Beyond the Identification of Transcribed Sequences: Functional, Evolutionary and Expression Analysis 12th International Workshop

October 25-28, 2002
Washington, DC

Sponsored by the U.S. Department of Energy

Home * List of Abstracts * Speakers

Meeting Objective

The 12th workshop in this series was held in Washington D.C., Friday evening, October 25 through Monday afternoon, October 28, 2002. Interested investigators actively engaged in any aspect of the functional, expression or evolutionary analysis of transcribed sequences were invited to send an abstract.

Topics 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 and genomic sequence analysis in model organisms (e.g. yeast, *C.elegans*, *drosophila*, zebrafish, pufferfish, chicken, mouse rat); construction and analysis of transgenic organisms; evolutionary comparisons; novel approaches for functional analysis of transcribed sequences; construction of full length and 5' specific cDNA libraries; cDNA array screening and analysis; RNA processing, including RNA editing, RNAi, antisense RNA and regulatory sequences; database construction, management and use in expression and functional analysis; and 2D-gel and mass spectrometry proteomics, analysis of protein structure, function and modification

Program Planning Committee:

- Tom Freeman, MRC Human Genome Programme, Hinxton, UK
- Katheleen Gardiner, Eleanor Roosevelt Institute, Denver, CO, USA
- Bernhard Korn, German Cancer Research Center, Heidelberg, Germany
- Blair Hedges, Penn State University, University Park, PA, USA
- Sherman Weissman, Yale University, New Haven, CT, USA
- Thomas Werner, Institute for Saeugertiergenetik, Oberschleissheim, Germany

For Questions or Additional Information, contact:

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

Nan Matthews, bits@themattgrp.com
303-843-6414

2002 Confirmed Speakers

Yoav Arava
Stanford University
Palo Alto, CA USA

Michael Bittner
NHGRI/NIH
Bethesda, MD USA

Tom Blumenthal
Univ of Colorado Health Sciences Center
Denver, CO USA

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

Alvis Brazma
European Bioinformatics Institute
HInxton, UK

Olof Emanuelsson
Stockholm Bioinformatics Center
Stockholm, Sweden

James Fickett
AstraZeneca
Boston, MA USA

Thomas Freeman
MRC Human Genome Mapping Programme
Hinxton, Cambridge UK

Katheleen Gardiner
Eleanor Roosevelt Institute
Denver, CO USA

Mark Gerstein
Yale University
New Haven, CT USA

Tom Gingeras
Affymetrix
USA

Takashi Gojobori
National Institute of Genetics
Mishima, Japan

Bernhard Korn
German Cancer Research Institute
Heidelberg, Germany

Sudhir Kumar
Arizona State University
Tempe, AZ USA

Michael Liebler
Raytest GmbH
Straubenhardt, Germany

Lynne Maquat
University of Rochester Medical School
Rochester, NY USA

Osamu Ohara
Kazusa DNA Research Institute
Kisara, Japan

Graziano Pesole
Universita di Milano
Milan, Italy

Arvind Raghavan
University of Minnesota
Minneapolis, MN USA

Martin Ringwald
The Jackson Laboratory
Bar Harbor, ME USA

Lee Rowen
Institute of Systems Biology
Seattle, WA USA

Alla Rynditch
National Academy of Sciences of Ukraine
Kiev, Ukraine

Naruya Saitou
National Institute of Genetics
Mishima, Japan

Michael Shi
Sequenom Inc.
San Diego, CA USA

Yoram Groner
The Weizmann Institute
Rehovot, Israel

Xun Gu
Iowa State University
USA

Blair Hedges
Penn State University
University Park, PA USA

Matthias Hentze
EMBL
Heidelberg, Germany

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

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

Winston Hide
South African National Bioinformatics Inst
Belleville, South Africa

Nancy Hopkins
MIT
Boston, MA USA

Sorin Istrail
Celera Genomics
Rockville, MD USA

Kirk Jensen
Rockefeller University
New York, NY USA

Thomas Joos
University of Tuebingen, NMI
Reutlingen, Germany

Jeremy Simpson
EMBL
Heidelberg, Germany

Michael Snyder
Yale University
New Haven, CT USA

Stefan Stamm
University of Erlangen-Nuremberg
Erlangen, Germany

Eugene Sverdlov
Russian Academy of Sciences
Moscow, Russia

Scott Tenebaum
Duke University Medical Center
Research Triangle Park, NC USA

Harald von Melchner
University of Frankfurt Medical School
Frankfurt, Germany

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

Ruth Wellenreuther
German Cancer Research Center
Heidelberg, Germany

Thomas Werner
Institute for Saeugertiergenetik
Oberschleissheim, Germany

Jinghua Yang
Yale University School of Medicine
New Haven, CT USA

Home * List of Abstracts * Speakers

List of Abstracts

AstraZeneca R&D Boston

- **Mapping the transcriptome**
  Thomas Freeman
  HGMP-Resource Centre
- **Computational proteomics: Genome-scale analysis of protein structure, function, & evolution**
  Mark Gerstein, P Harrison, J Qian, R Jansen, V Alexandrov, P Bertone, R Das, D Greenbaum, W Krebs, Y Liu, H Hegyi, N Echols, J Lin, C Wilson, A Drawid, Z Zhang, Y Kluger, N Lan, N Luscombe, and S Balasubramanian
  Yale University
- **Large scale transcriptional activity observed in chromosomes 21 and 22**
  Thomas Gingeras
  Affymetrix
- **Functional genomics of Runx3 and DRG neurogenesis**
  Yoram Groner
  Dept of Molecular Genetics The Weizmann Institute of Science
- **The natural history of human gene families: Genome-wide duplication(s) v.s. small-scale duplications**
  Xun Gu
  Iowa State University
- **Multiprotein evolutionary analyses of diverse eukaryotes**
  S. Blair Hedges
  Pennsylvania State University
- **Translational control at the interface between the genome and the proteome**
  Matthias W. Hentze
  EMBL
- **3' untranslated regions and selenocysteine incorporation: From modulation of translation to a novel single nucleotide polymorphism**
  Stephane Villette, G. Bermano, John R. Arthur and **John E. Hesketh**
  University of Newcastle
- **S-Gal™: An autoclavable, water-soluble dye for enhanced color-selection of cloned DNA inserts**
  Sara Borchardt and **Ken Heuermann**
  Sigma-Aldrich Corporation
- **Development of a novel high throughput directional cloning platform**
  Ron Hernan, Leanne Snyder, **Ken Heuermann**, and Brian Ward
  Sigma-Aldrich Corporation
- **Diversity in gene expression: Assessment of exon skipping and expression states**
  Winston Hide, Tzu-Ming Chern, Janet Kelso and Vladimir Babenko
  University Western Cape
- **Insertional mutagenesis in zebrafish identifies a diversity of genes required for vertebrate development**
  Nancy Hopkins
  Massachusetts Institute of Technology
- **From large scale mouse mutagenesis to gene function**
  Martin Hrabe de Angelis
  GSF Research Center for Environment and Health
- **Inferring gene transcription networks: The Davidson model**
  Sorin Istrail
  Celera Genomics/Applied Biosystems
- **Direct contribution of a transposable element to the protein diversity: A novel type bovine Bcnt**

protein that includes the endonuclease domain of RTE-1
Shintaro Iwashita, Tomohito Itoh, Kenshiro Oshima, Katsuyuki Hashimoto and Wojciech Makalowski
Mitsubishi Kagaku Institute of Life Sciences (MITILS)

- **Cross-species comparison of known and predicted GDF-9 and GDF-9B protein sequences**
  Luke Jeffery, David Mottershead & **Olli Ritvos**
  University of Helsinki
- **Defining functional networks of RNA-protein interactions in the nervous system**
  Kirk Jensen
  The Rockefeller University
- **Protein microarray technology**
  Thomas Joos
  Natural and Medical Sciences Institute at the University of Tubingen
- **RNA-binding proteins as reporters of operational mRNA networks in tumors and complex tissues**
  Jack Keene, Luiz O. Penalva, Michael Burdick, Hedi Sutterluety and Patrick J. Lager
  Duke University Medical Center
- **Systematic functional analysis of mouse genes in the region syntenic to human Xq28**
  Anja A. Kolb, Hanna Bausbacher, Annemarie Poustka
  German cancer research center (DKFZ)
- **Resources for functional research: Repository and application**
  Bernhard Korn
  RZPD
- **RNAs isolated from ejaculate spermatozoa provide a noninvasive means to investigate testicular gene expression**
  G. Charles Ostermeier, Robert J. Goodrich, Kary Thompson, John Rocket, Michael P. Diamond, Karen Collins, NICHD Reproductive Medicine Network, David J. Dix, David Miller and **Stephen A. Krawetz**
  Wayne State University
- **BEST: A computational approach for comparing gene expression patterns from early stages of *Drosophila melanogaster* Development**
  Sudhir Kumar
  Center for Evolutionary & Functional Genomics
- **Crucial elements in array data evaluation**
  Michael Liebler and Till Sorensen
  raytest GmbH
- **Modeling and predictions of estrogen response element sequences in the human genome reveal potential roles for Alu repeats in motif propagation and transcriptional regulation**
  Chin-Yo Lin
  Genome Institute of Singapore
- **The next step in transcriptomics: Data mining after global array analysis and design of a sub-array specific for cardiac fibrosis of renal failure in the rat model**
  Christian Maercker, Heidrun Ridinger, Effi Rees, Matthias Schick, Christiane Rutenberg, Eberhard Ritz, Gerhard Mall, Bernhard Korn, and Kerstin Amann
  Resource Center for Genome Research (RZPD)
- **Nonsense-mediated mRNA decay in mammalian cells: Evidence for a "pioneer" round of mRNA translation**
  Lynne Maquat
  University of Rochester Medical Center
- **Biological roles for A to I conversion in untranslated regions of mRNAs?**
  Daniel P. Morse, P. Joseph Aruscavage, Leath A. Tonkin, Lisa Saccomanno, Thomas Brodigan, Michael Krause, and **Brenda L. Bass**

United States Naval Academy

- **Kazusa cDNA project 2002: Of mice and men**
  Osamu Ohara
  Kazusa DNA Research Institute
- **Genome-wide analysis of the sequence region surrounding the transcription start site of human mRNAs**
  Michele Iacono, Carmela Gissi, Giorgio Grillo*, Flavio Mignone, Sabino Liuni and **Graziano Pesole**
  Universita di Milano
- **Global analysis of mRNA decay in T cells**
  Arvind Raghavan and Paul R. Bohjanen
  University of Minnesota
- **The Mouse Gene Expression Database (GXD) and the Gene Expression Notebook (GEN)**
  C.M. Smith, D.A. Begley, I.J. McCright, T.F. Hayamizu, D.P. Hill, S.R. Pretel, P. Mani, J.S. Beal, D.J. Dahmen, J.A. Blake, J.T. Eppig, J.A. Kadin, J.E. Richardson, and **M. Ringwald**
  The Jackson Laboratory
- **Big genes: What? Where? Why?**
  Lee Rowen, Brian Birditt, Ryan Nesbitt, Amardeep Kaur, Gustavo Glusman, and Lee Hood
  Institute for Systems Biology
- **Full domain transcript of alpha-globin genes is a component of the nuclear matrix**
  A.V. Rynditch, V.P. Smalko, S.V. Razin, E. Loudinkova, K. Scherrer
  National Academy of Sciences of the Ukraine
- **Molecular evolutionary analyses of HoxA gene cluster of human and apes**
  Naruya Saitou
  National Institute of Genetics
- **Point mutations occur at a faster rate in exons than in noncoding DNA in primate genomes**
  Subashchandran Sankarasubramanian and Sudhir Kumar
  Arizona State University
- **Two pancreas clone sets and PancChips, new resources for diabetes**
  Marie Scearce and Klaus Kaestner
  University of Pennsylvania
- **Applying genome-wide SNP analysis for complex disease gene discovery**
  Michael Shi
  Sequenom, Inc.
- **Large-scale subcellular localisation and functional analysis of novel proteins**
  Jeremy C. Simpson, Urban I. Liebel, Vytaute Starkuviene, Holger Erfle,? Annemarie Poustka, Stefan Wiemann and Rainer Pepperkok
  EMBL
- **Statistical validation of pattern prevalence in regulatory gene loci**
  Gautam Singh and **Stephen A. Krawetz?**
  Oakland University
- **Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae***
  Christine E. Horak, Nicholas M. Luscombe, Jiang Qian, Stacy Piccirrillo, Mark Gerstein and **Michael Snyder**
  Yale University
- **Regulated alternative splicing: A new dimension of the human genome**
  Stefan Stamm
  Institute of Biochemistry
- **Genome-wide identification of differences in the retroelement integrations between humans and chimpanzees**

E. Sverdlov, A. Buzdin, I. Mamedov, S. Ustugova, E. Gogvadze, Yu.
Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences

- **Array analysis of mRNP complexes reveals clustered mRNA subsets that may represent posttranscriptional operons**
  Scott Tenenbaum and Jack D. Keene
  Duke University Medical Center
- **Functional analysis of the mammalian genome by large scale gene trap mutagenesis**
  Harald von Melchner
  Laboratory for Molecular Hematology, University of Frankfurt am Main
- **Fine-tuning the clock — the relationship between variability in Per3 and circadian phenotype in mammals**
  Malcolm von Schantz, Donna Robilliard, Oliver Wood, Marcel Smits, Adrian Williams, Debra J Skene, Jo Arendt, and Simon N Archer
  University of Surrey
- **Status of application of tiling genomic fragment arrays to molecular biology of mammalian cells**
  C. Horak, X-H Pan, J. Rimm, R. Goetsch, P. Bertone, P. Miller, N. Luscombe, P. Harrison, M. Gerstein, **S.M. Weissman**, and M. Snyder
  Yale University School of Medicine
- **The German cDNA network: cDNAs for functional genomics and proteomics**
  Ruth Wellenreuther, Stephanie Bechtel, Petra Moosmayer, Regina Albert, Ingo Schupp, Vladimir Kuryshev, Detlev Bannasch, Dorit Arlt, Ulrike Korf, Rainer Pepperkok, Jeremy Simpson, Annemarie Poustka, Stefan Wiemann, and the German cDNA Consortium
  German Cancer Research Center
- **Intracellular trafficking of a few inflammation-inducible ADAR1 isoforms**
  Jing-HuaYang, Qingchuan Zhao, Yongzhan Nie, Yingjun Su, Koroush Kibir and Reuven Rabinovici
  Yale University School of Medicine
- **Experimental verification of predicted splice variants of human genes**
  Dorothea Zink, Stefan Haas, Oliver Heil, Mathias Schick, Melanie Roth, Eivind Coward, Bernhard Korn, and Martin Vingron
  RZPD - Ressourcenzentrum fur Genomforschung

Global Analysis of Translation in Yeast

Yoav Arava[1], Yulei Wang[1], John D. Storey[2], Patrick O. Brown[1,3] and Daniel Herschlag[1]
Departments of Biochemistry [1], Statistics [2] and HHMI [3], Stanford University

We have performed a global analysis of translation in yeast to better understand the global gene expression program and to obtain new insights into translation mechanisms. Quantitative microarray analysis of mRNAs from fractions across sucrose gradients has allowed us to assemble a database describing ribosome association for thousands of genes.

Mining the ribosomes association database revealed that, for nearly all genes, the majority of mRNA molecules are associated with ribosomes and presumably engaged in translation, indicative of a very efficient recruitment of newly synthesized molecules into the translating pool. The number of ribosomes associated with mRNAs increases with increasing open reading frame (ORF) length, consistent with longer transit times for ribosomes translating longer ORFs. Surprisingly however, the density of ribosomes (i.e., the number of ribosomes per unit ORF length) decreases with increasing ORF length, suggesting lower translation efficiency for longer ORFs. Possible origins for such an effect are slower initiation rates for longer mRNAs, incomplete processivity during translocation steps (i.e., ribosomes falling off of the mRNA during elongation) and slow termination rates. These models yield physically distinct predictions: an even distribution of ribosomes along the mRNA is expected in the model of slower initiation, lower density of ribosomes towards the 3 end of the mRNA is expected if ribosomes dissociate during translocation, and higher density at the 3 end will be observed if termination is slow.

To determine the ribosome density in different regions of a particular mRNA, we developed a system that involves specific cleavage using antisense oligonucleotides followed by separation on a sucrose gradient. We used it to compare the number of ribosomes on the 5 and 3 halves for several mRNAs. In each case both halves have a similar number of ribosomes. These results strongly suggest that elongation is highly processive and that termination is not generally rate-limiting. We propose that longer mRNAs have lower translation efficiency due to slower initiation rates, perhaps the result of a more distal 3 polyA tail.

Home * List of Abstracts * Speakers

Gene Identification and Study of Paralogy on the Human X Chromosome

Ian Barrett and **Mark Ross**
The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA UNITED KINGDOM
Telephone: +44 (0)1223 494732
Fax: +44 (0)1223 494919
Email: ipb@sanger.ac.uk

Gene identification and study of paralogy on the Human X chromosome Ian P. Barrett and Mark T. Ross The Human X Chromosome group, Department of Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK. Email: ipb@sanger.ac.uk The availability of human genomic sequence has enabled the systematic search for genes and the study of their genomic context. In addition, the rapid rate of sequence data production from other organisms allows comparative analysis to identify conserved regions and provide information about the evolution of loci in different species. We have used the genomic sequence to identify genes and pseudogenes in a 15 Mb region of the human X chromosome, Xq22-q23, using a combination of bioinformatics and experimental verification. Selected features from the region will be presented, including evidence of an almost complete insertion of the mitochondrial genome into the nuclear genome and evidence of an apparent gene-fusion involving the NXF2 gene and its subsequent inverted duplication. In order to study the syntenic region of the mouse genome, a sequence-ready BAC tile-path of the region was constructed and is being sequenced. Preliminary comparative analysis of the COL4A5 and COL4A6 genes will be presented. During the course of the gene-mapping on Xq22-q23, extensive paralogy in the region became apparent. Work now in progress to evaluate the tissue expression patterns of these genes and to analyse the syntenic region in mouse should shed further light on the evolution of the region. The gene-map of the region also provided evidence for a large-scale duplication involving genes in the Xq22-q23 region, and genes on Xp. To date there appear to be at least 14 pairs of paralogs involved. Work is now underway to locate the orthologs of these genes in a marsupial genome. Evidence for the duplication will be presented and discussed. The ability to study genes in their context in a large region has illuminated some interesting examples of gene architecture and evidence of paralogy. Availability of large regions of annotated, contiguous genomic sequence from human, and the syntenic region in mouse, will be useful for refining comparative analysis tools and will aid the assignment of orthologs in a complex region.

Home * List of Abstracts * Speakers

Systematic Cloning of Open Reading Frames into Expression Systems

Stephanie Bechtel, Anny Duda, Kerstin Hettler, Detlev Bannasch, Alex Mehrle, Annemarie Poustka, and Stefan Wiemann
German Cancer Research Center (DKFZ), Dept. Molecular Genome Analysis (H0600), Im Neuenheimer Feld 350, 69120 Heidelberg Germany
telephone: +49 6221 56 4783
fax: +49 6221 56 4782
email: s.bechtel@dkfz.de

The large-scale cDNA sequencing approaches of the German cDNA Consortium within the last 5 years have resulted in the generation and identification of approximately 1000 novel full-length human cDNAs encoding proteins of unknown function. To allow for their functional characterization, a large scale systematic cloning technology applied in a high-throughput manner is essential. To achieve this, we use the Gateway system from Invitrogen, which is based on cloning by site-specific recombination. We systematically amplify the ORFs by 2-step-PCR. The PCR products are cloned by recombination into an entry vector in a one-step reaction (BP reaction) and resulting entry clones are sequenced. Our pool of confirmed entry clones corresponds currently to about 600 ORFs. For functional analysis, the ORFs can be shuttled from one entry clone into any Gateway compatible expression vector in a single step reaction (LR reaction), again by recombination, e.g. for eukaryotic or prokaryotic protein expression in a native form or as fusion proteins. We currently clone the ORFs without a native stop-codon to enable the generation of both N- and C-terminal protein fusions. In the future, we intend to additionally create entry clones containing ORFs with a terminator triplet to express the proteins in “real” native form. This also offers the opportunity to make the cloning procedure independent of destination vector reading frames. In this way, we accommodate the expanding number of destination vectors which are used for comprehensive functional proteomic assays which are currently under development in our group.

For systematic subcellular localization studies of the novel proteins, the ORFs are subcloned into Gateway compatible EYFP and ECFP expression vectors for generation of C-terminal yellow fluorescent and N-terminal cyan fluorescent fusion proteins. These studies are already established in cooperation with the group of Rainer Pepperkok at the EMBL, Heidelberg. In order to allow monitoring of the whole clone construction process all data are entered into a database together with the results of GFP fusion protein localizations which are entered using a web-interface. The localization data are accessible via the internet (http://www.dkfz.de/LIFEdb/).

Home * List of Abstracts * Speakers

Drawing Inferences from Microarray Data

Michael Bittner
NIH/NHGRI
50 South Drive, Building 50, Room 5154
Bethesda, MD 20892 USA
Telephone: 301 496 7980
Fax: 301 402 3241
Email: mbittner@nhgri.nih.gov

From its inception, expression profiling has been proposed as a way to develop inferences about genes involvement in normal and pathological processes. Over the past few years, a wide variety of strategies for developing “candidate genes” from microarray data have been tested. Each of these strategies has inherent minimal requirements in terms of prior knowledge of the general way in which cells operate, the specifics of the experimental system being studied, the capabilities of the profiling system and the discriminating power of the analytical tools that can be employed. The surprising result that emerges is that even starting from widely varying levels of prior knowledge about the particular cellular process you are studying, it is still possible to draw very sharp inferences about the roles that particular genes play if you can satisfy these minimum requirements. This presentation will examine three studies that provide examples of the levels of prior knowledge, breadth of samples available, available alternative measurements of the phenotypes and coherence of data that tend to produce interpretable, testable conclusions.

Home * List of Abstracts * Speakers

Identification and Characterization of a Novel Thermostable RNA ligase

T. Blondal,[2]&#65533; O. Fridjonsson, [1]&#65533; S. Hjorleifsdottir, [1]&#65533; A. &#65533;varsson, [1]&#65533; S. Skirnisdottir,[1]&#65533; A.G. Hermannsdottir, [1]&#65533; G.O. Hreggvidsson, [1,3] Albert Vernon Smith[2] and J.K. Kristjansson[1,3]
[1] Prokaria ltd. Gylfaflot 5, 101 Reyjavik, Iceland, [2] DeCode Genetics Inc, Reykjavik Iceland, [3] University of Iceland, Grensasvegur 12, 107 Reykjavik, Iceland
Telephone: +345 5701900
Fax: +354 5701902
Email: thorarinn@decode.is

T4-RNA-ligase has the ability to ligate single stranded nucleic acids. It catalyses intra- and inter-molecular formation of phosphodiester bonds between 5'-phosphate and 3'-hydroxyl termini of single-stranded RNA or DNA, in an ATP dependent manner. &#65533;This enzyme is increasingly being used in various molecular biology applications, especailly in RNA-based protocols for retrieving genes or gene fragments (such as RACE). Furthermore it has use in single gene specific primer&#65533; PCR methods for retrieving flanking regions of gene fragments and for modifications of&#65533; 5&#65533;-ends of single stranded nucleic acids. RNA-ligases of this type have only been found in T4-like phages, but recently a homologous gene has been&#65533; identified in the plant virus *Autographa californica*.

We report the identification of&#65533; a thermostable RNA ligase in the genome of a T4-like phage, RM 378 that infects&#65533; the thermophilic bacterium *Rhodothermus marinus*. The gene has been cloned and expressed in high quantities and purified. Properties of this enzyme will be reported.

- Tessier, D.C., Brousseau, R. and Vernet, T. 1986. Ligation of single-stranded oligo-deoxyribonucleotides by T4 RNA-ligase. Anal.Biochem. 158:179-188.
- Troutt, A.B., McHeyzer-Williams, M.G., Pulendran, B. and Nossal, G.J.V. 1992. Ligation-anchored PCR: A simple amplification technique with single sided specificity. Proc.Natl.Acad.Sci.USA. 89:9823-9825.
- Apte, A.N. and Siebert, P.D. 1993. Anchor-ligated cDNA libraries: A technique for generating a cDNA library for the immediate cloning of the 5&#65533;ends of mRNAs. Biotechniques. 5:890-893
- Xing-Hai Zhang and Vincent L. Chiang. 1996. Single stranded DNA ligation by T4 RNA ligase for PCR cloning of 5&#65533;-noncoding fragments and coding sequences of specific gene. Nucleic Acids Reserch. 24:990-991.

Home * List of Abstracts * Speakers

C. elegans Operons: A Novel Tool to Find Functionally Related Genes

Tom Blumenthal
Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Denver, CO 80262

C. elegans and its close relatives appear to be unique among animals in having large numbers of polycistronic transcription units, not unlike bacterial operons. These operons may constitute an important gene finding tool. Although the genes in these operons are eventually represented on monocistronic mRNAs, they are transcribed as a polycistronic precursor that is processed into gene-length units by trans-splicing and 3 end formation. We have recently completed a microarray analysis of all predicted and known *C. elegans* genes to determine which genes are contained in operons (Blumenthal et al., 2002, Nature **417**:851). Our results indicate that ~15% of all *C. elegans* genes are in operons ranging from 2 to 8 genes long. We have been analyzing the list of ~1000 operons, containing ~2,600 genes to determine what types of genes they contain and, importantly, to what extent functionally-related genes are co-transcribed in operons.

Whereas some classes of genes are very highly represented in operons, others are almost never in operons, suggesting they are not random gene assemblages. For example, ~50% of genes whose products are destined for the mitochondria are transcribed in operons. Similarly, around half of genes that encode the basic transcription, splicing and translation machinery are in operons. In sharp contrast, transcription factor genes are virtually never in operons. Also, genes that are highly regulated in a particular tissue are generally excluded from operons. These include genes for collagens, sperm proteins, intermediate filament proteins, cytochrome P450s, immunoglobulin domain proteins and basement membrane proteins. It seems as if proteins can be transcribed as parts of operons if they are either not regulated or need to be globally regulated in response to certain kinds of signals.

Many of the C. elegans operons do not appear to make any kind of functional sense. On the other hand, many others do. There are numerous operons that contain more than one gene encoding proteins that are part of the basic transcription machinery, and this occurs far more often than would be expected by chance. This is also true of the splicing machinery and of mitochondrial proteins. Furthermore, there are operons that contain a subunit of one of the RNA polymerases and one of the basic factors that act with that polymerase. The operon list can suggest many previously unsuspected functional relationships among proteins, but whether most operons involve functionally related genes is not yet known. Many homologs of human disease genes are contained in operons, and determining whether the products of the other genes in these operons are related proteins is worth investigating.

Home * List of Abstracts * Speakers

Computational and In Vitro Analysis of Destabilized DNA Regions in the Interferon Gene Cluster: The Potential of Predicting Functional Gene Domains

S. Goetze, A. Gluch, C. Benham and **J. Bode**
GBF- German Research Center for Biotechnology/Epigenetic Regulation, Mascheroder Weg 1, D-38124 Braunschweig; GERMANY
Telephone: +49 531 6181 251
Fax: +49 531 6181 262
Email: juergenbode@gbf.de

Recent evidence adds support to a traditional concept according to which the eukaryotic nucleus is organized into functional domains by scaffold or matrix attachment regions (S/MARs). These regions have previously been predicted to have a high potential for stress induced duplex destabilization (SIDD). Here we report the parallel results of binding (re-association) and computational SIDD analyses for regions within the human interferon gene cluster on the short arm of chromosome 9 (9p22). To verify and further refine the biomathematical methods, we focus on a 10 kb region in the cluster with the pseudogene IFNWP18 and the interferon alpha genes IFNA10 and IFNA7. In a series of S/MAR binding assays we investigate the promoter and termination regions, and additional attachment sequences that were detected in the SIDD profile. The promoters of the IFNA10 and the IFNA7 genes have a moderate ~20% binding affinity to the nuclear matrix; the termination sequences show stronger association (70-80%) under our standardized conditions. No comparable destabilized elements were detected flanking the IFNWP18 pseudogene, suggesting that selective pressure acts on the physicochemical properties detected here. In extended, non-coding regions a striking periodicity is found of rather restricted SIDD minima with scaffold binding potential. By various criteria, the underlying sequences represent a new class of S/MARs , thought to be involved in a higher level organization of the genome. Together, these data emphasize the relevance of SIDD calculations as a valid approach for the localization of structural, regulatory and coding regions in the eukaryotic genome. New data show that they have the added potential to cover DNAse I hypersensitive and fragile sites that are either of regulatory relevance or the molecular cause of genome instability or the insertion of foreign DNA. This concept will be supported by halo-FISH analyses.

- J. Bode, C. Benham, E. Ernst, A. Knopp, R. Marschalek, R. Strick, and P. Strissel (2000) Fatal Connections: When DNA Ends Meet on the Nuclear Matrix. J. Cell. Biochem. Suppl. 35, 3-22 ; http://www3.interscience.wiley.com/cgi-bin/issuetoc?ID=82002725.
- C. Mielke, C. Benham, J. Bode and M. Breindl (2001) Nuclear matrix association potential of regulatory elements in the collagen 1a1 gene. J. Cell Biochem. 84: 484-496, 2002
- H. Engel, H. R&#65533;hl, C. J. Benham, J. Bode and S. Weiss (2001) Germ-line transcripts of the immunoglobulin J-C clusters in the mouse: Characterization of the initiation sites and regulatory elements. Mol. Immunol. 38, 289-302.
- I. Liebich, J. Bode, M. Frisch and E. Wingender (2001) S/MARt DB - A database on scaffold / matrix attached regions. Nucleic Acids Res.30, 372-374.
- A. Dietz-Pfeilstetter, N. Arndt, V. Kay and J. Bode (2001) Molecular structure and regulatory potential of a T-DNA integration site in petunia. Transgenic Res. (in press)
- I. Liebich, J. Bode, I. Reuter, E. Wingender (2002) "Evaluation of sequence motifs found in scaffold / matrix attached regions (S/MAR)". Nucleic Acids Res. 30,3433-3442.
- S. Goetze, A. Gluch, C. Benham and J. Bode (2002) Computational and in vitro analysis of destabilized DNA regions in the interferon gene cluster: the potential of predicting functional gene domains. Biochemistry, in press.

Functional Characterization of ETV6, a Candidate Tumor Suppressor Gene Associated with Childhood Acute Lymphoblastic Leukaemia

Gino Boily[1] and Daniel Sinnett[1,2]
[1]Research Center, Sainte-Justine Hospital; [2]Department of Pediatrics, University of Montreal, Montreal, QC Canada
Telephone: 514-345-4931-6142
Fax: 514-345-4731
Email: gino_boily@hotmail.com

Deletion of the 12p12 locus is a frequent genetic abnormality observed in childhood pre-B acute lymphoblastic leukemia (ALL), suggesting the presence of a tumor suppressor gene in the region.? In most cases, these deletions are associated with the translocation t(12;21) on the non-deleted allele, resulting in the expression of the fusion protein ETV6-AML1 and in the inactivation of the ETV6 gene.? ETV6 encodes a member of the Ets family of transcription factors and may play a role as a transcriptional repressor.? In the present study, we want to evaluate the functional impact of the expression of ETV6 by overexpression or knock down experiments.? ETV6 cDNA was stably transfected into a pre-B ALL cell line possessing the translocation t(12;21) and expressing the related fusion protein ETV6-AML1 but not ETV6.? We found that the doubling time of the ETV6-transfected cells was increased by 12% as compared to the vector alone-transfected cells (p = 0.008).? The heterogeneous ETV6-transfected population weakly expressed ETV6, and out of the 11 clones generated from this population, none strongly expressed ETV6. All together, these data suggest that a negative selection of ETV6 expressing-cells is operating and support the hypothesis that ETV6 may act as a tumor suppressor gene.? We are now generating cellular models, by the means of inducible systems and RNAi technology, in which we will be able to modulate the expression status of ETV6 in combination with the presence of ETV6-AML1 or not.? Our preliminary data show that ETV6 can be silenced efficiently in HeLa and NIH 3T3 cells by transfection of siRNA.? These cellular models will help us to understand the role of ETV6 and to eventually identify target genes of this transcription factor.

Home * List of Abstracts * Speakers

Reconstructing and analysing gene networks from microarray data

Alvis Brazma
European Bioinformatics Institute, EMBL-EBI, Hinxton, Cambridge, CB10 1SD UNITED KINGDOM
Telephone: +44-(0)1223 494658
Fax: +44-(0)1223 494468
Email: brazma@ebi.ac.uk

We discuss a variety of methods how to use genome scale gene expression data for reconstructing gene networks. In particular, we consider two types of gene networks in yeast: (1) transcription factor binding network for 37 experimentally known binding sites, and (2) the network dependencies between gene expression profiles in a dataset from yeast mutation studies. We compare these network and analyze their architectural properties. For instance, we look for 'important' genes, i.e., genes with high out-degree in the dependency graph, and genes with complex regulation, i.e., genes with high in-degree in the graph. We also study the modularity of these networks and compare gene expression with protein-protein interaction data and study their correlations.

Home * List of Abstracts * Speakers

Identifying Novel Transcripts/Novel Genes in the Human Genome by Using Novel SAGE Tags

Jianjun Chen*, Miao Sun[], Sanggyu Lee*, Guolin Zhou*, Janet D. Rowley*, **San Ming Wang***
*Department of Medicine and []Computer Science, University of Chicago,5841 S. Maryland, MC2115, Chicago, IL 60637
Telephone: (773) 702-6788
Fax: (773) 702-3002
Email: jchen@medicine.bsd.uchicago.edu

One of the goals of human genome studies is to identify all the genes in the human genome for further functional analysis of each gene. However, the number of genes in the human genome remains a controversial issue. Whereas most of the genes in the human genome are said to have been physically or computationally identified, many short cDNA sequences identified as tags by use of SAGE (serial analysis of gene expression) do not match these genes. By performing experimental verification of more than 1,000 SAGE tags and analyzing 4,285,923 SAGE tags of human origin in the current SAGE database, we examined the nature of the unmatched SAGE tags. Our study shows that most of the unmatched SAGE tags are truly novel SAGE tags that originated from novel transcripts not yet identified in the human genome, including novel alternatively spliced transcripts from known genes and potential novel genes. Our study also indicates that by using GLGI (generation of longer cDNA fragments from SAGE tags for gene identification) and 5' RACE (rapid amplification of 5' cDNA ends), novel SAGE tags can be converted back to their corresponding 3' cDNAs and full-length cDNAs. By using such approach, we should be able to identify efficiently many novel transcripts/novel genes in the human genome that are difficult to identify by conventional methods, thus the rate of discovery of novel transcripts/novel genes in the human genome should be significantly accelerated. The same approach should also be applicable to gene identification in other eukaryotic genomes.

Home * List of Abstracts * Speakers

Computational Screening for Peroxisomal Proteins Using New Methods, Old Methods, and Human Expertise

Olof Emanuelsson[1], Susana Cristobal[2], Arne Elofsson[1], **Gunnar von Heijne**[1]
[1]Stockholm Bioinformatics Center, Stockholm University, Stockholm, Sweden [2]Dept. of Cell and Molecular Biology, Biomedical Center, Uppsala University, Uppsala, Sweden
Telephone: +46-(0)8-55378574
Fax: +46-(0)8-55378214
Email: olof@sbc.su.se

The peroxisome is an important cellular compartment, responsible for carrying out many reactions essential to the organism. We have developed a sequence-based predictor as a part of an attempt to identify novel potential peroxisomal proteins (concentrating on the C-terminal PTS1 signal, -SKL). Our predictor was applied on seven eukaryotic genomes to find candidate proteins for peroxisomal localization, and the resulting initial set was further analyzed by phylogenetic profiling based on domain analysis using the Pfam database. Thus we were able to narrow down the set of potential peroxisomal proteins to a reasonably reliable set. Although more successful than other peroxisomal protein predictors publicly available, our method is also an illustration of the fact that some protein locations simply are harder to predict than other. In the peroxisomal case, this is largely accounted for by the low complexity of the PTS1 signal, making it hard to distinguish a PTS1 from similar or even identical motifs that do not target proteins to the peroxisome.

Home * List of Abstracts * Speakers

Pathways, Molecular Processes and Disease Association: Needs and Solutions in the Pharmaceutical Industry

James Fickett
AstraZeneca R&D Boston, Waltham, MA 02451 USA
Telephone: 781 839 4761
Email: james.fickett@astrazeneca.com

We will review a few high priority areas in the drug discovery process where pathway information plays an important role. For each, the requirements will be analyzed, current solutions examined, and suggestions made for products that may be of use in the future. We suggest that in some areas the need is not so much detailed pathway information as rather more general concept associations around the molecular pathophysiology of disease.

Home * List of Abstracts * Speakers

Mapping the Transcriptome

Thomas Freeman
HGMP-Resource Centre, Hinxton, CB10 1SB UNITED KINGDOM
Telephone: 44 1223 494552
Fax: 44 1223 494512
Email: tfreeman@hgmp.mrc.ac.uk

Complete sequencing of the human, and more recently the mouse genome, provides an invaluable resource in our efforts to understand many aspects of biological function in both health and disease. Expression profiling, in particular using high-density arrays of gene-specific DNA probes, has firmly been demonstrated to be one of the most powerful and direct ways of translating sequence data into functional information. However, in spite of the fact that this technology has been in use for a number of years, we still have very little appreciation of when, where and to what degree most genes are expressed, at either the tissue or cellular level. This information is vital for our ability to appreciate transcriptome diversity, and the regulation and function of many genes. For a number of years now we have been endeavouring to map systematically in high throughput assays the expression of genes over a wide range of normal tissues and cell types with the aim of stabling databases of gene expression profiles. Initially, our efforts to map the mammalian transcriptome employed techniques such as RT-PCR but more recently we have used microarrays. I will discuss our findings and our continuing efforts to establish an "atlas" of gene expression. I will also discuss other aspects of expression analysis technology, and its use and development at the HGMP-RC.

Home * List of Abstracts * Speakers

Computational Proteomics: Genome-Scale Analysis of Protein Structure, Function, & Evolution

Mark Gerstein, P Harrison, J Qian, R Jansen, V Alexandrov, P Bertone, R Das, D Greenbaum, W Krebs, Y Liu, H Hegyi, N Echols, J Lin, C Wilson, A Drawid, Z Zhang, Y Kluger, N Lan, N Luscombe, and S Balasubramanian
Molecular Biophysics & Biochemistry Department, Yale University, New Haven, CT
http://bioinfo.mbb.yale.edu

My talk will address two major post-genomic challenges: trying to predict protein function on a genomic scale and interpreting intergenic regions. I will approach both of these through analyzing the properties and attributes of proteins in a database framework. The work on predicting protein function will discuss the strengths and limitations of a number of approaches: (i) using sequence similarity; (ii) using structural similarity; (iii) clustering microarray experiments; and (iv) data integration. The last approach involves systematically combining information from the other three and holds the most promise for the future. For the sequence analysis, I will present a similarity threshold above which functional annotation can be transferred, and for the microarray analysis, I will present a new method of clustering expression timecourses that finds “time-shifted” relationships.

In the second part of the talk, I will survey the occurrence of pseudogenes in several large eukaryotic genomes, concentrating on grouping them into families and functional categories and comparing these groupings with those of existing “living” genes. In particular, we have found that duplicated pseudogenes tend to have a very different distribution than one would expect if they were randomly derived from the population of genes in the genome. They tend to lie on the end of chromosomes, have an intermediate composition between that of genes and intergenic DNA, and, most importantly, have environmental-response functions. This suggests that they may be resurrectable protein parts, and there is a potential mechanism for this in yeast.

- J Qian, B Stenger, C Wilson, J Lin, R Jansen, W Krebs, V Alexandrov, N Echols, S Teichmann, J Park, M Gerstein. “PartsList: a web-based system for dynamically ranking protein folds based on disparate attributes, including whole-genome expression and interaction information.” Nucleic Acids Res 29: 1750-64 (2001).
- P Harrison H Hegyi, P Bertone, N Echols, T Johnson, S Balasubramanian, N Luscombe, M Gerstein. “Digging for dead genes: an analysis of the characteristics of the pseudogene population in the Caenorhabditis elegans genome.” Nucleic Acids Res 29: 818-30 (2001).
- J Qian, M Dolled-Filhart, J Lin, H Yu, M Gerstein. “Beyond synexpression relationships: Local Clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions.” J Mol Biol 314: 1053-1066 (2001).
- R Jansen, D Greenbaum, M Gerstein. “Relating whole-genome expression data with protein-protein interactions.” Genome Research 12: 37-46 (2002).
- P Harrison, H Hegyi, P Bertone, N Echols, T Johnson, S Balasubramanian, N Luscombe, M Gerstein. “Molecular fossils in the human genome: Identification and analysis of pseudogenes in chromosomes 21 and 22.” Genome Research 12: 273-281 (2002).

Home * List of Abstracts * Speakers

Large Scale Transcriptional Activity Observed in Chromosomes 21 and 22

Thomas Gingeras
Affymetrix, Santa Clara, CA 95051 USA
Telephone: 408-731-5069
Fax: 408-481-0422
Email: tom_gingeras@affymetrix.com

Large-Scale Transcriptional Activity Observed in Chromosomes 21 and 22. Thomas R. Gingeras, Affymetrix, Inc. Santa Clara, Ca, 95051. The draft sequences of the human chromosomes 21 and 22 currently indicate that there are approximately 786 well-characterized Ensembl genes (v7.29a). Recently, we described a collection of empirically derived maps identifying active areas of RNA transcription on these chromosomes using cytosolic poly A+ RNA obtained from 11 human cell lines. (Kapranov, P., et al. 2002 Science 296:916). These unbiased maps were created using oligonucleotide arrays containing 25- mer probes spaced on average every 35 base pairs along these chromosomes. When compared to the sequence annotations available for these chromosomes it is noted that as much as an order of magnitude more of the genomic sequence is used for transcription than envisioned by the predicted and characterized exons. Specifically, 9.7% (98,231) of the total number of probes used to interrogate these chromosomes detected transcripts in at least 5 of the 11 cell lines (with estimated false positive rate of ~5%). Of these interrogating probes 2.6%, 24.9% and 72.7% were positioned within exonic, intronic and intergenic regions, respectively. Of the total number of probes used, only 11% detected transcripts within an annotated exon. The remaining positive probes (~89%) were located in intronic and intergenic regions and approximately half of these were >300 base pairs away from the nearest annotation. Validation of these data has focused on 30 distinct regions along both chromosomes, situated well away from any annotations. Copies of cDNA for these regions could be found in NCI supplied cDNA library (from Dr. L. Hong) made from cytosolic polyA+ RNA of a single cell line, HepG2. Given that we have ~60% verification rate from a single cell line, provides additional confidence that these novel transcripts will be joined by an increasing number as we scale up the verification effort. Additional information concerning the location of these novel transcripts and the conservation to non-human genomic sequences will be presented.

Home * List of Abstracts * Speakers

Functional Genomics of *Runx3* and DRG Neurogenesis

Yoram Groner, Yael Bernstein, David Bettoun, Catherine Harris-Cerruti, Ofer Fainaru, Dalia Goldenberg, Ditsa Levanon, Joseph Lotem, Varda Negreanu, Amir Pozner, Eilon Wolf, Cuiying Xiao, Merav Yarmus
Department of Molecular Genetics the Weizmann Institute Rehovot ISRAEL
Telephone: 972-8-9343-972
Fax: 972-8-9344-108
Email: yoram.groner@weizmann.ac.il

The mammalian runt-related transcription factors (Runx) belong to a small gene family of three genes (Runx1, Runx2 and Runx3). They all contain a highly conserved DNA binding domain designated "runt domain," which is also found in the *Drosophila* gene *Runt*.**?** *Runx1* and *Runx2* are associated with human diseases and act as master regulators of gene expression in hematopoiesis and osteogenesis.**?** We have cloned, sequenced, and elucidated the genomic structure of the human *RUNX1* and *RUNX3* genes. *RUNX1* resides on human chromosome 21 and *RUNX3* on chromosome 1. Chromosomal translocations involving *RUNX1* are prevalent in human leukemias. Disruption of *RUNX1* obliterates definitive hematopoiesis and impairs formation of vascular capillaries. The mammalian *RUNX* genes arose early in evolution and maintained extensive structural similarities between them. Sequence analysis suggested that *RUNX3* is the most ancient of the three mammalian genes, consistent with its role in neurogenesis of the monosynaptic reflex arc, the simplest neuronal response circuit, found in the most primitive animals, the Cnidarians.

The three RUNX proteins bind to the same DNA motif; thus, their pleiotropic functions are likely to result from a regulated spatial/temporal expression pattern.**?** Interestingly, *Runx1* and *Runx3* genes contain RUNX binding sites in their promoter region, raising the possibility of cross regulation (both positive and negative) between them. We used immunohistochemistry and **?**-galactosidase (LacZ) activity of targeted *Runx3* and *Runx1* loci to determine the expression pattern of Runx1 and Runx3 during mouse embryogenesis.**?** We found that Runx3 expression overlapped with that of Runx1 in the hematopoietic system, whereas in sensory ganglia, epidermal appendages, and developing skeletal elements, Runx3 expression was confined to different compartments.**?** The data provided new insights into the function of Runx3 and Runx1 in organogenesis and support the possibility that cross regulation between them plays a role in embryogenesis.

Of the three *RUNX* genes, *RUNX3* is the least studied. In adults, Runx3 is highly expressed in the hematopoietic system, but its biological function is largely unknown. We generated knockout (KO) mice with disrupted Runx3 alleles by inserting a LacZ-*neo* cassette into exon 2. Heterozygous Runx3-mutant mice appeared phenotypically normal, whereas homozygous mutant mice showed posture abnormalities and severe limb ataxia. To understand the biological significance of what seemed to be a neuronal defect, we first examined in great detail the expression pattern of Runx3 in the nervous system. We found that Runx3 is highly expressed during mouse development in a subset of sensory neurons within the dorsal root ganglia (DRG), which we subsequently identified as TrkC, group Ia proprioceptive neurons. These neurons form monosynaptic connections with both muscle spindles and motor neurons to generate the monosynaptic stretch reflex circuit. Interestingly expression of the other family member Runx1 was also detected in the DRGs, but was confined to the small diameter TrkA nociceptive neurons. To further evaluate the physiological defect of the Runx3 mutant, electrophysiological studies were performed. Measurements revealed complete disruption of monosynaptic connectivity between intraspinal afferents and MNs. Tracing experiments, using anterograde DiI labeling, demonstrated an absence of afferent projections in the spinal cord of the KO mice, and further analysis revealed marked reduction in the number of Ia neurons in the KO DRGs. Taken together, the data demonstrate that Runx3 is a neurogenic TrkC neurons specific transcription factor. In its absence TrkC neurons in the DRG do not survive long enough to extend their axons toward target cells, resulting in lack of connectivity and ataxia.**?** The data provide new genetic insights into DRGs neurogenesis and may help elucidate the molecular mechanisms underlying somatosensory-related ataxia in humans.

The Natural History of Human Gene Families: Genome-Wide Duplication(s) v.s. Small-Scale Duplications

Xun Gu
Iowa State University, AMES, IA 50011-3223 USA
Telephone: 515-294-8075
Email: xgu@iastate.edu

The classical hypothesis (2R) for vertebrate genome duplications postulates two successive polyploidizations prior to the origin of fishes, which is being seriously challenged. Since the debate is mostly between the big-bang mode (large-scale duplication) vs. the continuous mode (constant creation by small-scale duplications), we address this issue by testing whether a significant portion of paralogous genes in the contemporary human genome was indeed generated in the early stage of vertebrates. After extensive search in major databases, we have dated 1,739 gene duplication events from the phylogenetic analysis of 749 vertebrate gene families, which shows a pattern characterized by two waves (I, II) and an ancient component. While Wave I represents a recent gene family expansion by tandem or segmental duplications, Wave II, a rapid paralogous gene increase in the early stage of vertebrates, supporting the notion of genome duplication(s) (the big-bang mode). Further analysis indicates that large and small-scale gene duplications both have significant contributions during early stage of vertebrate evolution to building the current hierarchy of human proteome.

Home * List of Abstracts * Speakers

Multiprotein Evolutionary Analyses of Diverse Eukaryotes

S. Blair Hedges
Pennsylvania State University

[abstract not available]

Home * List of Abstracts * Speakers

Translational Control at the Interface Between the Genome and the Proteome

Matthias W. Hentze
EMBL, Heidelberg, D-69117 GERMANY
Telephone: +49-6221-387 501
Fax: +49-6221-387 518
Email: hentze@embl-heidelberg.de

Translation of messenger RNAs into polypeptides represents the interface between the genome and the proteome. Not surprisingly, this step in the gene expression pathway is frequently used for the regulation of biological responses. I will discuss principles by which structural features of mRNAs, regulatory RNA elements and RNA-binding proteins regulate protein synthesis. This will include findings that imply a new role of the poly(A) tail in the translation of messenger RNAs via IRESes (internal ribosome entry sequences), the understanding of an important switch in Drosophila development, and results from whole genome microarray analyses in yeast which suggest that signal-induced changes in the transcriptome are amplified at the translational level. These latter results unveil a novel, higher level of coordinated gene regulation, which we refer to as ?potentiation?.

Home * List of Abstracts * Speakers

3' Untranslated Regions and Selenocysteine Incorporation: From Modulation of Translation to a Novel Single Nucleotide Polymorphism

St�phane Villette[1], G. Bermano[3], John R. Arthur[2] and **John E. Hesketh**[1]*
[1]School of Cellular & Molecular Biosciences, University of Newcastle,Newcastle upon Tyne,NE1 7RU, UK. [2] Division of Cell Integrity, Rowett Research Institute, Greenburn Road, Bucksburn, Aberdeen AB21 9SB, Scotland, UK. [3] University of Aberdeen Medical School, Foresterhill, Aberdeen AB25 2ZD, Scotland, UK.
Telephone: 00-44-191-222-8744
Fax: 00-44-191-222-8684
Email: j.e.hesketh@ncl.ac.uk

The 3� untranslated region (3�UTR) of mRNAs has important regulatory functions. As well as determining mRNA translation, localisation and stability, in a limited number of mRNAs the 3�UTR is important for incorporation of Selenium (Se) into proteins as the 21st amino-acid selenocysteine. Se is an essential micronutrient for human health; limited intake has been linked to cancer and heart disease. The biological roles of Se are attributed to its presence in a range of 20-30 selenoproteins including the cytosolic, gastrointestinal and phospholipid hydroperoxide glutathione peroxidases (GPX1, GPX2 and GPX4) which protect cells from oxidative stress. In addition, it has been suggested that GPX1/GPX4 may play a role in regulation of leukotriene biosynthesis and thus inflammation. In eukaryotes Se incorporation requires a stem-loop (the SECIS element) within the 3�UTR. If supply of Se is limiting in the diet or culture medium then selenoprotein synthesis is modulated but not all selenoproteins are affected equally; there is a tissue-specific prioritisation of the available Se. Such prioritisation depends on 3�UTR sequences (see Bermano et al, 1997; Hesketh & Villette, 2002) . It is therefore possible that genetic variation between individuals in the gene regions corresponding to 3�UTR sequences could cause different patterns of Se incorporation and different susceptibility to disease. Recently we found that the region of the GPX4 gene corresponding to the 3�UTR contained a T/C variant at position 718. The distribution of this SNP in our population was 34% CC, 25% TT and 41% TC (Villette et al, 2002). Individuals of different genotype exhibited significant differences in the levels of lymphocyte 5- lipoxygenase total products, C718 showing increased levels of those products compared to T718 and T/C718 (36% and 44% increases respectively). The data suggest that the SNP718 that we have identified has functional consequences.

Hesketh JE & Villette,S (2002) Proc. Nutr. Soc 61, 1-10

Villette et al. (2002) Blood Cells. Mol. Disease. in press

We thank our colleagues who have contributed to the work reviewed in this paper and Food Standards Agency and Scottish Executive Environment and Rural Affairs Department (SEERAD) for financial support.

Home * List of Abstracts * Speakers

S-Gal®: An Autoclavable, Water-Soluble Dye for Enhanced Color-Selection of Cloned DNA Inserts

Sara Borchardt and **Ken Heuermann**
Sigma-Aldrich Biotechnology, St. Louis, MO
Telephone: 314-289-8496, x8610
Fax: 314-286-7645
Email: kheuermann@sial.com

S-Gal (3,4-cyclohexenoesculetin-beta-D-galactopyranoside) outperforms X-gal as a beta-galactosidase substrate for automated and non-automated molecular cloning applications involving color selection.® Insertion of a DNA fragment into a vector multiple cloning region embedded in the alpha-complement of the *lacZ* gene disrupts beta-galactosidase activity in the host, resulting in the formation of a colorless or cream-colored colony.® Dark black staining of a colony indicates uninterrupted expression of the alpha-complement (no insertion). Better contrast is observed between the stained and unstained colonies and the background of the plated medium when using S-Gal as opposed to X-gal. As a result, colonies representing recombinants can be distinguished from those containing the parental vector at an earlier timepoint, typically between 19 to 22 hours following plating, using pUC18 and DH5alpha host strain. Other standard E. coli strains (JM109, XL-1Blue, and NovaBlue) have been successfully tested using S-Gal. Additionally, pSTBlue-1 transformants produce black colonies in the presence of kanamycin.®

Autoclavable/microwavable S-Gal is blended in LB agar (with IPTG <isopropyl beta-D-thiogalactoside> and ferric ammonium citrate) at a final plated concentration of 300 micrograms per milliliter.® Kanamycin can be added to this blend prior to autoclaving or microwaving for antibiotic selection. Addition of ampicillin to the blend confers selection to microwaved medium.® A water-soluble S-Gal sodium salt derivative has been developed which allows for customized adjustment of dye concentration and addition of S-Gal to alternative medium formulations. As in case of the free base form found in the blended agar preparation, the sodium salt has been shown to be heat-stable. Both are light-stable (X-gal is not) and have been tested to be stable in prepared medium containing 100 micrograms per milliliter of ampicillin for up to one month, with no effect on color development or antibiotic selection.

Home * List of Abstracts * Speakers

Development of a Novel High Throughput Directional Cloning Platform

Ron Hernan, Leanne Snyder, **Ken Heuermann**, and Brian Ward
Sigma-Aldrich Biotechnology St. Louis, MO
Telephone: 314-289-8496, x8610
Fax: 314-286-7645
Email: kheuermann@sial.com

The cloning and expression of a large number of genes has been an enormous challenge for gene function studies in the post-genomics era. One of the major limitations has been the lack of a pair of universal restriction enzymes for generating the cohesive ends required for directional cloning. Traditionally, different enzymes are used for directional cloning of each gene, which restricts throughput and increases workload.&#65533; To overcome this problem, thionucleotides dATPaS and dGTPaS are incorporated during PCR amplification with restriction enzyme sites sequences engineered into the primers. Digestion of the amplified product with Exonuclease III produces 5&#65533; ends that are restriction enzyme site compatible and capable of ligation into vectors digested with corresponding restriction enzymes. With this approach, cloning efficiencies greater then 90% (transformants with positive inserts in the correct orientation). Because this method of cloning is universal, it is highly adaptable to robotic applications. Using this technology we have developed a system to both systematically and directionally clone open reading frames in a universal fashion without consideration of internal restriction enzyme sites.&#65533; We have cloned 384 *E. coli* ORFs using this technology with a 74 % success rate.&#65533; In addition, we applied this technology to successfully clone selected ORF&#65533;s from mammalian systems into expression vectors with comparable success.

Home * List of Abstracts * Speakers

Diversity in Gene Expression: Assessment of Exon Skipping and Expression States

Winston Hide, Tzu-Ming Chern, Janet Kelso and Vladimir Babenko
SA National Bioinformatics Institute, University Western Cape, Bellville, South Africa
Telephone: +27 21 959 3645; Fax: +27 21 959 2512
Email: winhide@sanbia.c.za

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

Home * List of Abstracts * Speakers

Insertional Mutagenesis in Zebrafish Identifies a Diversity of Genes Required for Vertebrate Development

Nancy Hopkins
Massachusetts Institute of Technology, Cambridge, MA 02138 USA
Telephone: 617-253-6414
Fax: 617-258-0258
Email: nhopkins@mit.edu

It has been estimated from large chemical mutagenesis screens in zebrafish that approximately 2400 genes can be mutated to yield a developmental defect visible by low power microscopy during the first 5 days of zebrafish life. In order to rapidly identify a substantial fraction of these essential genes and provide an unbiased view of the diversity of genes required for vertebrate development we performed a large insertional mutagenesis screen in zebrafish using mouse retroviral vectors as the mutagen. As was observed in chemical mutagenesis screens, mutations in about a third of these genes lead to relatively specific developmental defects involving one or a few organs, while mutations in the remaining two-thirds result in relatively non-specific phenotypes or syndromes. The latter may be genes required in many or all cell types, while the former may be genes required for the patterning, differentiation, and physiology of specific cell types or organs. We have isolated mutants in about 450-500 different genes, roughly 15-20% of those that can be identified by this approach, and have identified about 250 of the mutated genes. These genes encode proteins with a wide variety of biochemical functions, from "classical" developmental roles such as intercellular signaling and the control of cell-type-specific gene expression to roles in basic cell biological processes such as mitosis, protein synthesis, and vesicle trafficking. Furthermore, about 20% of the genes are novel, in that the biochemical function of the proteins they encode can not be predicted with certainty if at all from their amino acid sequences. All of the genes have orthologues or related sequences in human and mouse, often in fly and worm, and sometimes in yeast. Our screen demonstrates the power of forward genetics using insertional mutagenesis to assist in rapidly assigning in vivo functions to essential vertebrate genes, and it is providing a more comprehensive view of the genetic basis of vertebrate development.

Home * List of Abstracts * Speakers

From Large Scale Mouse Mutagenesis to Gene Function

Martin Hrab\u00e9 de Angelis
Institute of Experimental Genetics, GSF Research Center for Environment and Health, Neuherberg, Germany

Due to the similarity of genome organisation, developmental and biochemical pathways, the mouse has become one of the major model systems to study the pathogenesis of diseases. As one of the largest ENU mutagenesis programs in Europe, the ENU Mutagenesis Consortium Munich has provided a valuable source of novel mutants for gene function studies and models for human diseases. After concentrating on dominant traits during the first phase of the project, in the second phase, we put our main efforts on recessive phenotypes by producing 100 micropedigrees (20 G3/ micropedigree) per year. In parallel, we continue to produce about 2000 F1 animals to further isolate novel dominant alleles of known and novel genes.\u00e9

Currently, more than 30.000 mice have been investigated for dysmorphological and over 10.000 mice for blood based parameters. To date, more than 400 mutant lines have been isolated. Novel dominant or recessive phenotypes have been identified with specific abnormalities comprising congenital malformations, biochemical alterations, immunological defects and complex traits such as behaviour or predipositions to allergies.

For several mutant phenotypes causative mutations were identified and could be assigned to specific biological functions or to human diseases. Detailed results on the genetic, pathologic, and molecular characterization of e.g. a novel Delta1 gene allele, different eye mutants and deaf mutants will be presented.

- **Hrab\u00e9 de Angelis M**. et al. (2000) Genome wide large scale production of mutant mice by ENU mutagenesis. *Nature Genetics* 25(4):\u00e9 444-447
- **Kiernan AE** et al. , (2001) The Notch ligand Jagged1 is required for inner ear sensory development. *Proc Natl Acad Sci U S A*. 98(7):3873-8.
- **Nadeau JH** et al (2001) Sequence interpretation. Functional annotation of mouse genome sequences. *Science*;291(5507):1251-5.
- **Vreugde S**. et al. (2002) Beethoven, a mouse model for dominant, progressive hearing loss DFNA36. *Nat Genet*. Mar;30(3):257-8.

Home * List of Abstracts * Speakers

Inferring Gene Transcription Networks: The Davidson Model

Sorin Istrail
Celera Genomics/Applied Biosystems, Rockville, MD 20850 USA
Joint work with Vladimir Filkov (UC Davis)

In 2001 Eric Davidson published his “Genomic Regulatory Systems” book where he reports on 30 years of work, together with his colleagues, on *purple sea urchin*. Their work provided a general experimental framework for the study of a gene s cis-regulatory region (an upstream DNA sequence containing a series of consecutive binding sites). Their approach consisted of systematic, almost exhaustive, series of mutations of individual binding sites, together with the associated measurements of the transcription rates. By quantitative analysis, they were able to infer a complete set of minimal functional units of regulation and their interrelations. They proceeded hierarchically to uncover “modularity” and “hardwired information processing logic” of a gene s cis-region. Most of their work was focused on the endo16 gene. Their extraordinary technology and the inference of the underlying “network” for this gene resulted in the most completely understood transcriptional system to date.

It is quite remarkable how combinatorial and robust their approach is. We present an analysis and a mathematical formalism for the Davidson transcriptional network inference framework together with combinatorial problems and algorithms related to it.

Home * List of Abstracts * Speakers

Direct Contribution of a Transposable Element to the Protein Diversity: A Novel Type Bovine Bcnt Protein that Includes the Endonuclease Domain of RTE-1

Shintaro Iwashita[1], Tomohito Itoh[2], Kenshiro Oshima[3], Katsuyuki Hashimoto[4]and Wojciech Makalowski[5]
[1]Mitsubishi Kagaku Institute of Life Sciences, [2]Shirakawa Institute of Animal Genetics, [3]Hitachi Instruments Service Co.,Ltd., [4]National Institute of Infectious Diseases, and [5]Pennsylvania State University
Telephone: 81-42-724-6265
Fax: 81-42-724-6316
Email: siwast@libra.ls.m-kagaku.co.jp

Bovine p97Bcnt (after Baucentaur) is a unique molecule that includes a retrotransposable element-1 (RTE-1)-derived endonuclease domain followed by two copies of an IR module with 40-amino acids in the C-terminus. Human and mouse Bcnt proteins have no RTE-1-derived region but one IR and a highly conserved C-terminal region that is absent from bovine p97Bcnt. The chicken Bcnt gene structure suggests that the human/mouse gene is an ancestral structure and that the bovine p97Bcnt structure must represent a recent evolutionary alteration. Indeed, screening of a bovine genomic library revealed another copy of the bovine Bcnt gene whose structure resembles the ancestral gene that encodes an IR module and a highly conserved C-terminus region. Biochemical analysis showed that both h-type Bcnt and p97 Bcnt are expressed in many bovine tissues and that they are active proteins. Both genes are localized linearly on bovine chromosome 18, suggesting tandem duplication as a mechanism for the creation of the p97Bcnt gene. Consequently, one of the duplicated copies lost its 3' exons but instead recruited an RTE-1-derived endonuclease domain. In addition, an IR coding exon got duplicated and another genomic piece was exonificated to create a unique 3' part of the p97Bcnt mRNA. The phylogenetic analysis showed that the new gene is under a relaxed evolution that enables fast diversification of duplicated genes. The described evolution of the Bcnt locus enabled us to infer in detail how transposable elements contribute to the protein diversity.

References:

J. Biol. Chem. **272**, 2801 (1997), *Gene* **211**,387 (1998), *BBA* **1427,** 408 (1999), *Gene* **268**, 59 (2001)

Home * List of Abstracts * Speakers

Cross-Species Comparison of Known and Predicted GDF-9 and GDF-9B Protein Sequences

Luke Jeffery, David Mottershead & **Olli Ritvos**
Program of Developmental and Reproductive Biology, Biomedicum Helsinki
University of Helsinki
telephone : +358-9-191 25662
fax : +358-9-191 25664
email : luke.jeffery@emce.fi

Growth differentiation factors -9 and -9B (GDF-9 and GDF-9B) form a distinct homology subset of the transforming growth factor-b (TGF-b) superfamily, and are important during follicular development of vertebrate species. Sequences for these proteins have been previously identified for 5 vertebrate species.

We overlapped whole genome shotgun (WGS) sequence fragments and expressed sequence tags (ESTs) to infer sequences for the mature regions of GDF-9 and GDF-9B in a larger number of vertebrates (10 species for GDF-9 and 12 for GDF9-B). This enabled cross-species comparisons of the proteins.

The resulting sequence alignments enabled us to perform preliminary cross-species sequence observations and predicted structural analyses. Preliminary modelling was performed for the two human proteins using the SWISS-MODEL algorithm. As a result of these analyses, interesting patterns of conservation and specialization were observed at putative receptor binding sights, consistent with suggested differences between GDF-9 and GDF-9B.

Home * List of Abstracts * Speakers

Defining Functional Networks of RNA-Protein Interactions in the Nervous System

Kirk Jensen
The Rockefeller University, New York, 10021 USA
Telephone: (212) 327-7471
Fax: (212) 327-7109
Email: jensenk@mail.rockefeller.edu

The exploration of the role played by RNA binding proteins in the nervous system continues to be an exciting and productive area of research. Neuronal RNA binding proteins function as alternative splicing factors, regulate mRNA stability and protein translation, and can localize mRNA to dendrites; thus these proteins are key regulators of gene regulation in neurons. However, a major obstacle in the further understanding of how neuronal RNA binding proteins regulate gene expression has been the systematic identification of RNA substrates. We have developed a new methodology to comprehensively identify RNA-protein interactions that occur in vivo, and present here data using this technique to identify RNA targets of the Nova and Hu RNA binding protein families.

Home * List of Abstracts * Speakers

Protein Microarray Technology

Thomas Joos
Biochemistry Department, NMI - Natural and Medical Sciences Institute at the University of T�bingen, Reutlingen, 72770 GERMANY
Telephone: +49 7121 51530 844
Email: Joos@nmi.de

Biochip technology allows the simultaneous analysis of thousands of molecular parameters within a single experiment. Most of the current applications focus on DNA array technology for gene expression analysis or on the detection of single nucleotide polymorphisms. However, any kind of ligand binding assay that uses an immobilised capture molecule for the detection of the binding of analyte from a solution can be miniaturised. Within the last few years, methods based on microarray technology have been adapted to the analysis of proteins and novel applications emerged. Protein microarrays offer the fascinating possibility to study protein interactions in a massively parallel fashion, including protein-protein, enzyme-substrate, protein-DNA or protein-drug interactions. Theoretical advantages and limitations of a miniaturized ligand assay system will be discussed.To illustrate the opportunities created by this technology examples of assay systems developed at the NMI will be presented.

1. Joos, T.O. *et al.* 2000. A microarray enzyme-linked immunosorbent assay for autoimmune diagnostics. Electrophoresis *21*:2641-2650.

2. Joos, T.O., D. Stoll and M.F. Templin. 2002. Miniaturised multiplexed immunoassays. Curr Opin Chem Biol *6*:76-80.

3. Templin, M.F. *et al.* 2002. Protein microarray technology. Trends Biotechnol *20*:160-166.

Dr. Thomas Joos
Head, Biochemistry Department, NMI - Natural and Medical Sciences Institute at the University of T�bingen, Markwiesenstr. 55, 72770 Reutlingen Germany
Joos@nmi.d
Phone +49 7121 51530 844

Thomas Joos is Head of the Department of Biochemistry at the Natural and Medical Sciences Institute (NMI) at the University of T�bingen. The NMI is a research foundation (non-profit organisation) founded in 1985. The principal goal of the NMI is to perform applied research for for industrial clients by transferring results from basic sciences into new technology and product development. In a multidisciplinary approach, scientists from the fields of applied and theoretical physics, chemistry, physical chemistry, biology and biochemistry are working together to archive efficient and innovative solutions at the interface between life sciences and material sciences. Dr. Joos has been with the NMI since 1998, where he is responsible for DNA, protein and cell biochip technology.

Prior to joining the NMI, Dr. Joos did his postdoctoral research in the laboratory of Prof. Peter Hausen at the Max-Planck-Institute of Developmental Biology, Department of Cell biology, studying cell-cell and cell-matrix interaction during early embryogenesis of *Xenopus laevis*. Dr. Joos studied biochemistry at the University of Tuebingen. He received his Ph.D. degree in 1995 on integrin-a_5 during early embroygenesis of *Xenopus laevis*.

RNA-Binding Proteins as Reporters of Operational mRNA Networks in Tumors and Complex Tissues

Jack D. Keene, Luiz O. Penalva, Michael Burdick, Hedi Sutterluety and Patrick J. Lager
Department of Molecular Genetics and Microbiology and Center For RNA Biology, Duke University Medical Center, Durham, NC 27710 USA
Telephone: (919) 6845138
Fax: (919) 6848735
Email: keene001@mc.duke.edu

Tumors and tissues consist of multiple cell types that are interdependent and interact with one another using secreted factors and other modes of cell-cell communication. For example, angiogenesis involves the secretion of factors by endothelial cells and by cancer cells that allow them to collaborate in tumor formation. Current methods of genomic analysis account for the steady state levels of messenger RNAs expressed in the whole tumor, but do not measure the expression of mRNAs within each cell type. In order to understand how individual cells within a society of communicating tumor cells affect the gene expression of surrounding cells, we have devised a new approach to gene expression profiling. By analyzing RNA-binding proteins within specific cell types that are associated endogenously with messenger ribonucleoproteins we can detect populations of mRNAs that are specific to each cell-type. In addition, we have engineered cells with variously epitope-tagged mRNA-binding proteins using specific promoters and virus-specific receptors, thus allowing us to profile the gene expression state of distinct cells within a mixture of cell types. Experiments based upon mixed cell populations of tumor cells and endothelial cells will be described showing the recovery of cell type-specific mRNA populations. In addition, using cDNA arrays we have demonstrated that certain endogenous mRNA-binding proteins can interact with unique subpopulations of mRNAs containing transcripts that encode functionally-related proteins. These mRNA subsets may be analogous to the polycistronic mRNAs of prokaryotic operons because they can assemble monocistronic mRNAs and coordinate their expression at the posttranscriptional level (Keene and Tenenbaum, *Mol Cell,* 2002). We have proposed that these mRNA subsets can function as operational genetic networks in that they contain both unique, as well as overlapping mRNA populations. Most importantly, these subpopulations have the ability to reassort combinatorially, potentially giving rise to the production of complex phenotypes. In order to elucidate operational mRNA networks within specific cell types of a tumor or complex tissue, we have immunoprecipitated endogenous mRNA-binding proteins that are tumor cell specific and are not present in the surrounding cells. These methods should allow precise determination of the mRNA populations within operational networks that control the expression of proteins required for cell-cell communication during tumor formation and in response to anti-tumor therapies.

Home * List of Abstracts * Speakers

Systematic Functional Analysis of Mouse Genes in the Region Syntenic to Human Xq28

Anja A. Kolb, Hanna Bausbacher, Annemarie Poustka
Division of Molecular Genome analysis, German cancer research center (DKFZ), Heidelberg, Germany
Telephone: +49-6221-424747
Fax: +49-6221-423454
Email: a.kolb@dkfz.de

The subchromosomal region Xq28 is a very well characterized part of the human genome with a high gene density. It is of special interest because a large number of diseases have been linked to this region. Although most of the genes have been identified and sequenced, the extent of their characterization at the level of expression and function varies. Some of the Xq28 genes, such as NEMO or DKC1, which cause Incontinentia Pigmenti (IP2) and X-linked recessive Dyskeratosis Congenita (DKC) respectively, have been identified as disease causing genes by our group. 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 hybridization. More than 30 different genes have been examined so far, some with known function like BGN or ATP6s1, but our main focus is on genes with presently unknown function. Limited information has been gained from sequence analysis for some of these genes, e.g. the chloride channel-like gene CLIC2 or the muscle-specific serine kinase MSSK1, whereas others are completely uncharacterized. The information obtained from RNA in situ hybridization experiments will give us useful hints towards understanding the function of such genes. At this stage, most of the genes under investigation show an ubiquitous expression pattern, but for almost half of these genes enhanced hybridization signals were associated with specific nerve cells of brain tissue. Such genes will be potential candidates for brain-related disorders linked to Xq28, such as several known forms of mental retardation syndrome. Our final goal is to evaluate the expression patterns for the majority of genes in Xq28. This will be of value for the identification of remaining disease-associated genes in this genomic region and will also provide integrated knowledge about the functional and evolutionary aspects of this model region.

Home * List of Abstracts * Speakers

Resources for Functional Research: Repository and Application

Bernhard Korn
RZPD, Heidelberg, 69120 GERMANY
Telephone: +49 (0)6221 42 4700
Fax: +49 (0)6221 42 4704
Email: korn@rzpd.de

With the multinational Human Genome Project, molecular biology and medicine entered large scale and high throughput approaches and applications. A new era of technology development, information collection and sharing of material was started. One of the most challenging next steps will be the understanding of the functional roles for all proteins. RZPD has started to establish a sequence-verified collection of full open reading frames (ORF) cDNAs representing more than 3.000 human genes. A shuttle vector system is used that allows protein expression in a wide range of hosts as well as in vivo and in vitro functional studies at different scales. In order to find out more about the influence of domains and secondary structure to the ability to express proteins in different systems, we perform highly parallel expression of ORFs in vitro and in vivo. In addition, the proteins are used in a matrix Two-Hybrid approach to gain more insights into human protein networks. The identification of interaction partners will be valuable to assign potential functions to proteins that are not characterised otherwise yet. RZPD is seeking to provide genomic and proteomic resources to the community. Beside genomic and cDNA clones and arrays, the development of material that is capable to directly support functional studies will be established, controlled and distributed.

Home * List of Abstracts * Speakers

RNAs Isolated from Ejaculate Spermatozoa Provide a Noninvasive Means to Investigate Testicular Gene Expression

G. Charles Ostermeier[1], Robert J. Goodrich[1], Kary Thompson[2], John Rocket[2], Michael P. Diamond[1], Karen Collins[1], NICHD Reproductive Medicine Network[1], David J. Dix[2], [3]David Miller and **Stephen A. Krawetz**[1,4,5]
[1]Department of Obstetrics and Gynecology, [4]Center for Molecular Medicine and Genetics, [5]Institute for Scientific Computing, Wayne State University, Detroit, MI, USA. [2]Reproductive Toxicology Division, National Health and Environmental Effects Research Laboratory, Environmental Protection Agency, Research Triangle Park, NC, USA, [3]University of Leeds, U.K.
Telephone: 313-577-6770
Fax: 313-577-8554
Email: steve@compbio.med.wayne.edu

Idiopathic male infertility will affect at least 6% of all couples. This is often not diagnosed until several IVF cycles have been completed. This places couples at unnecessary risk. The recent demonstration that mature ejaculate spermatozoa contain RNAs, may provide a window from which to view the gene expression history of spermatogenesis. If true, ejaculate spermatozoa could be used as non-invasive surrogates for testis infertility investigations. To assess whether spermatozoal transcripts encapsulate spermatogenic gene expression, a set of 27,016 unique expressed sequence tags (ESTs) was interrogated using cDNAs from a pool of testes representing 19 trauma victims and cDNAs from a pool of 9 individuals ejaculate spermatozoal mRNAs. The testes cDNA hybridized to 7,157 ESTs and contained all 3,281 ESTs identified by the pooled-ejaculate probe. The veracity of this observation was assessed by interrogating the same set of ESTs with cDNAs derived from the spermatozoal RNA of a single individual s ejaculate. All 2,784 of the ESTs identified were contained within the pooled testes cDNA and 2,780 of these were found within the pooled ejaculate probe. These findings support the view that a spermatozoal mRNA fingerprint can encapsulate the gene expression of spermatogenesis. The data further show that spermatozoal transcript populations vary among men. Prior to spermatozoal RNA based diagnosis of infertility,the range of normal must be established. We have begun to address this issue by comparing spermatozoal transcript profiles among normal fertile men. This analysis has revealed an invariant universal core and range of gene transcripts within ejaculate spermatozoa among fertile men. We expect that those RNAs which are invariant represent a core set of messages necessary for the production of normal fertile spermatozoa. Acknowledgments: This work was funded in part by NIH grant HD36512 and WSU grant 95200 to SAK and NIH grant HD39005 to MPD. GCO is supported in part by a Wayne State University's School of Medicine Dean's Post-Doctoral Fellowship award and the NICHD CIR-LRP. The information in this document has been funded in part by the U.S. Environmental Protection Agency. It has been subjected to review by the National Health and Environmental Effects Research Laboratory and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

Home * List of Abstracts * Speakers

BEST: A Computational Approach for Comparing Gene Expression Patterns from Early Stages of *Drosophila melanogaster* Development

Sudhir Kumar
Center for Evolutionary & Functional Genomics, Tempe, AZ 85287-1501 USA
Telephone: 480-727-6949
Fax: 480-965-2519
Email: s.kumar@asu.edu

Embryonic gene expression patterns are an indispensable part of modern developmental biology. Currently, investigators must visually inspect numerous images containing embryonic expression patterns to identify spatially similar patterns for inferring potential genetic interactions. The lack of a computational approach to identify pattern similarities is an impediment to advancement in developmental biology research because of the rapidly increasing amount of available embryonic gene expression data. Therefore, we have developed a computational approach to automatically compare the expression patterns contained in images of early stage Drosophila melanogaster embryos (prior to the beginning of germ band elongation); similarities and differences in gene expression patterns in these early stages have extensive developmental effects. Here we describe (a) the Basic Expression Search Tool (BEST) to retrieve best matching expression patterns for a given query expression pattern and (b) a computational device for gene interaction inference using gene expression pattern images and information on their genotypes and probes. The usage and impact of BEST for gene expression patterns is akin to that of the BLAST search for finding similar sequences. Analysis of a prototype collection of Drosophila gene expression pattern images is presented to demonstrate the utility of these methods in identifying biologically meaningful matches and inferring gene interactions by direct image content analysis. These Computational Developmental Biology methodologies are likely to make the great wealth of embryonic gene expression pattern data easily accessible and accelerate the discovery of developmental networks.

Home * List of Abstracts * Speakers

Crucial Elements in Array Data Evaluation

Michael Liebler and Till S�rensen
raytest GmbH, Aichtal, D 72631 GERMANY
Telephone: +49 7082 9255 0
Fax: +49 7082 20813
Email: liebler@raytest.de

Gene profiling in array technology is a quantitative method, in which signal detection results in grayscale image-data. The following data processing with evaluation softwares produces large amounts of result-data. Quality and reproducability of these data depends strongly on some essential steps of the computing process like spot alignment and background subtraction. The actual existing evaluation softwares for array experiments use different strategies for these evaluation steps. Investigations of these different evaluation strategies on slide or membrane shown, that these crucial elements have to be adepted to the various array designs. Possibilities for a quality control in form of spot rating will be discussed.

Home * List of Abstracts * Speakers

Modeling and Predictions of Estrogen Response Element Sequences in the Human Genome Reveal Potential Roles for Alu Repeats in Motif Propagation and Transcriptional Regulation

Chin-Yo Lin
Genome Institute of Singapore
1 Science Park Road, The Capricorn #05-01
Singapore, 117528 SINGAPORE
Telephone: 65-6827-5229
Fax: 65-6827-5204
Email: gislincy@nus.edu.sg

Estrogen receptors are ligand-dependent transcription factors that mediate hormone signaling in vertebrate development and diseases.[?] They bind DNA either directly through conserved estrogen response elements (EREs) or indirectly by binding other nuclear proteins.[?] The consensus response element sequence currently in use is based on limited information and is neither descriptive nor predictive of receptor binding sites in newly discovered hormone-responsive genes.[?] Here we present a comprehensive computational model composed of position weight matrices and transitional probabilities derived from published data.[?] Rigorous testing of the model using genomic sequences from different functional regions and organisms and extended promoter regions of human estrogen-responsive genes identified in a microarray study indicate that the model is sensitive and specific in detecting known and novel ERE-like patterns.[?] In our analysis, we identified an Alu repeat-associated ERE pattern as the most prevalent receptor-binding motif in the 5[?] regulatory regions of human genes.[?] These findings suggest an important role for mobile repetitive sequences in the generation of transcriptional regulatory cassettes during evolution and have implications in the evolutionary conservation of hormone response.[??]

Home * List of Abstracts * Speakers

The Next Step in Transcriptomics: Data Mining after Global Array Analysis and Design of a Sub-Array Specific for Cardiac Fibrosis of Renal Failure in the Rat Model

Christian Maercker[1], Heidrun Ridinger[1], Effi Rees[1], Matthias Schick[1], Christiane Rutenberg[1], Eberhard Ritz[2], Gerhard Mall[3], Bernhard Korn[1], and Kerstin Amann[4]

[1]Resource Center for Genome Research, DKFZ H0600, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany, maercker@rzpd.de, [2]Dept. of Internal Medicine, University of Heidelberg, Bergheimerstr. 56a, 69115 Heidelberg, [3]Dept. of Pathology, Grafenstr. 9, 64283 Darmstadt, [4]Department of Pathology, University of Erlangen, Krankenhausstr. 8-10, D-91054 Erlangen
Telephone: +49 6221 424741
Fax: +49 6221 423454
Email: c.maercker@dkfz.de

Death from cardiac causes is the most common fatality in uremic patients. The cardiac alterations, to which belong changes in blood pressure, electrolyte and hormone concentration, develop very early in renal insufficiency. We follow the approach of a gene expression profiling analysis in the rat animal model to identify the pathogenesis of the lesions.

Sprague-Dawley rats were subjected to subtotal nephrectomy (SNX) or sham operation (SHAM) and followed for 2 and 12 weeks, respectively. Poly(A)$^+$ RNA was used for expression profiling on RZPD-own global rat cDNA arrays (Rat Unigene-1, containing about 27.000 gene and EST sequences). After primary data analysis, only genes were extracted which showed a reproducible up- or down-regulation in all experiments. For data mining, genes were grouped as follows: a, All genes belonging to the group of the 10% strongest expressed genes, which were at least 2-fold up- or down-regulated, respectively; b, genes at least 5-fold up- or down-regulated, respectively; c, genes repeatedly regulated, which belong to the following gene families: 1. renin-angiotensin system (RAS) as potential participating hormone system, 2. extracellular matrix (ECM), 3. cell junctions (cell surface receptors, structural proteins), 4. signaling molecules (e.g. G-proteins, MAP/ERK cascade, second messenger, growth differentiation markers), 5. cytoskeleton (structural proteins, motor proteins), 6. growth factors. Genes and ESTs belonging to these groups were classified according to the GeneCards database (Weizmann Institute, http://bioinformatics.weizmann.ac.il/cards/) the GeneOntology database (http://www.geneontology.org/), and literature.

We identified about 150 genes strongly up- or down-regulated. Among them are not further characterized ESTs, but also genes responsible for cytoskeletal organization, energy metabolism, transport, signal transduction, etc.. Altogether, about 400 genes regulated could be classified into the selected gene families potentially involved in the cardiac lesions. In more detail, we found an up-regulation of the endothelin-receptor B (ETB), which is predominantly located on endothelial cells and which is known to up-regulated in left ventricular hypertrophy in a G protein dependent manner. Therefore, as shown for the kidney, there also might exist ETB specific answers in heart. As expected, the renin-angiotensin system (RAS) obviously is up-regulated as a inflammation reaction immediately after operation. Looking for downstream effector systems behind ETB, we found up-regulation of two Rac clones, of rhoB and ESTs similar to rhoC and rhoA. We also identified one member of the cytoplasmic mitogen-activated protein kinase/extracellular-signal regulated kinases (MAP/ERK), p38 mitogen-activated protein kinase (p38α MAPK), to be up-regulated in the 2w sample, as we would expect it in a ETB dependent manner. Up-regulated second messengers are protein kinase C and associated molecules (19 clones), other kinases (9 clones), phosphatases (9 clones), phospholipase C and associated proteins (10 clones). This is the reason, why we believe that the phosphosphatidylinositol pathway plays a major role in regulation of left ventricular hypertrophy. Expression of myosin genes (30 clones) and of genes from associated proteins, but also

other motor proteins, is regulated in our experiments. Although we also find down-regulated genes, most of them are up-regulated and, most interestingly, exclusively in the 12 w sample. A number of structural cytoskeletal proteins is necessary for cell-ECM contact and ECM formation. And indeed, we also found members of this gene family in our experiments. 22 clones encoding collagen subunits or enzymes involved in collagen turnover were up-regulated, most of them after two weeks (2 w). Only single clones show changes after 12 weeks, among them collagenase (UMCase), showing, that collagen turnover might steadily be continued. Procollagen C-proteinase enhancer protein, PCOLCE) was up-regulated after 2 and 12 weeks.

Whereas the majority collagens and proteoglycans obviously is up-regulated during the first days after operation, the 12 w group is dominated laminins, together with integrins. Therefore, initiated by the activation of the renin angiotensin system (RAS) at least two pathways seem to be involved in ECM activation, one going directly via G-proteins and second messengers (short term signaling), the other via motor proteins, actins and integrin (long term signaling). Some of the profiling data already could be confirmed by immuno-histochemical assays and in situ hybridizations (e.g. up-regulation of preproendothelin 1, endothelin 1, fibroblast growth factor). More experiments in this direction are under way. Currently, we are going for a cDNA subarray on glass slides and nylon membranes, respectively, containing about 1,000 genes potentially involved in cardiac failure and also about 100 “housekeeping” genes as controls. With this array we want to confirm the data obtained so far, but also investigate samples from other time points after operation, samples from animals treated with certain drugs before operation and other experimental setups.

Our work is supported by the BMBF, Germany.

Home * List of Abstracts * Speakers

Nonsense-Mediated mRNA Decay in Mammalian Cells: Evidence for a "Pioneer" Round of mRNA Translation

Lynne Maquat
University of Rochester Medical Center, Rochester, NY 14610 USA
Telephone: 585-273-5640
Fax: 585-172-2683
Email: lynne_maquat@urmc.rochester.edu

Nonsense-mediated mRNA decay (NMD) is a quality control mechanism used by cells to eliminate mRNAs that prematurely terminate translation (Maquat LE and Carmichael GG, 2001, *Cell*, 104: 173-176). As a general rule, nonsense codons elicit NMD when located within mRNA more than 50-55 nucleotides upstream of an exon-exon junction. Therefore, intron position within pre-mRNA is an important determinant of NMD.

Newly synthesized mRNA in mammalian cells is generally characterized by a splicing-dependent complex of proteins located approximately 20-24 nucleotides upstream of exon-exon junctions. Components of this complex are thought to recruit Upf3, a nucleocytoplasmic shuttling factor that functions in NMD. Upf3 and Upf2, another NMD factor, are detected on mRNA bound by the major nuclear cap binding proteins CBP80 and CBP20 but not mRNA bound by the major cytoplasmic cap binding protein eIF4E. These and other data indicate that NMD targets primarily CBP80-bound mRNA during a “pioneer” round of translation (Ishigaki Y, Li X, Serin G and Maquat LE, 2001, *Cell* 106: 607-17).

Data indicate that nuclear CBP80 but not nuclear eIF4E is readily detected in association with intron-containing RNA as well as the C-terminal domain of RNA polymerase II (Lejeune F, Ishigaki Y, Li X and Maquat, LE, 2002, *EMBO J*., 21: 3536-3545). Consistent with this, components of the exon-exon junction complex, including RNPS1, Y14, SRm160, REF/Aly and TAP, as well as Upf3 and Upf2, are detected in the nuclear fraction on CBP80-bound but not eIF4E-bound mRNA. Each of these components is also detected on CBP80-bound mRNA in the cytoplasmic fraction, indicating a presence on mRNA after export. Data suggest that a pioneer round of translation generally elicits the NMD of CBP80-bound mRNA harboring a nonsense codon located more than 50-55 nucleotides upstream of an exon-exon junction. Otherwise, mRNP is remodeled and becomes immune to NMD. The dynamics of mRNP remodeling and additional factors involved in NMD will be discussed.

Home * List of Abstracts * Speakers

Biological Roles for A to I Conversion in Untranslated Regions of mRNAs?

Daniel P. Morse[1], P. Joseph Aruscavage[2], Leath A. Tonkin[2], Lisa Saccomanno[2], Thomas Brodigan[3], Michael Krause[3], and **Brenda L. Bass**[2]
[1]Department of Chemistry, United States Naval Academy; [2]Department of Biochemistry and Howard Hughes Medical Institute, University of Utah; [3]Laboratory of Molecular Biology, NIDDK, NIH
Telephone: 410-293-6612
Fax: 410-293-2218
Email: morse@usna.edu

A family of RNA editing enzymes called adenosine deaminases that act on RNA (ADARs) convert adenosine to inosine within double-stranded regions of metazoan RNA. Since inosine is read as guanosine by the translation machinery, editing in coding regions of mRNAs can result in codon changes. Editing of this type in mammals results in the production of multiple, functionally distinct forms of a glutamate receptor subunit (GluR-B) and the serotonin 2C receptor.

Using a largely unbiased search for inosine-containing mRNAs in *C. elegans*, we were surprised to find that most of the inosine occurred in untranslated regions or introns rather than in coding regions. We wondered whether this was also true in mammals so, using the same method, we searched for ADAR substrates in human brain. We identified 19 new human brain ADAR substrates, each of which was edited only in non-coding regions. These results suggest that most inosine occurs in non-coding regions and that codon changes are the exception rather than the rule.

To further our understanding of ADAR function, we generated worms with deletions in one or both of the *C. elegans* ADAR genes. The observed mutant phenotypes and the effect on RNA editing will be discussed. More work will be required to determine whether alterations of editing in non-coding regions of specific ADAR substrates are responsible for the mutant phenotypes.

Home * List of Abstracts * Speakers

Kazusa cDNA Project 2002: Of Mice and Men

Osamu Ohara
Kauzsa DNA Research Insitute/RIKEN Research Center for Allergy and Immunology,
Kisarazu, 292-0818 JAPAN
Telephone: +81-438-52-3913
Fax: +81-438-52-3914
Email: ohara@kazusa.or.jp

We have conducted a human cDNA sequencing project for identification of unknown human transcripts in the past 8 years. A distinctive point of our project from other cDNA sequencing projects is in that we have focused our sequencing efforts on long cDNA clones (>4 kb), particularly those encoding large proteins in brain. This approach has brought us a unique cDNA resource consisting of large cDNA clones for newly identified human genes, which are known as KIAA cDNAs. The number of KIAA cDNAs has exceeded 2000 this year (the total number of the sequenced nucleotide residues, about 10 Mb) and we have made them publicly available. Because the number of genes encoding large proteins (>100 kDa) is always smaller than 10% of the total number of genes in eukaryotes, the number of human genes encoding large proteins is likely less than 3000. If this estimate is correct, we must have certainly entered endgame of the identification of human transcribed sequences expressed in the brain. Thus, this has strongly urged us to move toward the second phase of our project, *i.e.*, toward functional analyses of the newly identified transcripts and their products through exploiting fully the accumulated cDNA resource. In this regard, we have already started to perform expression profiling of KIAA genes on nylon cDNA microarray and to analyze protein networks involving KIAA proteins on the basis of yeast two-hybrid screening.

One of our goals beyond the identification of KIAA transcripts is to understand the function of KIAA proteins *in vivo*. For this purpose, we consider it reasonable to perform analysis of KIAA proteins in mice rather than in human because isolation of human proteins from various tissues might raise a serious ethical issue and various types of genetically engineered mice are currently available. We have thus initiated to prepare a set of mouse KIAA cDNAs for analysis of their products *in vivo*. Besides the characterization of KIAA proteins *in vivo*, the isolation of mouse KIAA cDNAs has enabled us to experimentally evaluate the integrity of KIAA cDNA clones through comparison of human and mouse KIAA cDNAs. Furthermore, in this mouse KIAA cDNA project, we have also planned to prepare a set of antibodies against all the mouse KIAA proteins, which will provide us with a powerful tool to explore the structure and function of the mouse KIAA proteins *in vivo*. In this presentation, more details regarding the current status of human and mouse KIAA cDNA projects will be described.

Home * List of Abstracts * Speakers

Genome-Wide Analysis of the Sequence Region Surrounding the Transcription Start Site of Human mRNAs

Michele Iacono, Carmela Gissi, Giorgio Grillo*, Flavio Mignone, Sabino Liuni and **Graziano Pesole**
Dipartimento di Fisiologia e Biochimica Generali, Università di Milano, Italy
Telephone: +39-02-50314915
Fax: +39-02-50314912
Email: graziano.pesole@unimi.it
*Sezione di Bioinformatica e Genomica, Istituto Tecnollogie Biomediche, CNR, Bari, Italy

Gene expression is finely regulated at both the transcriptional and post-transcriptional level. Transcriptional control is mediated by transcription factors, RNA polymerase and a series of cis-acting elements located in the DNA, mostly in the core promoter region, i.e. in close proximity to the transcription start site (TSS).

The recent availability of the human genome draft as well as of a very large number of full length transcript sequences makes now possible to carry out an extensive and systematic study of the genomic context of the TSS.

We present here a comprehensive sequence analysis of the sequence region spanning from -500 to +100 with respect to the TSS of human genes to investigate the frequency of TATA-box, Inr and CpG islands also in correlation to their expression pattern. Furthermore, through the application of pattern discovery algorithms we searched for novel sequence elements, particularly downstream the TSS, possibly involved in transcription regulation.

Home * List of Abstracts * Speakers

Global Analysis of mRNA Decay in T Cells

Arvind Raghavan and Paul R. Bohjanen
University of Minnesota, Minneapolis, MN 55455 USA
Telephone: 612-624-0469
Fax: 612-623-0623
Email: ragah002@tc.umn.edu

Regulation of gene expression at the level of mRNA stability facilitates rapid, selective and temporally precise responses to activation stimuli. For example, rapid decay of cytokine and proto- oncogene transcripts induced by T cell receptor (CD3) and co-receptor (CD28) stimulation is essential for the normal differentiation and proliferation of activated T cells. We have used Affymetrix oligonucleotide arrays to profile mRNA decay rates of transcripts in purified human T cells that were either unstimulated or were stimulated for 3 hours with anti-CD3 and/or anti- CD28 monoclonal antibodies. By arresting transcription with Actinomycin D, we derived mRNA decay curves for each of approximately 6,000 transcripts expressed in T cells under conditions of rest and activation. While the majority of transcripts in resting T cells were stable (t1/2 > 360 min.), we identified many short-lived transcripts (t1/2 < 60 min.) that encoded important regulators of cell-surface signaling, transcription, and apoptosis. Also, T cell stimulation led to dramatic changes in transcript decay rates. For example, hundreds of transcripts were repressed and destabilized following T lymphocyte activation. We also identified transcripts, including cytokine transcripts, that showed CD28 co-stimulation-dependent stabilization, in accordance with previously published results. Although numerous transcripts with rapid decay rates were found to contain AU-rich element (ARE)- like sequences, many transcripts that are regulated at the level of mRNA stability contain no previously characterized stability determinants. We found that many transcripts encoding critical components of T cell receptor signaling pathways are coordinately destabilized and downregulated following T cell activation, suggesting that mRNA decay may provide a mechanism for coordinate regulation of gene expression.

Home * List of Abstracts * Speakers

The Mouse Gene Expression Database (GXD) and the Gene Expression Notebook (GEN)

C.M. Smith, D.A. Begley, I.J. McCright, T.F. Hayamizu, D.P. Hill, S.R. Pretel, P. Mani, J.S. Beal, D.J. Dahmen, J.A. Blake, J.T. Eppig, J.A. Kadin, J.E. Richardson, and **M. Ringwald**
The Jackson Laboratory, Bar Harbor, ME 04069, USA
Telephone: (207) 288-6436
Fax: (207) 288-6132
Email: ringwald@informatics.jax.org

The Gene Expression Database (GXD) is a community resource of gene expression information from the laboratory mouse. By combining different types of expression data, GXD provides increasingly complete information about expression profiles of transcripts and proteins in different mouse strains and mutants. Integration with the Mouse Genome Database (MGD) and the Mouse Sequence Database (MGS) enables a combined analysis of genotype, sequence, expression, and phenotype data. Continued development and use of controlled vocabularies and classification schemes, such as those being built as part of the Gene Ontology project, and extensive interconnections with other databases place the expression information in the larger biological and analytical context.

During the last year, we have increased the utility of GXD by refining query and display interfaces and by adding and making new data available on a daily basis. Further, we have developed a new and much extended version of the Gene Expression Notebook (GEN), a user-friendly and customizable tool to store and organize expression data and images in the laboratory and to submit data for inclusion in GXD. Recent developments will be presented and new ways to integrate and correlate expression and phenotype data will be discussed.

GXD is accessible through the Mouse Genome Informatics web site at http://www.informatics.jax.org/ or directly at http://www.informatics.jax.org/menus/expression_menu.shtml.

GEN is freely available for download at http://www.informatics.jax.org/mgihome/GXD/GEN/.

GXD is supported by NIH grant HD33745. The Gene Ontology project is supported by NIH grant HG02273.

Home * List of Abstracts * Speakers

Big Genes: What? Where? Why?

Lee Rowen, Brian Birditt, Ryan Nesbitt, Amardeep Kaur, Gustavo Glusman, and Lee Hood
Institute for Systems Biology, Seattle WA 98103
Telephone: 206-732-1272
Fax: 206-732-1299
Email: leerowen@systemsbiology.org

As more of the human genome is represented in multimegabase-sized contigs and genomes of other vertebrates get drafted and scaffolded, features of overall genomic architecture can be investigated with more specificity.�? Of interest to us are the big genes, that is, genes whose primary transcripts (pre-mRNA) span over 500 kilobases.? To date, we have identified about 200 such genes accounting for about 5% of the human genome.

The big genes fall into several functional classes, the most noteworthy being their involvement in the development or physiology of the central nervous system.? Examples include glutamate receptors and neurexins.? Many big genes are loci of chromosomal instabilities -- trinucleotide repeat expansions, large deletions, and translocations.? The alternative splicing of some of the big genes is highly complex, and it is not obvious how cryptic splicing within the introns is avoided or detected and repaired.

Gene size may be an important mechanism for controlling the expression of some genes -- if the time required for transcription is longer than the cell division time, then the proteins will presumably not be made.? On the other hand, gene size could simply be an outcome of vertebrate evolution and there may be little functional significance to intron size.? Interestingly, the trivial explanation that big genes are big because the introns are full of transposable elements (interspersed repeats) is not true, unless the repeats are so old that their sequences have diverged far away from the consensus sequences.? We will analyse a selected subset of big genes in the mouse and pufferfish genomes to address the issue of whether genes that are big in human are also big in other species.

Home * List of Abstracts * Speakers

Full Domain Transcript of Alpha-Globin Genes is a Component of the Nuclear Matrix

A.V. Rynditch[1], V.P. Smalko[1], S.V. Razin[2], E. Loudinkova[2], K. Scherrer[3]

[1]Institute of Molecular Biology and Genetics, National Academy of Sciences of the Ukraine, Kiev, Ukraine; [2] Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia; [3] Institut J. Monod, 2, Place Jussieu, 75251 Paris, France

The investigation concerns the nature and possible significance of the Full Domain Transcript (FDT) of the alpha-globin genes in erythroleukaemic cells where globin gene expression is abortive. Using Northern blot hybridisation with strand-specific probes recognizing alpha-globin genes and intergenic sequences, we demonstrate that the entire locus of alpha-globin genes spanning ~ 30 kb between DNaseI hypersensitive sites is transcribed. The 3'-end of the transcribed area is mapped about 1500 bp downstream of alpha A, the last of the 3 alpha-type globin genes. Using primers downstream of the last gene and amplification of the cDNA with primer pairs placed way upstream, RT-PCR shows that about 30 kb of the sequenced domain in transcribed into one continuous FDT. This extends from the upstream control region over the replication origin and DNA loop anchorage site of the domain to the enhancer/silencer located downstream of the adult alpha A gene. RT-PCR and in situ hybridization with globin genic and extra-genic probes demonstrated the presence of such globin FDTs as a component of the nuclear matrix. On dividing AEV-transformed cells, FDTs were found to accumulate around the nucleoli, and to be excluded from those in matrix preparations. We propose that globin FDTs are the part of the dynamic nuclear architecture in the processing pathway and a component of the matrix, turning over slowly in AEV cells where globin gene expression is blocked post-transcriptionally.

Home * List of Abstracts * Speakers

Molecular Evolutionary Analyses of HoxA Gene Cluster of Human and Apes

Naruya Saitou
National Institute of Genetics, Mishima, 411-8540 JAPAN
Telephone: 81-55-981-6790
Fax: 81-55-981-6789
Email: nsaitou@genes.nig.ac.jp

Molecular evolutionary analyses of HoxA gene cluster of human and apes Kim Choong-Gon, Kitano Takashi, and Saitou Naruya Division of Population Genetics, National Institute of Genetics, Mishima, Japan Nucleotide difference between human and chimpanzee, the closest living organism to human, is about 1.2%. Therefore, half (0.6%) of this difference accumulated after the human lineage diverged from the chimpanzee lineage. All the genetic changes responsible for "humanness" must reside in those differences. We thsu started small scale ape genome sequencing called "Silver Project" (http://sayer.lab.nig.ac.jp/~silver/). As pilot project, we conducted nucleotide sequencing of Hox A gene cluster for chimpanzee, gorilla, and orangutan. We first designed PCR primers based on nucleotide sequences of human HoxA cluster available from DDBJ/EMBL/GenBank International Nucleotide Sequence Database, then did direct sequencing those PCR-amplified product using ape genomic DNA as template. A total of 21,000 bp were determined for chimpanzee, gorilla, and orangutan, including HoxA4, HoxA5, HoxA6, and HoxA7 genes as well as intergenic regions. We also constructed 3.5x gorilla fosmid library. Using this gorilla library as well as chimpanzee BAC library, the whole HoxA cluster sequences of about 100 kb were determined for chimpanzee and gorilla. We took approach of comparing hominoid as a whole, instead of conventional pairwise comparison. By this way, we could identify several short DNA regions where only few nucleotide substitutions accumulated during hominoid evolution.

Home * List of Abstracts * Speakers

Point Mutations Occur at a Faster Rate in Exons than in Noncoding DNA in Primate Genomes

Subashchandran Sankarasubramanian and Sudhir Kumar
Arizona State University, Tempe, AZ 85281 USA
Telephone: 480-965-2091
Fax: 480-965-2519
Email: sankar@asu.edu

Currently point mutation rate in exons and noncoding (introns and intergenic) regions are assumed to be the same. However, comparative sequence analyses of synonymous substitutions in exons and that of long fragments of human and chimpanzee genomes reveal up to a 50% higher mutation rate in exons than in noncoding DNA. We propose a differential CpG content hypothesis to explain this fundamental, and seemingly unintuitive, pattern. Expectations and predictions of this hypothesis are confirmed in an analysis using 151 protein coding genes, 42 pseudogenes, 25 introns and 202 kb of intergenic region between human and chimpanzee, gorilla, orangutan, macaque and baboon comparisons. We find that the mutation rate in non-CpG sites is the same in noncoding DNA and in the synonymous sites of exons; the increased exonic rate is the result of the overabundance of CpG dinucleotides. The rate of CpG decay is approximately constant over time in human, apes and old world monkey genomes. These results explain the excessive human single nucleotide polymorphisms in exons compared to noncoding regions observed in a study by Cargill et al. (1999), suggest an underlying mechanism for the higher mutation rate in GC-rich genes, and provide a new tool for estimating divergence times between human populations and closely related primates.

Home * List of Abstracts * Speakers

Two Pancreas Clone Sets and PancChips, New Resources for Diabetes

Marie Scearce and Klaus Kaestner
University of Pennsylvania, Philadelphia, PA 19104 USA
Telephone: 215-743-6372
Fax: 215-573-5892
Email: mscearce@mail.med.upenn.edu

In order to investigate gene expression in the developing pancreas, we have compiled two pancreas-enriched clone-sets and corresponding microarrays. Our initial 3.4K pancreas clone set contains unique clones representing genes expressed in the pancreas as well as known genes of interest which were identified through preliminary array experiments with commercially available arrays and through a search of dbEST libraries for libraries derived from pancreatic tissue. Non-redundant clones (3139) from all sets were obtained from the IMAGE consortium these clones were re-sequenced by the Genome Sequencing Center and 1,898 of the clones were sequence verified. In addition to this clone set we have identified approximately 260 clones of interest from know signal transduction pathways and other genes of interest and added these to the clone set. Our second clone set is compiled from libraries prepared in the Melton and Kaestner laboratories. A unique set of 7227 clones from these libraries were identified, of which approximately 60% are either novel clones or without functional groups previously deposited in dbEST. These clones were re-sequenced by the IMAGE consortium. The combined clone set consists of approximately 11,000 cDNAs and ESTs representing over 10,000 non-redundant mRNAs expressed in mouse pancreas. Plasmid DNA from the clones was isolated and the inserts were amplified with PCR. The purified PCR products were printed on poly-L-lysine coated glass microscope slides. We have used PancChip 2, (from the 3.4K clone set) to study the expression of genes in the developing pancreas at pc14.5, pc16.5, pc18.5, newborn, 7-day-old and adult mice.

Home * List of Abstracts * Speakers

Applying Genome-Wide SNP Analysis for Complex Disease Gene Discovery

Michael Shi
Sequenom, Inc., San Diego, CA 92121 USA
Telephone: 858-202-9122
Fax: 858-202-9127
Email: mshi@sequenom.com

Although extensively investigated in the past decade, genetic linkage and association studies for dissection of complex genetic traits have not yet delivered many fruitful results. Recently, genome-wide single nucleotide polymorphism (SNP) association studies are generating new hopes and becoming the method of choice for common disease gene identification. However, great technical challenges exist for this kind of approach, including a requirement of high-density informative SNP map, cost-effective technology for large- scale genotyping as well as high quality clinical sample and phenotype collection. Sequenom has developed novel approaches to provide solutions to address every aspect of the challenges to modern human genetics. By utilizing an integrated bioinformatics and a chip-based MALDI-TOF (MassARRAYTM) genotyping technology platform, we have generated a SNP map consisting over 300,000 validated gene-based and evenly spaced SNPs with confirmed allele frequencies across the entire human genome. Through internal collection effort and the recent acquisition of Gemini Genomics, we have access to over 50,000 distinct human clinical samples in many disease areas with enriched clinical information. In addition to case-control and age stratified healthy populations, many of our collection are ideal for complex genetic trait analysis by limiting the environmental factor impact, including unselected monozygous and dizygous twins, family and sibs, founder population and longitudinal disease cohorts. By implementing genome-wide high-throughput SNP analysis in our well-characterized clinical patient samples, we have identified many potential disease genes and novel drug targets that have significant impact on the human health.

Home * List of Abstracts * Speakers

Large-Scale Subcellular Localisation and Functional Analysis of Novel Proteins

Jeremy C. Simpson[1], Urban I. Liebel[1], Vytaute Starkuviene[1], Holger Erfle[1],[?] Annemarie Poustka[2], Stefan Wiemann[2] and Rainer Pepperkok[1]

[1]Cell Biology and Biophysics Programme, EMBL, 69117 Heidelberg, Germany and [2]Molecular Genome Analysis, DKFZ, 69120 Heidelberg, Germany
Telephone: +49 6221 387232
Fax: +49 6221 387306
Email: simpson@embl-heidelberg.de

One of the greatest challenges facing biology today is the conversion of huge amounts of genomic data into more tangible functional information about the proteins encoded. For example, the large-scale cDNA sequencing project of the German cDNA Consortium is providing us with vast numbers of open reading frames (ORFs) encoding novel proteins of completely unknown function.As a first step towards their characterisation we have tagged over 300 of these with the green fluorescent protein (GFP), and examined the subcellular localisations of these fusion proteins in living cells. These data have allowed us to classify[?] these proteins into subcellular groups which in turn determines the next appropriate steps to characterise them. To make further use of these GFP-tagged constructs, a series of functional assays have been designed and implemented to assess the effect of these novel proteins on processes such as cell growth, organelle morphology and protein transport.

Functional assays with such a large set of molecules has only been made possible through the introduction of automation. In this respect we have developed a fully automatic microscope which is integrated with a robotic liquid handling station. Results from the first series of assays have already allowed us to identify proteins which localise to distinct organelles of the secretory pathway and appear to be new regulators of different steps in protein transport. Molecular interaction studies with these factors has extended our screens and identified novel and known interactors. Altogether, this approach will ultimately allow us to identify functional networks of proteins in a morphological context, and will greatly contribute to our understanding of whole cell function.[??]

Home * List of Abstracts * Speakers

Statistical Validation of Pattern Prevalence in Regulatory Gene Loci

Gautam B. Singh and Stephen A. Krawetz** ********
*Computer Science & Engineering Department, Oakland University, Rochester, MI 48309, USA; FutureSoft Corporation, Rochester Hills, MI 48309, USA; Michigan Center for Biological Information, WSU Node
gbs@futuresoft.org
** Michigan Center for Biological Information, WSU Node, Wayne State University School of Medicine, Detroit, MI 48202
steve@compbio.med.wayne.edu

Gene expression in cells is controlled by the presence of DNA sequences called Matrix Association Regions (MARs) or Scaffold Association Regions (SARs). These sequences are embedded within the majority of nuclear non-coding DNA.* MARs regions are found to be distributed throughout the chromatin and flank the ends of genic domains encompassing various transcriptional units. It has also been shown that MARs bring together the transcriptionally active regions of chromatin such that transcription is initiated in the region of the chromosome that coincides with the surface of the nuclear matrix. Specific patterns of nucleotide arrangements have been identified in these non-coding DNA sequences, with varying functions. These examples indicate that the *patterns* embedded in the eukaryote DNA play an important role in its viability. Examples of these patterns include the A+T or G+C rich regions, telomeric repeats of AGGGTTin human DNA, the rare occurrence or absence of dinucleotides, e.g., TA and GC, and tetranucleotides, e.g., CTAG, and the GNN periodicity in the gene coding regions.

In this talk we will discuss the prevalence of MAR specific patterns with in the set of sequences collected within the SMAR-DB, as well as an independently collected set of sequences from various Origins of Replication (ORI). Interesting statistics related to curved, kinked, AT-rich regions, and other MAR-related patterns are presented. The actual occurrence of patterns is compared with their expected occurrence with the result of the log-ratio test used for establishing whether the putative patterns are an evidence of the occurrence of MARs. The patterns that are positively linked with MARs are further analyzed for periodicity. This is the first step towards testing the hypothesis that MAR periodicity may be associated with DNA packaging. The statistical analysis is presented within the Bayesian context.

A MAR-Pattern Distributed Annotation Sources (DAS) will be linked to the ENSEMBL database as part of the Michigan Center for Biological Information (MCBI) Genome databases.* This will provide MAR-related pattern loci information within the context of the existing annotations that are targeted towards gene transcription. This will hopefully help us better appreciate the correlation between the transcription apparatus and the DNA structure.*

Home * List of Abstracts * Speakers

Complex Transcriptional Circuitry at the G1/S transition in *Saccharomyces cerevisiae*

Christine E. Horak, Nicholas M. Luscombe, Jiang Qian, Stacy Piccirrillo, Mark Gerstein and **Michael Snyder**
Department of Molecular, Cellular and Developmental Biology and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520-8103
Telephone: 203 432 6139
Fax: 203 432 3597
Email: michael.snyder@yale.edu

In the yeast *Saccharomyces cerevisiae*, SBF (Swi4-Swi6 cell cycle box binding factor) and MBF (MluI binding factor) are the major transcription factors regulating the START of the cell cycle, a time just before DNA replication, bud growth initiation and spindle pole body duplication. These two factors bind to the promoters of 235 genes, but bind less than a quarter of the promoters upstream of genes with peak transcript levels at the G1 phase of the cell cycle (Iyer et al. 2001). Several functional categories, which are known to be crucial for G1/S events, such as spindle pole body duplication/migration and DNA synthesis are underrepresented in the list of SBF and MBF gene targets. SBF binds the promoters of several other transcription factors including HCM1, PLM2, POG1, TOS4, TOS8, TYE7, YAP5, YHP1 and YOX1. Here, we demonstrate that these factors are targets of SBF using an independent assay. To further elucidate the transcriptional circuitry that regulates the G1 to S phase progression, these factors were epitope-tagged and their binding targets were identified by chIp-chip analysis. These factors bind the promoters of genes with roles in G1/S events including DNA replication, bud growth, spindle pole complex formation, as well as the general activities of mitochondrial function, transcription and protein synthesis. While functional overlap exists between these factors and MBF and SBF, each of these factors has distinct functional roles. Most of these factors bind the promoters of other transcription factors known to be cell cycle regulated or known to be important for cell cycle progression and differentiation processes indicating that a complex network of transcription factors coordinates the diverse activities that initiate a new cell cycle.

Home * List of Abstracts * Speakers

Regulated Alternative Splicing: A New Dimension of the Human Genome

Stefan Stamm
Institute of Biochemistry; Fahrstra◆e 17; 91054 Erlangen; Germany
e-mail: Stefan@stamms-lab.net

The human genome project demonstrated that alternative splicing of gene is more the rule than the exception. Missplicing events are an important cause and indication of human disease. Changing alternative splicing patterns in response to an external stimulus seems to a physiological process performed by many cells. Organisms regulate alternative splice site selection by changing the concentration and activity of splicing regulatory proteins.

The SAM68 like molecules SLM-1 and SLM-2 provide good examples of how phosphorylation influences splice site selection. Both nuclear proteins are phosphorylated by tyrosine kinases, such as fyn, abl and sik. Mass-spec analysis identified three tyrosine phosphorylated residues in SLM-2. One of these tyrosines seems to regulate the intracellular localization of rSLM-2. Using yeast three hybrid and immunoprecipitation assays, we showed that phosphorylation influences the binding of several interacting proteins. Finally, in vivo splicing assays demonstrated that the regulation of the survival of motoneuron gene 2 (SMN2) by rSLM-1 is influenced by phosphorylation.

Tau exon 10 provides another example of phosphorylation dependent splice site selection. Tau is a microtubule-associated protein whose transcript undergoes regulated splicing in the mammalian nervous system. Exon 10 of the gene is an alternatively spliced cassette that is adult-specific and encodes a microtubule-binding domain. Mutations increasing the inclusion of exon 10 result in the production of tau protein which predominantly contains four microtubule-binding repeats and were shown to cause frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17). Exon 10 usage is regulated by CDC2 like kinases CLK1, 2, 3 and 4 that phosphorylate serine-arginine rich proteins, which in turn regulate pre-mRNA splicing. Cotransfection experiments suggest that CLKs achieve this effect by releasing specific proteins from nuclear storage sites.

Our data showed that phosphorylation is a physiological mechanism to regulate splice site selection. Given the prevalence of alternative splicing, its regulation by external signals enhances the genetic information and offers a mechanism for adaptation according to cellular needs. Finally, our results could lead to new therapeutic approaches for spinal muscular atrophy and tauopathies.

References:

Stamm, S. (2002) Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome. *Hum. Mol. Genet.*, **11**, in press.

Daoud, R., Mies, G., Smialowska, A., Ol◆h, L., Hossmann, K. and Stamm, S. (2002) Ischemia induces a translocation of the splicing factor tra2-beta1 and changes alternative splicing patterns in the brain. *J. Neurosci.*, **22**, 5889-5899.

Stoss, O., Olbrich, M., Hartmann, A.M., Konig, H., Memmott, J., Andreadis, A. and Stamm, S. (2001) The STAR/GSG family protein rSLM-2 regulates the selection of alternative splice sites. *J Biol Chem*, **276**, 8665-8673.

Hartmann, A.M., Rujescu, D., Giannakouros, T., Nikolakaki, E., Goedert, M., Mandelkow, E.M., Gao, Q.S., Andreadis, A. and Stamm, S. (2001) Regulation of alternative splicing of human tau exon 10 by phosphorylation of splicing factors. *Mol Cell Neurosci*, **18**, 80-90.

Genome-Wide Identification of Differences in the Retroelement Integrations between Humans and Chimpanzees

E. Sverdlov, A. Buzdin, I. Mamedov, S. Ustugova, E. Gogvadze, Yu.
Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, 117997 RUSSIA
Telephone: 7 (095) 330 6529
Fax: 7 (095) 330 6538
Email: sverd@humgen.siobc.ras.ru

Lebedev Sequences derived from transposable elements transposing through RNA intermediates (retroelements, REs) - long interspersed elements (LINEs), short interspersed elements (SINEs), LTR retrotransposons - comprise 13, 20, and 8% of the genome, respectively. They may play an important role in speciation of Hominoids providing new regulatory modules capable of changing gene expression networks. The identification of human specific integrations in the vicinity of human genes could reveal candidate regulatory elements within REs capable of affecting gene regulation and thus influencing the split of the human and other hominoid lineages. We have developed a method of Targeted Genomic Difference Analysis (TGDA) for genome-wide detection of differences in integration sites of interspersed repeats between related genomes. The method includes two principal steps: (i) a whole genome selective amplification of the flanks adjacent to target interspersed repetitive elements in both genomic DNAs under comparison, and (ii) subtractive hybridization of the selected amplicons. Differences between the human and chimpanzee genomes in the integration sites of HERV-K (HML-2) retroviruses and related solitary long terminal repeats (LTRs) were analyzed. Of 55 sequenced clones randomly chosen from a library enriched with human specific integration (HSI) sites, 33 (60%) represented HSIs. Together with HSIs described by other authors, the number of characterized LTR HSIs is increased to 40. Using these 40 human-specific LTR sequences, we have derived a consensus sequence for an evolutionary young HERV-K LTRs named the HS family. In the human genome the HS family is represented by ~150 LTR sequences, 90% of them being human-specific. The family can be subdivided into two subfamilies of 5.8 and 10.3 Myr evolutionary ages. We found human-specific HERV-K LTRs integrations in introns of 17 human genes, most of which being oriented in the opposite orientation with regard to the direction of gene transcription. TGDA has been applied also to genome-wide screening of human specific L1 integrations and their polymorphisms. We obtained a library highly enriched in human specific L1 insertions, and identified 24 new such insertions. Many of these insertions are polymorphic in human populations. The results suggest that TGDA is a universal method that can be successfully used for the detection of evolutionary and polymorphic markers in any closely related genomes. The research was supported by INTAS-01-0759, RFBR 00-15-97945 and RFBR 02-04-48614 grants, and by contract 43.073.1.1.2508 of the Ministry of Industry, Science and Technology of the Russian Federation.

Home * List of Abstracts * Speakers

Array Analysis of mRNP Complexes Reveals Clustered mRNA Subsets that May Represent Posttranscriptional Operons

Scott A. Tenenbaum and Jack D. Keene
Department of Molecular Genetics and Microbiology and The Center for RNA Biology, Duke University Medical Center, Durham, NC 27710 USA
Telephone: 919 684-2714
Fax: 919 684-8735
Email: tenen001@mc.duke.edu

RNA-binding proteins are essential in regulating posttranscriptional gene-expression in eukaryotes and are responsible for generating much of the diversity of the proteome.[illegible] We have developed methods for purifying endogenously formed mRNP-complexes that, when coupled with microarray technologies, allow the rapid identification of multiple mRNA targets, quantitatively, and *en masse* (Tenenbaum et al. (2002) *Methods* 26, 191). The primary steps involved in characterizing mRNA subsets clustered by mRNA-associated proteins include; (1) isolation of endogenously formed mRNP complexes, (2) *en masse* identification of clustered mRNA subsets using array analysis, (3) identification of similar cis-elements among clustered mRNAs, and (4) determination of functional relationships among the protein products coded for by the mRNAs in a subset.[illegible] This *ribonomics* approach has been used to identify unique mRNA profiles for several RNA-binding proteins including ELAV/HuB (Tenenbaum et al. (2000) *PNAS* 97, 14085), FMRP (Brown et al. (2001) *Cell* 107, 477), and a recently identified autoantigen GW182 (Eystathioy et al., (2002) *MBC,* 13, 1338).[illegible]

The ribonomics approach to functional genomics has revealed three novel findings; (1) mRNA binding proteins are associated with unique subpopulations of messages, 2) the composition of these mRNA subsets can vary with cellular conditions and (3) the same mRNA species can be found in multiple mRNP complexes. Based on these data, we proposed a model of posttranscriptional gene expression in which mRNA-binding proteins regulate mRNAs as fluctuating subpopulations (Keene and Tenenbaum (2002) *Molecular Cell* 9, 1161. This model predicts that functionally related genes are regulated posttranscriptionally as subpopulations by specificmRNA-binding proteins that recognize sequence elements in common among the clustered mRNA transcripts. This suggests that an mRNP infrastructure exists within mammalian cells consisting of mRNAs that are networked by multi-targeted RNA-binding proteins.[illegible] If subpopulations of monocistronic mRNAs are coordinately regulated *en masse*, their protein products may participate in the same biological process or pathway, thereby providing a posttranscriptional analog to the polycistronic operon.[illegible] The combinatorial reassortment of mRNA transcripts associated with RNA-binding proteins (termed quasigenomes) can bring together the ingredients needed for a variety of complex functions and phenotypes while utilizing a relatively small number of genes.

Home * List of Abstracts * Speakers

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

Harald von Melchner
Laboratory for Molecular Hematology, University of, Frankfurt am Main, 60590 GERMANY

Gene trap mutagenesis in mouse embryonic stem (ES) cells is a complementary approach to the functional annotation of the mouse genome. In concert with chemical (ENU) mutagenesis, the approach enables the analysis of gene function in the context of the entire organism and thus furthers our understanding of human disease. We have established a Research Consortium (German Gene trap Consortium, GGTC) to carry out large scale gene trap mutagenesis in ES cells. Its goal is to contribute to the saturation mutagenesis of the mouse genome and to generate a mouse model for each gene in cooperation with the International Mouse Mutant Consortium (IMMC). Towards this goal, the GGTC has generated 16,000 mutant ES cell lines and has identified the gene trap integration sites in 7332 clones. Of the generated gene trap sequence tags (GTSTs), 4,530 informative sequences were obtained of which 2,463 corresponded to known genes, 794 to ESTs and 1,273 to putative novel genes. Of all integrations into previously characterized genes, 66 occurred in genes involved in human disease. Furthermore, we have generated over 70 germ line chimeras and could show that 62% of resulting homozygous mutants exhibit an obvious phenotype. As for most available mouse mutant strains the significance for human disease is uncertain because germline mutations can reveal only the earliest, non-redundant role of a gene, we have also developed a gene trap approach to induce conditional mutations in ES cells. Finally, data obtained from individual clones, such as GTSTs, expression patterns and phenotypes, are deposited in the GGTC's database which is publicly accessible via http://genetrap.de. Corresponding cell lines, stored frozen at the GSF in Munich, are available freely upon request.

Home * List of Abstracts * Speakers

Fine-Tuning the Clock ? The Relationship Between Variability in *Per3* and Circadian Phenotype in Mammals

Malcolm von Schantz[1], Donna Robilliard[1], Oliver Wood[1], Marcel Smits[2], Adrian Williams[3], Debra J Skene[1], Jo Arendt[1], and Simon N Archer[1]

[1]Centre for Chronobiology, School of Biomedical and Life Sciences, University of Surrey, Guildford GU2 7XH, UK; [2]Department of Neurology and Sleep-Wake Disorders, Hospital de Gelderse Vallei, 6710 Ede, the Netherlands; [3]Lane Fox Unit, St Thomas' Hospital, Lambeth Palace Road, London SE1 7EH, UK
Telephone: 01144-1483-686468
Fax: (503)213-7757
Email: m.von.schantz@surrey.ac.uk

Circadian rhythms are an attractive system for studying the relationship between genotype and phenotype because they are created by a set of genes that are mostly uniquely dedicated to the process. In addition to the limited number of genes involved, and the limited and mostly controllable number of confounding factors between genotype and phenotype, the amenability of this system is increased by the fact that phenotype can also be determined non-invasively and quantitatively through parameters such as free-running period length (*tau*)? and morning/evening (diurnal) preference.

The *Period (Per)* genes are a central part of the molecular machinery that creates circadian rhythms in animals. Initially identified in mutant *Drosophila* with abnormal circadian rhythms, *Per* genes were identified in mammals through RT-PCR with nested degenerated primers against putative conserved domains, and three orthologues have been identified in mammals, *Per1*, *2*, and *3*. *Period* genes have also been cloned from birds, zebrafish, and *Xenopus*.

Whilst the roles of *Per1* and *2* in mammals are clearly defined, by knockout mouse studies, the function of *Per3* remains harder to define. *Per3* knockouts display a circadian period length only half an hour shorter than the wildtype. However, polymorphisms in *Per3* have been reported in human subjects and they have been suggested to be associated with delayed sleep phase syndrome (DSPS).

We have analysed inter- and intraspecific variability in the *Per3* gene, and have identified a region with particular variability, a stretch of tandem amino acid cluster repeats that is present in human but partially absent in murine *Per3*. We demonstrated a polymorphism in this region between mouse strains displaying different *tau*. We also showed that a polymorphism in the number of amino acid cluster repeats associates with diurnal preference in human subjects. Bioinformatic analysis showed that this polymorphic region contains recognition patterns for phosphorylation both by casein kinase e, which was previously known to be intimately involved in the regulation of circadian rhythms, and glycogen synthase kinase 3, which was not. *In vitro* phosphorylation experiments showed that PER3 is phosphorylated by both kinases. We hypothesise that this region of *Per3* is important for the fine-tuning of circadian rhythms, both in terms of intraspecific variability and in terms of the genetic differences that makes species diurnal and nocturnal.

Home * List of Abstracts * Speakers

Status of Application of Tiling Genomic Fragment Arrays to Molecular Biology of Mammalian Cells

C. Horak[2,3], X-H Pan[1], J. Rimm[2,3], R. Goetsch[2], P. Bertone[2], P. Miller[4], N. Luscombe[3], P. Harrison[3], M. Gerstein[3], **S.M. Weissman**[1], and M. Snyder[2,3]
[1]Department of Genetics; [2]Department of Molecular, Cellular, and Developmental Biology; [3]Department of Molecular Biophysics and Biochemistry; [4]Department of Medical Anesthesiology, Yale University School of Medicine, P.O. Box 208103, New Haven, CT 06520-8103
Telephone: (203) 737-2282
Fax: (203) 737-2286
Email: sherman.weissman@yale.edu

Recently the Yale Center for Excellence in Genomic Sciences has constructed microarrays consisting of PCR fragments of average length about 800 base pairs covering all the unique DNA sequences of chromosome 22. We have begun to use these arrays in several types of experiments intended to analyze globally various aspects of cellular molecular biology. One type of study is chip analysis of the results of chromatin immunoprecipitation in which antibodies against specific transcription factors are used to precipitate cross linked chromatin, and DNA from the resulting precipitates is analyzed by hybridization to the genomic arrays. Our results with the beta globin locus and other results with E2F indicate the promise of this approach but also show its difficulties in cells of higher organisms. Current work is focused both on extending the results to additional transcription factors, and on refining the approach technically.

A second area that we and others have investigated is the use of these chips to define the transcribed regions of genomic DNA. There is qualitative agreement that the amount of transcriptional activity detected exceeds that predicted from known and suggested genes, although the significance of these additional transcripts remains to be determined. A third area in which the genomic arrays may be useful is in combination with global methods for detection of sequence variation. We have been engaged for some time in developing the use of immobilized enzymes for capturing DNA duplexes that contain internal mismatches. This approach has been applied to cDNA from a diploid lymphoblastoid cell line, with promising results. Initial tests suggest that genomic arrays will be valuable here also for identifying polymorphic mRNA templates.

Home * List of Abstracts * Speakers

The German cDNA Network:? cDNAs for Functional Genomics and Proteomics

Ruth Wellenreuther[1], Stephanie Bechtel[1], Petra Moosmayer[1], Regina Albert[1], Ingo Schupp[1], Vladimir Kuryshev[1], Detlev Bannasch[1], Dorit Arlt[1], Ulrike Korf[1], Rainer Pepperkok[2], Jeremy Simpson[2], Annemarie Poustka[1], Stefan Wiemann[1], and the German cDNA Consortium

[1] Molecular Genome Analysis, German Cancer Research Center, Im Neuenheimer Feld 506, D-69120 Heidelberg, Germany and [2] Cell Biology Program, European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany
Telephone: ++49 6221 56 4765
Fax: ++49 6221 56 4782
Email: r.wellenreuther@dkfz.de

We have formed a network within the German Genome Project aiming at the generation and sequencing of novel full-length cDNAs, and the comprehensive functional analysis of the encoded proteins. Over 6,200 cDNAs (covering >18 Mb) have been sequenced since. This set and greater 140,000 EST-sequenced clones is used to generate a minimal set of full-length cDNAs for use in subsequent functional analysis. In order to study the function of proteins within complex biological systems like cell cycle regulation, apoptosis, or protein secretion, a multitude of complementary approaches need to be followed that combine genomics and proteomics strategies, but also cell biology and computational biology. The availability of full-length cDNAs is elementary for most of these. Only the integration of data from many sources will help to eventually understand protein function and interaction in protein networks and complex biological systems.

Home * List of Abstracts * Speakers

Intracellular Trafficking of A Few Inflammation-Inducible ADAR1 Isoforms

Jing-HuaYang, Qingchuan Zhao[1], Yongzhan Nie, Yingjun Su, Koroush Kibir and Reuven Rabinovici
Department of Surgery, Yale University School of Medicine, New Haven, CT 06520
Email: jinghua.yang@yale.edu

ADAR1 is an editing enzyme catalyzing RNA-specific adenosine deamination in pre-mRNA. We have previously demonstrated that ADAR1 activity is induced in inflammatory cells during acute inflammation. Here we report that variable ADAR1 isoforms are produced and differentially localized in inflammatory and splenic cells. In *in vitro* activated splenocytes, intracellular localization of ADAR1 variants is shifted from the nucleus to the nucleolus and cytoplasm. This intracellular shift is proved due to the production of a few ADAR1 isoforms that lack nuclear localization signal, RNA recognition domains and/or Z-DNA binding domain through inflammation-inducible alternative splicing in exons 2 and 7, which are verified at both mRNA and protein levels. Studies in various cell types transfected with these ADAR1 variants demonstrate that the full-length ADAR1 and short variants deleting the C-terminals are localized in the cytoplasm. In contrast, short ADAR1 isoforms lacking the N-terminal sequences are detected in the nucleolus. Thus, ADAR1-mediated RNA editing is controlled at three different levels including inflammation-inducible transcription and alternative splicing as well as intracellular trafficking of ADAR1. It is indicated that A-to-I RNA editing may occur in the nucleus as well as the nucleolus and cytoplasm.

Introduction

A-to-I RNA editing is catalyzed by RNA-specific adenosine deaminase (ADAR) that converts adenosine to inosine and leads to the production of mRNA variants and protein isoforms. This process is ubiquitous and widely conserved as it was identified in multiple species including mammals (3, 14, 23, 26, 27), *Xenopus*(1), *Drosophilae*(28) and Zebrafish (37). To date, four A-to-I RNA editing enzymes, termed ADAR1, ADAR2, ADAR3 and ADAT1, have been cloned from mammals (3, 14, 23, 26, 27). ADAR1 and ADAR2 are widely expressed in a variety of cells and tissues (14, 27) with the highest expression in the brain and spleen. ADAR 3 was identified solely in the brain and its deaminase activity has not yet been established. ADAT1 targets tRNA and has been cloned from the human (23), mouse (24) and yeast (9).

ADARs are conserved in their adenosine deaminase domain but differ in their RNA binding domains. ADAR1 and ADAR2 contain two or three dsRNA binding domains (dsRBD) in addition to the adenosine deaminase domain. These two enzymes are capable of both non-specific editing of dsRNA and site-specific editing of glutamate receptor sub-unit B (gluR-B) pre-mRNA and serotonin receptor pre-mRNA (2, 40). ADAR2 selectively edits gluR-B at the Q/R site and serotonin at the D site whereas ADAR1 preferably targets gluR-B at the hot spot and serotonin at the A and C sites (2). Q/R site editing requires the formation of a base pairing structure around the editing site for specific substrate recognition by ADAR2.[?] The importance of the secondary structure in substrate recognition was confirmed in a study in which deletion of the stem loops around the Q/R editing site abolished the site-specific editing (40). ADAR3 and ADAT1 do not seem to edit these substrates.

A nuclear localization signal (NLS) and a Z-DNA binding domain are present near the N-terminal region of ADAR1 and are conserved in all species. This signal has also been identified as a nuclear export signal (NES) (32).[?] In addition, the human ADAR1 has an atypical NLS within its dsRBDIII and thus displays the characteristics of a shuttling protein (5). In contrast, the *Xenopus* ADAR1 contains a different NLS, which leads this enzyme to the nascent ribonucleoprotein matrix on the *Xenopus* lampbrush chromosomes and is specifically

associated with active transcriptional sites. These findings suggest that the editing activity of *Xenopus* ADAR1 is coupled with transcriptional events or that the *Xenopus* ADAR1 targets newly synthesized RNAs (4).

Functional consequences of A-to-I RNA editing have been observed in the central nervous system. In the mammalian brain, editing by ADAR2 of gluR-B pre-mRNA has been shown to alter calcium permeability of excitatory neurons (6, 21). The role of ARAR2 was further studied in mice homozygous for a targeted functional null allele. In *ADAR2*-/- mice, A-to-I RNA editing was substantially reduced in diverse mRNAs and was coupled with seizure activity and early death (12). In *Drosophila*'s brain, disruption of the dADAR gene (a homologue of ADAR2) abolished sodium (Para), calcium (Dmca1A), and chloride (DrosGluCl- alpha) channels (10, 22, 29). Mutants lacking dADAR exhibited extreme behavioral deficits and neurodegeneration (29). Furthermore, a dADAR mutant displayed prolonged recovery from anoxic stupor, vulnerability to heat shock, and increased O_2 demands (22). Thus, editing of ion channel pre-mRNAs by dADAR appears to be critical for the integrity and function of the central nervous system.

Several studies suggest that A-to-I RNA editing also plays a role in immune system. First, ADAR1 can be induced by interferon (IFN) in human amnion-U cells (31, 39) and pulmonary macrophages (33). Second, ADAR1 and ADAR2 can destroy dsRNA that may indirectly regulate the function of dsRNA binding proteins including IFN-induced PKR and 2', 5'-oligo(A) nuclease (31, 36). Third, studies in ADAR1 chimeric mouse embryos demonstrated that this editing enzyme affects embryonic erythropoiesis (38), suggesting that editing in this developmental stage is critical for normal erythrocyte proliferationand/or differentiation. We have recently reported that A-to-I RNA editing by ADAR1 is also involved in acute lung inflammation (33). Since acute inflammation is the underlying process of many critical illnesses including systemic inflammatory response syndrome (SIRS), multiple organ failure, sepsis, adult respiratory distress syndrome (ARDS), and ischemia/reperfusion injury (15), additional insight into the regulation of ADAR1-mediated RNA editing during inflammation could shed light into the pathogenesis of these conditions.

The present study further elucidates regulation and intracellular localization of ADAR1 enzyme in inflamed and activated splenocytes with special emphasis on the generation of several ADAR1 isoforms through inflammation-induced alternative splicing. In particular, different localizations of endogenous ADAR1 and transiently expressed isoforms in the cytoplasm, nucleus and nucleolus are demonstrated.

Materials and Methods

Animal model of acute inflammation. The model used was described previously in detail (13). In brief, endotoxin (LPS, sigma) at 15 mg/kg (LD_{50}) was injected into the peritoneal cavity of conscious adult (6 week-old, 25 g) male C57Bl/6 mice (Jackson Laboratories, Bar Harbor, Me). Tissues were harvested after anesthesia (pentobarbital, 30 mg/kg) at several time points and processed for analysis as described below. The Yale Animal Care and Use Committee approved all animal protocols. Typically, five mice were used for each time point or experiment (n=5) and the same tissues were mixed for analysis.

Preparation of normal, activated and inflamed splenocytes. Normal and inflamed splenocytes were prepared from fresh spleens of normal or endotoxin-challenged (typically 15 mg/kg of LPS, 24 hours) mice. Spleens were cut into slides and passed through a strainer (Falcon). Red blood cells were removed by adding ACK cell lysis buffer (0.15 M NH_4Cl, 10 mM $KHCO_3$, 0.1 mM EDTA, pH 7.3) and splenocytes were washed with PBS and collected by centrifugation. Typically $5x10^7$ splenocytes were obtained from each spleen. For activation, $2x10^7$ normal splenocytes were cultured in RPMI 1604 media in the presence of 100 units/ml of IL-2 or 2.5 mg/ml of Con-A for 48 hours. The activated splenocytes were collected by centrifugation.

RNA and protein isolation. TotalRNA and mRNA were isolated from tissues or splenocytes using Trizol (Trizol, Inc.) and OligoTex (Qiagen) for Northern blotting, RT-PCR and ADAR1 cloning. Total protein was isolated by homogenizing splenocytes in four volumes of editing buffer (20 mM Hepes, pH 7.9, 100 mM KCl, 5 mM EDTA, 0.5% NP-40, 10% glycerol). The lysate was sonicated for 30 seconds and centrifuged for 30 seconds at 4,000 g.[illegible] Protein concentration in the supernatant was determined using BioRad protein kit and adjusted to 10 mg/ml for Western blotting or editing analyses.

Northern blotting, RT-PCR and Western analysis. Equalamounts[illegible] (~2 mg) of mRNA were used for Northern blotting. ADAR1 was detected by hybridization of the blots with ^{32}P labeled antisense probes (1305-1265 and 3004-2966, AF291050).[illegible] Membranes were hybridized at 65[illegible]C over night and the final wash was with 0.1xSSC at 55[illegible]C for 10 min. For RT-PCR, 2 mg of total RNA were used for reverse transcription primed with poly(dT)$_{12-18}$. ADAR1 mRNA was determined by RT-PCR using primers that flank exons 6-8 (1975-2003 and 2436-2408, AF291050) or the entire coding region (1-24 and 3459-3435, AF291050). The relative expression of ADAR1 mRNA was estimated in comparison to GAPDH. For Western blotting, 60 mg of total protein from splenic cells were resolved on 10% SDS-PAGE and transferred on a PVDF membrane. The blots were detected using antibodies against different regions of mouse ADAR1. The N-terminal antibody was produced with a synthetic peptide derived from the N-terminal sequence of mouse ADAR1 (Santa Cruz). Three C-terminal antibodies from different resources were used in the study: 1) antibodies developed in our laboratory using a recombinant protein of mouse ADAR1 (2765-3459 of AF291050), 2) antibodies against the C-terminal sequence of human ADAR1 (gift from Brenda Bass, Utah University) and 3) antibodies derived from a synthetic peptide of the C-terminal sequence of mouse ADAR1 (Santa Cruz).

Cloning and sequencing of ADAR1 variants. Products of the full-length ADAR1 cDNA from normal or inflamed splenic cells were cloned into pCRII (Invitrogen). Diversity of individual ADAR1 cDNA clones was analyzed by E.coRI digestion. The junction sites of alternative splicing were mapped by sequencing and analyzed by sequence alignment. Four typical ADAR1 cDNAs with open reading frame, ADAR1*Sa* (AF291875), ADAR1*Sb* (AF291877), ADAR1*La* (AF291050) and ADAR1*Lb* (AF291876), were subcloned for further analysis.

In vitro translation and expression of ADAR1 isoforms in a baculovirus system. The cDNA of ADAR1*Sa*, ADAR1*Sb*, ADAR1*La* or ADAR1*Lb* in the pCRII was translated *in vitro* using TNT T7 quick translation system following the manufacturer[illegible]s protocol (Promega). For recombinant proteins, these variants were subcloned into pFastBac (Gibco) in-frame with a His taq at the N-terminals to generate recombinant ADAR1-Bacmids. After transfection of Bacmids into Sf9 insect cells, recombinant ADAR1 isoforms were expressed and proteins isolated using Nickel columns following the manufacturer[illegible]s protocol (Pharmacia).

RNA editing assay using dsRNA.[illegible] RNA editing activity was evaluated by measuring A-to-I conversion of synthetic dsRNA (41). In a typical dsRNA editing assay, 10 ml of cell extracts (100 mg total protein) were mixed with ^{32}P-labeled dsRNA (~5x10^3 cpm) and incubated at 37[illegible]C for 1 hour. The mixture was treated with an equal volume of a PK buffer (10 mM Tris.HCl pH 8.0, 300 mM NaCl, 0.2% SDS, 0.5 mg/ml proteinase K). The edited dsRNA was extracted with phenol, precipitated in ethanol, resuspended in 10 ml of RNase P1 mix (5 mM Tris.Cl pH 7.5, 10 mM NaCl, 1 unit of RNase P1), and incubated at 37[illegible]C for 2 hours. The converted inosine was analyzed on thin layer chromatography (TLC) and visualized by autoradiography after being developed in saturated $(NH_4)_2SO_4$-isopropanol (95:5) solution. The radioactivity of each spot was quantified by scintillation.

Localization analysis of ADAR1-EGFP chimera and immunofluorescence. A restriction enzyme site (BamH I) was added at both ends of ADAR1 cDNA (ADAR1*Sa*, ADAR1*Sb*, ADAR1*La* and ADAR1*Lb*) by PCR using the pCRII-ADAR1 plasmids as templates. The PCR product was cleaved with BamHI and directly subcloned into pEGFP-N1 vector (Clontech) in-framed with the N-terminal of EGFP. Editing activity of the transiently

expressed ADAR1-EGFP chimeras were tested in 293 cells and shown positive in converting adenosine to inosine on synthetic dsRNA. Different cells, including mouse fibroblasts (3T3), neuroblastoma (N18), monocytes (RAW264.7), or human HeLa cells,were transfected in triplets (n=3)with ADAR1-EGFP plasmids and fluorescence observed under microscope 6 hours after transfection. For immunofluorescence, cells were fixed on glass slides coated with poly-L-lysine and the specific proteins were hybridized with antibodies against ADAR1 or nucleolin and positive signals detected with fluorescence labeled second antibodies as described in the manufacturer's protocol (Santa Cruz).

Results

Differential localization of endogenous ADAR1 variants. As ADAR1 is upregulated during acute inflammation (33) and the enzymatic activity of ADAR1 is to modify RNA, it is very important to determine the intracellular localization of ADAR1, which will help us to identify the RNAs that are targeted for editing. We first examined the localization of endogenous ADAR1 in a few cell lines using immunoflorescence. Mouse monocytes (RAW 264.7) were grown and fixed on glass slide covers coated with poly-L-lysine. The expression of ADAR1 protein was detected with antibodies against the N-terminal or C-terminal end of ADAR1. The anti-N-terminal antibody was to identify ADAR1 variants with intact N-terminals and anti-C-terminal antibodies were used to detect ADAR1 isoforms with intact C-terminals. As shown in Figure 1A, positive signals were dominantly detected in the cytoplasm of RAW cells when anti-N-terminal antibodies were used. However, signals from nucleolus-like particles were observed when anti-C-terminal antibodies were applied. Since both antibodies should detect the full-length ADAR1 protein, these results indicated that the full-length ADAR1 and/or variants with intact N-terminals were localized in the cytoplasm, while only ADAR1 variants with intact C-terminals were condensed in the nucleolus-like particles.

Interestingly, localization of ADAR1 not only varied in different cell types, but also differentially regulated during *in vitro* activation of splenocytes. In na&#65533;ve mouse splenocytes, dominant signals were detected in the nucleus (Fig. 1B). However, an intracellular localization shift was observed when splenocytes were activated with IL-2 *in vitro*. Positive ADAR1 signals were found in the cytoplasm as well as the nucleolus-like particles in roughly 10% of splenocytes after activation. These findings indicated that the localization of endogenous ADAR1 was shifted from the nucleus to the cytoplasm and nucleolus in the activated splenocytes, which may reflect the production and differential localization of ADAR1 variants in inflammatory or splenic cells.

Multiple ADAR1 variants are induced in mouse splenocytes in response to inflammatory stimulation. To examine the expression of all possible ADAR1 variants, total mRNA from different mouse organs was analyzed by Northern blotting. RNA probes were prepared to hybridize with ADAR1 in the region of dsRNA-binding or catalytic domains to reduce non-specific signals. Two ADAR1 transcripts of approximately 7 and 5 kb in length were detected in all tested tissues (Fig. 2A). The expression of ADAR1 mRNA varied in different tissues with the highest signal in the brain and spleen and lowest in the liver. The ratio between the short and long transcripts also varied in a tissue-specific manner. The highest value was observed in the spleen and lowest in the brain (Fig. 2A). Similar results were obtained when the full-length ADAR1 cDNA was used as a probe. Further experiment demonstrated that both transcripts were significantly increased in the inflamed splenocytes from the endotoxin-challenged mice (Fig. 2B).

In order to characterize these transcripts, the cDNA fragment coding for the interferon-inducible ADAR1 protein (7, 8) was amplified from normal or inflamed mouse splenic cells using RT-PCR. Consistent with Northern blotting, a long and a short fragments were obtained, measuring 3.4 kb (termed ADAR1*L*) and 2.0 kb (termed ADAR1*S*), respectively (Fig. 2C). The expression of ADAR1*S* and ADAR1*L* was significantly increased after inflammatory stimulation. Especially, the expression of ADAR1*S*,that was twenty-fold less than that of ADAR1*L* in normal tissues,rapidly reached the level of ADAR1*L*. Careful analysis could reveal slightly different sizes of

ADAR1*S* or ADAR1*L* (Fig. 2C), indicating more variations might occur. To characterize these variants, RT-PCR products of ADAR1*L* from spleens of normal or endotoxin-stimulated mice were cloned and individual cDNAs were analyzed by restriction enzyme. Three fragments (I, II and III) were identified, each covering exons 1-2, 2-5 or 5-15 of ADAR1, respectively (Fig. 2D and Fig. 3). Insertion or deletion mutations were indicated in fragments II and III because their sizes were different. In normal mouse splenic cells, two major variations in fragment III were obtained in 14 clones (Fig. 2D), while in stimulated splenocytes, three in fragment II and five in fragment III were found to be different out of 16 clones. Thus, not only the quantity but also the diversity of ADAR1 mRNA was increased during endotoxin-stimulated acute inflammation.

ADAR1 mRNA variants are generated by inflammation-induced alternative splicing. ADAR1 mRNA variants were analyzed by cloning and sequencing. Alignment analysis revealed that all variants were identical at their 5' or 3' ends. The difference between ADAR1*L* and ADAR1*S* was due to alternative selection of the 3' splice site of intron 1, which resulted in a deletion of the entire exon 2 composed of approximately 1.4 kb of ADAR1 mRNA (Fig. 3A, C). Variations in fragment III resulted from alternative splicing of exon 7 due to selective usage of a few cryptic 5' splice sites (Fig. 3B and C). Three ADAR1 variants with slightly different sizes, termed *a*-, *b*- and *c*-forms, were generated (Fig. 3C). The *b*-form matched the previously identified mouse ADAR1 cDNA (33) and was the dominant form in normal splenocytes (9 out of 14 clones). The *a*-form, which added 78-bp in exon 7 due to a cryptic 5' splice site within intron 7, occurred less frequently (3 out of 14 clones).' This variation was also seen in the human ADAR1, in which a similar insertion was found to alter the specificity of editing on glutamate and serotonin receptor pre-mRNAs (19). The *c*-form, skipping the entire exon 7, was rarely noticed in normal mice. Interestingly, these variations in *a*-, *b*- and *c*-forms might alter substrate recognition or binding specificity because they all occurred in the dsRNA-binding domain III (RBDIII). In one clone with variation in fragment II, alternative splicing skipped the entire exon 3 (Fig. 3C), resulting in a deletion in the RBDII.'

'''''''' **Variations in the dsRBDIII are differentially regulated through inflammation-induced alternative splicing.** To study the regulation of the *a*-, *b*- and *c*-forms, RT-PCR was performed to amplify the fragment covering only exons from 5 to 8 of ADAR1. As shown in Figure 4, *a*- and *b*-forms were inversely regulated in mouse splenic cells after endotoxin stimulation. The *a*-form was progressively induced and sustained at a high level during the entire experiment, while the *b*-form was dominant in normal mice and gradually diminished 6 hours after endotoxin stimulation. These observations were in agreement with the cloning studies in which only the *b*-form was identified in healthy mice and more *a*-forms were found in the inflamed splenocytes (Fig. 2C). The levels of the *c*-form were relatively low in normal mice but upregulated in splenocytes after inflammatory stimulation (Fig. 4A). Because the *a*- and *b*-forms differ in their RNA binding domains, ADAR1 isoforms varying in this region may affect site-specific recognition and change the spectrum of RNA for editing in splenocytes during acute inflammation.

For comparison, expression of the *a*-, *b*- and *c*-forms in mouse brain was also analyzed under the same condition. Unlike in the spleen, the *a*-form instead of *b*-form was the dominant one in the brain (Fig. 4B) and the expression of the *a*-form in the brain was not regulated in response to endotoxin stimulation. However, the *c*-form was still upregulated in the similar way as that in the spleen. Therefore, the *a*- and *b*-forms are expressed in a tissue-specific manner and their productions are inversely regulated in the spleen but not in the brain.''

'''''''' **ADAR1 protein isoforms with deletions at both N- and C-terminals are identified in splenic cells.**Antibodies against the N- or three different C- terminal regions of the protein were used to analyze ADAR1 expression at the protein level. Theoretically, the identified ADAR1*L* and ADAR1*S* mRNAs should generate approximately 150 and 80 kD proteins, respectively, based on their sequences. The 80 kD protein was predicted from the open reading frame of ADAR1*S* mRNA starting at position 1555 (AF291050). As expected, a 150 kD protein was detected in splenocyte extracts by Western blotting using the anti-N- and anti-C-terminal

antibodies (Fig. 5A), indicating that it was the full-length ADAR1 or ADAR1*L*150. A similar 150 kD ADAR1 is also known in human tissues (30). Two small protein bands measuring 115 and 80 kD, termed ADAR1*S*115 and ADAR1*S*80 respectively, were detected only when the anti-C-terminal antibodies were used (Fig. 5A), suggesting that they are ADAR1 isoforms which lack the N-terminal sequences. Apparently, ADAR1*S*80 is the product predicted from ADAR1*S* mRNA (Fig. 5C) whereas ADAR1*S*115 is probably from undefined alternatively spliced mRNAs.

Because the ADAR1*L* mRNA can be induced and also alternatively spliced to produce ADAR1*S* (Fig. 3), the overall expression of ADAR1*L*150 might vary with different stimulations. Indeed, the expression of ADAR1*L*150 in cultured splenocytes was upregulated following IL-2 stimulation but downregulated after Con-A stimulation (Fig. 5B). The expression of ADAR1*S*115 and ADAR1*S*80 was upregulated with all tested stimuli. In consistent with Northern and RT-PCR results, these findings indicated that the expression of the long and short ADAR1 isoforms were induced and differentially regulated in splenocytes after activation or inflammatory stimulation. Comparing with ADAR1*L*150, ADAR1*S*80 lacks the N-terminal NLS, the Z-DNA binding domain (11) and dsRBDI whereas ADAR1*S*115 probably lacks NLS and Za. Thus, it is conceivable that the function of ADAR1*S*80 and ADAR1*S*115 isoforms will be different from that of ADAR1*L*150.

Please note that slightly slower migrating protein corresponding to ADAR1*S*115 or ADAR1*S*80 was detected by Western blotting using antibodies against the C-terminal of ADAR1 (Fig. 5A, B). This protein was likely the modified variant of ADAR1*S*115 (or ADAR1*S*80) because it was also inducible after longer period of stimulation (unpublished data). In addition, two proteins with p90 and p100 kD were detected only when the N-terminal antibody was used, suggesting deletions also occur near the C-terminal end of ADAR1. Although these variants are likely inactive for RNA editing because of deletions in the catalytic domain, they might be differently localized and still be able to compete RNA substrates because of their intact RNA/Z-DNA binding domains and localization signals.

Inflammation-inducible ADAR1 isoforms are functionally active *in vitro*. To test whether inflammation-inducible ADAR1*L* and ADAR1*S*80 isoforms with differentially regulated *a*- and *b*-forms (termed *La, Lb, Sa* and *Sb,* Fig. 6A) were functionally active, editing activity of these protein products was evaluated. *La, Lb, Sa* and *Sb* identical in their deaminase domain were translated *in vitro* and the sizes of their protein products were determined to correlate with the predicted sequences (Fig. 6B). To obtain sufficient quantities, the *La, Lb, Sa* and *Sb* proteins were produced in insect cells using a baculovirus expression system. The recombinant proteins containing a 6xhis tag at their N-terminals were purified and tested for editing of ^{32}P-adensoine labeled dsRNA. As predicted, all isoforms were active to edit dsRNA and the activity varied with the highest in *La* and decreased in the order of *Lb, Sa,* and *Sb* (Fig. 6C).

Transiently expressed long and short ADAR1 isoforms are differentially localized in the cytoplasm and nucleolus. As the variations between ADAR1*L*150 and ADAR1*S*80 include the presence or absence of NLS and Z-DNA binding domain, these isoforms may be differentially localized in cells. To test this notion, *La, Lb, Sa* and *Sb* isoforms fused with EGFP at their C-terminals were first transfected into 3T3 cells and editing activity on dsRNA was examined. Results showed that editing activity was maintained in all chimeras (unpublished data) in spite of a previous report that mutations or deletions near the C-terminal of human ADAR1 affect RNA editing activity (16). Next, the localization of different ADAR1-EGFP chimeras was determined in transfected cells by fluorescence microscopy. The long ADAR1 variants were found in the cytoplasm (Fig. 7*La, Lb*), consistent with the human ADAR1 (32).

In contrast, ADAR1*S*80 isoforms were not only localized in the cytoplasm, but also accumulated as several bright particles in the nucleus (Fig. 7*Sa, Sb*). As these particles vary in size and number, ADAR1*S*80 appeared to reside in the nucleolus. This observation was confirmed by multi-color immunofluorescence study (Fig. 8). In

this study, 3T3 cells transfected with ADAR1*S*80*a*-EGFP were fixed on glass slides and hybridized with antibodies against the nucleolar protein nucleolin (17). Visualization by fluorescence microscopy demonstrated co-localization of ADAR1-EGFP (green) and nucleolin (red), confirming the nucleolar localization of the ADAR1*S*80 isoforms. Considering the overall upregulation of ADAR1*L*150 and ADAR1*S*80 isoforms (Fig. 2 and 5), this result is in agreement with the increased cytoplasmic ADAR1 signal in the activated splenocytes (Fig. 1B). The accumulation of signals in the nucleolus (Fig. 1B) reflects the nucleolar binding capability of ADAR1*S*80 isoforms after stimulation.

The cytoplasmic localization ofADAR1*L*150 and nucleolar localization ADAR1*S*80*a*isoforms were examined in several cell types including mouse inflammatory cells (RAW 264.7), neuronal cell (N18) and human HeLa cell (Fig. 7B). In all cell types, ADAR1*L*150 was found in the cytoplasm and ADAR1*S*80*a* was distributed in the nucleolus and cytoplasm. Thus, the differential localization of ADAR1 isoforms occurred not only in fibroblasts but also in neuronal and inflammatory cells derived from the mouse and human. Further more, distribution pattern between the cytoplasm and nucleolus appeared to vary with cell types. In HeLa cells (Fig. 7B) ADAR1*S*80*a* was dominantly localized in the nucleolus, while in neuronal cells or fibroblasts significant cytoplasmic signal was observed.

Discussion

The present study demonstrates that that the localization of ADAR1 in mouse splenic cells is shifted from the nucleus to the nucleolus and cytoplasm after activation. This intracellular shift is proved due to the production of a few ADAR1 variants that lack nuclear localization signal, RNA recognition domains and/or Z-DNA binding domain.[illegible] In vitro studies further support that these functional ADAR1 variants are differentially localized in the nucleolus or the cytoplasm of all tested cell types. Differential localization of transiently expressed or endogenous ADAR1 isoforms delineates a novel mechanism to regulate RNA editing through coupling of inflammation-inducible transcription and alternative splicing and intracellular trafficking. Several important points are emphasized:

Production of multiple ADAR1 isoforms with distinct functional domains. The demonstration that diverse ADAR1 variants with distinct functional domains are produced in the inflamed splenocytes is a key finding of this study. More than a dozen ADAR1 mRNA variants, classified as ADAR1*L* or ADAR1*S* based upon their sizes, were found in the splenocytes from endotoxin-stimulated mice. The ADAR1*S* variants are produced by alternative splicing that deletes the entire exon 2 encoding the Z-DNA binding domain and NLS. At the protein level, a full-length ADAR1*L*150 and a few short isoforms including ADAR1*S*115 and ADAR1*S*80 were detected by Western blotting. The identified ADAR1*L*150 is comparable to the 150 kD human ADAR1, the product of the full length mRNA transcribed under the control of an interferon-inducible promoter (7). The mouse ADAR1*S*80 is likely the product of alternatively spliced ADAR1 mRNA that skips the entire exon 2 and deletes the NLS, the entire Z-DNA binding domain and the RBDI. This isoform was not only detected in mouse splenocytes, thymocytes and other mouse cell lines, but also found in human HeLa and gastric cancer cells (KATO-3) on Western blots (unpublished data), suggesting that ADAR1*S*80 is commonly expressed in mammals. However, the ADAR1*S*115 isoform may be comparable to the human 110-kD protein (30) that originates from a methionine 246 and retains the second half of the Z-DNA motif, all three dsRBDs and the catalytic domain. This variant is constitutively expressed and localized in the nucleus. The cDNA of mouse ADAR1*S*115 has not yet been cloned. It is either a product of different alternatively spliced mRNAs or a transcript from different promoters (7, 8).

The slower migrating bands near ADAR*S*115 or ADAR*S*80 proteins on Western blot are likely the different forms of ADAR*S*115 or ADAR*S*80 that are generated by post-translational modifications or alternative splicing. The present data prefers to the latter because the minor variation in the dsRBDIII will slightly change the size of

ADAR*S*115 or ADAR*S*80. From sequence analysis, the size variation in this region will not only affect the function of RBDIII, but also change the distance from the catalytic domain to the N-terminal functional motifs. Further investigation needs to elucidate how the variations in this region contribute to the properties of ADAR*S*115 or ADAR*S*80 isoforms.

Differential regulation of ADAR1 isoforms in inflamed and activated splenocytes. Another key finding from this study is that ADAR1 isoforms are differentially regulated in the inflamed and activated splenocytes, which may provide a secondary control of ADAR1-mediated RNA editing. Results from RT-PCR, Northern and Western blotting show that ADAR1*S*115 and ADAR1*S*80 isoforms are induced in the activated splenocytes. However, ADAR1*L*150 expression was upregulated in IL-2 activated splenocytes but downregulated in the same cells after Con-A activation. Its overall expression in the inflamed splenocytes will depend on transcription and alternative splicing of ADAR1 transcripts. In addition to it, there is a preferential production of the *a*-form ADAR1 with increasing amount of *c*-forms. In contrast, the dominant *b*-forms in normal animals are quickly downregulated after inflammatory stimulation. Consequently, this differential regulation of ADAR1 isoforms during splenocyte activation will result in an increase of ADAR1*L*150*a*, ADAR1*S*80*a*, ADAR1*L*150*c* and ADAR1*S*80*c* and a decrease of ADAR1*S*150*b* and ADAR1*S*80*b* isoforms. Previous studies demonstrate that deletion of dsRNA binding I, II or III affects the editing activity as well as the specificity of ADAR1 enzyme (25). A similar alternative splicing in the dsRBDIII of human ADAR1 affects the site specificity of serotonin receptor pre-mRNA editing (18-20, 41). Therefore, these differently regulated ADAR1 isoforms in splenocytes are conceivably to target different spectrum of mRNAs for editing and lead to functional consequences on splenic cells.

The mechanisms responsible for the differential regulation of ADAR1 isoforms are not yet clear. Nevertheless, alternative splicing factors that are required to enhance the skipping of the entire exon 2 and/or activate the multiple 5' splice sites of exon 7 may be induced during inflammation. The *a*- and *b*-forms appear to be generated through competitive selection of two different 5' splice sites of intron 7. This is supported by the reciprocal response of *a*- and *b*-forms to inflammatory stimulation.' Since the induction of the *c*-form appears to be less significant, mechanisms other than competition should be contemplated. For example, it is possible that the generation of the *c*- form is secondary to the induction of independent splicing factors or to cell specific alternative splicing. Another potential regulatory mechanism of ADAR1 isoform production could be autoregulation by ADAR1 itself. The possibility that ADAR1 regulates the diversity of its own gene products is supported by previous reports that ADARs are capable of editing their own mRNA (35), interacting with spliceosomes (34), and regulating pre-mRNA splicing (2). Thus, the coupling of alternative splicing and inflammation-induced RNA editing could be one unique feature of the ADAR editing enzymes.

Regulation of intracellular trafficking of ADAR1 isoforms. The most important finding from this study is that the inflammation-regulated ADAR1 isoforms are differentially localized in the cytoplasm and the nucleolus in the activated splenocytes, which could reflect an additional level of regulatory mechanism for ADAR1-mediated RNA editing. For instance, it may differentially channel the induced ADAR1 isoforms in splenocytes to specific intracellular sites during immune or inflammatory responses.''

It is demonstrated in a variety of cell types that ADAR1*L*150 is accumulated in the cytoplasm whereas ADAR1*S*80is localized in the cytoplasm as well as the nucleolus. The cytoplasmic localization of ADAR1*L*150 is in agreement with that of human ADAR1 (32), suggesting that this phenomenon could be consistent in mammals. The nucleolar binding capability of ADAR1*S*80 isoforms can be explained by the presence of a typical basic-residue motif for nucleolar localization near the dsRBDIII (unpublished data). This motif also exists in the human ADAR1 and overlaps with the nucleus import signal that was recently mapped in the dsRBDIII (5). Markedly, localization of the endogenous ADAR1 in splenocytes is different before and after the activation with IL-2. In the activated splenocytes, endogenous ADAR1 proteins dominantly reside in the cytoplasm and

nucleolus whereas in the na&#65533;ve cells ADAR1 signals are detected in the nucleus. Thus, an intracellular localization shift of endogenous ADAR1 from the nucleus to cytoplasm and nucleolus is demonstrated in response to IL-2 stimulation. The upregulation of ADAR1*L*150 and ADAR1*S*80 in IL-2 activated splenocytes supports the increased ADAR1 signal in the cytoplasm and nucleolus. The nuclear signal in the naive splenocytes could be caused by ADAR1*S*115 because the comparable human 110 kD ADAR1 is localized in the nucleus (30). However, it still remains unclear how the localization of the N-terminal ADAR1 fragments contributes to ADAR1 signals in the nucleus of naive splenocytes. Investigation is underway to clone the cDNA of ADAR1*S*115 and characterize its expression and localization regulation.

Identification of ADAR1 isoforms in the cytoplasm, nucleus and nucleolus indicates that A-to-I RNA editing could occur in different intracellular sites. Our observation that the short ADAR1 isoforms are localized in the nucleolus indicates possible occurrence of ADAR1-mediated RNA editing on nucleolar RNAs (*i.e.* pre-rRNA, snoRNAs), which has not yet been documented. The cytoplasmic localization of full-length ADAR1 isoforms reveals a potential editing of cytoplasmic RNA (*i.e.* mature mRNA, rRNA or viral RNAs). In contrast, RNA editing in the nucleus could be inferred from several previous observations. For instance, RNA editing by ADAR1 and ADAR2 of gluR-B and serotonin pre-mRNA requires intron sequences (2, 27, 40), suggesting the editing event is in the nucleus. Recently, the human but not the *Xenopus* ADAR1 has been shown to shuttle between cytoplasm and nucleus (5) and the *Xenopus* ADAR1 without NLS accumulated in the nucleus of oocytes (4). This *Xenopus* ADAR1 is localized in the nucleus that binds the nascent ribonucleoprotein matrix on lampbrush chromosomes and is specifically associated with transcriptionally active loops (4). Thus, the nucleus localization and its association with transcriptionally active sites are likely to facilitate editing of newly synthesized RNA precursors. Taken together, our findings indicated that during splenocyte activation or under inflammatory conditions the differentially regulated ADAR1 isoforms are capable of editing wider spectrum of RNAs in the nucleus as well as in the cytoplasm and nucleolus.

In summary, the present study sheds light into the role of inflammation-induced ADAR1 isoforms during splenocyte activation, which reveals a novel regulatory mechanism of ADAR1-mediated RNA editing through coupling of alternative spicing and intracellular localization during inflammatory or immune responses.

Acknowledgement

We thank Alfred Bothwell and Chou Hung for helpful comments.&#65533; Q. Zhao and Y. Nie contributed equally to this work.

This work was supported by NIH Grants GM-60426 to J.-H.Y. and HL57963-01to R.R. It is also partially supported by the Natural Science Foundation of China Grants 39870418, 39870176 and 30021002 and a Hellman Foundation Fellowship to J.-H.Y.

References

Figure Legends:

Fig. 1. Intracellular distribution of endogenous ADAR1 proteins. **A.** ADAR1 localization in mouse RAW 264.7 cells detected with antibodies against the N- or C-terminal of ADAR1, indicated as N-terminal-Ab and C-terminal-Ab, respectively. **B.&#65533;** ADAR1 localization in na&#65533;ve or IL-2 activated splenocytes that were fixed on

glass slide covers and stained with antibodies against the catalytic domain of ADAR1 protein. N, nucleus; No, nucleolus; C, cytoplasm. Bar, 5 mm.

Fig. 2. Diverse ADAR1 variants are produced and upregulated in immune organs after endotoxin stimulation. **A**. Northern blotting analysis. Two ADAR1 mRNAs are expressed in a tissue-specific manner in all tested mouse tissues. **B**.Northern blotting analysis. Two ADAR1 mRNAs are induced in spleens harvested 4 hours after endotoxin-stimulation (analysis by Northern blotting). **C**. RT-PCR analysis.Long (~3.4 kb, *L*) and short (~2.0 kb, *S*) ADAR1 variants are induced in spleens harvested 4 hours after endotoxin stimulation. RT-PCR was performed using primers covering the entire coding region. **D**. A variety of long ADAR1 isoforms are induced by inflammation. Individual cDNAs that were cloned from the spleens of sham (0h) and endotoxin stimulated (4h) mice were analyzed by *EcoRI* enzyme digestion. Please note that more variations in fragment II and III are detected in endotoxin-challenged tissues. v, pCRII vector. I, II and III, fragments I, II and III, respectively. b-actin and GAPDH, internal controls.

Fig. 3. Sequences and scheme of alternatively spliced ADAR1 variants. **A**. Sequence alignment of the long (ADAR1*La*: AF291050, ADAR1*Lb*: AF291876) and short (ADAR1*Sa*: AF291875, ADAR1*Sb*: AF291877) variants. Underlined, translation start codon; Dashed line, deleted region. **B**. Sequence alignment of the alternatively spliced forms *a, b* and *c*. Sequences are from ADAR1*La* (AF291050) and ADAR1*Lb* (AF291876).**C.** Schematic summary of alternative splicing in mouse ADAR1. *L/S*, alternative splicing that skips exon 2 and generates the long and short variants. *a, b* and *c*, alternative splicing that results in minor deletions or insertions in exon 7. I, II and III indicate *E.coRI* digested fragments I, II and III, respectively.

Fig. 4. Differential regulation of inflammation-inducible ADAR1 mRNAs with variations in the dsRBDIII region. **A** and **B.** RT-PCR analysis of mouse ADAR1 fragment from exon 5 to 8. Three fragments of ADAR1, termed *a*-, *b*- and *c*-forms, are produced and differentially regulated in spleens (**A**) or brains (**B**) of mice that are challenged with endotoxin for 0, 2, 4, 6, 16 and 20 hours. Note that the *a*- and *b*-forms are regulated differently in spleen or brain whereas the c-form is induced in both tissues. GAPDH is an internal control. **C**. Scheme of alternative splicing that generates *a*-, *b*- and *c*-forms of ADAR1 in normal (top) or inflamed (bottom) splenic cells. Three different 5 alternative splicing sites are indicated. The *b*-form is dominant under normal conditions, while the *a*-form becomes dominant and the *c*-form increases during endotoxin-induced inflammation.

Fig. 5. Production and differential regulation of the long and a few short ADAR1 protein isoforms in splenic cells. **A.** Western blotting analysis of the long (*L*) and a few short (*S*) ADAR1 proteins in cultured splenocytes. The full-length protein (150 kD) and a few small isoforms (~115 and 80 kD) lacking variable length at their N-terminals were detected with antibodies against the C-terminal of ADAR1 (C-term). Two fragments (100 and 90 kD) missing partial deaminase domain were detected with antibody against the N-terminal of ADAR1 (C-term). **B.** Differential expression of the full-length ADAR1 and the short isoforms in na ve (N) and in IL-2 or Con-A activated (48 hours *in vitro*) splenic cells. Proteins were detected with antibodies against the C-terminal of ADAR1. **C.** Scheme for alternative splicing of ADAR1*L*150 and ADAR1*S*80 isoforms in normal (bottom) or activated (top) splenic cells. Bold bar, RT-PCR primer.

Fig. 6. Editing activity of four ADAR1 isoforms. **A.** Scheme of ADAR1*L*150 and ADAR1*S*80 variants with *a*- or *b*-form in dsRBDIII, termed *La, Lb, Sa* and *Sb*, respectively. Stripped area, Z-DNA binding domain; shadowed area, dsRBDs; triangle, NLS or putative nucleolar binding signal; dots, the deaminase domain; open box, insert in dsRBDIII. **B.** SDS-PAGE analysis of *in vitro* translated *La, Lb, Sa* and *Sb.* Note that the size of all variants corresponds to their sequences. **C.** TLC analysis of dsRNA editing activity of the recombinant *La, Sa, Lb* and *Sb* expressed in Sf9 cells using baculovirus system. The ratio between pI (5-monophosphate inosine) and pA (5-monophosphate adenosine) represents A-to-I editing activity.

Fig 7. Cytoplasmic and nucleolar localization of transiently expressed ADAR1*L*150 and ADAR1*S*80. **A**. Localization of ADAR1*L*150 and ADAR1*S*80 with *a* and *b* variations. *La, Lb, Sa* and *Sb* variants were tagged with EGFP at the C-terminals and tested positive for dsRNA editing by transient transfection of fibroblasts. Localization of ADAR1-EGFP chimeras was then analyzed by transient transfection of the same cells and visualized by fluorescence microscopy. The long variants, *La* and *Lb,* were found in the cytoplasm. The short variants, *Sa* and *Sb,* were found in the nucleus and formed irregular sharp multiple nucleolus-like particles. **B.** Cytoplasmic or nucleolar localization of ADAR1 isoforms in various cell types. Transiently expressed of ADAR1*La* and ADAR1*S*80*a* was examined in neuronal (N18), monocytic (RAW 264.7) and HeLa cells and visualized by fluorescence microscopy. Note that ADAR1*L*150 and ADAR1*S*80*a* were localized in the cytoplasm and nucleolus, respectively. Bar, 10 mm.**ð**

Fig. 8. Nucleolar localization of transiently expressed ADAR1*S*80*a*. HeLa cells transfected with the short ADAR1*S*80*a*-EGFP were detected by fluorescence microscopy (A, green) and immunofluorescence using antibodies against the nucleolar protein nucleolin (C23) and TRITC-conjugated second antibodies (B, red). ADAR1 (green) and nucleolin (red) were digitalized and superimposed (C), confirming the nucleolar localization of ADAR1*S*80*a*. Bar, 5 mm.

1.**ðð** **Bass, B. L., and H. Weintraub** 1988. An unwinding activity that covalently modifies its double-stranded RNA substrate Cell. **55:**1089-98.

2.**ðð** **Burns, C. M., H. Chu, S. M. Rueter, L. K. Hutchinson, H. Canton, E. Sanders-Bush, and R. B. Emeson** 1997. Regulation of serotonin-2C receptor G-protein coupling by RNA editing [see comments] Nature. **387:**303-8.

3.**ðð** **Chen, C. X., D. S. Cho, Q. Wang, F. Lai, K. C. Carter, and K. Nishikura** 2000. A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains [In Process Citation] Rna. **6:**755-67.

4.**ðð** **Eckmann, C. R., and M. F. Jantsch** 1999. The RNA-editing enzyme ADAR1 is localized to the nascent ribonucleoprotein matrix on Xenopus lampbrush chromosomes but specifically associates with an atypical loop J Cell Biol. **144:**603-15.

5.**ðð** **Eckmann, C. R., A. Neunteufl, L. Pfaffstetter, and M. F. Jantsch** 2001. The human but not the Xenopus RNA-editing enzyme ADAR1 has an atypical nuclear localization signal and displays the characteristics of a shuttling protein Mol Biol Cell. **12:**1911-24.

6.**ðð** **Feldmeyer, D., K. Kask, R. Brusa, H. C. Kornau, R. Kolhekar, A. Rozov, N. Burnashev, V. Jensen, O. Hvalby, R. Sprengel, and P. H. Seeburg** 1999. Neurological dysfunctions in mice expressing different levels of the Q/R site-unedited AMPAR subunit GluR-B Nat Neurosci. **2:**57-64.

7.**ðð** **George, C. X., and C. E. Samuel** 1999. Characterization of the 5'-flanking region of the human RNA-specific adenosine deaminase ADAR1 gene and identification of an interferon- inducible ADAR1 promoter Gene. **229:**203-13.

8.**ðð** **George, C. X., and C. E. Samuel** 1999. Human RNA-specific adenosine deaminase ADAR1 transcripts possess alternative exon 1 structures that initiate from different promoters, one constitutively active and the other interferon inducible Proc Natl Acad Sci U S A. **96:**4621-6.

9.**ðð** **Gerber, A., H. Grosjean, T. Melcher, and W. Keller** 1998. Tad1p, a yeast tRNA-specific adenosine deaminase, is related to the mammalian pre-mRNA editing enzymes ADAR1 and ADAR2 Embo J. **17:**4780-9.

10.**Hanrahan, C. J., M. J. Palladino, B. Ganetzky, and R. A. Reenan** 2000. RNA editing of the Drosophila para Na(+) channel transcript. Evolutionary conservation and developmental regulation Genetics. **155:**1149-60.

11.**Herbert, A., J. Alfken, Y. G. Kim, I. S. Mian, K. Nishikura, and A. Rich** 1997. A Z-DNA binding domain present in the human editing enzyme, double- stranded RNA adenosine deaminase Proc Natl Acad Sci U S A. **94:**8421-6.

12.**Higuchi, M., S. Maas, F. N. Single, J. Hartner, A. Rozov, N. Burnashev, D. Feldmeyer, R. Sprengel, and P. H. Seeburg** 2000. Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2 [In Process Citation] Nature. **406:**78-81.

13.**Kabir, K., J. P. Gelinas, M. Chen, D. Chen, D. Zhang, X. Luo, J. H. Yang, D. Carter, and R. Rabinovici** 2002. Characterization of a murine model of endotoxin-induced acute lung injury Shock. **17:**300-3.

14.**Kim, U., Y. Wang, T. Sanford, Y. Zeng, and K. Nishikura** 1994. Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing Proc Natl Acad Sci U S A. **91:**11457-61.

15.**Krafft, P., P. Fridrich, T. Pernerstorfer, R. D. Fitzgerald, D. Koc, B. Schneider, A. F. Hammerle, and H. Steltzer** 1996. The acute respiratory distress syndrome: definitions, severity and clinical outcome. An analysis of 101 clinical investigations Intensive Care Med. **22:**519-29.

16.**Lai, F., R. Drakas, and K. Nishikura** 1995. Mutagenic analysis of double-stranded RNA adenosine deaminase, a candidate enzyme for RNA editing of glutamate-gated ion channel transcripts J Biol Chem. **270:**17098-105.

17.**Li, Y. P., R. K. Busch, B. C. Valdez, and H. Busch** 1996. C23 interacts with B23, a putative nucleolar-localization-signal- binding protein Eur J Biochem. **237:**153-8.

18.**Liu, Y., R. B. Emeson, and C. E. Samuel** 1999. Serotonin-2C receptor pre-mRNA editing in rat brain and in vitro by splice site variants of the interferon-inducible double-stranded RNA- specific adenosine deaminase ADAR1 J Biol Chem. **274:**18351-8.

19.**Liu, Y., C. X. George, J. B. Patterson, and C. E. Samuel** 1997. Functionally distinct double-stranded RNA-binding domains associated with alternative splice site variants of the interferon-inducible double-stranded RNA-specific adenosine deaminase J Biol Chem. **272:**4419-28.

20.**Liu, Y., and C. E. Samuel** 1999. Editing of glutamate receptor subunit B pre-mRNA by splice-site variants of interferon-inducible double-stranded RNA-specific adenosine deaminase ADAR1 J Biol Chem. **274:**5070-7.

21.**Lomeli, H., J. Mosbacher, T. Melcher, T. Hoger, J. R. Geiger, T. Kuner, H. Monyer, M. Higuchi, A. Bach, and P. H. Seeburg** 1994. Control of kinetic properties of AMPA receptor channels by nuclear RNA editing Science. **266:**1709-13.

22.**Ma, E., X. Q. Gu, X. Wu, T. Xu, and G. G. Haddad** 2001. Mutation in pre-mRNA adenosine deaminase markedly attenuates neuronal tolerance to O2 deprivation in Drosophila melanogaster J Clin Invest. **107:**685-93.

23.**Maas, S., A. P. Gerber, and A. Rich** 1999. Identification and characterization of a human tRNA-specific adenosine deaminase related to the ADAR family of pre-mRNA editing enzymes Proc Natl Acad Sci U S A. **96:**8895-900.

24.**Maas, S., Y. G. Kim, and A. Rich** 2000. Sequence, genomic organization and functional expression of the murine tRNA-specific adenosine deaminase ADAT1 Gene. **243:**59-66.

25.**Maas, S., T. Melcher, A. Herb, P. H. Seeburg, W. Keller, S. Krause, M. Higuchi, and M. A. O'Connell** 1996. Structural requirements for RNA editing in glutamate receptor pre-mRNAs by recombinant double-stranded RNA adenosine deaminase J Biol Chem. **271:**12221-6.

26.**Melcher, T., S. Maas, A. Herb, R. Sprengel, M. Higuchi, and P. H. Seeburg** 1996. RED2, a brain-specific member of the RNA-specific adenosine deaminase family J Biol Chem. **271:**31795-8.

27.**Melcher, T., S. Maas, A. Herb, R. Sprengel, P. H. Seeburg, and M. Higuchi** 1996. A mammalian RNA editing enzyme Nature. **379:**460-4.

28.**Palladino, M. J., L. P. Keegan, M. A. O'Connell, and R. A. Reenan** 2000. dADAR, a Drosophila double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing Rna. **6:**1004-18.

29.**Palladino, M. J., L. P. Keegan, M. A. O'Connell, and R. A. Reenan** 2000. A-to-I pre-mRNA editing in Drosophila is primarily involved in adult nervous system function and integrity Cell. **102:**437-49.

30.**Patterson, J. B., and C. E. Samuel** 1995. Expression and regulation by interferon of a double-stranded-RNA- specific adenosine deaminase from human cells: evidence for two forms of the deaminase Mol Cell Biol. **15:**5376-88.

31.**Patterson, J. B., D. C. Thomis, S. L. Hans, and C. E. Samuel** 1995. Mechanism of interferon action: double-stranded RNA-specific adenosine deaminase from human cells is inducible by alpha and gamma interferons Virology. **210:**508-11.

32.**Poulsen, H., J. Nilsson, C. K. Damgaard, J. Egebjerg, and J. Kjems** 2001. CRM1 mediates the export of ADAR1 through a nuclear export signal within the Z-DNA binding domain Mol Cell Biol. **21:**7862-71.

33.**Rabinovici, R., K. Kabir, M. Chen, Y. Su, D. Zhang, X. Luo, and J. H. Yang** 2001. ADAR1 is involved in the development of microvascular lung injury Circ Res. **88:**1066-71.

34.**Raitskin, O., D. S. Cho, J. Sperling, K. Nishikura, and R. Sperling** 2001. RNA editing activity is associated with splicing factors in lnRNP particles: The nuclear pre-mRNA processing machinery Proc Natl Acad Sci U S A. **98:**6571-6.

35.**Rueter, S. M., T. R. Dawson, and R. B. Emeson** 1999. Regulation of alternative splicing by RNA editing Nature. **399:**75-80.

36.**Samuel, C. E.** 1998. Reoviruses and the interferon system Curr Top Microbiol Immunol. **233:**125-45.

37.**Slavov, D., M. Clark, and K. Gardiner** 2000. Comparative analysis of the RED1 and RED2 A-to-I RNA editing genes from mammals, pufferfish and zebrafish Gene. **250:**41-51.

38.**Wang, Q., J. Khillan, P. Gadue, and K. Nishikura** 2000. Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis Science. **290:**1765-8.

39.**Weier, H. U., C. X. George, K. M. Greulich, and C. E. Samuel** 1995. The interferon-inducible, double-

stranded RNA-specific adenosine deaminase gene (DSRAD) maps to human chromosome 1q21.1-21.2 Genomics. **30:**372-5.

40.**Yang, J. H., P. Sklar, R. Axel, and T. Maniatis** 1995. Editing of glutamate receptor subunit B pre-mRNA in vitro by site- specific deamination of adenosine Nature. **374:**77-81.

41.**Yang, J. H., P. Sklar, R. Axel, and T. Maniatis** 1997. Purification and characterization of a human RNA adenosine deaminase for glutamate receptor B pre-mRNA editing Proc Natl Acad Sci U S A. **94:**4354-9.

Home * List of Abstracts * Speakers

Experimental Verification of Predicted Splice Variants of Human Genes

Dorothea Zink[1, 2], Stefan Haas[3], Oliver Heil[1], Mathias Schick[1], Melanie Roth[1], Eivind Coward[4], Bernhard Korn[1], and Martin Vingron[3]
[1]RZPD - Ressourcenzentrum f&#65533;r Genomforschung, INF 506, D-69120 Heidelberg, Germany; [2]German Cancer Research Centre, Department of Molecular Genome Analysis, INF 506, D-69120 Heidelberg, Germany; [3]Max-Planck Institute for Molecular Genetics, Ihnestr. 73, D-14195 Berlin, Germany; [4]Bioinformatic group, Department of Informatics, University of Bergen, PB. 7800, N-5020 Bergen, Norway
Telephone: ++49-6221-424712
Fax: ++49-6221-424704
Email: d.zink@dkfz-heidelberg.de

Public EST databases currently contain more than 4 million human EST sequences, representing at least 30.000 different human genes. Within these data exists a large redundancy. We take advantage of this redundancy by analysing the differences of sequences belonging to the same gene. The EST sequences are clustered and assembled to a consensus sequence. However, many clusters cannot be assembled into a single consensus sequence. The EST sequences then fall into multiple consensus sequences (contigs) within one cluster. The differences might be due to imperfect sequence data (e.g. partially unspliced sequence templates, sequencing errors) or due to alternative splicing. The visualizing of predicted spliced human genes from EST data is collected in a public database called SpliceNest (http://splicenest.molgen.mpg.de). Instead of a gene coding for a single mRNA leading to a single protein, alternative splicing of transcripts may lead to different mRNA species and therefore also to potentially different proteins. Splice variants are often due to alternative exon usage, which we verify by RT-PCR. We have set up a medium throughput strategy that does allow us to screen expression of genes in 66 different human cell lines and tissues of multiple stages. The experimental procedure of large scale RT-PCR analysis has been automated on a Biomek 2000 station. We initiated this pipeline to i) verify the nature of the human EST data, ii) qualitatively analyse alternative splicing in human mRNAs on a genome level. Our results indicate, that the theoretical data represented in EST databases can be verified in many cases by our experimental design. Moreover, we do find additional splice products that are not defined by any EST sequence. In order to gain more insight, we re-sequence PCR products in question, to confirm their origin and nature. Nevertheless, in more than 35% of the cases, we cannot experimentally support EST data by RT-PCR. We currently expanded analyses by applying GeniomOne chip technology for alternative splicing and establish a database containing the EST data and the experimental result. In a first step we want to confirm the RT-PCR results by GeniomOne chip hybridisation. In a second step we will verify predicted splice events on the human genome level by chip hybridisation.

Home * List of Abstracts * Speakers
