated with the Mouse Genome Database (MGD) to enable a combined analysis of genotype, expression, and phenotype data. In collaboration with Flybase, the Saccharomyces Genome Database, and MGD we continue to build shared controlled vocabularies to describe biological processes, molecular functions and cellular components, and to assign those terms to genes and their products. These classification schemes and skilled data curation provide important new search parameters for expression data. Extensive interconnections with sequence databases and with databases from other species further extend GXD's utility for analysis of gene expression information.

*Edinburgh collaborators: J. Bard, R. Baldock, D. Davidson, M.Kaufman
GXD is supported by NIH grant HD33745.

Speakers * Abstracts * Home

Gene Targeting By Homologous Recombination in *Drosophila*

Yikang Rong
University of Utah
Department of Biology
257 S. 1400 E. Rm. 201
Salt Lake City, UT 84112 USA
telephone: (801)585-5208
fax: (801)581-4668
email: goliclab@bioscience.utah.edu
prestype: Platform
presenter: Yikang Rong

Yikang Rong and Kent G. Golic
Department of Biology, University of Utah, Salt Lake City, UT 84112

We have developed a method of homologous gene targeting in *Drosophila*. By using a pair of yeast enzymes involved in DNA metabolism, we were able to efficiently generate, in the *Drosophila* germ cells, a donor DNA molecule for homologous recombination which had a double strand break within the gene we wanted to target. In an attempt to repair this break, the cellular machinery pairs the extrachromosomal donor with the endogenous locus and carries out homologous recombination, leading to the integration of the donor at the target locus. Targeting was quite efficient at the yellow locus with an estimated frequency of one event in 500 female gametes. Recently, we have used a modified targeting scheme to recover the first directed knock-out in *Drosophila*, making it clear that this technique will be generally useful. It enables us to take full advantage of the newly sequenced *Drosophila* genome by allowing us to generate mutations in essentially any gene starting with only the DNA sequence of that gene and without relying on prior knowledge of the mutant phenotype.

Speakers * Abstracts * Home

Gene Functions on the X-Chromosome of *Drosophila melanogaster*

Dr. Ulrich Schäfer
Max-Planck-Institut für biophysikal. Chemie
Abt. Molekulare Entwicklungsbiologie
Am Fassberg 11
D-37077 Göttingen Germany
telephone: (0551) 201-1798
fax: (0551) 201-1755
email: uschaef@gwdg.de
prestype: Platform
presenter: Dr. Ulrich Schäfer

Ulrich Schäfer, Annette Peter, Petra Schöttler, Meike Werner, Nicole Beinert, Gordon Dowe, Lore Dentzer, Foteini Mourkioti, Peter Burkert and Herbert Jäckle
Max-Planck-Institut für biophysikalische Chemie, Abt. für Molekulare Entwicklungsbiologie, D-37070 Goettingen

The genome of *Drosophila melanogaster* has recently been completely sequenced. Only 10% of the predicted 13,600 genes in this well studied model organism were up to now characterized at the functional level. To provide the tools for a functional analysis we performed a systematic screen in which vital genes of the *Drosophila* X-chromosome were tagged by the enhancer trap vector P-lacW. Towards this aim, we have so far screened over 50,000 individual females containing a new site of a P-lacW insertion in the genome. Out of 38,066 vials with progeny 501 lines were kept as l(1)G### (1.3%). In addition, we isolated 74 viable lines with a mutant phenotype from the same pool (0.2%). Sequences flanking the insertion site have been isolated from almost all the lines. BLAST searches have in most cases identified the affected gene. We found insertions in well analyzed *Drosophila* genes (e. g. Notch, armadillo), in known genes without a previously described mutation (e. g. dWnt-5, innexin-2) as well as in genes defined by a mutation where a molecular analysis has not been performed yet (e. g. trol, stardust). More importantly, there are many insertions in genes predicted by the Drosophila Genome Projects that may or may not have orthologues in other genomes.

It is known that over 60% of all the genes will not produce an easily scorable phenotype when mutated and thus will not be detected by a loss-of-function screen. Hence, we are complementing our ongoing work with a gain-of-function screen employing a specific P vector. We want to saturate the chromosome with insertions of P elements containing single upstream activating sequence (UAS). The UAS sites are oriented to transcribe flanking genomic sequences in response to a transgene-derived yeast transcription factor (GAL4) expressed in a tissue- and stage-dependent manner. This way we force GAL4-dependent misexpression, overexpression or even antisense expression of the tagged gene to cause mutant phenotypes that reveal "phenotypically silent" transcription units in addition to those carrying vital functions. The first results of this approach will be presented. Our collection of X chromosomal lethal lines has already proven to be a valuable tool for the Drosophila community. Based on the high degree of sequence conservation between human and fly genes one can expect that the data generated will also have an impact on the functional analysis of the human genome.

Supported by grants of the Deutsches Humangenomprojekt to H. J.

Speakers * Abstracts * Home

Reverse Engineering of Gene Regulatory Networks: A Finite State Linear Model

Thomas Schlitt and Alvis Brazma
European Bioinformatics Institute
EMBL outstation Hinxton, Wellcome Trust
Genome Campus,
Hinxton, Cambridge CB10 1SD, UK
telephone:
fax:
email: schlitt@ebi.ac.uk
prestype: Poster
presenter: Thomas Schlitt and Alvis Brazma

We propose a new model for describing gene regulatory networks that can capture discrete (boolean) and continuous (differential) aspects of gene regulation. We are giving some illustrations of the model, we study the problem of the reverse engineering of such networks, i.e., how to construct a network from gene expression data. We prove that for our model there exists an algorithm finding a network compatible with the given data. An informal extension of the model can be used as a basis for a language for depicting real-world gene regulatory networks elements. We demonstrate such semi-formal language extension by describing lambda-phage decision making circuit.

Speakers * Abstracts * Home

KIAA0537, A New Tumor Suppressor Gene Candidate, is Inactivated in Human Brain Tumors

Mathias Schmid, M.D.
University of Ulm
Internal Medicine III
Robert-Koch Strasse 8
89081 Ulm, Germany
telephone: 0049-731-502-4396
fax: 0049-731-502-4393
email: mathias.schmid@medizin.uni-ulm.de
prestype: Poster
presenter: Mathias Schmid

Mathias Schmid, Sylvia Linortner and Hartmut Döhner
University of Ulm, Internal Medicine III, Robert-Koch Strasse 8, 89081 Ulm, Germany

Next to deletion and mutation, hypermethyltion of CpG-dinucleotides in the promoter area is another mechanism of gene inactivation. Using the SAGE database (http://www.ncbi.nlm.nih.gov/SAGE), we identified KIAA0537 as a gene with a CpG-rich promoter that is predominantly expressed in normal but not in malignant brain tissue. Multiple tissue northern blot experiments showed that KIAA0537 is also expressed in heart and lung. It belongs to the family of serine/threonine kinases and has a highly conserved catalytic domain. Phylogenetic analysis revealed a high homology to a C.elegans protein with so unknown function and to the human cdc25C phosphatase, a gene involved in cell cycle control. In four matched RNA glioma specimens (normal vs. tumor tissue from the same patient), KIAA0537 RNA was found exclusively in the normal but not in the tumor tissue. Furthermore, gene expression could be induced by the demethylating agent 5´-aza- 2´deoxycytidine in 2/5 glioma cell lines and none of the lines had a deletion at the KIAA0537 gene locus. Inhibition of histone deacetylase by Trichostatin A had no effect on KIAA0537 RNA expression. Finally, the full length KIAA0537 cDNA was cloned into a mifepristone-regulated mammalian expression vector and transfected into the human glioma cell line GMS-10. Overexpression of the KIAA0537 protein induced cell cycle arrest in G2/M phase. These data suggest that KIAA0537 might be a new tumor suppressor gene candidate that is frequently inactivated in human gliomas. In addition, our work puts emphasis on the importance of publicly accessible databases.

This work was supported by a grant from the Deutsche Krebshilfe (Schm 1083-2/1) to MS

Speakers * Abstracts * Home

Identification of Novel Genes Involved in Physiological Networks Using Differential Display RT-PCR

Michel Simonneau
INSERM E9935 Hôpital Robert Debré 48 Boulevard Sérurier 75019 Paris, France
telephone: +33 1 40 03 19 23
fax: + 33 1 40 03 19 03
email: simonaeu@infobiogen.fr
prestype: Platform
presenter: Michel Simonneau

Michel Simonneau, Fabien Guimiot, Christophe Mas, Hassan Al Halabiah, Jean-Marie Moalic, Béatrice Levacher, Francine Bourgeois
Neurogénétique, INSERM E9935, Hôpital Robert Debré, 48 Boulevard Sérurier 75019 Paris, France

Although the number and level of expression of the entire set of genes expressed in a given eukaryotic cell type can be assessed by high-thoughput differential gene-expression technologies (see for instance, Velculescu et al., 1999), the identity of a large number of the transcripts remains unknown.

Our group is working on the identification of novel genes involved in proliferation and migration of stem cells in the embryonic future cortex (dorsal telencephalon), which occur at early stages of embryonic development (embryonic days 10, in the mouse). To identify such genes, we used differential display RT-PCR, which can detect rare transcripts and generate fingerprints of mRNA populations.

We report three set of experiments:

> (i) analysis of genes involved in the proliferation of stem cells by comparing proliferating cells to embryonic postmitotic cells from embryonic telencephalon;
>
> (ii) analysis of genes involved in the migration of neuroblasts by comparing transcripts from transgenic embryos having mutations in genes involved in migration of neuroblasts with those from normal embryos;
>
> (iii) analysis of genes regulated by a 570 kb YAC transgene from human chromosome 21 (125F7; a line generated in E.M. Rubin's laboratory) which contains four Down syndrome critical genes (TTC3, DYRK1A, DSCR5, DSCR3).

More than half the differential transcripts identified in experiments (i) and (ii) have unknown functions. Some of them have no database match. We describe the characterisation of some of these transcripts.

As differential display RT-PCR allows multiple comparison, a developmental profile can be obtained for low-abundance transcripts which is not attainable by northern blotting. Furthermore, the sensitivity of the RT-PCR technique allows is discussed for the detection of small variations in low-abundance transcripts at the difference of microarray technologies.

Speakers * Abstracts * Home

Human Chromosome 21 Gene Annotation II: Refining Gene Structures and Expression Analysis

Dobromir Slavov
Eleanor Roosevelt Institute
1899 Gaylord Street
Denver, CO 80206 USA
telephone: 1-303-336-5684
fax: 1-303-333-8423
email: slavov@eri.uchsc.edu
prestype: Poster
presenter: Dobromir Slavov

Dobromir Slavov, Alla Rynditch and Katheleen Gardiner
Eleanor Roosevelt Institute

The 225 genes and gene models recently reported within human chromosome 21 provide an approximately complete set of candidates for relevance to Down syndrome and other chromosome 21 associated diseases. However, annotation of the genomic sequence and the genes/models is not complete and remains an ongoing process. We are focusing on the generation of complete cDNAs, the analysis of gene-specific genomic structures , and the development of expression patterns for a number of chromosome 21 genes. These aims are being pursued by: i) sequencing of ESTs, and RACE and RT-PCR products; ii) analysis of genomic sequence annotation; iii) qualitative and quantitative RT-PCR in >30 human tissues and cell lines; and iv) creation of an expression database containing dbEST- derived information on tissues of expression. Results of these investigations include: discovery of previously unannotated genes, correction of previous gene models, completion or increased length of cDNA sequences, and expression analysis of >40 genes that complements data from dbEST and identifies genes with very restricted expression. Gene- specific features of particular interest include: i) organization and expression patterns of a glutamate receptor gene and its apparent antisense RNA; ii) a gene spanning >500 kb with ubiquitous expression, comprising >8 exons and at least 4 splice variants, but as yet showing no significant open reading frame; iii) an apparently complete cDNA with multiple, short, sequential open reading frames; and iv) 5' and 3' overlapping gene structures. These data also show that thorough and complete annotation of human genomic sequence remains a time consuming and gene-specific process.

Speakers * Abstracts * Home

Gene Expression Profiling in Renal Cell Carcinoma

Dr. Holger Sültmann
Deutsches Krebsforschungszentrum
Abt. Molekulare Genomanalyse
Im Neuenheimer Feld 280
D-69120 Heidelberg
telephone: ++49-6221-424705
fax: ++49-6221-424704
email: h.sueltmann@dkfz.de
prestype: Platform
presenter: Dr. Holger Sültmann

H. Sültmann[1], W. Huber[1], J. Boer[1,4], F. Wilmer[1], L. Füzesi[3], B. Gunawan[3], S. Haas[2], A. v. Heydebreck[2], M. Vingron[2], A. Poustka[1]
[1]Abt. Molekulare Genomanalyse, [2]Theoretische Bioinformatik, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, D-69120 Heidelberg
[3]Zentrum Pathologie, Georg-August-Universität, Robert-Koch-Str. 40, D-37075 Göttingen
[4]Leiden Univ. Medical Center, Wassenarseweg 72, Leiden 2333AL, NL

Cancer cells show altered gene expression compared to normal cells. Knowledge of the changes in gene expression for certain types and stages of tumors can give insight into the molecular changes involved in tumor development and progression and provide molecular markers for tumor diagnosis and prognosis.

We use the cDNA array hybridization technology as a high throughput method to determine the expression levels of 32,000 different cDNAs spotted in duplicate onto nylon membranes. These represent known human genes and expressed sequence tags (ESTs). Normal tissue and primary tumor tissue are used to isolate poly (A)+ RNA which is reverse transcribed into 33P-labelled single stranded cDNA. The hybridization of both cDNA populations is performed on different membranes using a standardized protocol. The membranes are exposed to phosphoimage plates, and expression profiles are calculated through spotwise quantification of the signal distribution.

We have collected array expression data for 37 renal cell carcinoma samples (predominantly clear cell types) of different tumor stages and differentiation grades, and for the corresponding normal tissues of the same patients. More than 1700 genes were identified with statistical significance to be expressed at different levels between normal and tumor tissues. Among these were several genes which had been known to be differentially expressed in renal carcinoma, e.g. vimentin, VEGF, haptoglobin, metallothionein, and kininogen. This confirms the utility of the experimental approach of typing expression levels of thousands of genes simultaneously. In addition to the genes known to be associated with kidney cancer, many other genes and ESTs were found. Our data allow the definition of genes that are significantly transcribed only in certain tumor stages (e.g. in metastases). A detailed analysis of the correlation of gene expression with tumor progression is currently being performed.

The renal cell carcinoma specific genes, as well as a selection of genes which are known to have oncogenic potential in other cancer types, are being amplified by PCR and spotted on membranes or glass slides to set up a kidney tumor specific gene array. With this, we plan to conduct a further focused investigation on the differential transcription of genes in renal cell carcinoma.

A queryable database combining expression data for all genes on the array with histopathological and clinical

follow-up information for the tumor material as well as tools to mine these large data sets, are under development. This database will eventually be open for the public.

Speakers * Abstracts * Home

Towards Understanding Functional Significance of a Great Many of Solitary LTRs in the Human Genome

Eugene D. Sverdlov
Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russ. Acad. Sci
Miklukho-Maklaya 16/10
117871 Moscow Russia
telephone: +7 095 3306529
fax: +7 095 3306538
email: eds@glasnet.ru
prestype: Platform
presenter: Eugene D. Sverdlov

T. Vinogradova, L. Nikolaev, Y. Lebedev, G. Monastyrskaya, E. Sverdlov
Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russ. Acad. Sci; Miklukho-Maklaya 16/10; 117871 Moscow Russia

The human genome contains tens of thousands of remnants of ancient retroviral infections including solitary retroviral LTRs. The LTRs retain structural elements of enhancers, promoters, polyadenylation signals etc. thus being potential transcription regulators. The involvement of LTRs in transcription is usually considered in connection with neighboring genes regulation. However, many of the LTRs are localized very far from any known or predicted genes. We are interested in possible functions of such "orphan" LTRs in the genome regulatory network. Using a number of different approaches we have demonstrated that various solitary LTRs are characterized by different tissue specific abilities to initiate or terminate transcription of adjacent genomic areas. Their activity is changed in tumor tissues as compared to normal ones. The transcriptional regulation activity of evolutionary younger LTRs tends to be higher than that of older LTRs. The meaning of these findings will be discussed from the point of view of the "orphan" LTRs involvement in the genome functioning.

Speakers * Abstracts * Home

Differential Proteome Analysis of Germ-Line Maturaturation-Dependent Proteins of *C. elegans*

Prof. Dr. Hans-Juergen Thiesen
Proteome Center Rostock
Medical Faculty
University of Rostock,
Joachim-Jungius Str. 9
18059 Rostock
telephone: +49 381 4059687
fax: +49 381 4059686
email: hans-juergen.thiesen@med.uni-rostock.de
prestype: Platform
presenter: Hans-Juergen Thiesen

H.-J. Thiesen[1], M. Bantscheff[1], B. Ringel[1], R. Schnabel[2] and M.O. Glocker[1]

[1]Proteome Center Rostock, Medical Faculty, University of Rostock, Joachim-Jungius Str. 9, 18059 Rostock, Germany

[2]TU Braunschweig, Biozentrum / Genetik A, Spielmannstr. 7, 38106 Braunschweig, Germany

Caenorhabditis elegans (*C. elegans*) is a small (about 1 mm long) soil nematode found in temperate regions and may be handled as a microorganism. Although *C. elegans* is a primitive organism it nonetheless shares many of the essential biological characteristics that are central problems of human biology (1). The development and function of this diploid organism that has been fully sequenced (2-4), is encoded by an estimated 17,800 distinct genes. *C. elegans* is regarded as a model system to facilitate the purification of low abundant proteins involved in regulatory protein networks. Our initial investigations are focussed on studying germ line formation during maturation of *C. elegans* as a model system for regulatory and developmental processes in multicellular organisms. In this project we are studying protein expression patterns of the temperature-sensitive *C. elegans* mutant glp-1(e2144) (5). This mutant presents a normal phenotype when grown at 15 °C. However, when grown at 25 °C formation of germ line cells is lacking.

Differential proteome analyses (6-8) of the *C. elegans* mutant grown at different temperatures were performed and characteristic alterations in protein expression were observed by 2D-gel electrophoresis (2-DE). The protein extracts (1.5 ml) contained a total of 4.1 mg protein, each. 2D-gel electophoresis (30 x 23 cm2) was performed with 15 µl aliquots, applying high resolution conditions (9). Isoelectric focussing was carried out using a pH gradient ranging from 4 to 10. After staining with silver approx. 3000 spots were detected by computer assisted image analysis. The protein spots were excised using a picking robot. From the approx. 100 protein spots that were analyzed and identified by MALDI-TOF MS up to now, about 15 spots / spot families were found reproducibly different in abundance. While for many of them the function is still unknown, others are associated with germ cell function. The major sperm protein, for instance, was found in high abundance only in the 2-DE gels from the culture grown at 15 °C. This finding correlates well with normal germ cell development at this temperature. By contrast, in the culture grown at 25 °C vitellogenins were present in much higher abundance in the corresponding 2-DE gels than in the gels from the 15 °C culture. This can be explained by the lack of germ cells at 25 °C, as these proteins after production in the intestine are taken up by the healthy gonads (10).

Hence, mass spectrometric analyses of in-gel digested proteins and subsequent data base searches are applicable for the identification of protein candidates associated with the observed phenotypes. The obtained informations lay the grounds for detailed biochemical and cell biological investigations for studying regulatory and

developmental processes in multicellular organisms.

References

1. J.D. Hogan, J.E. Lew-Smith, A.M. Fancher, M.E. Cusick, C.A. Lingner, B.P. Davis, C. Lengieza, M. Tillberg and K.J. Roberg-Perez (1999) Worm Breeder´s Gazette 16, 1-12.
2. The *C. elegans* Sequencing Consortium (1998) Science 282, 2012-2018.
3. S. A. Chervitz et al. (1998) Science 282, 2022-2028.
4. E. Peninsi (1998) Science 282, 1972-1974.
5. R. Schnabel, H. Hutter, D. Moermann and H. Schnabel (1997) Dev. Biol. 184, 234-265.
6. L. Bini, H. Heid, S. Liberatori, G. Geier, V. Pallini, R. Zwilling (1997) Electrophoresis 18, 557-562.
7. M. Wilm, A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis and M. Mann (1996) Nature 379, 466-469.
8. M.O. Glocker, M. Bantscheff, M. Kalkum, P. Lorenz, H.-J. Thiesen and M. Przybylski (1999). Proceedings of the 47th Conference of the American Society for Mass Spectrometry, 636-637.
9. J. Klose and U. Kobalz (1995) Electrophoresis 16, 1034-1059.
10. J. Spieth, K. Denison, S. Kirtland, J. Cane and T. Blumenthal (1985). Nucleic Acid Res. 13, 5283-5295.

Speakers * Abstracts * Home

The *C. elegans* Protein Interaction Mapping Project

Marc Vidal
Dana-Farber Cancer Institute and
Department of Genetics, Harvard Medical School
Smith Bldg., 858A
44 Binney Street
Boston, MA 02115 USA
telephone: +1-617-632-5180
fax: +1-617-632-2425
email: Marc_Vidal@dfci.harvard.edu
prestype: Platform
presenter: Marc Vidal

Until recently, mostly classical genetics and biochemistry were used to investigate how organisms develop, reproduce, behave, age, etc. However, the availability of complete genome sequences suggests additional approaches into biological research. Our goal is to generate and interpret comprehensive protein-interaction maps, particularly in *C. elegans*. Early versions of protein interaction maps and other functional maps, such as those based on global expression profiling, suggest they will help our understanding of the processes that control the biology of living organisms.

Speakers * Abstracts * Home

Expression Profiler: An Integrated Tool for Gene Expression and Sequence Analysis

Jaak Vilo, Alvis Brazma
European Bioinformatics Institute EBI
EMBL Outstation
Wellcome Trust Genome Campus
Hinxton, Cambridge CB10 1SD
United Kingdom
telephone: +44 1223 494633
fax:
email: vilo@ebi.ac.uk
prestype: Poster
presenter: Jaak Vilo

Jaak Vilo, Alvis Brazma, Alan Robinson
European Bioinformatics Institute EBI

Analysis suites for the analysis of vast amounts of microarray gene expression data are becoming more popular as a part of many initiatives of public microarray expression databases. The integration of the databases and analysis and visualization tools will become the neccessity for all major public databases.

We are developing a set of Internet tools called collectively Expression Profiler (see http://www.ebi.ac.uk/microarray/) that will allow users to browse and query microarray data stored in a microarray database ArrayExpress at EBI as well as from other databases on the web [1]. The main challenge is the integration of different types of data and presentation of these data in useful form for biologists to perform the analysis and study the complex relationships in the data. Expression Profiler consists of four major components.

EPCLUST, the expression profile clustering and analysis tool allows users to perform cluster analysis and visualization of expression data. Main methods for analysis include the hierarchical and K-means clustering with different distance measures and clustering parameters (choice of hierarchical clustering method, choice of starting points for K-means etc.). Visualization of expression profiles is based on methods developed by Mike Eisen [2].

The cluster-analysis of gene expression data is only the first step of many. GENOMES, the tool for retrieving the gene annotations and sequences (e.g. upstream regulatory sequences), fulfills the need for species-specific databases to be able to handle queries for sets of genes simultaneously, as users would like to have "executive summaries" about the clusters they have discovered from the expression data.

Due to large numbers of tools and analysis methods on-line, we have developed a "middle layer", an URLMAP tool, which allows to send data about cluster contents to various other databases, as for example to ask "in which pathways are the genes from my cluster participating in?". Ideally each on-line database should provide mechanisms to pre-fill-in the contents of cluster contents to their tools.

The discovery of putative transcription factor binding sites [3] is an application of microarray expression data. We have developed a pattern discovery tool SPEXS that is able to perform a rapid exhaustive search for a priori unknown statistically significant sequence patterns of unrestricted length. The statistical significance is determined for a set of sequences in each cluster in respect to a set of background sequences allowing the detection of subtle regulatory signals specific for each cluster in comparison to the background distribution.

These four tools form currently the core of the Expression Profiler analysis suite. The tools can be used in the manner facilitating an automatic discovery of potential regulatory signals in genomes as described in [3].

Integration of data from high-throughput microarray experiments with other types of data, for example the genomic sequences, protein-protein interactions, metabolic pathways and signaling data will open new ways for genomic-scale analysis methods. We are exploring the development of new techniques as well as technology transfer from marketing and telecommunications domains, e.g. application of visualisation, data mining and statistical analysis. Technologies such as CORBA and XML provide ways to interact with other tools and databases over the Internet. This will allow external users to use the data stored at EBI databases or integrate the analysis tools developed at EBI with their tools.

References

[1] A. Brazma, A. Robinson, G. Cameron and M. Ashburner One stop shop for microarray data. Nature 2000 vol. 403 (pp. 699-700)

[2] M. Eisen, P. T. Spellman, D. Botstein and P. O. Brown Cluster Analysis and Display of Genome-wide Expression Patterns. PNAS 1998 volume 95 (pp. 14863--14867)

[3] J. Vilo, A. Brazma, I. Jonassen, A. Robinson, E. Ukkonen Mining for Putative Regulatory Elements in the Yeast Genome Using Gene Expression Data. ISMB-2000 August 2000. AAAI press. (pp. 384--394)

Speakers * Abstracts * Home

Searching for the Protein Coding Genes in the Human Genome Sequence

Jean Weissenbach
Genoscope
91057, Evry Cedex France
telephone: 33 1 60 87 25 02
fax: 33 1 60 87 25 32
email: jsbach@genoscope.cns.fr
prestype: Platform
presenter: Jean Weissenbach

Jean Weissenbach, Hugues Roest Crollius, Olivier Jaillon, Alain Bernot, Lucie Friedlander,Abel Ureta Vidal, Gabor Gyapay, William Saurin
Genoscope and CNRS-FRE 2231, Evry, France

Despite the availability of most of the human genome sequence, the identification of genes on the DNA sequence remains a difficult task. We have built a search tool (called Exofish, for Exon FInding by Sequence Homology) that combines a specific setting of TBLASTX, output filtering and a collection of random DNA sequences reads representing at present a third of the genome of the pufferfish *Tetraodon nigroviridis* (closely related to Fugu). Exofish detects sequence matches in 2/3 of several sets of human genes with a backround of false positive matches below 1%.

Exofish has been successively applied to the December 99 and June 2000 versions of the "working draft" of the human genome. The latter analysis indicates that the protein coding gene number is now around 27,000-28,000, somewhat below our earlier estimates of 28,000-34,000.

Exofish analysis of the Unigene set of human ETSs indicates that about 50% of the coding fraction of the human genome is still missing in the public sequence databanks.

About 15% of the total number of exons detected by Exofish on human chromosome 22 fall outside annotated genes. A more detailed analysis of this annotation using new full length cDNA sequences, suggests however, that (1) most of the Exofish detected exons falling outside annotations actually belong to annotated genes, (2) many of the annotated genes are not yet accurately delimitated and (3) a number of these genes will merge together.

All these observations indicate that a valuable annotation of the human genome sequence still requires enlarged sets of additional sequence data (cDNAs, related genomes) for comparison purposes. In addition, since any sequence analysis method suffers some limitations, it is essential to rely on a panel of tools that are as diverse as possible.

Speakers * Abstracts * Home

Generation of Full Length cDNA Libraries in the Course of the German Human Genome Project

Dr. Ruth Wellenreuther
German Cancer Research Center (DKFZ)
Division of Molecular Genome Analysis
Im Neuenheimer Feld 506
69120 Heidelberg
telephone: ++49-6221-42 4702
fax: ++49-6221-42 4704
email: r.wellenreuther@dkfz.de
prestype: Platform
presenter: Ruth Wellenreuther

Ruth Wellenreuther, Stefan Wiemann, Daniela Heiss, Annemarie Poustka
Department of Molecular Genome Analysis, German Cancer Research Center, Im Neuenheimer Feld 280/506, 69120 Heidelberg, Germany

We generate human cDNA libraries that are enriched in full-length clones i.e. from the translation start to the polyA 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 known genes.

Libraries are created by directional cloning of cDNAs into plasmid vectors. For first strand cDNA synthesis, both conventional and 5'-selective methods are used.

In the conventional cloning strategy, intensive size selection of cDNAs before cloning reduces entry of incomplete cDNAs and complete small cDNAs (which are already overrepresented in existing libraries) into the libraries.

As 5'-selective method, we use Clontech's SMART technology for first strand synthesis. 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 sole source for the cDNA sequencing effort of the German Human Genome Project that is carried out by a consortium of nine laboratories. Over 1,800 full-length cDNAs could be identified by the consortium within the last two years.

The libraries generated in this project are available through the Resource Center of the German Human Genome Project.

Speakers * Abstracts * Home

In Silico Prediction of Matrix Attachment Regions in Large Genomic Sequences

Thomas Werner
Genomatix Software
Karlstr. 55
D-80333 München
telephone: +49-89-5490839-0
fax: +49-89-5490839-9
email: genomatix@gsf.de
prestype: Platform
presenter: Dr. Thomas Werner

Matthias Frisch[1], Kornelie Frech[1], Andreas Klingenhoff[1], Kerstin Quandt[1], Ines Liebich[2], and Thomas Werner[1,3]

[1]Genomatix Software GmbH, Karlstr. 55, D-80333 München, Germany.
[2] Research Group Bioinformatics, Gesellschaft für Biotechnologische Forschung mbH, Mascheroder Weg 1, D-38124 Braunschweig, Germany.
[3] Institute of Experimental Genetics, GSF-National Research Center for Environment and Health, Ingolstädter Landstrasse 1, D-85764 Neuherberg, Germany.

Matrix attachment regions (MARs) are essential regulatory DNA elements of eukaryotic cells. They are major determinants of locus control of expression and shield gene expression from position effects. Experimental detection of MARs requires substantial efforts not suitable for large-scale screening of genomic sequences. In silico prediction of MARs can provide a crucial first selection step to reduce the amount of candidates. We used 34 experimentally defined MARs as training set and generated a library of 97 MAR-associated, AT-rich patterns described as weight matrices. We developed a new tool, SMARTest, identifying potential MARs in genomic sequences. SMARTest carries out a density analysis based on the MAR matrix library. The SMARTest approach does not depend on the sequence context and is suitable to analyse long genomic sequences up to the size of whole chromosomes on a workstation. To demonstrate the feasibility of large-scale MAR prediction we analysed the recently published chromosome 22 sequence and found 1198 MAR candidates.

Speakers * Abstracts * Home

Sequencing and Analysis of Full Length cDNAs in the German cDNA Network

Stefan Wiemann
German Cancer Research Center
Dept. Molecular Genome Analysis
Im Neuenheimer Feld 506
69120 Heidelberg, Germany
telephone: +49 6221 42 4702
fax: +49 6221 42 4704
email: s.wiemann@dkfz.de
prestype: Platform
presenter: Stefan Wiemann

Stefan Wiemann[1], Bernd Weil, Ruth Wellenreuther, Sabine Krieger, Wilhelm Ansorge, Michael Böcher, Helmut Blöcker, Helmut Blum, Andreas Düsterhöft, Jürgen Lauber, Andreas Beyer, Karl Köhrer, Christian Gruber, Hans-Werner Mewes, Brigitte Obermaier, Birgit Ottenwälder, Dagmar Heubner, Rolf Wambutt, Jeremy Simpson, Rainer Pepperkok, Annemarie Poustka

[1]Molecular Genome Analysis, German Cancer Research Center, Im Neuenheimer Feld 506, D-69120 Heidelberg Germany, and the German cDNA Network

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 3,500 cDNAs (> 8.6 Mb) have been sequenced since.

We use the set of fully sequenced clones in combination with the EST-sequenced clones to generate a minimal set of full-length clones for employment in subsequent functional analysis. All sequences are first analyzed for possible protein function in silico. To systematically characterize function of the encoded proteins in vivo, we initially determine the subcellular localization of the proteins. A progress report of the network activities and the achievements will be presented.

In future projects cellular assays will be applied to comprehensively unravel the role of most or all proteins in view of their cellular function(s), the pathways they are involved in and possible disease relations.

Speakers * Abstracts * Home

Molecular Genetics and Evolution of Ultraviolet Vision in Vertebrates

Shozo Yokoyama
Department of Biology
Syracuse University
130 College Place
Syracuse, New York 13244 USA
telephone: 315-443-9166
fax: 315-332-2012
email: syokoyam@mailbox.syr.edu
prestype: Platform
presenter: Shozo Yokoyama

In many fish, bird, amphibian, reptilian, and mammalian species, ultraviolet (UV) vision has profound effects on the evolution of organisms by affecting such behaviors as mating preference and foraging strategies. UV vision is determined by visual pigments, each of which consists of the chromophore, 11-cis-retinal, and a transmembrane protein, an opsin. The UV pigments have the maximum wavelengths of absorption (lambda-max) at around 360 nm, whereas evolutionarily closely related violet (or blue) pigments have lmax values at 390- 420 nm.

We have successfully interchanged the color-sensitivities of the mouse ultraviolet pigment (lambda-max = 358 nm) and the human blue pigment (lambda-max = 414 nm) by introducing forward and reverse mutations at five sites. This unveils for the first time the general mechanism of ultraviolet vision. Most contemporary ultraviolet pigments in vertebrates have maintained their ancestral functions by accumulating no more than one of the five specific amino acid changes. The avian lineage is an exception, where the ancestral pigment lost ultraviolet-sensitivity but some descendants regained it by one amino acid replacement at another site.

Speakers * Abstracts * Home
