Table of Contents

Floor Plan of the Conference Venue

Kazusa Akademia Hall　　Okura Akademia Park Hotel

General Information

Program Planning Committee

Osamu Ohara
Kazusa DNA Research Institute, Chiba, Japan

Alvis Brazma
EMBL-EBI, Cambridge, UK

Bernhard Korn
German Cancer Research Centre, Heidelberg, Germany

Tom Freeman
HGMP, Cambridge, UK

Katheleen Gardiner
Eleanor Roosevelt Institute, Denver, CO, USA

Martin Ringwald
The Jackson Laboratory, ME, USA

Geoff Hicks
University of Manitoba, MB, Canada

Sherman Weissman
Yale University , New Haven, CT, USA

Thomas Werner
Genomatrix Software GmbH, München, Germany

Sponsors

Kazusa DNA Research Institute

The U.S. Department of Energy

Chiba Industry Advancement Center
JST-CREATE program

Industrial Support

Luncheon Seminar
Applied Biosystems Japan Ltd.
PerkinElmer Japan Co., Ltd.

Exhibition
Affymetrix Japan K.K.
Applied Biosystems Japan Ltd.
Chiba Industry Advancement Center (JST-CREATE program)
FUJI PHOTO FILM CO., LTD.
Ikeda Scientific, Co., Ltd.
Illumina K.K.
Promega K K
Roche Diagnostics K.K.
SHIMADZU CORPORATION

Venue Information

Kazusa Akademia Center
2-3-9 Kazusa-Kamatari, Kisarazu,
Chiba 292-0818, Japan
http://www.kap.co.jp/

Workshop Secretariat

Kazusa Akademia Park Co., Ltd.
http://www.kap.co.jp/bits2004
E-mail: bits2004@kap.co.jp

Notes

Program of BITS2004

Friday October 29

SESSION 1 Chair: Michio Oishi

13:00-13:10 **Opening Remarks**
Michio Oishi, Kazusa DNA Research Institute

13:10-13:35
Kazuo Shinozaki
RIKEN Genomic Sciences Center
Gene networks involved in drought and cold stress responses

13:35-14:00
Joseph R. Ecker
The Salk Institute
Genome-Wide Discovery of Transcription Units and Functional Elements in Arabidopsis

14:00-14:25
Richard A. Jorgensen
University of Arizona
RNA silencing as a tool for plant functional genomics: mechanisms, strategies, and applications

14:25-14:50
Daisuke Shibata
Kazusa DNA Research Institute
A large-scale identification of metabolite-related genes of *Arabidopsis thaliana* with "omics" approaches

14:50-15:20 Coffee Break

SESSION 2 Chair: Thomas Werner

15:20-15:45
Osamu Ohara
Kazusa DNA Research Institute
From KIAA cDNAs to KIAA proteins: Filling the gap between transcriptome and proteome by collection of antibodies against KIAA proteins

15:45-16:10
Harald v on Melchner
University of Frankfurt Medical School
The German Gene Trap Consortium's mutant ES cell line resource

16:10-16:35
Geoffrey G. Hicks
Manitoba Institute of Cell Biology and the University of Manitoba
ESdb: Genome-Wide Sequence Analysis and Annotation of Induced Mutations in Embryonic Stem Cells

16:35-17:00
Derrick E. Rancourt
University of Calgary
Genomic Analysis of the Implantation Serine Proteinases

-18:00 Poster Presentation/Exhibition

18:00-20:00 *Welcome Reception (at Okura Akademia Park Hotel)*

Saturday October 30

SESSION 3 Chair: Winston Hide

8:35-9:00
Martin Ringwald
The Jackson Laboratory

The Mouse Gene Expression Database (GXD): Updates and Enhancements

9:00-9:25
Andrew I. Su
Genomics Institute of the Novartis Research Foundation

Using Common Laboratory Mouse Strains for *in silico* QTL Mapping

9:25-9:50
Tom Freeman
MRC Rosalind Franklin Centre for Genomics Research

Development of High Density Reverse Transfection Microarrays for Use in The Functional Screening of Mammalian Genes

9:50-10:15
Srinka Ghosh
Affymetrix Inc.

Analysis of Functional Classes via an Unbiased Mapping of The Human Transcriptome

10:15-10:40

Coffee Break

SESSION 4 Chair: Martin Ringwald

10:40-11:05
Yoshihide Hayashizaki
RIKEN, Genomic Science Center

Dynamic transcriptome analysis in FANTOM3

11:05-11:30
Gabriella Rustici
EBI

Periodic gene expression program of the fission yeast cell cycle

11:30-11:55
Thomas Werner
Genomatix Software GmbH

From microarray back to biology: Elucidating HIV Nef effects on astrocytes by promoter analysis and literature mining

11:55-12:20
Satoru Miyano
The University of Tokyo

Computational Challenges in Systems Biology

12:20-13:30

Luncheon seminor (Appllied Biosystems)
Poster presentation/Exhibition

SESSION 5 Chair: Bernhard Korn

Saturday October 30

13:30-13:55
Kazuho Ikeo
National Institute of Genetics

Comparative analysis of gene expression of camera eye between octopus and human

13:55-14:20
Giorgio Bernardi
Stazione Zoologica Anton Dohrn

Functional genomics: Some lessons from isochores

14:20-14:45
John Hesketh
University of Newcastle

***Cis*- and *trans*-acting factors that determine perinuclear localisation of *metallothionein-1* mRNA**

14:45-15:10
Winston Hide
South Africa National Bioinformatics Institute
University of the Western Cape

Preliminary analysis of normal human gene expression using a transcript database

15:10-15:35

Coffee Break

SESSION 6 Chair: Tom Freeman

15:35-16:00
Michael Boutros
German Cancer Research Center

Genome-wide RNAi Approaches to Dissect Signaling Networks

16:00-16:25
Kevin P. White
Yale University School of Medicine

Genomic Regulatory Networks in Drosophila

16:25-16:50
Stefan Wiemann
German Cancer Research Center

From ORFeome to Biology: A Functional Genomics Pipeline

16:50-17:15
Bernhard Korn
RZPD German Resource Center for Genome Research Heidelberg

Systematic comparison of human protein expression *in vivo* **and** *in vitro*

-18:30

Poster presentation/Exhibition

19:00-21:00

Excursion to Sea food restaurant (optional)

Sunday October 31

SESSION 7 Chair: Katheleen Gardiner

8:30-8:55
Sherman Weissman
Yale University School of Medicine
"Hematomics": Genomic scale analyses in mammalian cells with emphasis on the hematopoietic system:

8:55-9:20
Dixie Mager
British Columbia Cancer Agency
Role of transposable elements in evolution of mammalian gene promoters

9:20-9:45
Yoram Groner
The Weizmann Institute of Science
Transcription regulation of TGFβ-mediated dendritic cells development and function by Runx3

9:45-10:10
Bento Soares
The University of Iowa
Identification and Initial Analysis of the Full-Open Reading Frames of 1,500 Transcripts Expressed in the Developing Mouse Nervous System

10:10-10:30 Coffee Break

SESSION 8 Chair: Sherman Weissman

10:30-10:55
Katheleen Gardiner
Eleanor Roosevelt Institute
A database for pathway analysis in Down syndrome

10:55-11:20
Martina Muckenthaler
University of Heidelberg
Towards a systems biology of mammalian iron metabolism From 'IronChips' to 'IronBase' to the 'Iron Regulatory Network'

11:20-11:45
Lynne E. Maquat
University of Rochester Medical Center
Nonsense-mediated mRNA decay in mammalian cells: The exon junction complex-dependent pathway, and a new exon junction complex-independent pathway that involves the RNA binding protein Staufen 1

11:45-12:10
Stefan Stamm
University of Erlangen
The brain specific snoRNA HBII-52 missing in Prader-Willi syndrome patients regulates alternative pre-mRNA splicing of the Serotonin receptor 5HTR2C gene

12:10-13:20 *Luncheon seminor (PerkinElmer)*
Poster presentation/Exhibition

Sunday October 31

SESSION 9 Chair: Alla Rynditch

13:20-13:45
Ruth Brack-Werner
GSF-National Research Center for Environment and Health

Systematic identification of candidate cellular HIV control factors

13:45-14:10
David E. Hill
Dana-Farber Cancer Institute

From Genome Sequences Back to (Systems) Biology: Metazoan ORFeome collections and the construction of protein interaction networks

14:10-14:35
Juergen Bode
German Research Centre for Biotechnology

Dominant genomic structures: Detection and signal functions

14:35-15:05 Coffee Break

SESSION 10 Chair: Osamu Ohara

15:05-15:30
Tom Wehrman
Stanford University

Complex pools of shRNA vectors created using REGS (Restriction Enzyme Generated siRNAs)

15:30-15:55
Akhilesh Pandey
Johns Hopkins University

Proteomic Databases for Systems Biology

15:55-16:20
Stefan Dübel
Technical University of Braunschweig

The German Proteomics Antibody Initiative 'Antibody Factory'

16:20-16:45
Makio Tokunaga
National Institute of Genetics

Single Molecule Imaging and Quantitative Analysis of Molecular Interactions Inside Cells

16:45-17:10
Joshua LaBaer
Harvard Medical School

Harnessing the Human Proteome

Closing

Notes

Speaker Abstracts

Gene networks involved in drought and cold stress responses

Kazuo Shinozaki[1,3], Kazuko Yamaguchi-Shinozaki[2], Motoaki Seki[1,3], Yuriko Osakabe[2], Ri-ichiro Yoshida [1,3], Kyonoshin Maruyama [2], Taishi Umezawa [3]

[1]Plant Functional Genomics Research Group, RIKEN Genomic Sciences Center, Yokohama, Japan; [2]Biological Resources Division, Japan International Research Center for Agricultural Sciences (JIRCAS), Tsukuba, Japan; [3]Laboratory of Plant Molecular Biology, RIKEN Tsukuba Institute, Tsukuba, Japan

Plants respond and adapt to drought and high salinity to survive in severe stress conditions. A variety of genes are induced in response to these stresses at transcriptional level. Their gene products are thought to function in stress tolerance and response. We present recent progress on global analysis of gene networks in response to abiotic stress and functions of stress-inducible genes. We have collected more than 18,000 Arabidopsis full-length cDNAs not only for the correct annotation of expressed genes but also for further functional analyses of their gene products (*Science* **296**: 141-145). We have used a 7K cDNA microarray to analyze changes in gene expression profiles in response to environmental stimuli and hormone treatments.
We have identified at least four independent regulatory systems in stress-responsive gene expression, two are ABA-dependent and two are ABA-independent (*Curr Opin Plant Biol* **6**: 410-417). In one of the ABA-independent pathways, a *cis*-acting element (DRE/CRT) and its binding proteins, DREB1/CBF and DREB2, are important *cis*- and trans-acting elements in stress-responsive gene expression. Based on microarray analysis, many DREB1A/CBF-target genes that function in stress tolerance have been identified. In two ABA-dependent pathways, bZIP transcription factors (AREB/ABF) and MYC/MYB transcription factors are involved in stress-inducible gene expression. Recently, we showed that one of the NAC transcription factors functions in stress-responsive gene expression.
We have analyzed signal transduction cascades in osmotic stress and ABA response and identified two types of protein kinases that are involved in ABA signaling. They are a receptor like kinase and a SnRK2 protein kinases. The roles of these protein kinases will be discussed in ABA signaling.

Ref: Shinozaki et al. *Curr Opin Plant Biol* **6**: 410-417, 2003

GENOME-WIDE DISCOVERY OF TRANSCRIPTION UNITS AND FUNCTIONAL ELEMENTS IN ARABIDOPSIS

Joseph R. Ecker *et al.*

Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037

Complete genome sequences are now available for a wide variety of organisms and many more are on the way! In order to carryout functional analysis in these organisms, accurate determination of gene structures and complete gene inventories will be essential. Computational gene prediction methods are improving but alone are inadequate for new gene discovery and accurate annotation of genomes, in particular for certain gene classes such as non-coding RNA genes. New approaches are required to identify the entire complement of transcription units (protein coding and non-coding), and their associated regulatory elements (e.g. TF/chromatin binding locations and sites of DNA methylation). We are pursuing empirical approaches to decode this information using the genome sequence of the reference plant Arabidopsis, enabling more rapid assessment of the biological functions of the ~30,000 predicted genes. Unbiased mapping of the transcription units is being carried out using third-generation Affymetrix whole genome tiling array (WGA) technology. The high-resolution transcription unit location information is being used to guide the construction of a complete, expression-ready, gene inventory- "the plant ORFeome". We have also begun to utilize WGAs as a "universal" data-gathering platform for capturing a variety of types of genome-scale information, including the chromosomal locations of DNA methylation sites (and identification of the methylases that target these sites), chromatin/transcription factor binding sites and for discovery of genome-wide allelic variations among geographically isolated Arabidopsis accessions. When coupled with transcriptome mapping data, these unbiased sets of genome-wide regulatory information will begin to allow the construction of an integrated set of cellular/molecular connectivity maps for Arabidopsis.

RNA silencing as a tool for plant functional genomics: mechanisms, strategies, and applications

Richard A. Jorgensen, Chonglie Ma, and Natalie Doetsch

Department of Plant Sciences, University of Arizona, Tucson, Arizona, USA

RNA silencing of endogenous plant genes can be triggered by three types of transgenes: (a) inverted repeat transgenes producing double-stranded RNA (dsRNA) transcripts, (b) antisense transgenes, and (c) sense overexpression transgenes. dsRNA transgenes (referred to as RNA interference transgenes) are usually the most efficient of these; however, single-copy sense transgenes can also be reasonably efficient when engineered for high level overexpression, a means of triggering RNA silencing referred to as sense cosuppression. Sense cosuppression requires that a transgene be driven by a strong promoter and produce a translatable transcript, which is recognized and copied by an endogenous RNA-dependent RNA polymerase to produce dsRNA, the immediate trigger of RNA silencing. Normally, antisense transgenes are relatively inefficient triggers of RNA silencing; however, these too can be engineered to trigger sense cosuppression at high efficiency by removal of nonsense codons to create a long open reading frame in the antisense strand.

dsRNA transgenes are useful tools for reverse genetics, but their usefulness may depend on transgene copy number. Single-copy integrants typically show greater reduction of target RNA levels than do multi-copy integrants, probably because transcriptional silencing of the dsRNA transgene's promoter is common in multi-copy integrants, especially in sexual progeny. Interestingly, the level of reduction of target RNA is target-specific: single copy transgenics targeting the same RNA reduce RNA levels to the same extent, but this level of reduction varies among target sequences by several orders of magnitude. For some targets, RNA levels are not reduced at all. This does not necessarily mean that the target gene has not been silenced because the primary effect of RNA silencing can be translational arrest rather than RNA degradation. If there is not a direct correlation between target transcript levels and gene silencing, this is problematic for reverse genetics, where the phenotype of the target gene mutation is not known and the absence of any detectable phenotypic change is a common outcome. This problem can be resolved by assaying the level of protein product, but this option is usually not available. A potentially attractive alternative approach is to target promoters, rather than transcripts, with dsRNA constructs that will produce transcriptional silencing; in this case, the degree of silencing is directly proportional to the reduction in transcript level, which is easily assayed by RNA blot or RT-PCR analysis. However, effectiveness of promoter silencing by dsRNA constructs has been demonstrated for only a small number of genes. It should be noted that this approach may not be as useful for multi-copy gene families because promoter sequences tend to be less conserved than transcribed sequences.

Sense cosuppression is a potentially useful tool for both reverse genetic and forward genetic analyses, especially in polyploid plant species, which exhibit extensive gene redundancy. It also has the potential to be targeted to specific cell or tissue types, as will be described. A major limitation of dsRNA constructs is that each must be made individually in order to obtain a clone with two inverse copies of the same target sequence. Thus, it is would not be feasible to produce a library of dsRNA constructs corresponding to an RNA population from a specific cell type. Sense cosuppression, on the other hand, requires only a single copy of the target sequence, allowing populations of RNA silencing constructs to be made from normalized cDNA populations. By using such populations of constructs, we are attempting to target mutagenesis by sense cosuppression to specific cell and tissue types to produce large populations of plants that can be screened for phenotypes in forward genetic program. Such a population of plants should possess RNAi-induced mutations in only those genes represented in the cDNA population used to produce the sense cosuppression construct population and so should be highly enriched for mutations affecting the cell or tissue type from which the cDNA population was obtained.

A large-scale identification of metabolite-related genes of *Arabidopsis thaliana* with “omics” approaches

Daisuke Shibata

Kazusa DNA Research Institute, 2-6-7 KazusaKamatari, Kisarazu, Chiba 292-0818, Japan

We have started a project for identifying metabolite-related genes in plants since October 2002, which has been supported through a grant from the New Energy and Industrial Technology Development Organization (NEDO). The project is designed to create a large number of transgenic plant cell lines that carry either introduced metabolite-related genes under a strong promoter for over-expression or an RNAi construct for suppressing internal metabolism genes. A key issue for successful analyses is how to manage a large number of transgenic cell lines. Thus, we developed a protocol for cryopreservation of the suspension-cultured Arabidopsis cell line T87, by which transgenic cell lines can be stored for a long time period in liquid nitrogen before metabolite and transcript analyses. To analyze metabolites of transgenic lines, we equipped our lab with state-of-the-art mass spectrometers (MSs), capillary electrophoresis (CE)-MS, liquid chromatography (LC)-MS and Gas Chromatography Time-of-Flight (GC-TOF)-MS. The transgenic cell lines are also subjected to transcriptome analysis with oligonucleotide arrays. We have been developing a web-based plant pathway viewer for integrating both metabolome and transcriptome data on metabolite maps. We produced more than 800 gene constructs with Arabidopsis full-length cDNA clones, most of which were subjected to transformation of the T87 cell line. The current situation of the project will be represented.

From KIAA cDNAs to KIAA proteins: Filling the gap between transcriptome and proteome by collection of antibodies against KIAA proteins

Koga, H.,[1,2] Okazaki, N.,[1] Kikuno, R.F.,[1] Nagase, T.,[1] and **Ohara, O.**[1,3]

[1]2-6-7 Kazusa-Kamatari, Kisarazu, Department of Human Gene Research, Kazusa DNA Research Institute, Chiba 292-0818, Japan:[2] 2-6 Nakase, Mihama-ku, Chiba Industry Advancement center, Chiba261-7126, Japan:[3] 1-7-22 Suehiro, Tsurumi-ku, Laboratory for Immunogenomics, RIKEN Research Center for Allergy and Immunology, Yokohama 230-0045, Japan.

Our group has been characterized long cDNA clones of human unidentified genes for nearly 10 years. The number of human cDNAs we newly identified exceeds 2000, and these cDNAs are known as KIAA cDNAs because we systematically designate them using KIAA plus 4 digit number. What is unique in our project is in largeness of KIAA cDNAs and their encoded proteins; the average sizes of KIAA cDNAs and the predicted KIAA proteins reaches 5.1 kb and 949 amino acid residues.

Because our major interest is in functions of these gene products, it is a serious concern how to link the KIAA mRNAs with KIAA proteins *in vivo*. To address this, we initiated another project to collect mouse KIAA-homologous cDNAs and antibodies against their encoded proteins under the support from Japan Science and Technology Agency. In this project, antibodies are used as reagents for linking the mRNAs with their encoded proteins. We have almost finished collection and characterization of mouse KIAA-homologous cDNA clones and steadily increased the number of antibodies against mouse KIAA proteins. The resultant polyclonal antibodies have been characterized in various applications and also examined the performance of an antibody array system using them.

Because we have thus accumulated many lines of information regarding KIAA genes as described above, it is time to consider how to integrate these lines of information to extract important biological insights from them. As the first step toward this end, we have recently launched a new database, InGap (http://www.kazusa.or.jp/ingap/index.html), which will enable us to address biological functions of KIAA genes more efficiently and convincingly. In this presentation, we will describe more details of these approaches.

The German Gene Trap Consortium's mutant ES cell line resource

Harald von Melchner for the German Gene Trap Consortium

Laboratory for Molecular Hematology, University of Frankfurt Medical School, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany

Gene trapping is a high-throughput approach that can be used to introduce insertional mutations across the genome in mouse embryonic stem (ES) cells. Gene trap vectors simultaneously mutate and report the expression of the endogenous gene at the site of insertion and provide a DNA tag for the rapid identification of the disrupted gene. The generation of mutant mice from a large collection of ES cell lines harboring gene trap insertions could be applied to large-scale functional analysis of the ~30,000 mammalian genes. An update of the existing gene trap ES cell line resources will be given with special emphasis on the resource generated by the German Gene Trap Consortium with a novel conditional gene trap vectors.

ESdb: Genome-Wide Sequence Analysis and Annotation of Induced Mutations in Embryonic Stem Cells.

Yanglong Mou, Luke DeLange, Deborah Tsuyuki, Djula Arapovic, Songyan Liu, **Geoffrey G. Hicks**.

Manitoba Institute of Cell Biology and the University of Manitoba, Winnipeg, Canada.

We established a novel form of gene trapping called Tagged-Sequence Mutagenesis that greatly increases functional analysis of the mammalian genome *in vivo*. Through random insertion of the gene-trap vector into the genome, genes are both tagged and invariably mutated. Each mutation will be studied first by direct DNA sequencing. Comparison of the sequence tags (PSTs) with the existing databases identifies disruptions in known genes or genes that may be related by homology or functional domains. As a result, 10 - 20,000 ES cell lines bearing disruptions in specific genes of particular interest will be immediately available for making mutant mice where functional analysis of human disease genes can be assessed within the context of the intact organism. To organize information on thousands of new ES cell clones and improve the sequence analysis strategy, a web-based application, ESdb, was developed. ESdb will automate data entry, data validation, submission to NCBI, Blast searches, and storage of PST sequences, keywords, and blast results. When possible, gene-specific identifier such as accession number, LinkOut, taxonomy, pubmed ID, OMIM ID for target gene, genomic, cDNA and protein sequences (actual or predicted) will be annotated and stored in the database. Access to ESdb is available through our webpage at www.EScells.ca and includes Blast interface. Furthermore, integration of gene trap tags from the entire International Gene Trap Consortium (IGTC) is now available through the NCBI GenBank server as well as the Ensembl genome viewer. Finally, the role of ESdb, IGTC and the newly formed Canadian Mouse Consortium to the recently announce international Knockout Mouse Project will be presented.

Genomic Analysis of the Implantation Serine Proteinases

Rancourt DE, Tang L, Xu H, Liu S and O'Sullivan CM

Dept of Biochemistry and Molecular Biology, University of Calgary, Calgary, Canada.

The S1 serine protease family is one of the largest gene families known. Within this family there are several subfamilies that have been grouped together as a result of sequence comparisons and substrate identification. The grouping of related genes allows for the speculation of function for newly found members by comparison and for novel subfamilies by contrast. Analysis of the evolutionary patterns of genes indicates whether or not orthologs are likely to be identified in other species as well as potentially indicating that hypothesized orthologs are in fact not. Looking at subtle differences between subfamily members can reveal intricacies about function and expression. Previously, we have described genes encoding two novel serine proteinases, ISP1 and ISP2, which are most closely related to tryptases. The ISP1 gene encodes the embryo-derived enzyme strypsin, which is necessary for blastocyst hatching and invasion *in vitro*. Additionally both ISP1 and ISP2 are co-expressed in the endometrial gland during the time of hatching, suggesting that they may also both participate in zona lysis from within the uterine lumen. Here, we demonstrate that the ISPs are tandemly linked within the tryptase cluster on 17A3.3. We suggest that remarkable similarities within the 50-untranslated and first intron regions of ISP1 and ISP2 may explain their intimate co-regulation in uterus. We also suggest that ISP genes have evolved through gene duplication and that the ISP1 gene has also begun to adopt an additional new function in the murine preimplantation embryo. This work is supported by the Canadian Institute of Health Research and the Alberta Heritage Foundation for Medical Research.

THE MOUSE GENE EXPRESSION DATABASE (GXD): UPDATES AND ENHANCEMENTS

D.P. Hill, D.A. Begley, I.J. McCright, J.H. Finger, T.F. Hayamizu, C.M. Smith, J.A. Blake, C.J. Bult, J.T. Eppig, J.A. Kadin, J.E. Richardson, and **M. Ringwald**

The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA

The Gene Expression Database (GXD) collects and integrates different types of gene expression data from the laboratory mouse, and thus provides information about expression profiles of transcripts and proteins in different mouse strains and mutants. Integration with the other Mouse Genome Informatics (MGI) Databases enables a combined analysis of genotype, sequence, expression, and phenotype data. Participation in the Gene Ontology project provides standardized search terms to query for genes involved with specific molecular functions, biological processes and cellular components. 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 adding new query features and by making new data available on a daily basis. A particular emphasis has been on closer integration of expression and sequence data. In collaboration with the other MGI projects, all mouse sequences from GenBank, RefSeq, SWISS-PROT, TrEMBL, the public mouse genome assembly, and the TIGR, DoTS, and NIA mouse gene indices have been integrated with the rich biological information about mouse genes and strains contained in MGI. In addition, information about the source of the sequences, such as strain, tissue, and library, has been carefully translated into controlled vocabularies to ensure accurate and complete query results. This work enables new querying capabilities and provides an important basis for the integration and analysis of microarray-based expression data.

GXD is accessible through the Mouse Genome Informatics web site at http://www.informatics.jax.org/. GXD is supported by NIH grant HD33745. The Gene Ontology project is supported by NIH grant HG02273.

Using common laboratory mouse strains for *in silico* QTL mapping

Philip McClurg, Mathew T. Pletcher, Tim Wiltshire, **Andrew I. Su**

Genomics Institute of the Novartis Research Foundation, San Diego, CA, USA

The development and application of high-throughput technologies for biological discovery is a central goal of our research. Toward this goal, we have developed an approach to identify quantitative trait loci (QTLs) in the mouse genome using large-scale genotyping and phenotyping of common laboratory strains of mice. Our algorithm accounts for variable haplotype structure, uses bootstrap models to non-parametrically compute significance, and realistically corrects for multiple testing error. We have validated this algorithm on simple monogenic traits and a well-characterized complex trait, and leveraged our 10K SNP set to map novel candidate QTLs in several biomedically relevant phenotypes, including heart disease, bone density, and gallstones.

DEVELOPMENT OF HIGH DENSITY REVERSE TRANSFECTION MICROARRAYS FOR USE IN THE FUNCTIONAL SCREENING OF MAMMALIAN GENES

Ella Palmer and **Tom Freeman**

Microarray Programme, MRC Rosalind Franklin Centre for Genomics Research, Hinxton, Cambridge, CB10 1SB.

Reverse transfection microarray technology was first described by Ziauddin and Sabatini in 2001 (*Nature* **411**:107-110). It involves growing cells over the surface of a glass slide on to which full-length open reading frames (ORFs) in expression constructs have been printed. With the aid of lipid-based transfection reagent, gene transfection is induced in cells growing over each spot. When printed in a microarray format, many transfection events can be monitored in a single study. We have previously optimised the methodology, examining the use of Gateway expression vectors and the effect of N- and C-terminal GFP tagging on the localisation and expression of the fusion-proteins (*Comp. Funct. Genomics* **5:** 342-353 2004). Following these pilot studies we wished to increase the content of the reverse transfection arrays to enable to us carry out high-density gene screens. Due to the potential of gene tagging to introduce errors in the ORF during construction and to disrupt the normal localisation of the protein, we have chosen to adopt the use of un-tagged full-length clone resources. We have prepared plasmids from 2,796 MGC clones, of which approximately two thirds yielded enough DNA for subsequent assays. A high-density reverse transfection array was then constructed with 1,959 MGC clones which were printed in grids surrounded by GFP expression constructs to enable the co-ordinates of the spots to be easily identified under a microscope. These high-density arrays have been used to screen for genes that induce apoptosis or increase tyrosine kinase activity. We have been able to identify and verify genes that do both. We have subsequently performed microarray expression analysis of 3 of the apoptosis-inducing genes identified, analysing transcriptional changes at various time points following transfection. In each case, these analyses have revealed an expression signature commensurate with cells undergoing apoptosis, but are also potentially revealing differences in the way that over expression of these genes results in cell death.

Analysis of functional classes via an unbiased mapping of the Human Transcriptome

[1]**Srinka Ghosh**, [1]Dione Kampa, [2]Edward A. Sekinger, [1]Philipp Kapranov , [2]Heather Hirsch, [1]Jill Cheng, [1]Stefan Bekiranov, [1]Antonio Piccolboni, [1]Victor Sementchenko, [1]Sandeep Patel, [1]Jorg Drenkow, [1] Shane Brubaker, [1]Simon E. Cawley, Huck Ng, [1]Gregg Helt, [1]Hari Tammana, [1]Jeff Long, [1]Ian Bell, [2]Kevin Struhl and [1]Thomas R. Gingeras

[1]Affymetrix, Inc., Santa Clara, CA 95051; [2]Harvard Medical School, Boston, MA 02115

An unbiased mapping of the human transcriptome through interrogation at a very high resolution is critical for exploring the coding regions but especially for understanding the biological significance of transcripts with reduced coding capacity. Such transcripts have been referred to as transcripts of unknown function (TUFs). These RNAs are integral to the cell's transcriptome and may be functionally important based on their shared characteristics with well characterized protein coding transcripts. A significant number of these TUFs are present as overlapping transcripts to well characterized coding genes with many being intronic and intergenic in origin. A majority of the transcripts being unannotated (i.e. not accounted for by any current set of human genome annotations) are not considered as candidates for matching biological phenotypes to underlying molecular causes. The apparent number of these TUFs are several folds higher than coding transcripts, appear to be largely cell type specific and they share many molecular characteristics with the coding RNAs, including being temporarily regulated. Computational methods both parametric and non-parametric have been developed and implemented to ascribe significance to the regulation and differential modification of TUFs in response to retinoic acid. Frequent co-regulation of overlapping pairs of protein-coding and TUFs have also been observed (Cawley,*et al*. 2004). Greater than a 65% RT-PCR verification rate has been observed for these unannotated transcripts. Analysis of the possible regulatory elements that may control the expression of these unannotated TUFs has revealed that about 75% of the binding sites for cMyc, Sp1 and p53 in two cell lines are located in sites other than at the 5'ends of well characterized coding genes. These findings thus far, underscore not only a complex landscape of the human transcriptome but indicate a non-linear architecture, where the fragment of a given transcript is potentially implicated in multiple overlapping transcriptions. The primary goal of our efforts is to construct *ab-intio*, empirical maps of the functionally important sequences in the human genome. The functionally important regions of the genome include sites of: (i) RNA transcription; (ii) Transcription factor binding, (iii) chromatin modification (iv) DNA methylation; (v) origins of replication and (vi) of binding by RNA binding proteins. Such empirically generated maps are being created using high density oligonucleotide arrays containing 25- mer probes spaced on average as low as 5 to as high as 35 base pairs along the chromosomes. The use of these array tools, the analysis software used to garner and interpret data from these arrays and the biological implications of the data from this approach will be discussed.

Dynamic transcriptome analysis in FANTOM3

Yoshihide Hayashizaki

Genome Exploration Research Group, Riken Genomic Sciences Center, Riken Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045 Japan

We have established the comprehensive mouse full-length cDNA collection and sequence database to cover as many genes as we can, named Riken mouse genome encyclopedia, organizing FANTOM consortium. This comprehensive analysis gave a lot of striking information sketching out the mammalian transcriptome. One of the most significant results was the discovery of unexpectedly large amount of non-coding RNAs. Also, unexpected number of alternative spliced transcripts was found.
With this new knowledge in hand, we now step in a new era of transcriptomics. The next stage that life science should step into is analysis of expression regulatory regions, expression profiles, protein-RNA interactions, thus ultimately unraveling the ‘Genome Network’.

As the first step, we are establishing a large-scale system to identify transcriptional starting sites and promoter region, named CAGE. CAGE provides the enormous information of high-throughout promoter activities and profiling of transcriptional starting site including promoter usage analysis. More than 10,000,000 CAGE tags were analyzed to discover new promoters and genes, thus transcriptional network. With this approach, we found more than 150,000 new promoters in FANTOM3 activities. It's interesting to note that many TSSs (transcription starting sites) of the same gene can be varied as the stages of differentiation. Highlight of FANTOM3 will be presented.

The comprehensive database of various functions of biomolecules, such as protein-DNA interactions, protein-protein interactions and interacellular localization of transcripts is going to be established

The genome network to connect causative genes to symptoms and to connect the target molecule of drugs to drug effects will be clarified using these genome structure and functional database. Unveiling the Genome Network will provide multiple new drugs. Furthermore, it allows prescription of drug to fit for each patient.

Periodic gene expression program of the fission yeast cell cycle

Rustici G*, Mata J, Kivinen K, Lio P, Penkett CJ, Burns G, Hayles J, Brazma A, Nurse P, Bahler J.

**EBI, , Hinxton, Cambridge CB10 1SD, UK*

Cell-cycle control of transcription seems to be universal, but little is known about its global conservation and biological significance. We report on the genome-wide transcriptional program of the *Schizosaccharomyces pombe* cell cycle, identifying 407 periodically expressed genes of which 136 show high-amplitude changes. These genes cluster in four major waves of expression. The forkhead protein Sep1p regulates mitotic genes in the first cluster, including Ace2p, which activates transcription in the second cluster during the M-G1 transition and cytokinesis. Other genes in the second cluster, which are required for G1-S progression, are regulated by the MBF complex independently of Sep1p and Ace2p. The third cluster coincides with S phase and a fourth cluster contains genes weakly regulated during G2 phase. Despite conserved cell-cycle transcription factors, differences in regulatory circuits between fission and budding yeasts are evident, revealing evolutionary plasticity of transcriptional control. Periodic transcription of most genes is not conserved between the two yeasts, except for a core set of approximately 40 genes that seem to be universally regulated during the eukaryotic cell cycle and may have key roles in cell-cycle progression

From microarray back to biology: Elucidating HIV Nef effects on astrocytes by promoter analysis and literature mining

Susanne Kramer-Hämmerle[1], Alexander Hahn[2], Ruth Brack-Werner[1], **Thomas Werner[2]**

[1]GSF-Institute of Molecular Virology, 85764 Neuherberg, Germany, [2]Genomatix Software GmbH, Landsbergerstrasse 6, D-80339 München, Germany

The challenge of microarray analysis is to unveil the biological mechanisms behind the chip data. Due to the sometimes-counteracting influences of *de novo* transcription, RNA processing and degradation, the discovery of any particular mechanism is difficult. Since regulatory networks underlying transcription are crucial for orchestrating the cellular response to multiple challenges monitored by microarray experiments it is important to elucidate transcriptional control mechanisms. The aim is to trace back the molecular basis of transcriptional mechanisms to frameworks of transcription factor binding sites in gene promoters. These frameworks can then be used to elucidate the functional context of the genes. We illustrate for an example how this approach can be applied to elucidate known and unknown transgene effects via gene networks.
Human astrocytes infected by HIV limit HIV expression mainly to early viral regulatory factors, like the HIV Nef protein. HIV infected astrocytes can be expected to persist in the brains of infected individuals for many years and are not eliminated by antiviral therapies. Microarray studies were carried out to elucidate the effects of constitutive HIV Nef expression on the transcriptome of astrocytes. Statistical evaluation of microarray results revealed small clusters of genes specifically upregulated by native Nef protein but not in astrocytes expressing a. non-myristoylated Nef variant, which is functionally defective. Two of these clusters were analyzed by comparative promoter analysis. Four resulting promoter frameworks were then used to scan the Genomatix human promoter database (GPD) for additional putatively co-regulated genes. These candidates were then checked for functional association and for known connections using ElDorado and BiblioSphere (Genomatix). The resulting networks agree with observed phenotypic changes as well as indicate additional not directly observable intranuclear processes. Another lesson from the study was that gene networks could only be deduced by comparing cells harboring the native Nef with cells producing a a defective form of Nef of, while comparison of gene expression patterns of cells with and without Nef failed to yield meaningful results in this analysis.

Computational Challenges in Systems Biology

Satoru Miyano

Human Genome Center, Instituite of Medical Science, The University of Tokyo 4-6-1 Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Advances in measurement technology have enabled genome-wide biological data production. Our challenge in this scope is comprised of the following two approaches.

The first is “how to extract gene network information” from heterogeneous genome-wide data. We developed a computational method for estimating gene networks from microarray data obtained from various perturbations such as gene disruptions, gene overexpressions, drug responses, etc. The method combines the Bayesian network approach with nonparametric regression, where genes are regarded as random variables and the nonparametric regression enables us to capture from linear to nonlinear structures between genes. As a criterion for choosing good networks, we introduced the BNRC (Bayesian network and Nonparametric Regression Criterion) score. Naturally, the sole use of microarray data has limitations on gene network estimation. For improving the biological accuracy of estimated gene networks, we have made a general framework by extending this method so that it can employ genome-wide other biological information such as sequence information on promoter regions, protein-protein interactions, protein-DNA interactions, localization information, subcelluar localization, and literature. Computational experiments were conducted with yeast data and they show that cascades of gene regulations were effectively extracted.

The problem of finding an optimal Bayesian network is known computationally intractable. For example, in order to find an optimal Bayesian network of 20 genes from 100 microarray data, the brute force algorithm employing all computing resources in the world even requires the time exceeding the life time of the solar system. Our recent computational challenge has made possible to search and enumerate optimal and suboptimal Bayesian networks in feasible time on supercomputers. Computational experiments with this search algorithm have provided evidences of the biological rationality of our computational strategy.

Gene networks are applied to find drug target genes. By exploring gene networks estimated from over 500 yeast microarray data based on gene disruptions and antifungal drug doses, a novel drug target gene was identified and validated. For this purpose, we newly developed a software for visualizing and analyzing gene networks which played an important role in discovery. This suggests that our gene network approach can be a strong tactics for searching drug target genes.

The second is “how to model and simulate gene networks”. Obviously, an important challenge is a creation of a platform with which scientists in biology/medicine can comfortably model and simulate dynamic causal interactions and processes in the cell(s) such as gene regulations, metabolic pathways, signal transduction cascades, etc. For this direction, we defined a notion called Hybrid Functional Petri Net with extension (HFPNe) which is a graphical programming language for describing concurrent processes. It was implemented in our software Genomic Object Net (http://www.genomicobject.net/) (commercial version: *Cell Illustrator*^{TM}). Since Genomic Object Net equips a biology-oriented GUI, modeling of very complex biological processes with HFPNe can be performed in a simply way. Its effectiveness was demonstrated by modeling biological processes such as alternative splicing, frameshifting, Huntington's disease model, p53 modifications, which are hard to model with the former architecture. Although Petri net has been studied independently of biology, its affinity to biological process modeling is surprisingly good.

Comparative analysis of gene expression of camera eye between octopus and human

Kazuho Ikeo, Atushi Ogura, and Takashi Gojobori

Center for Information Biology and DDBJ, National Institute of Genetics, Mishima Japan

Although the camera eye of the octopus is very similar to that of humans, phylogenetic and embryological analyses have suggested that their camera eyes have been acquired independently. It has been known as a typical example of convergent evolution. To study the molecular basis of convergent evolution of camera eyes, we conducted a comparative analysis of gene expression in octopus and human camera eyes. We sequenced 16,432 ESTs of the octopus eye, leading to 1052 nonredundant genes that have matches in the protein database. Comparing these 1052 genes with 13,303 already-known ESTs of the human eye, 729 (69.3%) genes were commonly expressed between the human and octopus eyes. On the contrary, when we compared octopus eye ESTs with human connective tissue ESTs, the expression similarity was quite low. To trace the evolutionary changes that are potentially responsible for camera eye formation, we also compared octopus-eye ESTs with the completed genome sequences of other organisms. We found that 1019 out of the 1052 genes had already existed at the common ancestor of bilateria, and 875 genes were conserved between humans and octopuses. It suggests that a larger number of conserved genes and their similar gene expression may be responsible for the convergent evolution of the camera eye.

FUNCTIONAL GENOMICS: SOME LESSONS FROM ISOCHORES

Giorgio Bernardi

Laboratory of Molecular Evolution, Stazione Zoologica Anton Dohrn 80121 Naples, Italy

The first part of this presentation will briefly introduce the essential points discovered in our laboratory about the organization of the vertebrate genome: 1) the compositional compartmentalization; 2) the compositional patterns of DNA and coding sequences in both cold- and warm-blooded vertebrates; 3) the compositional correlations between coding and non-coding sequences, as well as among codon positions; 4) the distribution of genes in the genome and in chromosomes; and 5) the correlations of gene distribution with structure (compositional heterogeneity, intron and UTR size, chromatin structure) and function (gene expression, replication timing, recombination levels).

The second part will explore the correlations between gene density and chromatin compactness in prometaphase human chromosomes and in interphase nuclei.

The third part will deal with a recent re-discovery of the latter findings.

A general reference for the presentation is: *"Structural and Evolutionary Genomics. Natural Selection in Genome Evolution"* by G. Bernardi, Elsevier, Amsterdam, 2004.

Cis- and *trans*-acting factors that determine perinuclear localisation of *metallothionein-1* mRNA.

Herve Chabanon[1], Ian Mickleburgh[1] David Nury[1], Brian Burtle[1], Zofia Chrzanowska-Lightowlers[2] and **John Hesketh[1]**.

[1]Institute for Cell and Molecular Biosciences and [2] School of Neurology, Neurobiology and Psychiatry, The Medical School, University of Newcastle, Newcastle-upon-Tyne, UK..

There is increasing evidence that mRNA localisation is a mechanism for synthesising proteins close to where they will function. mRNAs encoding proteins such as MYC, FOS and metallothionein-1 (*MT1*) are found localised in the perinuclear cytoplasm and associated with the cytoskeletal-bound polyribosomes. This targeting of mRNAs to the cytoskeleton and the perinuclear cytoplasm is due to signals within the 3'untranslated regions (3'UTRs) of the mRNAs. Our recent work has been focussed on studies to define the *cis*-acting localisation signal in the 3'UTR of *MT1* mRNA and the *trans*-acting protein factors that bind to this region. *In situ* hybridisation of cells expressing either *MT1* constructs with specific deletions within the 3'UTR or chimaeric reporter *ß-globin* constructs linked to regions of the *MT1* 3'UTR show that nucleotides 45-76 of the 3'UTR, in particular nucleotides 66-76, are required for perinuclear localisation of the transcripts. This section of the 3'UTR contains a CACC repeat. Gel retardation and UV crosslinking assays show that a 50kDa protein binds to the RNA transcripts corresponding to the *MT1* 3'UTR. Competitive assays show that deletion of nucleotides 66-76 abrogates binding but that deletion of 76-86 has no effect, indicating that the binding is specific. The correlation between binding data and localisation suggests that the 50kDa protein plays a role in the perinuclear localisation of *MT1* mRNA. A combination of RNA-affinity techniques and mass spectrometry showed the 50kDa protein to be the translation factor elongation factor 1α. Using mutagenesis, protein binding assays and *in situ* hybridisation, we are testing our hypothesis that the perinuclear localisation signal in *MT1* mRNA is formed by a combination of the CACC repeat and neighbouring structural features.

This work was supported by grants from BBSRC (13/C13737 and 13/C14236)

Preliminary analysis of normal human gene expression using a transcript database

Winston Hide, Nicki Tiffin, Simon Cross and Janet Kelso

South African National Bioinformatics Institute, University of the Western Cape, Private Bag x17, Bellville, South Africa.

Gene expression varies widely in expression level, timing and location of expression. Gene expression events are captured using high throughput sampling systems such as microarrays, EST sequencing from cDNA libraries, SAGE, MPSS and more recently, iAFLP and CAGE. The description of the source material used to generate these samplings is valuable if it can be associated with the underlying sample source. Large scale mapping of gene expression events onto the eVOC ontologies has resulted in the ability to array known human full length transcripts from the H-INVdb against their expression in 381 anatomical sites. We demonstrate that for the spatial expression vector, a human anatomy of expression can be compiled across several thousands of experiments to result in a combined atlas of normal human gene expression. Initial findings demonstrate that only a few hundred genes are described as having a ‘tissue specific’ or ‘unique term’ expression state. An analysis of the initial atlas of human gene expression will be presented.

Genome-wide RNAi Approaches to Dissect Signaling Networks

Michael Boutros

Boveri-Group Signaling and Functional Genomics, German Cancer Research Center, Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany

A crucial aim upon completion of whole genome sequences is the functional analysis of all predicted gene products. Drosophila is one of the best-studied genetic model systems and has been instrumental to the identification of conserved pathway components with important roles from flies to humans.
We have developed an RNAi library containing 21306 fragments that target every predicted gene in the Drosophila genome. Treatment of cultured Drosophila cells with dsRNA leads to the efficient depletion of the corresponding transcript and the generation of specific and penetrant phenotypes. We have applied a high-throughput RNAi screen in macrophage-like cells to characterize the function of nearly all predicted Drosophila gene in cell growth and viability. We found several hundred dsRNAs that identified essential genes, among which 80% lacked mutant alleles *in vivo*. Quantitative analysis showed that mutant phenotypes allowed classification of phenotypes into distinct functional classes.
We are now applying the combination of a genome-wide RNAi library and massively parallel phenotyping to systematically dissect cellular pathways. The genome-wide RNAi library is adaptable for screening for many different cellular pathways and processes, which should ultimately facilitate the understanding of complex cellular networks. We will present results from genome-wide RNAi screens to comprehensively identify signaling pathway components.

http://www.dkfz.de/signaling/

Genomic Regulatory Networks in Drosophila

Kevin P. White

Department of Genetics, Yale University School of Medicine, New Haven CT, USA

With the completion of the genome sequencing projects for human and the major genetic model organisms, a fundamental challenge in genome science is to identify the functional elements of these genomes and to assemble genes and their products into biological networks. Towards meeting that challenge using *Drosophila* as a model system, we have developed whole genome DNA microarrays that contain sequences from all non-repetitive regions of the *Drosophila* genome. We used these arrays to characterize the expressed non-coding sequences in the genome, to assay coding gene transcription exon-by-exon, to identify thousands of novel splice forms, and to map protein-DNA interactions genome wide. We provided evidence for the majority of predicted genes that previously had no EST or other evidence of expression, and we found that a large proportion of intergenic and intronic expressed sequences also appear to be developmentally regulated. By mapping transcription factors to their *in vivo* binding sites in the genome we identified the direct targets of important developmental control factors such as the pair-rule homeodomain proteins and the ecdysone receptor. By combining expression profiling data, DNA-protein mapping and large-scale protein interaction datasets, we are assembling developmental networks genome-wide.

From ORFeome to Biology: A Functional Genomics Pipeline

Stefan Wiemann

Molecular Genome Analysis, German Cancer Research Center, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany (s.wiemann@dkfz.de)

Among the greatest challenges facing biology today is the generation and exploitation of huge amounts of genomic data, and their conversion into functional information about the proteins encoded. The large-scale cDNA sequencing project of the German cDNA Consortium is providing open reading frames (ORFs) which encode novel proteins of completely unknown function. We aim to put these proteins into their cellular context and to model their mode of action in the living cell.

As a first step towards the functional characterisation of newly identified proteins we have tagged over 1,200 proteins with the green fluorescent protein (GFP), and examined the subcellular localisations of the fusion proteins in cells. 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, cell death, and development. Functional assays with a large set of molecules is only possible by automation. Therefore, we have developed, and adapted cellular assays for use by robotic liquid handling and reading stations. Integrating the functional information that is generated in the assays with data from large scale expression profiling experiments and further functional genomics and proteomics approaches (e.g. yeast two hybrid, protein and antibody arrays), will ultimately allow us to analyse functional networks of proteins in a morphological context, and will greatly contribute to our understanding of cell function in health and disease processes. While the DNA sequences are submitted to the EMBL/GenBank/DDBJ databases, the clones are distributed by the RZPD resource center (www.rzpd.de), and functional data are published through www.lifedb.de

Systematic comparison of human protein expression *in vivo* and *in vitro*

Bernhard Korn

RZPD – Resource Center for Genome Research,
Im Neuenheimer Feld 580, 69120 Heidelberg, Germany, www.rzpd.de

The growing field of proteomics is resulting in an ever increasing demand for the large-scale production and purification of recombinant proteins for structural und functional studies. Therefore we now have full length clones for 10,681 different human genes – based on unigene clustering. More than 3.500 of these transcripts have been subcloned using a shuttle vector systems, to get access to the coding region only. A subset has been translated using multiple expression vectors and in vitro systems, making use of different fusion tags (His6, GST, MBP).
Although *E.coli* is the most frequently used prokaryotic *in vivo* expression system for the high-level production of recombinant proteins, its use harbours some disadvantages, such as the production of misfolded protein in insoluble inclusion bodies or problems with the expression of toxic gene products. In light of these difficulties, cell-free *in vitro* expression systems based on *E. coli* and wheat germ lysate were employed to overcome these obstacles. The systems offer rapid protein synthesis, compatibility with PCR-generated templates and plasmids, the possibility to express toxic gene products, and to easily optimise codon usage and RNA secondary structure.
A systematic approach of comparative protein expression *in vitro* and *in vivo* and of expression optimisation will be discussed.

“Hematomics”: Genomic scale analyses in mammalian cells with emphasis on the hematopoietic system:

Yuval, Kluger,Koya, Jin Lian, Zheng Lian, MengMin Liu, Mioind Mahajan, Yasuhiro Nakayama, XingHua Pan, Ranjana Poddar, **Sherman Weissman**, Yukio Yasukochi,
and
Paul Bertone, Ghia Euskirchen, Mark Gerstein, Tom Gingeras, Roland Green, Dione Kompe, Rebecca Martone, John Rinn, Rebecca Selzer, Michael Snyder, Viktor Stolc

We have been exploring the use of various platforms and methodologies for the study of human cell function at a genomic scale, with emphasis on cells of the hematopoietic system. For various experiments we have used freshly isolated normal cells representing a number of hematopoietic lineages. Alternatively we have used the human promyelocytic cell line NB4, and partially matured neutrophils or monocytes derived from this line, the murine promyelocytic MPRO cell line, and the multipotent murine hematopoietic precursor EML cells. The approaches we have been using include conventional mRNA analysis, transcript analysis using genomic tiling arrays, chromatin immunoprecipitation studies combined with analyses on various microarrays, methods for enriching polymorphic or mutant DNA fragments prior to microarray analysis, methods for fine structure analysis of copy number of genomic segments, and methods for assessing the presence and uniformity of methylation across genomic tracts.
Cells of different lineages show cell type specific relative levels of expression of a large fraction of expressed genes, including many transcription factors, and also lineage specific patterns of DNA methylation. In some cases, extra-genic transcripts that have been observed in recent genomic scale expression analyses also show lineage specific patterns of expression. The expression of these transcripts is not always proportional to that of adjacent genes. Correlation of transcript expression patterns with the genomic distribution of unmethylated CpG islands is currently underway.

Role of transposable elements in evolution of mammalian gene promoters

Dixie Mager[1], Louie van de Lagemaat[1], Catherine Dunn[1], Mark Romanish[1] and Patrik Medstrand[2].

[1]Terry Fox Laboratory, British Columbia Cancer Agency and Dept. of Medical Genetics, University of British Columbia, Vancouver, Canada and [2]Dept. of Cell and Molecular Biology, Lund University, Lund, Sweden.

Transposable elements (TEs) comprise at least 45% of the human and 40% of the mouse genome. This fact alone indicates that TEs have had a major influence on genome organization and evolution. While the negative ramifications of TEs in causing mutations in individuals are well recognized, their most significant effect within the host species may be their ability to induce changes in gene regulation without destroying existing gene function. The goal of our work is to determine the impact of human and mouse TEs on cellular gene expression. Using bioinformatics, we analyzed TE prevalence in untranslated regions (UTRs) of human and mouse mRNAs and found evidence suggesting that TEs affect expression of many genes through donation of transcriptional regulatory signals. Furthermore, we found that recently expanded gene classes, such as those involved in immunity or response to stimuli, have transcripts enriched in TEs, whereas TEs are excluded from mRNAs of highly conserved genes with basic functions in development or metabolism. We have also analyzed selected genes that have adopted an ancient retroviral long terminal repeat (LTR) as a promoter. LTRs can serve as one of several alternative promoters or they can act as the gene's dominant or only known transcriptional promoter. These results support the view that TEs play a significant role in the creation and evolution of transcriptional promoters in mammals.

Transcription regulation of TGFβ-mediated dendritic cells development and function by Runx3

Ofer Fainaru, Eilon Wolf, Ori Brenner*, Varda Negreanu, Yael Bernstein, Dalia Goldenberg, Joseph Lotem, Steffen Jung[#], Ditsa Levanon and **Yoram Groner**

*Departments of Molecular Genetics, *Veterinary Resources, and [#]Immunology. The Weizmann Institute of Science, Rehovot Israel.*

The mammalian runt-related transcription factors (RUNX) belong to a small gene family of three genes *(RUNX1, RUNX2 and RUNX3)* that were recently linked to human autoimmune diseases. Runx3 regulates cell lineage decisions in thymopoiesis and neurogenesis. Runx3 knockout (KO) mice develop spontaneous eosinophilic lung inflammation associated with airway remodeling and mucus hypersecretion. We found that Runx3 is expressed in dendritic cells (DC) and is upregulated during their maturation. The function of Runx3 in DC is to mediate cellular response to TGFβ. When Runx3 is lost the KO DC become insensitive to TGFβ induced maturation inhibition. As a result the DC maturation is accelerated and accompanied by increased efficacy to stimulate T cells and by enhanced propensity to migrate toward the draining lymph nodes. The later is mediated by an elevated expression of a specific chemokine receptor. The abnormalities in DC function constitute a primary immune system defect, which leads in the KO to the spontaneous development of lung inflammation. The data provide new information about the molecular mechanisms underlying the pathogenesis of allergic airway inflammation in humans. Significantly, the clinical symptoms characterizing the lung inflammation in Runx3 KO mice are known hallmarks of asthma in humans. The genetics of inheritance of asthma and allergy in humans is largely unknown, but assumed to involve both general allergy genes and organ specific genes. Intriguingly, human RUNX3 resides in a region on chromosome 1p36.1 known to contain susceptibility genes for asthma and hypersensitivity against environmental antigens. Thus, RUNX3 deficiency may constitute an asthma and allergy risk factor in humans.

Identification and Initial Analysis of the Full-Open Reading Frames of 1,500 Transcripts Expressed in the Developing Mouse Nervous System

[1]Bonaldo, M.F., [2]Bair, T.B., [10]Scheetz, T.E., [1]Snir, E., [1]Akabogu, I., [1]Bair, J.L., [1]Berger, B., [1]Crouch, K., [1]Davis, A., [1]Eyestone, M.E., [1]Keppel, C., [1]Kucaba, T.A., [1]Lebeck, M., [3]Lin, J-L., [1]Melo, A.I.R., [1]Rehmann, J., [3]Reiter, R.S., [1]Schaefer, K., [1]Smith, C., Tack, D., [1]Trout, K., [1, 4]Sheffield, V.C., [3]Lin, J. J-C., [2, 8, 9]Casavant, T.L. and [1, 5, 6, 7]**Soares, M.B.**

Departments of [1]Pediatrics, [2]Electrical and Computer Engineering, [3]Biological Sciences, [4]Howard Hughes Medical Institute, [5]Biochemistry, [6]Physiology and Biophysics, [7]Orthopaedics, [8]Biomedical Engineering, [9]Center for Bioinformatics and Computational Biology, [10]Ophthalmology and Visual Sciences. The University of Iowa, Iowa City, Iowa, USA.

As part of the trans-NIH Mouse Brain Molecular Anatomy Project (BMAP), and in close coordination with the NIH Mammalian Gene Collection Program (MGC), we conducted a large-scale project to clone, identify, and sequence the complete open reading frame (ORF) of transcripts expressed in the developing mouse nervous system. To date, we have identified and determined the complete and accurate sequence of approximately 1,500 full-ORFs, obtained from fifty-one full-length-enriched cDNA libraries that we constructed for this project. These libraries were derived from size-fractionated cytoplasmic mRNA isolated from brain and eye tissues obtained at several embryonic stages, and postnatal days. Altogether, including the full-ORF MGC sequences derived from these libraries by the MGC Sequencing Team, NIH_BMAP full-ORF sequences correspond to about 20% of all transcripts currently represented in mouse MGC. It is noteworthy that 54% of our full-ORF-containing cDNA sequences are longer than 4 kb, and that they amount to 45% of all mouse MGC sequences in the 4-9 kb size range. Preliminary analysis has revealed several transcripts that are exclusively or predominantly expressed in brain and eye tissues, many of which encoding yet uncharacterized proteins.

A database for pathway analysis in Down syndrome

Oleksii Nikolaienkov[1], Cao Nguyen[2], Linda Crnic[2], Krzysztof Cios[2] and **Katheleen Gardiner[1]**

[1]*Eleanor Roosevelt Institute at the University of Denver,*
[2]*University of Colorado at Denver and Health Sciences Center, Colorado*

Down syndrome is due to an extra copy of chromosome 21 and the increased expression of genes encoded within it. Most recent genomic sequence annotation identified ~170 protein coding genes on chromosome 21 with orthologues in mouse. A challenge in Down syndrome research now is correlating the functions of these genes with features of the cognitive and behavioral phenotypes seen in Down syndrome and the major mouse models. A database providing function and expression information on chromosome 21 genes and their orthologues in model organisms would be a useful tool in this effort. To create such a database, information is being collected from specialized databases and the literature, including large and small scale projects that generate knockouts, RNAi and protein-protein interaction maps, in zebrafish, Drosophila, *C. elegans* and yeast. Additional information includes interactors, substrates and targets of chromosome 21-encoded proteins. The database is designed to allow: (i) basic searches on individual genes to obtain details of their structure, function and conservation, (ii) advanced searches to identify all chromosome 21 genes with specified characteristics, e.g. all transcription factors/regulators, all mammalian-specific proteins, all proteins encoding pleckstrin domains, etc, and (iii) viewing of graphical representations of orthologous protein interaction networks in model organisms and conservation of network components in mammals. For maximum utility, the database requires manual curation. Additional experimental data relevant to MAP kinase, calcineurin and sumoylation pathways will be discussed.

Towards a systems biology of mammalian iron metabolism From 'IronChips' to 'IronBase' to the 'Iron Regulatory Network'

Martina Muckenthaler[1,2], Cindy Roy[3], Nancy Andrews[3], Alvis Brazma[4], Evgenij Vainshtein[2] and Matthias W. Hentze[2]

[1]*Department of Pediatric Oncology, Hematology and Immunology University of Heidelberg, Germany*
[2]*European Molecular Biology Laboratory, Heidelberg, Germany*
[3]*Harvard Medical School and HHMI, Boston, USA*
[4] *European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Great Britain*

The iron regulatory network is a prototype multi-dimensional network combining biochemical aspects like oxidative stress and protein synthesis, with cellular functions like immune response, replication and apoptosis as well as, on the whole organism level, organ specific iron absorption and retention regulated by iron hormones. In order to define and ultimately visualize a complete iron regulatory network we produce and collect specific signature profiles derived from cell based and whole animal experiments and correlate them with biochemical, phenotype and physiological information.
Our '*in silico*' iron regulatory network is based on expression profiles recorded on 'IronChips', cDNA microarrays that accurately define mRNA expression patterns of genes that encode iron transporters, storage proteins and regulators together with genes from interlinked pathways (e.g. NO metabolism, redox pathway, stress responses as well as acute phase and innate immunity). We use this tool to monitor compensatory gene expression changes in selected organs of a series of knock-out mice in response to deletions of central players in iron metabolism.
The iron regulatory network not only harbours information about neurodegeneration, inflammation and immune response but could also help to understand two frequent disorders of iron metabolism: Hereditary Hemochromatosis (HH) and Anemia of Chronic Disease (ACD). In this presentation we will focus on our findings that HFE, the gene mutated in HH, plays an important role in the regulation of the iron-hormone and β-defensin like peptide Hepcidin.

Nonsense-mediated mRNA decay in mammalian cells: The exon junction complex-dependent pathway, and a new exon junction complex-independent pathway that involves the RNA binding protein Staufen 1

Yoon Ki Kim[1], Fabrice Lejeune[1], Shang-Yi Chiu[1], Aparna Ranganathan[1], Luc Furic[2], Luc DesGroseillers[2] and **Lynne E. Maquat**[1]

[1]Department of Biochemistry and Biophysics, School of Medicine and Dentistry, 601 Elmwood Avenue, Box 712, University of Rochester, Rochester, New York, USA; [2]Département de Biochimie, Université de Montréal, C.P. 6128, succursale Centre Ville, Montréal, (Qué) Canada H3C 3J7

Nonsense-mediated mRNA decay (NMD) in mammalian cells is generally a splicing-dependent mechanism by which cells recognize and degrade mRNAs that prematurely terminate translation. The dependence on pre-mRNA splicing reflects the deposition of exon junction complexes (EJCs) on mRNA (see, e.g., Lejeune et al., 2002). These EJCs then recruit the Upf NMD factors.

The bulk of cellular proteins derive from translation of mRNA that is bound at the 5' cap by eukaryotic translation initiation factor (eIF)4E. However, our studies of NMD indicate that mRNA that is bound by cap binding protein (CBP)80, which is a precursor to eIF4E-bound mRNA, can also be translated during what we call a "pioneer" round of translation (Ishigaki et al., 2001; Chiu et al., 2004; Lejeune et al., 2004). Consistent with our kinetic studies indicating that NMD targets only newly synthesized mRNA, the EJC is detected on CBP80-bound but not eIF4E-bound mRNA, and NMD appears to be restricted to CBP80-bound mRNA (Ishigaki et al., 2001; Lejeune et al., 2002). We have demonstrated that NMD degrades mRNA from both 5' and 3' ends by recruiting decapping and 5'-to-3' exonuclease activities as well as deadenylating and 3'-to-5' exonuclease activities (Lejeune et al., 2003).

We have recently uncovered a novel mRNA decay mechanism that involves mammalian Staufen (Stau)1, the Upf1 NMD factor, and a termination codon. Unlike NMD, this mechanism does not involve pre-mRNA splicing or NMD factors Upf2 or Upf3X. We show that Stau1 binds directly to Upf1 and can elicit mRNA decay when tethered downstream of a termination codon. We also show that Stau1 interacts with the 3' untranslated region of ADP-ribosylation factor (Arf)1 mRNA. Accordingly, down-regulating Stau1, while of no detectable consequence to the splicing-dependent NMD of either β-globin or glutathione peroxidase 1 mRNA, increases Arf1 mRNA stability. These findings suggest that Arf1 mRNA is a natural target for Stau1-mediated decay, and data indicate that other mRNAs are also natural targets. We discuss this new pathway as a means for cells to down-regulate the expression of Stau1-binding mRNAs.

Chiu S-Y, Lejeune F, Ranganathan A., and Maquat LE (2004) The pioneer translation initiation complex is functionally distinct from but structurally overlaps with the steady-state translation initiation complex. *Genes & Dev*, **18**:745-754.

Ishigaki Y, Li X, Serin G, and Maquat, LE (2001) Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* **106**:607-617.

Lejeune F, Ishigaki Y, Li X, and Maquat LE (2002) The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *EMBO J.* **21**:3536-3545.

Lejeune F, Li X, and Maquat LE (2003) Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. *Mol. Cell* **12**:675-87.

Lejeune, F, Ranganathan, AC, and Maquat, LE (2004) eIF4G is required for the pioneer round of translation in mammalian cells. *Nat. Struct.& Mol. Biol*., in press.

The brain specific snoRNA HBII-52 missing in Prader-Willi syndrome patients regulates alternative pre-mRNA splicing of the Serotonin receptor 5HTR2C gene

Shivendra Kishore and **Stefan Stamm**

Institute of Biochemistry, University of Erlangen, Fahrstraße 17, 91054 Erlangen

SnoRNAs (small nucleolar RNAs) are small non-protein coding RNAs that reside in introns. After these introns have been spliced out of the pre-mRNA, the snoRNAs are released from the introns by processing through endonucleases. To data, the only known function of snoRNAs is their guidance in editing site recognition in snRNAs and rRNAs. HBII-52 is a brain-specific C/D box snoRNA located on IC-SNURF-SNRPN, an imprinted gene on human chromosome 15. The antisense box of HBII-52 is complementary to the alternative exon VIb of the serotonin-receptor 5-HT2RC, located on the X-chromosome. Exon VI of the serotonin-receptor contains three alternative 5' splice sites, giving rise to exon VIa, VIb and VIc isoforms. Exons VIb and VIc are predominantly included in all brain regions, except in the choroid plexus, where VIa predominates. Likewise, the HBII-52 snoRNA is expressed in all brain areas, except choroid plexus. We investigated 5-HT2RC regulation by constructing minigenes containing the alternative exon VI and its flanking constitutive exons. Skipping of exon VIb and VIc is the predominant form in all cells analyzed; even when the 5' splice site of VIb is mutated into a perfect U1-consensus. An increase of the HBII-52 snoRNA stimulates VIb and VIc inclusion, as does mutating the four editing sites present in VIb from A->G. However, HBII-52 mediated exon inclusion does not cause RNA editing at these sites. The effect of HBII-52 depends on its binding to exon VIb, as shown by mutation of its antisense box. Mutations in the C and D boxes abolish the effect, demonstrating that an intact snoRNA is necessary for function. To elucidate the mechanism of snoRNA action, we performed RNAse H mapping experiments and demonstrate that HBII-52 displaces U1 binding at the 5'splice site between exon VIa and VIb. The data indicate that HBII-52 inhibits the recognition of the normally preferred splice site. To test whether these findings are relevant *in vivo*, we analyzed RNA from Prader-Willi patient brains. We found a statistical significant decrease of the non-edited version of the 5-HT2RC splice variant that includes exon VIb, which is in agreement with our model. The protein part encoded by exonVIb is part of the intracellular loop that couples to G proteins. Editing changes the amino acid composition of this loop, which decreases the coupling to G proteins. Therefore, lack of HBII-52 snoRNA reduces the amount of 5-HT2RC serotonin receptor with the highest serotonin potency in Prader-Willi patients.
Our data provide the first evidence that snoRNAs can influence tissue-specific alternative splicing, acting on a gene located on a different chromosome. HBII-52 snoRNA acts most likely by a steric effect on U1snRNA binding. It also explains why Prader-Willi patients respond to serononin-reuptake inhibitors.

Systematic identification of candidate cellular HIV control factors

Susanne Kramer-Hämmerle, Horst Wolff, Francesca Ceccherini-Silberstein, Christian Bickel, Volker Erfle, Thomas Werner* and **Ruth Brack-Werner**[1]

[1]*Institute of Molecular Virology, GSF-National Research Center for Environment and Health, Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany; brack@gsf.de.*
**Genomatix Software GmbH, Landsbergerstr. 6, D-80339 München, Germany*

The human immunodeficiency virus (HIV) establishes a life-long infection in humans and causes acquired immunodeficiency syndrome (AIDS). Current antiviral therapies reduce the levels of free infectious virus but do not eradicate the virus from an infected individual. HIV replicates to very high levels in CD4+ T-cells, which are killed as a result. However, HIV also invades cells that suppress HIV replication, shielding the virus from the host immune system and enabling survival of the infected cells. These cells can transmit HIV by various means and therefore are reservoirs for replication-competent HIV in the infected host.

The central nervous system constitutes a major reservoir for HIV. We are analyzing HIV functions in various brain cell types to identify viral targets underlying cellular control of HIV. The HIV regulatory factor Rev has a crucial role for virus production because it stimulates the nuclear export and translation of HIV-mRNAs that encode structural components of HIV. Our studies demonstrate that human astrocytes suppress the activity of the HIV Rev protein and alter its nucleocytoplasmic shuttling behavior.

This suggests that the human genome encodes factors capable of inhibiting Rev activity. To test this hypothesis our laboratory is pursuing various approaches to identify Rev modulating factors in human cDNA libraries. In one, we are testing the influence of randomly isolated cDNAs on Rev activity using a dual-fluorescent Rev-reporter assay. This approach has led to identification of several candidate Rev inhibitory factors. In a second approach we are screening human cDNA libraries for gene products that interact with Rev. In a yeast-two-hybrid screen of a human cDNA library we identified a novel human Rev interacting factor. Interaction with Rev was confirmed in human cells by mammalian two hybrid analysis and by colocalization studies. Potential coding sequences for this factor are contained in cDNAs in various databases, including the HUGE database. No function had been described for this factor previously. We demonstrate that this factor localizes to the cytoplasm, is exported from the nucleus by CRM1 and contains a novel nuclear export signal different from that in Rev. In overexpression assays, this factor showed an inhibitory effect on the transactivation function of Rev. Conversely, RNAi-mediated inhibition of expression of the endogenous factor stimulated Rev activity.

We conclude that the human genome encodes factors capable of modulating HIV gene expression by interfering with the functionality of Rev. These endogenous factors may contribute to cell-specific control of HIV replication and may eventually be exploitable for development of novel anti-HIV therapies.

From Genome Sequences Back to (Systems) Biology: Metazoan ORFeome collections and the construction of protein interaction networks.

David E. Hill

Department of Cancer Biology and Center for Cancer Systems Biology, Dana-Farber Cancer Institute, 1 Jimmy Fund Way, SM858 Boston, MA 02115

The availability of complete genome sequences suggests new approaches for biological research to complement conventional genetics and biochemistry. In this context, our goals are to generate comprehensive protein-protein interaction, or interactome, mapping strategies for humans and for *C. elegans*, and develop new concepts to integrate the resulting interactome data with other functional maps such as expression profiles (transcriptome), global phenotypic analysis (phenome), localization of expression projects (localizome), etc. Such integrated maps should be valuable for a systems biology approach to development and for understanding the dynamic functioning of the proteome. *C. elegans* is a particularly useful model organism because of the current availability of a complete genome sequence coupled with the fact that the major signalling pathways (i.e., Ras, Akt, Rb, p53, and DNA damage response) are conserved in the worm. More importantly, novel interactions identified among worm proteins can be validated through forward and reverse genetics approaches to suggest function for novel, and in many cases, previously unknown genes. A human interactome map is an essential element that will complement the genomic sequence and transcriptional profiling data already available.

The starting point for our work is the creation and use of the *C. elegans* ORFeome, a physical collection of all expressed, protein-encoding genes cloned as full length open reading frames (ORFs). An immediate consequence of the ORFeome is that we are able to both verify the genome annotation and create a resource to functionally characterize the proteome. Approximately ~12,500 ORFs have been successfully cloned, of which ~4,000 correspond to genes that have remained untouched by any cDNA or expressed sequence tag. Unexpectedly, more than 50% of predicted genes needed corrections in their intron-exon structures.

Importantly, these *C. elegans* proteins are being expressed under many conditions and characterized using various high-throughput strategies, including large-scale interactome mapping. Starting with a subset of metazoan-specific proteins, more than 4000 interactions were identified from high-throughput yeast two-hybrid (Y2H) screens. Independent co-affinity purification assays experimentally validated the overall quality of this Y2H data set. Together with already described Y2H interactions and interologs predicted *in silico*, the current version of the Worm Interactome (WI5) map contains ~5500 interactions. From this initial map, we are able to i) discern various topological and biological features of the *C. elegans* interactome network, ii) explore the possibilities of integrating interactome data with phenome and transcriptome data sets, and iii) develop strategies to improve the completeness, coverage and quality of the *C. elegans* interactome map.

Our efforts at generating a human interactome map are an extension of our recently completed Human ORFeome, version 1.1 that is comprised of over 8,000 full length ORFs derived from the existing set of unique coding sequence available from the NHGRI Mammalian Gene Collection.

Dominant genomic structures: Detection and signal functions

Martin Klar, Erik Stellamanns, Prashanth AK and **Juergen Bode**

GBF, German Research Center for Biotechnology/Epigenetic Regulation, Mascheroder Weg 1, D-38124 Braunschweig, UC Davis Genome Center Davis, CA 95616-8816

Higher genomic structures are formed in a hierarchical manner, the nucleosome constituting the bottom level. There are regions where nucleosomes are positioned according to dominant genomic structures (1). We have identified these signals according to their Stressed-Induced-Duplex-Destabilization (SIDD-) properties (2). Duplex destabilization is a prerequisite for secondary structure formation. Our primary models are the type I interferon gene clusters on human chromosome 9 (9p21-22) and murine chromosome 4.

The mechanism of human interferon-β expression (i.e. the induction of the IFNB1 gene) seems to be well established and has almost reached the status of a textbook paradigm. However, all known features (repression and induction) have been localized to the immediate upstream region (i.e. up to position 0.21 kb) of a gene domain which extends over 14 kb. The dominant signals we derive from an SIDD profile, however, are localized at positions –0.5, -2.0 and –3.4 kb where they coincide strikingly with prominent DNAse I hypersensitive sites (HS-3, -2 and –1, resp.) as they are detected in intact nuclei (3). We have shown before that HS-3 and HS-2 function as nucleosome-positioning signals. During gene induction a nucleosome adjacent to HS-2 is released (1) hinting at a regulatory role of this region. Deletion experiments underline the relevance of these HSs for gene induction and transcriptional levels.

We have applied several strategies to identify the responsible proteins. An EMSA signal-scanning procedure with 30-50 bp probes has so far led to the detection of five specific factor binding sites; typically, binding occurred at a flank of an SIDD-signal and in one case at multiple sites throughout the destabilized region. Our studies are most advanced in case of two YY1 sites at HS-1 and HS-2 which are the only strong sites for this factor over at least 4kb. Mutation of either of these sites leads to a 4-fold decrease of the induced interferon level while YY1 overexpression results in a partial de-repression of the IFNB1 gene combined with a loss of its inducibility.

Why does a strong YY1-binding core motif only occur at destabilized sites? Two explanations have emerged that are not necessarily mutually exclusive: i - the binding factor needs both the specific motif as well as the destabilization to bind. The difference between the G(x) (free energy) level required for strand separation at the destabilized sites versus the average level amounts to about 9 kcals/mol. This difference makes it about 2,000,000 fold easier to separate the duplex at those locations than the average, at equilibrium. If strand opening is the rate-limiting step, this would explain the conservation of such features during evolution; ii – Alternatively, the factor needs the motif to bind whereby it causes the transfer of destabilization to another site, as a consequence of coupling (via superhelical stresses) of strand opening behaviours of basepairs. This site where the destabilization is transferred to then becomes functional

In summary, we have demonstrated the existence of remote-control mechanisms participating in the induction of the IFNB1 gene. In addition we are in the process of developing a rational procedure of specifying remote control mechanisms which have the potential of a wide applicability.

1 - Bode, J., Goetze, S., Heng, H., Krawetz S. A.& Benham C. (2003) From DNA structure to gene expression: Mediators of nuclear compartmentalization and -dynamics. *Chromosome Res*. **11**(5), 435-445 and references quoted here

2 - Goetze, S. Gluch, A. Benham, C. & Bode, J. (2003) Computational and *in vitro* analysis of destabilized DNA regions in the interferon gene cluster: the potential of predicting functional gene domains. *Biochemistry*, **42**, 154-166.

3 - J. Bode, H. J. Pucher and K. Maaß (1986) *Eur. J. Biochem*. **158**, 393-401 Chromatin Structure and Induction-Dependent Conformational Changes of Human Interferon-ß Genes in a Mouse Host Cell.

Complex pools of shRNA vectors created using REGS (Restriction Enzyme Generated siRNAs)

Tom Wehrman, George Sen, Jason Myers, and Helen Blau

Dept. of Microbiology and Immunology, Stanford University, Stanford, CA USA

Small interfering RNA (siRNA) technology has greatly expanded the ability to study loss of function phenotypes in tissue culture cells and animal models. The ability to express these molecules from plasmid or viral vectors enables stable knockdown of virtually any gene of interest to be studied in mammalian cells. However generating multiple siRNA vectors using oligonucleotides is a slow and costly process that is generally not amenable to the construction of siRNA libraries by the individual researcher or academic laboratory. To overcome these limitations we developed a method of generating numerous siRNA constructs using a series of restriction enzyme digests and ligations (REGS – Restriction Enzyme Generated siRNAs) that creates approximately 30 unique shRNA constructs per kilobase of sequence. This approach enables the isolation of multiple functional siRNA vectors from a single reaction almost ensuring success in finding a functional siRNA vector. We used the REGS process to generate the first genome-wide library of siRNA vectors, created from a complex pool of cDNA. The ability to create a wide variety of siRNA vectors to a given gene affords several advantages over traditional techniques that will be discussed. We have improved the system to generate libraries of higher complexity and streamlined the process such that the entire procedure can be completed in 2-3 days.

Proteomic Databases for Systems Biology

Shubha Suresh[1], Goparani Mishra[1], Babylakshmi Muthusamy[1], G. Hanumanthu[1], R. Reshmi[1], S. Sriranjini[1], Naveen Talreja[1], L. Karthick[1], Vidya Niranjan[1], B. M. Vrushabendra[1], Salil Sharma[1], Neelanjana Ghosh[1], B. P. Rashmi[1], C. Krishna Prasad[1], Shibani Panda[1], Nandan Deshpande[1], Krishna S. Deshpande[1], T.K.B. Gandhi[1], Raghothama Chaerkady[1], S. Sujatha Mohan[1], **Akhilesh Pandey[2]**

[1]Institute of Bioinformatics, International Tech Park Ltd., Bangalore 560 066, India
[2]McKusick-Nathans Institute of Genetic Medicine and Department of Biological Chemistry and Oncology, Johns Hopkins University, Baltimore, MD 21205, USA

There are several protein databases that provide information about certain aspects of proteins. In order to get maximal possible information about proteins, integration of the existing information is required. With the explosion in the availability of protein information, sifting through this data has now become a major challenge. We have developed the **Human Protein Reference Database (HPRD) (http://www.hprd.org)** as a comprehensive biological database of protein information. It has already become the largest repository of protein-protein protein interactions and post-translational modifications for human proteins. The protein-protein interaction data is compliant with HUPO's Proteomics Standards Intitiative while the annotations for molecular function and subcellular localization are compliant with the Gene Ontology Consortium terms. The data contained in HPRD has allowed us to carry out a detailed analysis of the human interactome and to compare it with large interaction datasets reported recently for the worm and fly proteomes. We have now developed the **Plasma Proteome Database (PPD) (http://www.plasmaproteomedatabase.org)** as part of HUPO's **Plasma Proteome Project**. This database resource brings together data that was generated by the project, and enriched with curated data on each and every protein that was identified as being resident in the plasma including protein isoforms and coding single nucleotide polymorphisms. **Distributed Annotation System (DAS)** is a client-server model in which a single client integrates information from the multiple servers that allows several groups scattered around the globe to annotate the same centralized entry. We have developed a prototype of a **Protein Distributed Annotation System** for **Human Protein Reference Database** using ZOPE and XML-RPC (XML-remote procedure call). This should allow users to see the HPRD data that is enriched further by individual proteomics and molecular biology laboratories. We hope that this will spur the scientific community to become intimately involved in the annotation of data thereby strengthening the building blocks upon which any systems biology approaches will be based.

The German Proteomics Antibody Initiative ‘Antibody Factory’

Stefan Dübel

Biotechnology Department, Technical University of Braunschweig,

Braunschweig, Germany, www.antibody-factory.de

Antibodies are key reagents for functional characterization of the large number of potential gene products (ORF) yielded by the various genome projects. A huge number of antigens arose from the human genome analysis alone, with similar needs from animal models such as mouse. Here, animal based antibody generation has to be supplemented by other approaches, as it is in many cases limited by the bottleneck of the *in vivo* immune response. In contrast, *in vitro* methods do not have these limitations. Phage display, the selection principle employed in this program, has been validated for more than 10 years and yielded most of the human antibodies present today worldwide. The method allows the generation of binders to highly conserved (non-immunogenic) molecules. Further, as the biochemical conditions during the very selection process can be controlled. As a result, antibodies to conformation mutants, e.g. after cofactor binding, can be obtained. In addition, minute variations in antigen structure can be discriminated - a fine-tuning of specificity is possible, e.g. by adding a competitor during *in vitro* selection. Antibodies to post-translational modifications (e.g. phosphorylation) may be made. We will present the structure and goals of the antibody initiative established within the proteomics section of the german NGFN2 program.

Reference:

Hust, M. and Dübel, S. (2004) Mating antibody phage display to proteomics.
Trends in Biotechnol. **22**(1), 8-14

Single Molecule Imaging and Quantitative Analysis of Molecular Interactions Inside Cells

Makio TOKUNAGA[1,2]

1. Structural Biology Center, National Institute of Genetics, Mishima, Shizuoka, Japan
2. Research Center for Allergy and Immunology, RIKEN, Yokohama, Kanagawa, Japan

Objective-type total internal reflection fluorescence microscopy enabled us to visualize single molecules on or just below the cell surfaces. We have further developed a new single-molecule method to visualize individual molecules inside cells, Highly Inclined and Laminated Optical sheet (HILO) microscopy. Illumination with a thin-layered laser beam minimized background, and allowed us to visualize single fluorescent molecules up to the depth of about 10 μm from the glass-medium interface.

Using this, we have clearly visualized single molecules of GFP-tagged importin beta, a carrier protein, and GFP-tagged cargo protein during transport on the nuclear envelope. Single molecule mages inside cells were almost as clearly as those on cover glass surfaces. Image analysis of single nuclear pores showed that two point resolution of 70 nm was achieved. Kinetic parameters of the interactions between translocating molecules and NPCs were obtained through quantitative analysis. Both visualization of single pores and single molecules enabled us to analyze images quantitatively at the molecular level and to obtain kinetic parameters in cells, i.e., retention time at the pore, number of molecules bound to pores, and binding constant.

Two types of binding were found, weaker binding, which gathers up to ~100 molecules/NPC, and stronger binding, with an affinity that changes drastically upon translocation ability. Accessibility of importin beta to the stronger binding site is critical for NPC translocation. Based on these findings, which correspond well with translation rates examined, we propose a novel model of NPC translocation.

Our recent results of single molecule imaging in living cells confirmed our model. Translation rates obtained by cellular level image analysis showed good coincidence with the results of single molecule imaging.

Thus, single molecule imaging has opened a new way to quantify the dynamics, molecular interactions, and kinetics inside cells.

Harnessing the Human Proteome

Joshua LaBaer

Harvard Institute of Proteomics 320 Charles St Cambridge, MA 02141-2023

One of the most compelling steps in the post-genomic era will be learning the functional roles for all proteins. The Harvard Institute of Proteomics (HIP) has initiated a project to create a sequence-verified collection of full-length cDNAs representing all human coding regions in a recombinational vector system that allows the immediate in-frame transfer of all coding regions into virtually any protein expression vector. These transfers allow the addition of peptide tags to either or both end of the proteins. This repository, called the FLEXGene Repository (for **F**ull-**L**ength **Ex**pression-ready), will enable the high-throughput (HT) screening of protein function for the entire set (or any customized subset) of human genes using any method of *in vitro* or *in vivo* expression.

HIP is taking a multidisciplinary approach to build and exploit this exciting resource. HIP has developed an online LIM system that tracks all stages of clone assembly and storage. A novel database called MedGene identifies all genes/proteins that have been linked to a disease in rank order. Through the development of automation, the process of clone capture can now be achieved at a pace of 500-1000 genes per week with fidelity of greater than 90%. As a result, we have captured and sequence verified over 5100 genes from *S. cerevisiae*, 3000 human cDNAs (including nearly half of all human kinases and a large subset of genes related to breast cancer) and over 5400 *Ps. Aeroginosa* genes await sequencing.

The most exciting part of this project has been the ease with which the clones from the repository can be rapidly incorporated in HT biological experimentation. Using automated gene transfer methods and protein purification it is now possible to purify thousands of proteins. Using HT retroviral methods, proteins capable of driving cell migration, altering the morphogenesis of normal epithelial structures, and affecting substrate dependent growth have been identified. Novel technologies to exploit the rich resource of the FLEXGene collection, including a novel form of protein microarray and improved cell free protein expression systems are also underway.

Poster Abstracts

1
Automated annotation of DNA-binding proteins using electro-topological properties

Shandar Ahmad and Akinori Sarai

Kyushu Institute of Technology, Iizuka, Fukuoka-ken, Japan

Growing number of available genomic data necessitates the discovery of rapid and reliable methods of automatic annotation of proteins with specific functions. DNA-binding proteins, playing vital roles in different aspects of gene regulation and expression form one of the most important categories of these proteins. We have developed automated methods to annotate these proteins by looking at the electrotopological properties such as net charge, dipole moment and quadrupole moments of such proteins. These properties have been shown to successfully distinguish all DNA-binding proteins from others at a very high accuracy. Thus, these methods can be applied to develop powerful tools for filtering potential DNA-binding candidate proteins. We also developed a method to visualize geometry of charge distribution in these proteins which is helpful in distinguishing structurally similar proteins into the binding and non-binding ones.

2
ARM: Carefree Metabolic Map Editor

Masanori Arita

University of Tokyo

It is not straightforward to trace metabolic pathways manually because, in each reaction, molecular structures of substrates are rearranged into products; a beginner may easily lose the track of the target structure by not considering molecular symmetry or chirality. To aid such an analysis of metabolic network, Atomic Reconstruction of Metabolism (ARM) software is designed to explore and visualize logical traces of arbitrary metabolic pathways in a defined network. The search engine can output atomic-level traces (i.e. tracer result) of any reaction steps and the visualizer can graphically display searched pathways through drag-and-drop manipulations. We introduce our latest ‘Carefree Metabolic Map Editor’, which can draw and analyze metabolic maps as in Microsoft PowerPoint.

3
Evolution of mammalian microRNA genes and their regulatory targets

Roberto Barrero[1], Taku Tamura[2], Hitomi Sakurai[1], Shiho Hayakawa[1], Yoshio Tateno[1], Kazuho Ikeo[1], Tadashi Imanishi[3] and Takashi Gojobori[1,3]

[1]Center for Information Biology-DDBJ, National Institute of Genetics, Japan; [2]BITS, Co., Ltd., Japan; [3]Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, Japan

MicroRNAs (miRNAs) form a novel class of small RNA genes of 21-25 nucleotides derived from highly conserved hairpin precursors (pre-miRNAs) present from mollusc to mammals. miRNAs act as post-transcriptional repressors of target transcripts via mRNA degradation or translation inhibition. Based on miRNA conserved features we identified 152 novel miRNAs on both mouse and human genomes. MiRNAs are found distributed on the genome as single genes or clusters encoding up to six pre-miRNAs. Nearly, 25% of the identified miRNAs were found overlapping intronic regions of host mRNAs in both human and mouse genomes. In addition, we present a novel large scale RNA-affinity screening tool combined with the evaluation of RNA secondary structures to identify putative microRNA recognition elements (MREs) on target mRNAs. We identified 475 MREs for 29 known mammalian miRNAs. Interestingly, more than 90% of the MREs presented sequence variation between human and mouse indicating that compensatory mutations may have occur during the evolution to maintain miRNA-driven expression regulation. Top predicted MREs for Let-7a, miR-20, miR-97 and miR-182a were tested for posttranscriptional down-regulation activity using a reporter assay. As a result, MREs mediated translational inhibition of GFP protein suggesting that our target prediction algorithm, RNAFFY, identifies highly reliable miRNA targets involved in a wide variety of cellular and developmental processes. Thus, miRNAs may constitute a crucial component of the gene expression regulation network that may have evolved to not only mediate cell fates onto particular cell-types, but also to coordinate normal development in animals as well as in plants.

4
Structural studies of ultra-stable higher ordered aggregates generated by G (guanine)-rich sequences.

Manish Biyani and Koichi Nishigaki

Rational Evolutionary Design of Advanced Biomolecules, Saitama Small Enterprise Promotion Corporation, SKIP city, 3-12-18 Kamiaoki, Kawaguchi, Saitama 333-0844, Japan, Department of Functional Materials Science, Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama-shi, Saitama 338-8570, Japan

Structural properties of single-stranded DNA oligomers that have repetitive tracts of Guanine-rich oligonucleotides (GROs) are currently of strong interest since they exist in specific regions of genomes such as telomeric ends of chromosomes and regulatory regions of several important oncogenes. Guanine is unique among the four DNA bases in its four hydrogen bonding sites to form quadruplex structures and thus have a propensity to show an amazing polymorphism. Additionally, GROs are also known to show a strong tendency of self-aggregation based on G-quartets and thus form a family of stable higher-ordered structures. Thus, with the current status of structural diversity, GROs are gathering substantial interest for not only therapeutic applications but also nanomaterials. Herein we investigated the supermolecular self-assembling behavior of a series of simple GRO sequences by gel electrophoresis as a function of the number of contiguous guanine residues with our without an intervening non-guanine base. The occurrence of aggregates was exhibited by a minor modification in the base sequence of GROs. Next, the most striking property of aggregates was observed with their extreme high-stability against thermal and/or chemical denaturation (8 M urea and 40% formamide) at the boiling temperature and even the treatment with an excess amount of nucleases, which was confirmed by assay of electrophoretic mobility. Based on the facts thus revealed, we have found some rules for the aggregation phenomenon and built a model for developing an advanced nanomaterial with GROs. Currently, high resolution atomic force microscopy (AFM) is planned to be employed to demonstrate feasibility of the model.

5
Systematic analysis of nuclear receptor responsive elements in a genome scale

Hidemasa Bono

Research Center for Genomic Medicine, Saitama Medical School

In order to study gene expression regulation network by nuclear receptors, nuclear receptor responsive elements (NREs) in genomes were explored *in silico*. By mapping predicted binding sequences to available genomes, I analyzed statistical features for the distribution of NREs in conjunction with the functional feature of target genes. In addition, DAS server hosting the genomic position information for those NREs is set up for those who are interested in these transcription factors with corresponding NREs. It makes us possible to integrate NREs with other genome annotations in DAS compatible genome browsers for further functional inference of those sites.

6
Snake venom proteomics: identification and comparison of two major protease families from venoms of crotalid snakes by a proteomic approach

Yen-Shan Chen and Shyh-Horng Chiou

Institute of Biochemical Sciences, National Taiwan University, and Institute of Biological Chemistry, Academia, Taipei, TAIWAN

It is well known that snake venoms contain complex mixtures of pharmacologically active peptides and proteins. To date no definitive identification or classification of families of proteins present in total venoms of any particular snake species has been reported, *i.e.* there is still a lack of detailed proteome maps for snake venom proteins. In this study, we have identified two major protease families from the venoms of Taiwan habu (*Trimeresurus mucrosquamatus*) and American-western diamondback rattlesnakes (*Crotalus atrox*) in the *Crotalidae* family. The global distribution of protein components of total crude venom proteins from these two crotalid snakes were first unveiled by high-resolution two-dimensional gel electrophoresis (2-DE) followed by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry. Further peptide mass fingerprinting by in-gel spot digestion coupled with mass spectrometry and bioinformatic database searches revealed two major groups of proteolytic enzymes with multiple isoforms of distinct charge and size heterogeneities. Venom proteins of Taiwan habu showed kallikrein-like serine proteases of near neutral p*I* and metalloproteinases of slightly basic properties. In contrast, proteins of American-western rattlesnakes showed abundant acidic proteins of kallikrein-like proteases and metalloproteinases. A multidimensional protein analysis protocol by employing conventional open-column ion-exchange chromatography coupled with high-resolution 2-DE and sensitive MALDI-TOF MS may constitute a systematic and practical proteomic approach for large-scale preparation and microscale analysis of proteins to achieve an efficient comparative classification of complex snake venom proteins from diverse snake species.

7
Proteomic analysis of proteins expressed by *Helicobacter pylori* under oxidative stress

Shyh-Horng Chiou[1], Ming-Hong Chuang[1], Ming-Shiang Wu[2], Jou-Town Lin[2]

[1] *Institute of Biochemical Sciences, National Taiwan University, Taipei, Taiwan,* [2] *Division of Gastroenterology, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan*

Helicobacter pylori is a spiral, slow growing Gram-negative microaerophilic (about 5% O_2) bacterium. It has been shown to be the etiologic agent of gastroduodenal diseases, such as chronic gastritis, gastric and duodenal ulcers, and gastric cancer, *etc*. To address the influence of oxidative stress and its underlying mechanisms, we have compared proliferation, urease activity and protein expression profile of *H. pylori* incubated under normal microaerophilic (5% O_2) and aerobic stress (20% O_2) conditions. Oxidative-stress cells displayed coccoid morphology and time-dependent decrease in proliferation. The urease activity was completely abrogated after 32 hours. We have further compared the protein expression profiles of *H. pylori* under normal growing and oxidative-stress conditions by a global proteomic analysis, which includes high-resolution two-dimensional gel electrophoresis (2-DE) followed by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry and bioinformatic databases search/ peptide-mass comparison. The results revealed that more than ten proteins were differentially expressed under oxidative stress. Most notably, the protein expression level of UreE (urease accessory protein E, an essential metallochaperone for urease activity) and alkylhydroperoxide reductase (AhpC) with antioxidant potential are greatly down-regulated under stress conditions. Measurements of messenger-RNA (mRNA) transcription level by performing reverse transcription-polymerase chain reaction (RT-PCR) on total mRNA also confirmed that gene expressions for these two proteins are consistently repressed under oxygen tension. These changes form a firm basis to account for the loss of urease activity and anti-oxidative ability of *H. pylori* after long-term exposure to reactive oxygen. Conceivably, UreE and AhpC may thus be listed as potential targets for the development of therapeutic drugs against *H. pylori*.

8
Cell-free protein synthesis system using insect cell extract prepared with freeze-thawing

Toru Ezure[1], Shoken Higashide[2], Eiichi Shintani[2], Kohki Endo[2], Takashi Suzuki[1], Shinichiro Kobayashi[1], Taro Masuda[1], Masamitsu Shikata[1] and Masaaki Ito[1]

[1]Life Sci. Lab., Shimadzu Co., Nakagyo-ku, Kyoto, Japan [2]Central Lab., Rengo Co., Ltd., Fukusima-ku, Osaka, Japan

We established a novel cell-free translation system derived from *Trichoplusia ni* (High Five™, Invitrogen) insect cell by simple extraction method. Firefly luciferase, β-galactosidase, and green fluorescent protein (GFP) were synthesized in this system with active forms. We analyzed and optimized, (1) preparation method of insect cell extract, (2) concentration of the reaction components, (3) mRNA structure.
The extract was prepared with freeze-thawing insect cell suspended in the extraction buffer. This preparation method was a simple and superior method compared with conventional method using Dounce homogenizer. Furthermore, protein synthesis efficiency was improved by addition of 20% (v/v) glycerol to extraction buffer. Concentrations of the reaction components, such as Mg^{2+}, K^{+}, ATP, GTP, and the energy regenerating system composed of creatine phosphate and creatine kinase, were optimized to increase protein synthesis efficiency. Moreover, mRNAs containing 5'-untranslated regions (5'-UTRs) derived from baculovirus polyhedrin genes showed high protein synthesis activity. Especially, the leader composition of *Ectropis obliqua* nucleopolyhedrovirus polyhedrin gene showed highest enhancement activity among the six 5'-UTRs tested.
As these results, in a batch reaction approximately 40 mg of GFP was synthesized per milliliter of reaction volume at 25˚C for 6h. In a dialysis reaction, approximately 140 mg/mL of GFP was synthesized at 25˚C for 20h. This synthesized GFP was easily purified by HPLC. Human ferritin, which 24 subunits of H- and L-chain composed, was synthesized in this cell-free system. The human ferritins with various molecular sizes were formed by the addition of H- and L-chain mRNAs with different ratios. Moreover, this method for establishment of cell-free translation system was applied also to *Spodoptera frugiperda* 21 (Sf21) insect cell. The productivities of these systems were sufficient to perform gene expression analyses such as measurement of enzyme activity and western blotting. Thus, these cell-free systems may be useful tools for the simple synthesis of post-genomic studies as novel protein production methods.

9
CellMontage: Reality for finding homologous gene expression profiles

Wataru Fujibuchi and Paul B. Horton

National Institute of Advanced Industrial Science and Technology (AIST) 2-43 Aomi, Koto-ku, Tokyo, Japan

A large amount of gene expression profile data has been produced throughout the world, and much of it is publicly available. However, there is a lack of algorithms or systems for searching for homologous, i.e. similar in a meaningful way, gene expression profiles.
We have developed a system called CellMontage which enables users to search data from various platforms for similar expression profiles to their query. Our preliminary analysis using Gene Expression Omnibus data found that a profile for a human kidney from one platform successfully matched with the best score to other kidney cell expression profiles among 16 cell types, even for different platforms.

10
G-compass: A Web Tool for Comparative Genomics Analysis

Yasuyuki Fujii[1,2], Takeshi Itoh[2,3], Ryuichi Sakate[1,2], Akihiro Matsuya[1,4], Kanako O. Koyanagi[5], Takuya Habara[1,2], Kaori Yamaguchi[1,2],Yayoi Kaneko[1,2],Tadashi Imanishi[2], and Takashi Gojobori[2,6]

[1]Japan Biological Information Research Center, Japan Biological Informatics Consortium, Tokyo, Japan, [2]Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan, [3]National Institute of Agricultural Science, Tsukuba, Japan, [4]Hitachi Co., Ltd., Tokyo, Japan, [5]Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan, [6]Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Mishima, Japan

The alignment of complete genomic sequences is invaluable for the progress of evolutionary biology and comparative genomics. However, the genome sizes of species such as human and mouse are extremely large, and dynamic and complicated rearrangements of genome sequences took place during evolution. Therefore, we need a special tool that can cover a wider range of genomic segments and describe the detailed nature of each alignment. We thus developed a new web tool for comparative genomics analysis called G-compass. In the current version of G-compass, nearly 820,000 aligned segments of the human and mouse genomes are shown in its main view. The genome alignments were generated from a pair of 3Gb sequences that are completed genomes of human and mouse. The large-scale computation schema was implemented into PC-cluster system to fully accelerate the generation of the sequence alignments. The computation time spent was only 3days. By further computation, 13,601 orthologue candidates of human and mouse were identified based on the gene set that was found on conserved genomic regions. In the main panel of G-compass, the set of orthologue candidates as well as their exon-intron organization are displayed clearly. Additionally to these functions, G-compass has a dynamic window analysis system to compute nucleotide diversity, proportion of insertions/deletions in each alignment, CpG density and GC contents of each genomic sequence. The results of window analysis are immediately displayed in graphics alongside the genomic sequences. Hence G-compass was designed not only as a genome alignment viewer but also as a user-friendly platform for studying comparative genomics. In addition, G-compass is widely applicable to various researches of post-genomic era such as extraction of conserved genomic regions. G-compass will be integrated into a human gene annotation database H-InvDB (http://www.h-invitational.jp/) that is freely accessible on the world wide web. Integration of G-compass will make the annotation of human genes in H-InvDB more valuable by providing information on evolutionary conservation of genes in these two species.

11
Computational approach to evaluate coding potential in mammalian cDNA sequence

Masaaki Furuno, Carol Bult

The Jackson Laboratory, Bar Harbor, Maine, USA

In recent years, a number of high-throughput cDNA (HTC) sequencing projects have generated and published large data sets consisting of tens of thousands of mRNA sequences for human, mouse and rat. The sequences of these transcripts are essential for genome annotation and for predicting the amino acid sequences of proteins. The ability to correctly predict coding regions from transcript sequences can be complicated by a combination of cloning errors and sequence quality issues. For example cloning artifacts such as internal priming can result in truncated coding region; sequence quality issues due to base-calling errors can introduce false stop codons.

We have developed a computational algorithm called mTRANS to estimate protein coding potential in cDNA sequences. The mTRANS program judges coding potential based on the presence of specific sequence features of transcript sequences and by the analysis of EST and genomic sequence information. Specifically, the mTRANS program estimates the coding potential of cDNA by evaluating such sequence attributes as the completeness of 5' and 3' ends of transcripts, intron retention, and the presence of a signal for nonsense mediated decay (NMD). The algorithm in mTRANS attempts to compensate for base-calling errors by creating a virtual cDNA sequence if sequence differences are detected between the actual transcript sequence and its corresponding genomic sequence. The output of mTRANS includes a prediction of the coding sequence (CDS) region for each cDNA that is classified as protein-coding as well as a classification of the transcripts (e.g., complete CDS, immature, 3'-truncated, etc.). To test mTRANS we evaluated more than 60,000 full-length cDNA clones for the mouse that were generated by the RIKEN Genome Sciences Center in Japan. We compared the CDS predictions and classifications generated by mTRANS to the results obtained for the same data set based on manual curation. The results of this comparison will be discussed.

12
Role of Inter and Intra-molecular Interactions to Protein-DNA Recognition

M. Michael Gromiha(1,*), Joerg G. Siebers(2), Samuel Selvaraj(3), Hidetoshi Kono(4) and Akinori Sarai(2)

(1)Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Aomi Frontier Building 17F, 2-43 Aomi, Koto-ku, Tokyo, 135-0064, Japan (2)Department of Biochemical Engineering and Science, Kyushu Institute of Technology (KIT), 680-4 Kawazu, Iizuka 820-8502, Japan. (3)Department of Physics, Bharathidasan University, Tiruchirapalli 620024, Tamilnadu, India (4)Neutron Science Research Center and Center for Promotion of Computational Science and Engineering, Japan Atomic Energy Research Institute (JAERI), 8-1, Umemidai, Kizu-cho, Souraku-gun, Kyoto 619-0215 Japan

Protein-DNA recognition plays an essential role in the regulation of gene expression. Regulatory proteins are known to recognize specific DNA sequences directly through atomic contacts (intermolecular readout) and/or indirectly through the conformational properties of the DNA (intra-molecular readout). However, little is known about the respective contributions made by these so called direct and indirect readout mechanisms. We addressed this question by making use of information extracted from a structural database containing many protein-DNA complexes. We quantified the specificity of intermolecular (direct) readout by statistical analysis of base-amino acid interactions within protein-DNA complexes. The specificity of the intra-molecular (indirect) readout due to DNA was quantified by statistical analysis of the sequence-dependent DNA conformation. Systematic comparison of these specificities in a large number of protein-DNA complexes revealed that both intermolecular and intra-molecular readouts contribute to the specificity of protein-DNA recognition, and that their relative contributions vary depending upon the protein-DNA complexes. We demonstrated that combination of the intermolecular and intra-molecular energies derived from the statistical analyses lead to enhanced specificity, and that the combined energy could explain experimental data on binding affinity changes caused by base mutations.

Reference: M. Michael Gromiha et al. (2004) *J. Mol. Biol.* 337, 285-294.

13
Genome profiling: An approach towards comparative genome analysis

Keiichi Hamano[1], Yoshito Sadaie[1,2], Motofumi Suzuki[3], Yasushi Hirano[1,4], Koichi Nishigaki[1,5]

[1]Rational Evolutionary Design of Advanced Biomolecules, Saitama, Saitama, Japan [2] Department of Molecular Biology, Saitama University, Saitama, Japan [3]Microbe Division /Japan Collection of Microorganisms RIKEN BioResource Center, Saitama, Japan [4] Saitama Pref. Agriculture-and-Forestry Research Center, Saitama, Japan [5] Department of Functional Materials Science, Saitama University, Saitama, Japan

Comparative genome research performs comparisons among genomes of various organisms at the levels of structure and function, leading to disclosure of universality, versatility, mutual dependence, and continuity of living organisms. Therefore, the whole genome sequencing of a single species is not sufficient and it is definitely important to compare many genomes of different-species. To adopt the whole genome sequencing is easy to assume but difficult to carry out even in the age when 165 species of bacteria have been sequenced. Since more than 99% of the whole bacteria are assigned to be‘unculturable’, it is also difficult to collect a necessary amount of purified DNA for whole genome sequencing. Therefore, it cannot be strategically reasonable nor effective to pursue the whole genome sequencing for all purposes and for any occasions. Thus, a methodology is awaited which enables us to obtain a necessary and sufficient amount of genome information on any kind of organism with ease and speed. Genome profiling, invented by Nishigaki and collaborators at Saitama Univ., is the very technology that can meet all the requirements imposed on genome analysis, which is now widely applied to a diversity of organisms so as to collect necessary information on the genomes of those organisms dealt in Saitama Bio Project (REDS). In this paper, we will present GP (genome profiling) data on the genomes of rice and bacteria which visually show the similarity or distance in genome between any pair of closely-related species through rapid and convenient processing inherent to GP.

14
Identification of the 3' end of the human mu opioid receptor mRNA and comparison of human and mouse 3'-untranslated regions

Wenhua Han [1)], Shinya Kasai [1)], Soichiro Ide [2)], Junko Hasegawa [1)], Ichiro Sora [3)], Kazutaka Ikeda [1)]

1)Dept Mol Psychiatry, Tokyo Inst Psychiatry, Tokyo, Japan, 2)Dept Pharmaceut Sci, Hiroshima International Univ, HiroshimaKura, Japan, 3)Dept Psychobiol Tohoku Univ Grad Sch Med, Sendai, Japan,

Mu opioid receptor (MOR) is a major site for the analgesic action of most opioid drugs such as morphine, methadone and heroin. The structure of the human MOR gene has been partially defined and the sequence information is available in Genbank. It is now apparent that 3'-untranslated region (UTR) can regulate the rate of translation and degradation of mRNA through 3'-UTR sequence-binding proteins. Northern blot analysis suggests that MOR gene possesses an especially long 3'-UTR. However, the 3' end of the human MOR mRNA has not been reported until now. In the present study, we identified the 3' end of the human MOR mRNA by the 3'-rapid amplification of cDNA ends (3'RACE)-PCR method. We also compared the human and murine 3'-UTR sequences. We detected polyA signal 13633bp downstream from the stop codon. Scatterplot comparisons of these species' 3' sequences clearly show the area of high cross-species conservation at the regions from +1 to +2264, from +8418 to+9643 and from +11792 to +13633. Polymorphisms from these regions would thus be likely candidates for functional variations.

15
Genetic analysis of Green rice leafhopper-resistant Japonica rice by RDA.

Yoshinori Hatakeyama[1], Yasushi Hirano[2], Yoshio Akiba[2]

[1]*Saitama Pref. Collab.of Regional. Entities for the Adv. of Tech. Excellence, JST, Saitama, Japan.*
[2]*Saitama Pref. Agri. and Fores. Res. Cent. Saitama, Japan*

Green rice leafhopper (Grh), Nephotettix cincticeps, is one of the major serious insect pest in Asian rice fields which causes direct damage to the rice plant by sucking the plant sap or by transmitting the rice dwarf virus diseases. For this purpose, although a safe and nutritionally adequate cultivation is desired, extermination of Grh has thus so far been performe by Agrochemicals but not completely since the appearance of agrochemical resistant-Grh. In this stream, the genetically modified organisms (GMOs) technology which use to trick plants into producing their own pesticides, can mitigate agrochemicals but the rate of adoption for GMOs is low and declining. Therefore, it seems to be necessary to create the Grh-resistant rice without using GMOs technologies.
Several Indica species of rice has been identified with Grh-resistant gene but not yet any Japonica rice varieties. Moreover, it was reported to Indica rice from genetic analysis that there were five kinds of Grh resistance factors independently. The functions of these resistant factors are unknown. Therefore, there have been efforts to created Grh-resistant gene in Japonica rice by crossing them with Indica rice.
With this approach, here, we created 104 lines of F2 strains by the mating between Grh-resistant Japonica rice (SAINOKAGAYAKI) and Grh-non resistant Japonica rice (TAMAKEI No.86). Next, we extracted DNA from selected 5 resistant and 5 non-resistant strains and then analyzed their difference by using Representational Difference Analaysis (RDA) (Lisitsyn et al., 1993). Furthermore, we detected 18 specific DNA fragments from Grh-resistant strain and also some common DNA fragments in both strains.

Reference: Lisitsyn N, Lisitsyn N, Wigler M. Cloning the differences between two complex genomes. *Science*. 1993 Feb **12**;259(5097):946-51

16
Alternative Splice Variants Encoding Unstable Protein Domains Exist in the Human Brain

Keiichi Homma[1,2], Reiko F. Kikuno[3], Takahiro Nagase[3], Osamu Ohara[3,4] and Ken Nishikawa[1]

[1] *National Institute of Genetics, Mishima, Japan*
[2] *Japan Science and Technology Corp., Kawaguchi, Japan*
[3] *Kazusa DNA Research Institute, Kisarazu, Japan*
[4] *RIKEN Research Center for Allergy and Immunology, Yokohama, Japan*

Alternative splicing has been recognized as a major mechanism by which protein diversity is increased without significantly increasing genome size in animals and has crucial medical implications, as many alternative splice variants are known to cause diseases. Despite the importance to know what structural changes alternative splicing introduces to the encoded proteins for the consideration of its significance, the problem has not been adequately explored. Therefore we systematically examined the structures of the proteins encoded by the alternative splice variants in the HUGE protein database derived from long (>4 kb) human brain cDNAs. Limiting our analyses to reliable alternative splice junctions, we found alternative splice junctions to have a slight tendency to avoid the interior of SCOP domains and a strong statistically significant tendency to coincide with SCOP domain boundaries. These findings reflect the occurrence of some alternative splicing events that utilize protein structural units as cassette. However, 50 cases were identified in which SCOP domains are disrupted in the middle by alternative splicing. In 6 of the cases, insertions are introduced at the molecular surface, presumably affecting protein functions, while in 11 of the cases alternatively spliced variants were found to encode pairs of stable and unstable proteins. The mRNAs encoding such unstable proteins are much less abundant than those encoding stable proteins and tend not to have corresponding mRNAs in non-primate species. We propose that most unstable proteins encoded by alternative splice variants lack normal functions and are evolutionarily dead end.

17
WoLF PSORT Protein Localization Prediction Tool

Paul Horton, Keun-Joon Park

AIST, Computational Biology Research Center

WoLF PSORT builds upon the work of the k nearest neighbor predictor PSORTII. Like PSORTII, WoLF PSORT uses the PSORT localization features developed by Nakai *et al*. In addition WoLF PSORT considers amino acid content and 4 features taken from iPSORT (Bannai et al.) WoLF PSORT offers three improvements relative to PSORTII. 1). WoLF PSORT uses a new dataset derived from SWISS-PROT 44. 2). WoLF PSORT includes an algorithm for selecting and weighting features. 3) WoLF PSORT handles multiply localized proteins in a natural way.

Cross-validation studies show that WoLF PSORT improves the prediction accuracy by about 10% relative to PSORTII. We also investigate the relationship between sequence similarity based prediction and the more *ab initio* method used by WoLF PSORT. We show that for sequences without close BLAST hits, WoLF PSORT performs better than sequence similarity.

18
Cloning, expression, and genomic analysis of small heat shock proteins from the nematode *Caenorhabditis elegans*

Meng-Yin Hsieh, I-Ju Yeh and Shyh-Horng Chiou

Institute of Biochemical Sciences, National Taiwan University, Taipei 10617, Taiwan

The *Caenorhabditis elegans* genome project was completed in 1998, making it the first multicellular organism for which the genome has been sequenced. Here, we use *C. elegans* as a model to investigate the structure and function of small heat shock proteins (sHSPs). sHSPs form a diverse family of proteins that are produced under various stresses in all organisms. The common structural features for sHSPs comprise the N-terminal domain and a C-terminal tail, which flank the evolutionarily conserved alpha-crystallin domain, which exists ubiquitously as a structural domain of the major eye lens protein, i.e. alpha-crystallin. Thus sHSP and alpha-crystallin of animal eye lenses have been proposed to constitute a superfamily of related molecular chaperones with similar secondary structure and aggregation property. In this study we have made an attempt to characterize and compare the physical-chemical property and chaperone function of sHSP12.2 and sHSP12.3 from *C. elegans*, two smallest known members of sHSPs to date. We have amplified cDNAs of sHSP12.2 and sHSP12.3 from *C. elegans* at a mixed stage under heat-shock by RT-PCR, using the designed primers based on the cDNA sequences (http://www.wormbase.org). We have purified these small heat shock protein members by conventional ion-exchange and gel-filtration chromatographies in large quantity suitable for the physicochemical characterization and activity comparison of these proteins with eye lens α-crystallin. Based on the ultracentrifugation analysis and cross-linking assay, these two proteins were both found to form a tetrameric complex. Besides, sHSP12.3 (Tm>90˚C) is more thermostable than sHSP12.2 (Tm=62˚C). However, in contrast to alpha-crystallin and other sHSP of larger sizes, both sHSP12.2 and sHSP12.3 possess very weak chaperone activity. A proteomic analysis by 2D-electrophoresi coupled with MALDI-TOF mass spectrometry has identified sHSP12.2 and sHSP12.3 as two distinct protein spots in the proteome map from *C. elegans* under heat stress. Further investigation on the biological roles of these small sHSPs *in vivo* is currently in progress.

19
Functional role of mouse KIAA genes on neural differentiation -Chronological analysis for gene expression by mKIAA oligonucleotide microarray-

Kazuhide Imai[1,2], Makoto Kawai[1,2], Takahiro Nagase[3], Hisashi Koga[1,2,3]

[1] CREATE, [2] Chiba Industry Advancement Center, [3] Kazusa DNA Research Institute

Our project begun in December 2001 was to collect and characterize cDNAs that encode mouse counterparts of human KIAA proteins (mKIAA) which had been characterized by Kazusa DNA Research Institute. Comprehensive approach, such as DNA microarray, has become strong tools in further characterization of mKIAA genes. Therefore, we have also begun to generate cDNA and oligonucleotide microarrays consisting of originally cloned mKIAA genes. We have already examined and reported the optimal condition of the glass-base oligonucleotide microarray. Since mKIAA genes are predominantly expressed in the brain and most of them have not yet been functionally elucidated, we thus examined functional role of mKIAA genes in the brain using conventional model for neural differentiation (a mouse neuroblastoma cell line after treatment with all-trans-retinoic acid [ATRA]) and mKIAA oligonucleotide microarray. After mRNA extraction from the cells exposed to ATRA for 1, 3, 6, 12 and 24 hours, the mRNAs were labeled with Cy5 by reverse transcription primed by random nonamer. The labeled cDNAs were hybridized on the microarray and the resulting quantified data were then analyzed on a microcomputer (Windows PC) using Visual Mining Studio (Mathematical Systems) modified for our project. Chronologically significant change of gene expression was detected in several genes, and the changes were confirmed by RT-PCR technique. In these genes, half of them were functionally unidentified. From the structural information of encoded proteins, it seems that these genes participate in neural differentiation through pathways of cell damage and signal transduction related to IP3 or G protein. We will present the detail of these data and discuss the functional role of mKIAA genes on neural differentiation.

This project was performed by the support of Japan Science and Technology Agency (JST).

20
Use of mammalian expression vectors derived from ORF trap clones to evaluate quality of antibodies comprehensively raised against mKIAA proteins

Susumu Inamoto[1, 2, 3], Toshiyuki Nankumo[1], Mari Watanabe[1], Kiyo Shimada[2, 3], Mihoko Nagano[2, 3], Hisashi Koga[2, 3, 4], Daisuke Nakajima[4], Takahiro Nagase[4], Osamu Ohara[4], Akiyuki Ozaki[1, 5]

[1]*Institute of Research and Innovation, 1201 Takada, Kashiwa, Chiba 277-0861, Japan*
[2]*Chiba Industry Advancement Center, 2-6 Nakase, Mihama-ku, Chiba 261-7126 Japan*
[3]*JST ―CREATE*
[4]*Kazusa DNA Research Institute, Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan*
[5]*NEDO fellow*

cDNAs for KIAA genes were isolated as human long cDNAs (>4kb) of unknown functions. To facilitate the functional analysis of KIAA genes, we have isolated and determined structures of cDNAs for mouse orthologues of KIAA genes (mKIAA genes). Furthermore, we comprehensively raised antibodies against mKIAA proteins. The quality of antibodies has been evaluated by various methods including Western blots using extracts derived from various mouse cell lines, by which more than 50% of antibodies gave bands with expected molecular weight. Interestingly, however, specific bands with unexpected molecular weight were also detected with some antibodies. To reveal the molecular nature of these unexpectedly migrating bands and to improve evaluation of specificity of antibodies, mammalian expression vectors to express C-terminally FLAG epitope-tagged proteins were constructed from ORF trap clones by an *in vitro* recombination-assisted method in a high-throughput fashion. Extracts from HEK 293 cells transfected with these vectors were detected with both anti-FLAG and anti-mKIAA antibodies. The results of improved evaluation of specificity of antibodies as well as detection of possible posttranslational modification of mKIAA proteins will be presented.

21
Construction of Comprehensive Gene Expression Data Base for Cell Lineage Fate Analysis by *in vitro* ES Cell Differentiation System: Analysis of Endothelial Differentiation Pathway

Martin Jakt, Mitsuhiro Okada, Muneaki Miyata, Mari Kono, Shigenobu Oshima, Ritsuko Fujii and Shin-Ichi Nishikawa

Riken Center for Developmental Biology and The Foundation for Biomedical Research and Innovation, Kobe, Japan

Embryonic stem (ES) cells are derived from the inner cell mass of blastocysts and retain their pluripotency during *in vitro* culture. ES cells are a particularly attractive model system for studying the initial developmental decisions that occur during embryogenesis and the associated mechanisms underlying these decisions. As an initial study, we have focused on the vascular cell lineage using a previously established endothelial differentiation system. The high extent of synchronisation of these *in vitro* cultures allows relatively large cell numbers at defined stages of differentiation to be obtained, enabling the identification and analysis of the processes driving differentiation towards the endothelial cell type.

We have used Affymetrix oligonucleotide arrays and *in vitro* differentiation ES derived Flk+ endothelial progenitor cells to identify the molecules and processes underlying vascular development (Nishikawa et al. 1998). These progenitor cells were derived by the differentiation of ES cells on collagen in the absence of LIF and purified by FACS. These cells can give rise to at least three different cell types; endothelial, smooth muscle and hematopoietic lineages (Nishikawa et al 1998, Yamashita et al 2000). In the presence of VEGF, these cells give rise to a mixture of endothelial and smooth muscle cells within 24 hours. We measured the transcriptional profiles of the Flk+ progenitors and VeCadherin+ endothelial cells derived from these by up to 3 days of differentiation. This data was compared against a library of data from a wider set of cell types to identify genes that are induced during the differentiation process. Genes induced during endothelial differentiation fall into 4 distinct categories depending on their prior characterisation; a) known endothelial markers, b) genes which have been shown to be induced during angiogenesis but which are not normally expressed in the vasculature, c) genes which have been characterised as having some other specific expression pattern and d) genes which have been only superficially or not at all characterised. The discrepancies which we observe for genes in group c we believe are largely due to the failure of prior identification of vascular expression for these genes either because they may only be expressed in short lived intermediate cell types, or because the signal may be masked by stronger signals in surrounding tissues.

We will present data that sheds light on the origin of the endothelial cell and the processes which lead to the formation of the vascular network in addition to details of our data analysis system based on gene profiling data base.

References:

Shin-Ichi Nishikawa, Satomi Nishikawa, Masanori Hirashima, Norihisa Matsuyoshi and Hiroaki Kodama (1998) Progressive lineage analysis by cell sorting and culture identifies FLK1+VE-cadherin+ cells at a diverging point of endothelial and hemopoietic lineages

22
Large-scale subcellular localization analysis of human proteins encoded in FLJ cDNA

Yasutomo Kisu[1,2], Takushi Togashi[3], Saki Sono[3], Hiromi Mochizuki[3], Kei Yamaguchi[3], Namiko Takada[3], Hisashi Kuwayama[3], Susumu Matukura[3], Akira Yoshida[3], Tomoe Seki[3], Kohei Murao[3,4], Shuji Suda[3,4], Hideji Ohtani[3,4], Masatoshi Naruse[5], Atsushi Miyawaki[6], Naoki Goshima[1], Sumio Sugano[1,7], Nobuo Nomura[1]

1) *Biological Information Research Center, AIST*
2) *Biotechnology Center, NITE*
3) *Japan Biological Information Research Center, JBIC*
4) *BioIT Business Development Group, Fujitsu Ltd.*
5) *Fujitsu OSL Ltd.*
6) *Brain Science Institute, RIKEN*
7) *Graduate School of Frontier Sciences, University of Tokyo*

With finishing the genome sequence analysis, the human genomic study has shifted to focus on comprehensive functional analysis of gene products. As the subcellular localization of human proteins is one of the most important biological information in the post-genomic era, we have begun the analysis using FLJ cDNAs.

In order to pursue large scale analysis efficiently, we have developed the high-throughput analyzing system, which consists of high-throughput preparation of expression clones using Gateway technology, DNA transfection with an automated robot, and cell imaging using optical microscopes. In this system, fluorescent protein (Venus)-tagged proteins of FLJ cDNAs were expressed in HeLa cells and analyzed efficiently.

So far, over 7,000 clones have been transfected. In about 95% among tested clones, fluorescent images have been captured, and about half of the transfected cDNAs have been annotated to be located in a specific organelle.

23
Effect of phorbol ester on subcellular transcriptome of U937 cells

Hiroshi Kitamura, Tomoko Nakagawa, Michiyo Takayama, Yayoi Kimura, and Osamu Ohara

Laboratory for Immunogenomics, Research Center for Allergy and immunology (RCAI), The Institute of Physical and Chemical Research (RIKEN)

A human histomonocytic cell line U937 differentiates into macrophage-like cells in response to phorbol 12-miristate 13-acetate (PMA). We examined mRNA expression profiles using subcellularly fractionated RNAs before and after PMA stimulation (32nM, 48h). A nylon membrane array contains 2,701 cDNAs represented that 69 and 43 transcripts were up- and down-regulated in the cytoplasm by the stimulation, respectively. In the nuclear fraction, 78 transcripts were also altered in its expression levels. Interestingly, only 28 transcripts were affected in the both fractions. Intracytoplasmic distribution was further investigated using Agilent-made microarray slides. Polysomal/total cytoplasmic expression ratio, representing integration efficiency of transcripts into polysome, became diversified after PMA stimulation. PMA modulated 528 out of 15,017 transcripts in their expression levels in the polysome, while only 303 transcripts were altered in the cytoplasm. Amongst 528 transcripts affected in the polysome, 361 (68%) transcripts were changed only in the fraction. Eventually, we identified 105 transcripts which are altered in the expression level in a single fraction, simultaneously changed in the polysomal/cytoplasmic expression ratio. This set contains several immune-related transcripts, for example αV-integrin and caspase recruitment domain family member 6 (CARD6). Real-time PCR and western blotting confirmed the array results. All these results collectively indicated that PMA-induced differentiation remarkably influenced mRNA distribution as well as its expression level in U937 cells.

24
Expression and functional analysis of a small heat-shock protein (HSP16.1) from the nematode *Caenorhabditis elegans*

Chun-Jen Lin, Chung-Cheng Shie and Shyh-Horng Chiou.

Institute of Biochemical Sciences, National Taiwan University, Taipei 10617, Taiwan

Small heat shock proteins (sHSPs) form a diverse family of proteins that are produced under various stresses in all organisms. It has been shown that they have chaperone-like activity, which can bind unfolded or misfolded proteins and maintain them in a folding-competent state. The common structural features of small heat shock proteins comprise an N-terminal domain and a C-terminal tail, which flank the evolutionarily conserved α-crystallin domain. However, α-crystallin, a major protein class of most animal eye lenses, was also found to possess chaperone-like activity similar to small heat shock proteins. Thus sHSPs and α-crystallin constitute a superfamily of related molecular chaperones with similar chaperone-like activity. In this study, we have cloned, overexpressed, and characterized the chaperone-like activity of HSP16.1 from *Caenorhabditis elegans*. The cDNA sequence encoding HSP16.1 was amplified using reverse transcriptase/ polymerase chain reaction (RT-PCR) based on the two primers designed according to the nucleotide sequences obtained from the genome database of *C. elegans*. The overexpressed HSP16.1 with high purity was then characterized with regard to its native molecular size, thermostability and chaperone-like activity assay. Although the midpoint temperature for protein aggregation (Tm) of HSP16.1 is about 86˚C much higher than mammalian α-crystallins, its secondary structure gradually changes when the temperature is over 50˚C, accompanied by the decrease of chaperone-like activity. Moreover, in contrast to porcine α-crystallin, the oligomeric complexes and chaperone-like activity of HSP16.1 do not change after preheating treatment. These results suggested that HSP16.1 may be a thermostable protein with refolding potential, and its secondary structure is important to its chaperone-like activity. Crystallographic study of this novel recombinant HSP16.1 is currently in progress.

25
Finding genes based on the human-mouse genome comparison

Kazuharu Misawa[1] and Reiko F. Kikuno[2]

[1]CREATE, Chiba Industry Advancement Center, [2]Kazusa DNA Research Institute

Identifying protein coding regions (CDS) on the genome DNA sequences is one of the most important aspects in the post-genome era. To develop a new method to find CDS, we focused on the fact that CDS usually show a characteristic substitution pattern because of the functional constraints on the protein sequences when compared with non-coding region, such that, synonymous substitutions are more frequently observed than nonsynonymous substitutions. We constructed a codon-to-codon substitution matrix by comparing CDS from the orthologous gene pairs between mouse and human. By using the matrix, we made a scoring scheme that gives higher scores for CDS and lower scores for non-CDS. We implemented this scoring scheme into BLAST program and conducted a cross-species comparison by applying this program. We searched for the homologous DNA sequences on the mouse genome by using the human cDNA sequences coding for the known proteins as queries and *vice versa*. We detected candidates of CDS, and calculated the protein-coding probability for each candidate. This algorithm is expected to detect short genes as well as short exons which are not found before. We will show the results of the cross-species comparison and discuss the reliability of this method.

26
Construction of expression-ready clones for mammalian long cDNAs encoding relatively large proteins and analysis of subcelluler localization of the tagged proteins

Daisuke Nakajima[1], Kenji Saito[2], Hisashi Yamakawa[1], Noriko Okazaki[1], Reiko Kikuno[1], Manabu Nakayama[1,2], Reiko Ohara[1], Hisashi Koga[1,3,4], Takahiro Nagase[1], Osamu Ohara[1,2,5]

[1]Kazusa DNA Res. Inst., [2]Chiba Univ., Grad. Shc. of Pharm. Sci., [3]JST/CREATE, [4]Chiba Industry Advancement Cent., [5]RIKEN Res. Cent. for Allergy and Immunology

For comprehensive functional analysis of human gene products, some groups have started collecting human cDNA clones. Since 1994, we have been conducting a human cDNA sequencing project and already predicted the protein coding sequences of over 2000 genes previously unidentified. We have focused our sequencing efforts on the analysis of long cDNAs encoding large proteins since these products were likely to play an important role in higher multicellular organisms. To create a resource for functional analysis of these mammalian large gene products, we attempted to prepare Gateway-based protein expression clones in forms of native and tagged proteins using human and mouse KIAA full-length cDNAs and cDNAs for human genes encoding relatively large proteins. A homologous recombination system in *Escherichia coli* was applied for direct cloning of protein-coding sequences into entry clones to avoid laborious manipulation and to prevent mutations in the protein-coding sequences. More than 2800 entry clones in native and fusion forms have been confirmed to transfer correctly by one-pass sequencing. The sizes of recombinant products produced in an *in vitro* transcription/translation system have been estimating by SDS polyacrylamide gel electrophoresis to evaluate their integrity having no obvious deletions and/or insertions instead of re-sequencing of their entireties. The subcellular localization of the exogenous GFP-fusion proteins has been analyzing by fluorescent microscope using 293 cells. We are also trying to prepare the stable cell lines which harbor one-copy of exogenous gene at particular chromosomal locus and express their proteins under the control of Tet repressor. These resources are ready to use for functional analysis of recombinant proteins and will provide us invaluable information of gene functions from proteomic point of view.

27
A knowledge base for the protein localizome.

Mitsuteru Nakao, Keun-Joon Park and Paul Horton

Computational Biology Research Center (CBRC), National Institute of Advenced Industrial Science and Technology (AIST), Aomi Frontier Building 17F, 2-43 Aomi, Koto-Ku, Tokyo, 135-0064, Japan

We are developing a novel knowledge base, CBRC Localizome Knowledge Base (CLKB), integrating protein localization information from various resources from computer predictions to experiments. CLKB is designed to assign subcellular localization annotation to protein reference sequences, and to retrieve protein sequences based on controlled evidence levels.
Recently, there are many studies to determine large-scale protein localization based on high-throughput methods. Small-scale determination of protein localization is also a typical step in functional analysis of proteins. We are integrating such various data resources into the CBRC Localizome Knowledge Base. Localizome information is stored on the framwork of the Gene Ontology database. This allows us to manage complex data queries.
CLKB enables users to fetch protein sequences for a set of annotations with certain evidence in specific organisms. This functionality provides a pipeline to flexibly construct a training dataset for developing protein localization prediction methods. CLKB also provides a platform for mining protein localization information to understand the molecular mechanism of protein subcellular localization and to discover new localization signals.

28
Collection of phenotypes of KIAA genes encoding extremely large proteins using gene-targeted mice

Manabu Nakayama[1], Midori Iida[2], Haruhiko Koseki[2] and Osamu Ohara[1]

[1]Department of Human Gene Research, Kazusa DNA Research Institute, Kisarazu, Japan [2]Laboratory for Developmental Genetics, Research Center for Allergy and Immunology, The Institute of Physical and Chemical Research (RIKEN), Yokohama, Japan

Given that thousands of genes exist in the mammalian genome, criteria are needed to prioritize their functional analysis and to decrease the likelihood of producing gene-targeted mice that lack overt phenotypes. Initial analysis efforts are likely to be fruitful if focused on genes encoding large proteins, since at least some large proteins serve as frameworks for intricate assembly of protein complexes, and their inactivation would render definitive, observable phenotypes. Here, we describe the functional characterization of the murine homologues of five human KIAA genes (KIAA1409, KIAA1440, KIAA1447, KIAA1768, KIAA1276) that encode large proteins. Gene-targeted mice showed phenotypic and developmental defects resulting from the functional deletion of three of these five genes. Mice with targeted disruption of KIAA1409 lacked the ability to drink milk and those with targeted disruption of KIAA1447 displayed hind leg motor dysfunction. Disruption of KIAA1440 led to embryonic lethality at the blastocyst stage. The high success rate of our approach demonstrates the rationale for the genome-wide functional examination of large proteins in mice using reverse genetics. Moreover, our approach brings a new perspective to the world of biology: a protein complex can be thought of as being organized hierarchically in terms of function. Some proteins are essential for the proper functioning of and integrity of the complex; others are relatively less important, taking on modifier roles. We propose that extremely large proteins are the key element of a protein complex because they function as the framework around which the complex is assembled.

29
Chk tyrosine kinase-induced inhibition of cell proliferation

Yuji Nakayama, Asae Igarashi, Akiko Kawana, Naoto Yamaguchi

Department of Molecular Cell Biology, Graduate School of Pharmaceutical Sciences, Chiba University

The Src-family of non-receptor-type tyrosine kinases consists of proto-oncogene products and related proteins, and is involved in a variety of signal transduction events. Csk-family tyrosine kinases, Csk and Chk, are known to suppress the activities of Src kinases by tyrosine phosphorylation at the C-terminal tail. Although Csk and Chk show 54% homology and have at least partly redundunt functions in regulating Src-family kinase activities, they exhibit distinctive features. To investigate the role of Chk in cell proliferation, various Chk mutants were expressed and effects on cell proliferation was examined.
Expression of the kinase active Chk induced inhibition of HeLa cell proliferation accompanied by cell death. Chk uniquely localizes to the centrosome and the nuclear envelope besides the nucleus and the cytoplasm. In situ binding experiments substantiated the involvement of the N-terminal domain of Chk in its unique localization. Induction of tyrosine phosphorylation was detected on the centrosome and the nuclear envelope. Our findings provide the first evidence for tyrosine phosphorylation of proteins present on the centrosome and the nuclear envelope possibly contributing to inhibition of cell proliferation.
The expression of Chk in COS-1 cells induced dynamic multi-lobulation of the nucleus. The nuclear localization of Chk augmented the multi-lobulation in a kinase activity-independent manner and prolonged S phase of the cell cycle. Cell sorting analysis revealed that nuclear multi-lobulated cells were enriched in late S phase. Multi-lobulated nuclei were surrounded with lamin B1 that was particularly concentrated in concave regions of the nuclei. Furthermore, treatment with nocodazole or taxol disrupted multi-lobulation of the nucleus. These results suggest that Chk-induced multi-lobulation of the nucleus in late S phase, which is dependent on polymerization and depolymerization of microtubules, partly contribute to the inhibition of cell proliferation.
Our findings suggest that the expression of Chk-induced inhibition of cell proliferation involves the kinase-activity-dependent and –independent mechanisms.

30
Genome-wide RNAi in the cricket *Gryllus bimaculatus* as a new model system to study gene functions

Taro Mito, Katsuyuki Miyawaki, Yohei Shinmyo, Hideyo Ohuchi, and **Sumihare Noji**

Department of Biological Science and Technology, Faculty of Engineering, University of Tokushima, Tokushima, Japan.

With the sequencing of the human genome and the genomes of most major model organisms completed, the systematic characterization of gene functions remains a key challenge. During the past few years, RNA interference (RNAi) has become a powerful tool to silence the expression of genes and analyze their loss-of-function phenotype. Genome-wide RNAi screens against all predicted genes have been successfully used to dissect a variety of biological processes in *Caenorhabditis elegans*. We have developed a new model system to study development and regeneration of a non-*Drosophila* insect, the cricket *Gryllus bimaculatus*. Recently, we found that we can use three different types of RNAi, embryonic RNAi, nymphal RNAi, and parental RNAi in *Gryllus* for loss-of-function analyses. With these RNAi methods, *Gryllus* becomes a model system ideally suited for the study of gene functions. If we could obtain a genome-wide library of double-stranded RNAs, that target every gene in the *Gryllus* genome, we could perform high throughput assays for various gene functions with embryos, nymphs, and adults. In this poster, recent advances will be summarized.

We isolated about 40 *Gryllus* cDNAs and performed RNAi to study development of *Gryllus*. The molecular mechanisms directing anteroposterior patterning of the *Drosophila* embryo (long-germband mode) are well understood. However, how these mechanisms differ in other modes of insect embryogenesis remains largely unknown. We have examined functions of *caudal* (*cad*) in the intermediate-gemband insect *Gryllus bimaculatus*. Reduction of *Gryllus cad* (*Gb'cad*) levels by RNA interference resulted in embryos with only anterior heads. We found that *Gb'cad* regulates the expression of the gap genes, *hunchback* and *Krüppel*, in the gnathal and thoracic regions. Furthermore, *Gb'cad* was found to be involved in the posterior elongation, acting as a downstream gene in the Wingless/Armadillo signaling pathways. These findings indicate that *Gb'cad* does not function in such a manner as in *Drosophila*, suggesting that functional changes of *cad* occurred during insect evolution. Since Wnt/Cdx pathways are involved in the posterior patterning of vertebrates, such mechanisms may be conserved in animals that undergo sequential segmentation from the posterior growth zone. We also performed RNAi in *Gryllus* nymphs to study molecular mechanisms of regeneration of amputated legs. We found that EGFR is involved in regeneration of the legs. Based on these results, screening strategies and applications as a route to comprehensively characterizing gene functions will be discussed.

31
A cDNA library construction method from a small amount of RNA: Adaptor ligation approach for two-round cRNA amplification using T7 and SP6 RNA polymerase

Reiko Ohara, Reiko F. Kikuno, Hiroshi Kitamura, and Osamu Ohara

Human Gene Research, Kazusa DNA Research Institute, Kisarazu, Japan (R.O., R.K.,O.O.); Research Center for Allergy and Immunology, RIKEN, Yokohama, Japan (H.K.,O.O.); Graduate School of Pharmaceutical Sciences, Chiba University, Chiba, Japan (O.O.)

In this study, we developed a method for cDNA library construction from a small amount of RNA without causing serious size-bias in cDNA population. Nowadays, it becomes a high priority goal to establish a method for preparing a high-quality cDNA library from a small amount RNA source. This is because researchers are often interested in hypothetical genes which are supposed to be expressed only in certain types of cells or tissues under particular conditions such as those seen in pathological samples and developing embryos. For this purpose, we adopted two-round cRNA amplification by T7 and SP6 RNA polymerases in the method of cDNA library consruction. The first-round cDNAs flanked by the phage promoter sequences of T7 and SP6 RNA polymerases were synthesized from 1 microgram total RNA and then subjected to two rounds of cRNA amplification. Comparison of the sizes of the first- and the second-round cRNAs indicated that the size-bias effect caused by two-round cRNA synthesis was not serious. Then the cDNA library was constructed using the second-round cRNAs as template. As the result, we obtained the cDNA library with an efficiency of 1.2×10^{11} colony formation unit/microgram of starting total RNA by using *in vitro* lambda phage recombination system. Characterization of the resultant cDNA library in terms of the insert size, clone redundancy and integrity of 3 prime-ends of cDNAs indicated that the amplified library was as good as a library constructed by a conventional method, though large cDNAs tend to be slightly truncated in the amplified library. This method might enable us to construct a library from a small amount of RNA, even from a single cell, at least on calculation.

32
Novel in situ hybridization methods with InSitu Chip and Sheet-Chip to construct expression databases for all mouse genes

Hideyo Ohuchi, Taro Adachi, Yasunori Hayashibara, Masauki Mitsumori*, Takeshi Ohata, Yoshimasa Kawano[#], Michio Ogasawara**, and Sumihare Noji

*CLUSTER Project by Tokushima Industrial Promotion Organization, University of Tokushima, Tokushima, Japan. *Aloka Co., Ltd. Tokyo, Japan. **Department of Biology, Faculty of Science, Chiba University, Chiba, Japan. [#] Nanbu works Inc., Tokushima, Japan.*

Gene expression profiling is an important component of functional genomics. In order to make a database for 3-dimentional (3D) expression patterns of mouse genes, we have been developing novel methods for in situ hybridization (ISH) analysis. We present a time and cost efficient high-throughput in situ technique to perform a large-scale gene expression analysis in various animal tissues.

For whole-mount ISH, we have developed a rapid and stable system using a two-filters-attached column named InSitu Chip (Aloka Co., Ltd.) In the InSitu Chip system, specimens and reaction buffer are constantly held in the reaction space between the lower and upper filters. Therefore, addition of a buffer, removal of a buffer, and drain-stop are performed by "one-step" of buffer-loading onto the InSitu Chip. With this chip, we can perform the high-throughput (HTP) analysis of gene expression patterns by whole-mount ISH.

For tissue-section ISH, we are developing a sheet-chip method using an adhesive sheet made of a thin plastic film. A specially prepared sheet-chip is adhered to the cut surface of the sample in order to support the section and cut with a disposable blade. With the sheet-chip, we will be able to make automatically sequential sections of a whole mouse. We have performed in situ hybridization with sections on the sheet-chip and obtained intense signals as observed with slides. With this sheet-chip, we will be able to construct a 3D image of expression pattern of a particular gene.

With these methods, we will construct a 3D-expression database of all mouse genes with RNA probes. Now, full-length cDNA libraries for various tissues were made and approximately 10,000 independent clones were selected. RNA probes were synthesized with the mouse clones. A robot to make RNA probes was also made with Aloka Co., Ltd.

Partial cDNA sequences and expression patterns are documented with BLAST results, cluster analysis, images and descriptions, respectively. This information will be entered collectively into a web-based database. As a collaborative work, we will perform ISH with mouse samples.

33
Phylogenetic comparative genomics of photosynthetic eukaryotes and prokaryotes

Naoki Sato(1,2), Masayuki Ishikawa(2), Makoto Fujiwara(1) and Tsuyoshi Saito(2)

(1) Department of Life Sciences, Graduate School of Arts and Sciences, University of Tokyo, Komaba, Meguro-ku, Tokyo 153-8902, Japan (2) Department of Molecular Biology, Saitama University, Saitama 338-8570, Japan

Photosynthesis is a good target of phylogenetic comparative genomics, as chloroplasts are definitely descendents of a cyanobacteria-like endosymbiont. Genomic data of many photosynthetic organisms are now available, including complete genomic sequences of ten cyanobacteria, a red alga Cyanidioschyzon merolae, and a flowering plant Arabidopsis thaliana, as well as draft sequences of rice, a green alga Chlamydomonas reinhardtii, and a diatom Thalassiosira pseudonana. In addition, the genomic sequence of nucleomorph of Guillardia theta, a nuclear remnant of endosymbiotic red alga, is also available. We have developed a software GCLUST to extract clusters of homologous proteins encoded in these genomes. GCLUST uses all-against-all BLASTP data and extracts homolog groups by progressively increasing the threshold E-value. Except for the largest cluster that includes multidomain proteins, the homolog groups represent phylogenetically related families of proteins. We used such homolog groups to analyze relationship between cyanobacteria and photosynthetic eukaryotes, relationship of various photosynthetic eukaryotes, as well as the phylogenetic status of the nucleomorph. In addition, We have extracted 40 homolog groups of unknown function that are conserved in cyanobacteria, a red alga and a plant, for further functional genomic studies. The results indicated that most of the 56 Arabidopsis proteins that belong to these groups are targeted to chloroplasts and regulated by light, and may constitute new chloroplast proteins related to photosynthesis.

34
Genome scale functional classification of GPCR genes detected from bacteria, archer, and eukaryote genomes

Makiko Suwa, Yukiteru Ono

Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-41-6 Aomi, Koto-ku, Tokyo 135-0064, JAPAN

We have developed an automated system for discovering G-protein coupled receptors (GPCR) genes in the whole human genome by integrating tools such as HMM based gene finding, sequence alignment, motif and domain assignment, and transmembrane helix prediction, followed by gene quality refinement. The integration enabled us to detect multi-exon sequences and remote homologues that cannot be detected by only using conventional sequence search tools. By carefully assessing each component, we predicted GPCR candidate genes in four confidence levels ranging from 1,032 candidates at the highest specificity to 1,700 at the highest sensitivity.

The overview of these datasets suggested that chromosome 11 has the largest number of GPCR most of which are olfactory/gustatory receptors, while chromosome 21 and Y show the smallest number of candidates. Almost all of olfactory receptor is single exon gene and they are highly concentrated on the narrow regions in which repeat sequences (LINE) are frequently observed. This fact suggests that the olfactory genes are amplified by frequent gene duplication.

Since the protocol of GPCR gene finding is tuned as applicable for all other species, we applied this method to 134 kinds of bacteria, archer genomes and 8 kinds of eukaryote genomes. It is interesting that no GPCR gene is found from bacteria and archer genomes, while they increased rapidly in case of eukaryote genomes. Each genome has different family distributions, which are characteristic to species. These sequences are summarized in SEVENS database (http://sevens.cbrc.jp/1.20/). Additional information (such as expression data, tertiary structure data, promoter region data etc.) will be included in database with every update chance.
Based on this database we are planning to perform structural and functional classification of GPCR candidates, by our original classification program, based on such as SVM, HMM method with physico-chemical parameters. We hope these tools can contribute to assign functions to more than hundreds of orphan receptor genes which are detected from human genome.

35
Cell-free protein synthesis of human long cDNAs using extracts from insect cells

Takashi Suzuki[1], Takahiro Nagase[2], Toru Ezure[1], Masamitsu Shikata[1], Masaaki Ito[1], and Osamu Ohara[2]

[1] Life Sci. Lab., Shimadzu Co., Nakagyo-ku, Kyoto, Japan and [2] Kazusa DNA Research Institute, Kisarazu, Chiba, Japan

We developed two cell-free translation systems with the extracts from *Trichoplusia ni* (HighFive™, invitrogen) and *Spodoptera frugiperda* 21 (Sf21) insect cells. These systems are able to synthesize firefly luciferase with the yield of about ten-fold higher than rabbit reticulocyte cell-free system. In order to evaluate the performance of these translation systems, expression analysis with human cDNAs obtained at Kazusa DNA Research Institute was performed. We mainly constructed the protein expression vectors for cDNA clones encoding large proteins (average length of amino acid residues is 1,100) by using homologous recombination of *E. coli*. Seventy-three human cDNAs were subcloned and transcribed to mRNAs using T7 promoter. The desired proteins, which were labeled with FluoroTect™ GreenLystRNA (Promega) were successfully expressed in HighFive™ extract, Sf21 extract, and rabbit reticulocyte lysate. The products less than 1,200 amino acid residues expressed in HighFive™ or Sf21 extract were clearly detected. Most of proteins expressed in insect cell-free extracts showed stronger fluorescent intensity than those expressed rabbit reticulocyte lysate. Moreover, non-specific protein bands were not observed in fluorescence analysis. In conclusions, the cell-free translation systems from insect cells would be useful tools for functional analysis of human proteins, and would accelerate biochemical approaches in array-based screens.

36
The Human Anatomic Gene Expression Library (H-ANGEL), the H-Inv Integrative Display of Human Gene Expression across Disparate Technologies and Platforms

Motohiko Tanino[1,2], Marie-Anne Debily[3], Takuro Tamura[1,4], Teruyoshi Hishiki[5], Osamu Ogasawara[6], Katsuji Murakawa[1,7], Shoko Kawamoto[8], Kouichi Itoh[6], Shinya Watanabe[9], Sandro José de Souza[10], Sandrine Imbeaud[3,11], Esther Graudens[3,11], Eric Eveno[3,11], Phillip Hilton[1,2], Yukio Sudo[12], Janet Kelso[13], Kazuho Ikeo[6], Tadashi Imanishi[2], Takashi Gojobori[2,6,14], Charles Auffray[3,11],Winston Hide[13], and Kousaku Okubo[5,6]

[1]Integrated Database Group, Japan Biological Information Research Center, Japan Biological Informatics Consortium, Tokyo, Japan, [2]Integrated Database Group, Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan, [3]Genexpress-CNRS-Functional Genomics and Systemic Biology for Health, Villejuif, France, [4]BITS Co., Ltd, Shizuoka, Japan, [5]Functional Genomics Group, Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan, [6]Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Shizuoka, Japan, [7]Hitachi Science Systems Co., Ltd., Tokyo, Japan, [8]Nara Institute of Science and Technology, [9]Human Genome Center, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan, [10]Ludwig Institute for Cancer Research, Sao Paulo, Brazil, [11]Sino-French Laboratory in Life Sciences and Genomics, Shanghai, China, [12]Ashigara Research Laboratories, Fuji Film Co., Ltd., Kanagawa, Japan, [13]South African National Bioinformatics Institute, University of Western Cape, Bellville, South Africa, [14]Department of Genetics, Graduate University for Advanced Studies, Shizuoka, Japan.

The Human Anatomic Gene Expression Library (H-ANGEL) is a resource for information regarding the anatomical distribution of human gene transcripts. The tool contains protein expression data from multiple platforms which has been associated with both manually annotated full-length cDNAs from H-InvDB and RefSeq sequences. 18,897 of the H-Inv predicted genes have associated expression data generated by at least one platform. It utilizes categorized mRNA expression data from both publicly available and unpublished preliminary sources. H-ANGEL incorporates data generated by three types of methods from seven different platforms. The data is provided to the user in the form of a web based viewer with numerous query options. H-ANGEL is updated with each new release of cDNAs and genome sequence build. In future editions we will incorporate the capability for expression data updates, from existing and new platforms. H-ANGEL is accessible at http://www.jbirc.aist.go.jp/hinv/h-angel/.

37
Analysis system for time-series data of gene expressions

Daisuke Tominaga

Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, Aomi 2-43, Koto, Tokyo 135-0064, Japan

Quantitative time-series data of gene expression levels are being more precisely, and having higher throughput by recent observation technologies. By transfection arrays or cell arrays, for example, expression level of hundreds of genes in living cells can be observed simultaneously in real time. We developed processing and analysis system for quantitative time-series data of gene expression levels as a web based server-client application. By the system, clusters of genes and sequences of changes of genes suggesting dependencies/control sequence of genes are easily obtained.
In the analysis system, outlier data in input data are eliminated at first. When data are obtained by multiple observations, AIC value is calculated for all elimination combinations among observations on each sampling time point. Elimination is decided as a combination which makes AIC maximum value. In second, average values are calculated for each sampling time point of each gene, and then polynomial curves are determined for each gene as best fitted to calculated average values. An order of polynomial curves are determined by AIC, also, compared total sum of fitting errors with the order of the polynomial. Input data without outlier data, calculated average values and fitted polynomial curves are plotted for each gene. Average values and fitted polynomical of all genes are also plotted.
User can divide whole observation time range of gene expression level observation into several time stages. Clustering based on correlation coefficient and introducing AIC to determine threshold of correlation coefficients is done on each stage. Determined clusters, dependencies among genes are shown as clear pictures on a web browser.

38
Human Transcriptome Analysis: variants of splicing and transcription start sites of full-length cDNAs constructed by oligo-capping method.

Ai Wakamatsu[1], Tetsuo Nishikawa[1,2], Kouichi Kimura[2], Jun-ichi Yamamoto[1], Sumio Sugano[3], Nobuo Nomura[4], Takao Isogai[1]

1) Reverse Proteomics Research Institute, Kisarazu, Chiba, Japan; 2) Central Research Laboratory, Hitachi Ltd., Kokubunji, Tokyo, Japan; 3) Department of Medical Genome Sciences, Graduate school of Frontier Sciences, University of Tokyo, Minato-ku, Tokyo, Japan; 4) Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, Koto-ku, Tokyo, Japan.

Full-length cDNA clones are important materials for experimental elucidation of gene function. About 30 thousand of human full-length cDNAs were sequenced by our NEDO human cDNA sequencing project, and were deposited to DDBJ/GenBank/EMBL. From the analysis of human genome sequences, the number of human genes is predicted to be approximately 30K-40K. And the number of mRNA transcribed from the human genome is predicted to be approximately 100K. From these results, a lot of variant cDNAs including splicing and transcription start site variations could be obtained. In order to search splicing variant cDNA efficiently, 30 thousand full-length cDNAs of our project, 42 thousand cDNAs of other cDNA projects and RefSeq were mapped to 28 thousand loci of human genome. We found that 12 thousand loci that consisted of two or more splicing patterns had 41 thousand of splicing patterns. Then the variations of the splicing patterns that affected ORFs were analyzed.
We are analyzing and cloning variant cDNAs including splicing and transcription start site variations from about 1.4 million cDNAs obtained from about 100 kinds of cDNA libraries that consist of human tissues and cells constructed by oligo-capping method. About 500 bp of all 5’-end of the cDNAs were sequenced by our NEDO human cDNA sequencing project. The full-length ratio of 5’-end of those cDNAs was about 90%. We mapped those cDNA sequences to human genome sequences, and are selecting variant cDNAs that affect ORFs for cloning of the selected cDNAs. As variations, we obtained not only splicing variations, but also transcription start site variations. Some of those variations had tissue specific expression profiles.
This work was supported by a grant from New Energy and Industrial Technology Developmental Organization (NEDO) project of the Ministry of Economy, Trade and Industry of Japan.

39
A simple and robust method for preparation of cDNA nylon microarray

Hisashi Yamakawa[1], Satoru Yokoyama[1], Takeo Hirano[1,2], Hiroshi Kitamura,[3] and Osamu Ohara[1, 3]

Department of Human Gene Research, Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan[1], Pharmaceutical Research Department, Sato Pharmaceutical Co., Ltd, 8-5 Higashi-ohi, 6-Chome, Shinagawa-ku, Tokyo 140-0011, Japan[2], and Laboratory of Immunogenomics, RIKEN Research Center for Allergy and Immunology, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan[3]

The DNA array technology has made remarkable progresses in recent years and become an indispensable tool in molecular biology. Although various types of DNA arrays become commercially available, how to prepare high-quality custom-made DNA arrays at a reasonable cost is still an important concern because we cannot abandon the use of a DNA array system designed for specific purposes. To address these problems, we here report the use of rolling circle amplification products of cDNA plasmids dissolved in 80% formamide as DNA probe immobilized on nylon membrane. Because formamide is practically non-volatile under ambient conditions of working space and well dissolves nucleic acids, the use of formamide as DNA solvent eliminated a risk of change in concentration of the DNA solution during arraying, which often takes several hours to a day depending on the numbers of DNA spots and arrays to produce. The use of rolling circle amplification technology greatly reduced a labor for preparation of the spotted DNA. The results in this study demonstrated that the introduction of these two modifications in preparation of nylon DNA array greatly improved its quality.

40
Chronic psychostimulant treatment induces stable changes in some gene expressions in a region specific manner

Hideko Yamamoto[a,*], Yukio Takamatsu[a], Kazuhide Imai[b], Etsuko Kamegaya[a], Makiko Kishida[a], Yoko Hagino[a], Yasuhiro Hara [b,e], Kiyo Shimada [b,e], Toshifumi Yamamoto[a,c], Mika Saito[a], Wenhua Han[a], Ichiro Sora[a,d], Hisashi Koga [b,e], Kazutaka Ikeda[a]

[a]*Dept. of Molecular Psychiatry, Tokyo Inst. Psychiatry, Tokyo, Japan.*
[b]*Chiba Industry Advancement Center, Chiba, Japan.*
[c]*Mol. Recognition, Yokohama City Univ., Yokohama, Japan.*
[d]*Dept. of Neuroscience, Division of Psychobiology, Tohoku Univ., Sch. Med., Japan.*
[e]*Dept. of Genome Informatics, Laboratory for Genome Informatics, Kazusa DNA Research Institute, Japan.*

Repeated methamphetamine (MAP) administration results in behavioral sensitization that remains during long periods of withdrawal. Perpetuation of behavioral sensitization related to abuse and/or relapse may be associated with stable changes in gene expression. To explore the participated genes, in the present study, we used a newly developed cDNA array system comprising mouse KIAA (mKIAA) cDNA clones. We examined the changes in gene expression levels 24 h or 21 days (long-term withdrawal period) after chronic MAP (2 mg/kg, i.p.) treatment for 2 weeks. mRNAs were purified from pooled hippocampus, frontal cortex or nucleus accumbens of 20 mice per group. ^{33}P-labeled DNA probes were synthesized from the mRNAs as templates and hybridized with cDNA array containing approximately 1700 mKIAA clones.

MAP-induced changes (more than 30 %) in gene expression were found in 855 clones in the hippocampus, 100 clones in the frontal cortex, and 445 clones in the nucleus accumbens 24 h after chronic MAP administration. After long-term withdrawal, stable changes remained in 14 clones in the hippocampus, 21 clones in the frontal cortex, and 69 clones in the nucleus accumbens. In these comparisons, their gene expression profiles markedly differ. Among the most populated, the genes are grouped into broad functional categories with synaptic elements, receptor/signal transduction components, ion channel/transport proteins and unknown genes. In the hippocampus and the frontal cortex, expressions of some genes related to opioid receptors were remarkably changed, although they were differently changed depending on the regions. For example, delta opioid receptor gene expression was increased in the hippocampus but decreased in the frontal cortex 24 h after chronic MAP treatment. The characteristic patterns of gene expression were verified by SDS/PAGE and Western blot. Of interest, in the frontal cortex of mice with long-term withdrawal period, mu-opioid receptor gene and protein expressions were markedly decreased. These results suggest that each opioid receptor in the hippocampus or in the frontal cortex has played different roles in development of drug abuse or relapse after long-term withdrawal period.

Furthermore, in the nucleus accumbens, several changed genes could be grouped into categories with cytoskeletal components/synaptic elements, signal transduction molecules/kinases and unknown genes. The alteration in expression of these genes might be also involved in an ongoing synaptic plasticity after repeated MAP treatment and long-lasting neuronal responses.

The present study demonstrated that customized cDNA array could provide information regarding stable changes of gene expressions after long-term withdrawal period. Thus, these genes can be the best candidates for therapeutic targets.

41
H-Invitational Database (H-InvDB): Integrated Database of the Human Transcriptome

Chisato Yamasaki[1], Kanako O. Koyanagi[2], Yasuyuki Fujii[1,3], Takeshi Itoh[1,4], Barrero A. Roberto[5], Takuro Tamura[3,6], Yumi Yamaguchi-Kabata[1], Motohiko Tanino[1,3], Jun-ichi Takeda[1], Satoshi Fukuchi[5], Satoru Miyazaki[7], Nobuo Nomura[1], Sumio Sugano[1,7], Tadashi Imanishi[1,2], and Takashi Gojobori[1,5,9*]

1, Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan; 2, Hokkaido University, Hokkaido, Japan; 3, Japan Biological Information Research Center, Japan Biological Informatics Consortium, Tokyo, Japan; 4, National Institute of Agrobiological Sciences, Ibaraki, Japan; 5, Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Shizuoka, Japan; 6, BITS Co., Ltd., Shizuoka, Japan; 7, Tokyo University of Science, Chiba, Japan; 8, The University of Tokyo, Tokyo, Japan; 9, Department of Genetics, The Graduate University for Advanced Studies, Shizuoka, Japan.

H-Invitational Database (H-InvDB; http://www.h-invitational.jp/) is a human transcriptome database, containing integrative annotation of 41,118 full-length cDNA clones mapped into 21,037 cDNA clusters. The data consists of descriptions of their gene structures, novel alternative splicing isoforms, functional RNAs, functional domains, subcellular localizations, metabolic pathways, predictions of protein 3D structure, mapping of SNPs and microsatellite repeat motifs in relation with orphan diseases, gene expression profiling, and comparisons with mouse full-length cDNAs in the context of molecular evolution. All the data in H-InvDB are shown in two main views, the cDNA view and the Locus view, and five auxiliary databases with web-based viewers; DiseaseInfo Viewer, H-ANGEL, Clustering Viewer, G-integra and TOPO Viewer, and provided as flat files and XML files. H-InvDB is a product of the H-Invitational project, an international collaboration to systematically and functionally validate human genes by analysis of a unique set of high quality full-length cDNA clones using automatic annotation and human curation under unified criteria. This unique integrative knowledge-based platform for conducting in silico data mining with functions such as keyword search, sequence similarity search, and cDNA and genome browser, represents a substantial contribution to resources required for the exploration of human biology and pathology.

Notes

Industrial Support

Luncheon Seminar 12:20-13:05 Saturday 30 October

Applied Biosystems Japan Ltd.

Speaker : **Masahide Habu** (Applied Biosystems Japan Ltd. General Manager)

Title: Integrated Science: Operationalizing iScience at Applied Biosystems

iScience: To better understand the complex interaction of biological systems, life scientists are developing revolutionary approaches to discovery that unite technology, informatics, and traditional laboratory research. In partnership with our customers, Applied Biosystems provides the innovative products, services, and knowledge resources that make this new, Integrated Science possible.

4-5-4, Hatchobori, Chuo-ku, Tokyo, Japan
Phone: 81-3(5566)6100
Fax: 81-3(5566)6501
http://www.apliedbiosystems.co.jp

Applied Biosystems Headquarters
850 Lincoln Centre Drive Foster City, CA 94404 U.S.A.
Phone: 650-638-5800, 800-327-3002
Fax: 650-638-5884
http://www.appliedbiosystems.com

Luncheon Seminar

12:10-12:40 Sunday 31 October

PerkinElmer Japan Co., Ltd.

Territory Product Line Leader: **KAZUHISA MIURA**

Abstract: Recently we, PerkinElmer Life and Analytical Sciences, have started to introduce our innovative tools for the array based proteomics research from this April, which is generally called “Protein Array Technology”. We believe that the fluorescence based Protein Array analysis is now being one of the key technology same as DNA arrays, in order to analyze the protein functions, protein-protein or protein-small molecules interactions, post-translational modifications, antibody-profiles, and biomarkers screening.

Therefore we developed a series of products for preparing Protein Arrays on slide glasses (Piezorray™), processing the Protein Arrays (ProteinArray Workstation™), amplifying of fluorescence signals (Tyramide Signal Amplification), and visualizing of those arrays.

And we are also offering our patented 3D-gel slide glass for Protein Arrays (HydroGel™), which is the best substrate for immobilizing the spotted proteins simply but effectively.

Exhibition Information

Affymetrix Japan K. K.
The Affymetrix GeneChip® System provides a systems biology perspective for both expression profiling and DNA analysis. This integrated system gives you the flexibility to view the genome at a global level or focus on a specific subset of genes. The key to the GeneChip advantage is that each high-density array provides multiple, independent measurements for each transcript and genotype call. Multiple probes mean that you get a complete data set with accurate, reliable, reproducible results from every experiment. By integrating gene expression and DNA analysis on a single platform, the GeneChip System provides a total solution to elucidate complex mechanisms and networks underlying biological process and diseases. Please visit http://www.affymetrix.co.jp/.

Applied Biosystems Japan Ltd.
Applied Biosystems 3730xl DNA Analyzer
The 96-capillary 3730xl DNA Analyzer reduces the time, cost, and complexity of large-scale sequencing projects.
- Higher optical sensitivity and advanced polymers enable you to obtain higher-quality sequencing data for less.
- Multiple automation features decrease costly human errors.
- Optimized polymers increase your productivity without compromising your results.
- New assays with instrument and software upgrades enable new applications in SNP Discovery and Genotyping.

Proteomics Discovery
The 4700 Proteomics Discovery System is the definitive solution for the identification of proteins--from automated acquisition to analysis. This comprehensive system includes a MALDI tandem time-of-flight mass spectrometer with TOF/TOF™ optics, sophisticated bioinformatics, and enabling chemistries.
The 4000 Q TRAP® LC/MS/MS System is a high-performance hybrid triple quadrupole/linear ion trap mass spectrometer for drug discovery, drug development, and proteomics.
The iTRAQ™ Reagent Kit is the first set of amine-specific, stable isotope reagents that can label all peptides in up to four different biological samples simultaneously, enabling relative and absolute quantitation from MS/MS spectra.

Chiba Industry Advancement Center (JST-CREATE program)
There are two remarkable R&D achievements in our project. They are: (1) construction of **InGaP** database (**In**tegrative **G**ene **a**nd **P**rotein expression database) and (2) development of a prototype of DNA and protein microarrayer. The **InGaP** database is opened to the public through our website. 2,000 of mouse homologues of human KIAA genes (>4kbp) as well as related 800 anti-mKIAA antibodies are accumulated through our research, of which 127 mKIAA genes and protein expression profiles are viewed in the **InGaP**. Particular thing with the InGaP is that it contains the results of cDNA microarray analysis, anti-mKIAA antibody analysis such as Western blotting, immunohistochemical staining and subcellular localization patterns of the GFP-mKIAA fusion proteins. Our DNA and/or protein microarrayer with unique sample cooling and pin spotting systems is capable of spotting samples on both glass and membrane bases with higher accuracy.

FUJI PHOTO FILM CO., LTD.
Fujifilm launched an innovative automatic nucleic acids isolation system “QuickGene-800”, along with three types of “isolation kits.” This system offers faster operation, higher purity, and a higher yield than existing nucleic-acid isolation systems. Our technology is the original advanced polymer membrane, it can capture a high yield of nucleic acids in its extremely thin 80-micrometer body. We will further reinforce our R&D capabilities and sales force for the bio-tool business in the life science field.
Please access our home page http://lifescience.fujifilm.com.

Ikeda Scientific, Co., Ltd.
Ikeda Scientific, Co., Ltd. will have an exhibition booth in BITS2004. We are looking forward to seeing you in our booth.

Ikeda Scientific, Co., Ltd http://www.ikedarika.co.jp
Kanda KS building, Kaji- cho 1-8-6, Chiyoda- ku, Tokyo 101-0011

Illumina K.K. (Illumina, Inc.)
◆◇◆◇◆◇◆Flexibility and scalability of BeadArray™ technology◆◇◆◇◆◇◆
Illumina develops tools for the large-scale analysis of genetic variation and function, the results of which will pave the way for personalized medicine, a key goal of genomics and proteomics. Our technology platform also extends to other applications such as high-throughput screening of pharmaceutical compounds and chemical detection.
◆◇◆◇◆Contact us：03-5252-7771(TEL), Sales@illuminakk.co.jp(E-mail) ◆◇◆◇◆

Promega K K
Promega is a worldwide leader in applying biochemistry and molecular biology to development of innovative, precision-designed systems and reagents for molecular and cellular approaches. Promega KK provides optimized products for techniques such as bioluminescence and non-isotopic reporter studies, transfection, in vitro transcription and translation, amplification, cloning, mutagenesis, RNA and DNA purification, cellular signaling studies, apoptosis detection and cell viability determination. For more information, please visit our web site www.promega.co.jp or www.promega.com.

Roche Diagnostics K.K.
Rapid Translation System (RTS) for cell-free protein expression
RTS is a scalable *in vitro* transcription/translation protein expression platform that produces sufficient amounts of protein for characterization studies, functional assays, or structural analysis. RTS offers the advantages of a cell-free protein expression system, eliminating the need for laborious up- and down-stream steps typically associated with cell-based expression systems. Easily adapt your specific applications with a truly "open" system.
For detail procedures and product recommendations, please visit www.proteinexpression.com.

SHIMADZU CORPORATION
SHIMADZU BIOTECH provides total support for proteome research, from sample preparation to MALDI-TOFMS analysis.
We have three types of sample preparation apparatus (Xcise, CHIP-1000, AccuSpot) and two types of MALDI-TOFMS.
AXIMA QIT is a new generation of MALDI-TOFMS. By cleaving the target molecule in several stages in a quadrupole ion trap, it is possible to obtain detailed molecular structure information for many types of modified proteins and sugar chain, etc.

Roster of Participants

Invited Speakers

Bernardi, Giorgio
Laboratory of Molecular Evolution
Stazione Zoologica Anton Dohrn
Villa Comunale, Naples
80121 Italy
Tel: +39 081 5833 300/402
Fax: +39 081 2455807
bernardi@szn.it

Bode, Juergen
Molecular Biotechnology
German Research Centre for Biotechnology
Mascheroder Weg 1, Braunschweig
D-38124 Germany
Tel: +49 531 6181 251
Fax: +49 531 6181 262
jbo@gbf.de

Boutros, Michael
Boveri-Group Signaling and Functional Genomics
German Cancer Research Center
Im Neuenheimer Feld 580, Heidelberg
69120 Germany
Tel: +49 6221 421951
Fax: +49 6221 421959
m.boutros@dkfz.de

Brack-Werner, Ruth
Institute of Molecular Virology
GSF-National Research Center for Environment and Health
Ingolstaedter Landstr. 1, Oberschleisshem
85764 Germany
Tel: +49-(0)89 3187 2923
Fax: +49-(0)89 3187 3329
brack@gsf.de

Dübel, Stefan
Biotechnology
Technical University of Braunschweig
Spielmannstr. 7, Braunschweig
38106 Germany
Tel: +49-531-391-5731
Fax: +49-531-391-5763
s.duebel@tu-bs.de

Ecker, Joseph R.
Plant Biology Laboratory
The Salk Institute
10010 North Torrey Pines Road
LA Jolla, California
92037 USA
Tel: +1- 858-453-4100
Fax: +1- 858-558-6379
ecker@salk.edu

Freeman, Tom
MRC Rosalind Franklin Centre for Genomics Research,
Wellcome Trust Genome Campus, Hinxton,
Cambridge, CB10 1SB, UK
Tel: 01223 494552
Fax: 01223 494512
tfreeman@hgmp.mrc.ac.uk

Gardiner, Katheleen
Eleanor Roosevelt Institute
University of Denver
1899 Gaylord
Denver, Colorado
80206 USA
Tel: +1- 303-336-5652
Fax: +1- 303-333-8423
KGARDINE@DU.EDU

Ghosh, Srinka
Affymetrix Labs _ Transcriptome Group
Affymetrix Inc.
6550 Vallejo Street, Suite 100, Emeryville
94608 USA
Tel: +1-510.428.8502
Fax: +1-510.428.8585
srinka_ghosh@affymetrix.com

Groner, Yoram
Molecular Genetics
The Weizmann Institute of Science
202 Hertzel St., Rehovot
76100 Israel
Tel: 972-8-9343972
yoram.groner@weizmann.ac.il

Invited Speakers

Hayashizaki, Yoshihide
Genome Exploration Research Group
RIKEN, Genomic Science Center
1-7-22, Suehiro-sho, Tsurumi-ku, Yokohama
230-0045 Japan
Tel: +81-45-503-9222
Fax: +81-45-503-9216
yosihide@gsc.riken.jp

Hesketh, John
Institute for Cell and Molecular Biosciences
University of Newcastle
Agriculture Building, Kings Rd
Newcastle-upon-tyne
NE1 7RU UK
Tel: +44-191-222-8744
Fax: +44-191-222-8684
j.e.hesketh@ncl.ac.uk

Hicks, Geoffrey G.
Cell Biology
Manitoba Institute of Cell Biology and the
University of Manitoba
675 McDermot Ave, Room ON5029
Winnipeg, Manitoba
R3E 0V9 Canada
Tel: 204-787-2133
Fax: 204-787-2190
hicksgg@cc.umanitoba.ca

Hide, Winston
South African National Bioinformatics Institute
University of the Western Cape
Modderdam Road
Bellville, 7535 South Africa
Tel: +27 21 959 3645
Fax: +27 21 959 2512
winhide@sanbi.ac.za

Hill, David E.
Center for Cancer Systems Biology and
Department of Cancer Biology
Dana-Farber Cancer Institute
44 Binney, Boston, MA
02115 USA
Tel: +1- 617-632-3802
Fax: +1- 617-632-5739
David_hill@dfci.harvard.edu

Ikeo, Kazuho
Center for Information Biology and DDBJ
National Institute of Genetics
1111 YATA, MISHIMA
411-8540 Japan
Tel: +81-559-816851
Fax: +81-559-816848
kikeo@genes.nig.ac.jp

Jorgensen, Richard A.
Plant Sciences
University of Arizona
Tucson
85721-0036 USA
Tel: 5206269216
raj@ag.arizona.edu

Korn, Bernhard
RZPD German Resource Center for Genome
Research Heidelberg
Im Neuenheimer Feld 580, Heidelberg
69120 Germany
Tel: (49)6221-424700
Fax: (49)6221-424704
korn@rzpd.de

LaBaer, Joshua
Harvard Institute of Proteomics
Harvard Medical School
320 Charles St, Cambridge
02141-2023 USA
Tel: 617-324-0827
Fax: 617-324-0824
mmore@hms.harvard.edu

Mager, Dixie
Terry Fox Laboratory
British Columbia Cancer Agency
601 West 10th Avenue
Vancouver, British Columbia
V5Z1L3 Canada
Tel: 604-877-6070 ext. 3185
Fax: 604-877-0712
dmager@bccrc.ca

Invited Speakers

Maquat, Lynne E.
Biochemistry and Biophysics
University of Rochester Medical Center
601 Elmwood Avenue, Box 712
Rochester, NY
14642 USA
Tel: 585 273 5640
Fax: 585 271 2683
lynne_maquat@urmc.rochester.edu

Miyano, Satoru
Human Genome Center, Institute of Medical Science
The University of Tokyo
4-6-1 Shirokanedai, Minatoku, Tokyo
108-8639 Japan
Tel: +81-3-5449-5615
Fax: +81-3-5449-5442
miyano@ims.u-tokyo.ac.jp

Muckenthaler, Martina
Department of Pediatric Oncology, Hematology and Immunology
University of Heidelberg
Im Neuenheimer Feld 153, Heidelberg
69120 Germany
Tel: 0049 - 06221 566923
Fax: 0049 - 06221 564580
martina.muckenthaler@med.uni-heidelberg.de

Ohara, Osamu
Department of Human Gene Research
Kazusa DNA Research Institute
2-6-7 Kazusa-Kamatari, Kisarazu, Chiba
292-0818 Japan
Tel: +81-438-52-3915
Fax: +81-438-52-3914
ohara@kazusa.or.jp

Oishi, Michio
Director of the Institute
Kazusa DNA Research Institute
2-6-7 Kazusa-Kamatari, Kisarazu, Chiba
292-0818 Japan
Tel: +81-438-52-3944
Fax: +81-438-52-3911
oishi@kazusa.or.jp

Pandey, Akhilesh
McKusick-Nathans Institute of Genetic Medicine
Johns Hopkins University
733 N. Broadway, Broadway Research Bldg. Room 569
Baltimore, MD
21205 USA
Tel: 410-502-6662
Fax: 410-502-7544
Pandey@jhmi.edu

Rancourt, Derrick E.
Biochemistry and Molecular Biology
University of Calgary
3330 Hospital Dr. NW, Calgary
T2N 4N1 Canada
Tel: 403-220-2888
Fax: 403 283-8727
rancourt@ucalgary.ca

Ringwald, Martin
The Jackson Laboratory
600 Main Street, Bar Harbor
ME 04609 USA
Tel: +1 (207) 288-6436
Fax: +1 (207) 288-6132
ringwald@informatics.jax.org

Rustici, Gabriella
Microarray Bioinformatics Group
EBI
Hinxton, CB101SD UK
Tel: +441223494679
gabry@ebi.ac.uk

Rynditch, Alla
Molecular oncogenetics
Institute of molecular biology and genetics
150 Zabolotnogo, Kiev
03143 Ukraine
Tel: 380-44-266-3498
rynditch@yahoo.com

Invited Speakers

Shibata, Daisuke
Department of Plant Gene Research
Kazusa DNA Research Institute
2-6-7 Kazusa-Kamatari, Kisarazu, Chiba
292-0818 Japan
Tel: +81-438-52-3947
Fax: +81-438-52-3948
shibata@kazusa.or.jp

Shinozaki, Kazuo
Plant Functional Genomics
RIKEN Genomic Sciences Center
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama
230-0045 Japan
Tel: +81-45-503-9625
Fax: +81-45-503-9586
sinozaki@rtc.riken.jp

Soares, Bento
Department of Pediatrics, Biochemistry,
Physiology and Biophysics, Orthopaedics
The University of Iowa
375 Newton Road-4184
Medical Education
Biomedical Research Facility
Iowa city IA 52242 USA
Tel: +1 319 335 8250
Fax: +1 319 335 9565
bento-soares@uiowa.edu

Stamm, Stefan
Insitute for Biochemistry
University of Erlangen
Fahrstrasse 17, Erlangen
91054 Germany
Tel: +49 9131 8524622
Fax: +49 9131 8524605
Stefan@stamms-lab.net

Su, Andrew I.
Scientific Computing
GNF
10675 John Jay Hopkins Dr.
San Diego, CA
92037 USA
Tel: 858-812-1656
Fax: 858-812-1630
asu@gnf.org

Tokunaga, Makio
Structural Biology Center
National Institute of Genetics
Yata 1111, Mishima
411-8540 Japan
Tel: +81-55-981-6863
Fax: +81-55-981-6865
mtoku@lab.nig.ac.jp

von Melchner, Harald
Laboratory for Molecular Hematology
University of Frankfurt Medical School
Theodor Stern Kai 7
Frankfurt Am Main
60590 Germany
Tel: +49-69-63016696
Fax: +49-69-63016930
melchner@em.uni-frankfurt.de

Wehrman, Tom
Baxter Labs _ Microbiology and Immunology
Stanford University
269 Campus Dr., Stanford
94305-5174 USA
Tel: 650-218-8064
Wehrman@stanford.edu

Weissman, Sherman
Genetics
Yale University School of Medicine
300 Cedar Street, TAC S-323
New Haven, CT
06520 USA
Tel: 203-737-2281/2282
Fax: 203-737-2286
sherman.weissman@yale.edu

Werner, Thomas
Genomatix Software GMBH
Landsberger Strasse 6, Münich
80339 Germany
Tel: +49 89 599766 0
Fax: +49 89 599766 55
info@genomatix.de

Invited Speakers

White, Kevin P.
Genetics
Yale University School of Medicine
333 Cedar Street
New Haven, CT
06510 USA
Tel: +1 (203) 785-5572
Fax: +1 (203) 785-6350
kevin.white@yale.edu

Wiemann, Stefan
Molecular Genome Analysis
German Cancer Research Center
IM Neuenheimer Feld 580, Heidelberg
69120 Germany
Tel: +49 6221 42 4702
Fax: +49 6221 4252 4702
s.wiemann@dkfz.de

Corresponding Authors and Presenters for Poster Presentations

Ahmad, Shandar (P-1)
Kyushu Institute of Technology
Iizuka, Fukuoka-ken, Japan
shandar@bse.kyutech.ac.jp

Arita, Masanori (P-2)
University of Tokyo
arita@k.u-tokyo.ac.jp

Barrero, Roberto (P-3)
Center for Information Biology-DDBJ, National Institute of Genetics, Japan
rbarrero@genes.nig.ac.jp

Biyani, Manish (P-4)
Rational Evolutionary Design of Advanced Biomolecules, Saitama Small Enterprise Promotion Corporation, SKIP city, Saitama, Japan; Department of Functional Materials Science, Saitama University, Saitama, Japan
sts101@saitama-j.or.jp

Bono, Hidemasa (P-5)
Research Center for Genomic Medicine, Saitama Medical School
bono@saitama-med.ac.jp

Chen, Yen-Shan (P-6)
Institute of Biochemical Sciences, National Taiwan University, and Institute of Biological Chemistry, Academia, Taipei, TAIWAN

Chiou, Shyh-Horng (P-7)
Institute of Biochemical Sciences, National Taiwan University, and Institute of Biological Chemistry, Academia, Taipei, TAIWAN
shchiou@gate.sinica.edu.tw

Ezure, Toru (P-8)
Life Sci. Lab., Shimadzu Co., Nakagyo-ku, Kyoto, Japan

Fujibuchi, Wataru (P-9)
National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan
fujibuchi-wataru@aist.go.jp

Fujii, Yasuyuki (P-10)
Japan Biological Information Research Center, Japan Biological Informatics Consortium, Tokyo, Japan; Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan
yfujii@jbirc.aist.go.jp

Furuno, Masaaki (P-11)
The Jackson Laboratory, Bar Harbor, Maine, USA
furuno@informatics.jax.org

Gojobori, Takashi (P-3, 10, 41)
Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan; Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Shizuoka, Japan; Department of Genetics, The Graduate University for Advanced Studies, Shizuoka, Japan
tgojobor@jbirc.aist.go.jp

Gromiha, M. Michael (P-12)
Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan
michael-gromiha@aist.go.jp

Hamano, Keiichi (P-13)
Rational Evolutionary Design of Advanced Biomolecules, Saitama, Japan
sts106@saitama-j.or.jp

Han, Wenhua (P-14)
Dept Mol Psychiatry, Tokyo Inst Psychiatry, Tokyo, Japan

Hatakeyama, Yoshinori (P-15)
Saitama Pref. Collab.of Regional. Entities for the Adv. of Tech. Excellence, JST, Saitama, Japan
sts105@saitama-j.or.jp

Homma, Keiichi (P-16)
National Institute of Genetics, Japan Science and Technology Corporation
khomma@lab.nig.ac.jp

Corresponding Authors and Presenters for Poster Presentations

Horton, Paul (P-17)
AIST, Computational Biology Research Center
horton-p@aist.go.jp

Hsieh, Meng-Yin (P-18)
Institute of Biochemical Sciences, National Taiwan University, Taipei 10617, Taiwan

Ikeda, Kazutaka (P-14, 40)
Dept Mol Psychiatry, Tokyo Inst Psychiatry, Tokyo, Japan
ikedak@prit.go.jp

Imai, Kazuhide (P-19)
CREATE; Chiba Industry Advancement Center

Inamoto, Susumu (P-20)
Institute of Research and Innovation, Chiba, Japan; Chiba Industry Advancement Center, Chiba, Japan; JST-CREATE
inamots@iri.or.jp

Ito, Masaaki (P-8)
Life Sci. Lab., Shimadzu Co., Kyoto, Japan
ma-itou@shimadzu.co.jp

Jakt, Martin (P-21)
Riken Center for Developmental Biology and The Foundation for Biomedical Research and Innovation, Kobe, Japan

Kisu, Yasutomo (P-22)
Biological Information Research Center, AIST, Biotechnology Center, NITE
ykiss@nifty.com

Kitamura, Hiroshi (P-23)
Laboratory for Immunogenomics, Research Center for Allergy and immunology (RCAI), The Institute of Physical and Chemical Research (RIKEN)
ktmr@rcai.riken.go.jp

Koga, Hisashi (P-19)
CREATE; Chiba Industry Advancement Center; Kazusa DNA Research Institute
hkoga@kazusa.or.jp

Lin ,Chun-Jen (P-24)
Institute of Biochemical Sciences, National Taiwan University, Taipei 10617, Taiwan

Misawa, Kazuharu (P-25)
CREATE, Chiba Industry Advancement Center
misawa@kazusa.or.jp

Nagase, Takahiro (P-26)
Department of Human Gene Research, Kazusa DNA Research Institute
nagase@kazusa.or.jp

Nakajima, Daisuke (P-26)
Department of Human Gene Research, Kazusa DNA Research Institute
nakajima@kazusa.or.jp

Nakao, Mitsuteru (P-27)
Computational Biology Research Center (CBRC), National Institute of Advenced Industrial Science and Technology (AIST), Aomi Frontier Building 17F, 2-43 Aomi, Koto-Ku, Tokyo, 135-0064, Japan

Nakayama, Manabu (P-28)
Department of Human Gene Research, Kazusa DNA Research Institute
nmanabu@kazusa.or.jp

Nakayama, Yuji (P-29)
Department of Molecular Cell Biology, Graduate School of Pharmaceutical Sciences, Chiba University
nakayama@p.chiba-u.ac.jp

Nishikawa, Tetsuo (P-38)
Reverse Proteomics Research Institute, Chiba, Japan; Central Research Laboratory, Hitachi Ltd., Tokyo, Japan
nisikawa-t@reprori.jp

Noji, Sumihare (P-30)
Department of Biological Science and Technology, Faculty of Engineering, University of Tokushima, Tokushima, Japan
noji@bio.tokushima-u.ac.jp

Corresponding Authors and Presenters for Poster Presentations

Ohara, Reiko (P-31)
Human Gene Research, Kazusa DNA Research Institute, Kisarazu, Japan
ohreiko@kazusa.or.jp

Ohuchi, Hideyo (P-32)
CLUSTER Project by Tokushima Industrial Promotion Organization, University of Tokushima, Tokushima, Japan.

Okada, Mitsuhiro (P-21)
Riken Center for Developmental Biology and The Foundation for Biomedical Research and Innovation, Kobe, Japan
mitu@cdb.riken.jp

Sato, Naoki (P-33)
Department of Life Sciences, Graduate School of Arts and Sciences, University of Tokyo, Tokyo, Japan; Department of Molecular Biology, Saitama University, Saitama, Japan
naokisat@bio.c.u-tokyo.ac.jp

Suwa, Makiko (P-34)
Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-41-6 Aomi, Koto-ku, Tokyo 135-0064, JAPAN
m-suwa@aist.go.jp

Suzuki, Takashi (P-35)
Life Sci. Lab., Shimadzu Co., Kyoto, Japan
t-suzuki@shimadzu.co.jp

Tanino, Motohiko (P-36)
Integrated Database Group, Japan Biological Information Research Center, Japan Biological Informatics Consortium, Tokyo, Japan; Integrated Database Group, Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan
mtanino@jbirc.aist.go.jp

Tominaga, Daisuke (P-37)
Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan
tominaga@cbrc.jp

Ai Wakamatsu (P-38)
Reverse Proteomics Research Institute, Kisarazu, Chiba, Japan

Yamakawa, Hisashi (P-39)
Department of Human Gene Research, Kazusa DNA Research Institute, Chiba, Japan
yamakawa@kazusa.or.jp

Yamamoto, Hideko (P-40)
Dept. of Molecular Psychiatry, Tokyo Institute of Psychiatry, Tokyo, Japan
yamahide@prit.go.jp

Yamasaki, Chisato (P-41)
Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan