TRANSCRIPTOME 2002: From Functional Genomics to Systems Biology

March 10-13, 2002
Seattle, Washington, USA

Poster Abstracts

P-01
Definition of Sources of Variability and Noise Factors in Affymetrix Expression Profiling
Marina Bakay, Yi-Wen Chen, Rehannah Borup, Po Zhao, Kanneboyina Nagaraju, and Eric P Hoffman

P-02
Image Analysis and Signal Extraction for cDNA Microarray Images
Bergemann, T.L., Laws, R.J., and Zhao L.P., Fred Hutchinson Cancer Research Center, Seattle WA

P-03
A Superparamagnetic Particle-Based Kit for the Purification of Fluorescent Dye-Labeled cDNA
Lisa R. Booth, Dwayne T. Campogan, Karin A. Hughes, Douglas A. Spicer, Amy L. Springer, Prolinx, Inc., Bothell, WA
Poster PDF file

P-04
Development and Production of a Redundant Oligonucleotide Musclechip: Use for Verification and Prioritization of ESTs
Rehannah HA Borup, Stefano Troppo, Yi-Wen Chen, Tanya M Teslovich, Gerolamo Lanfranchi, Giorgio Valle, and Eric P Hoffman

P-05
Paracel TranscriptAssembler: Identification of Splice Variants for Gene Discovery and Gene Expression Profiling
Cecilie Boysen, Stephanie S. Pao, Jun Qian, Charles P. Smith, Lingyan Zhu, and Joseph A. Borkowski

P-06
Downstream of an RNA Regulatory Protein Causing a Common Human Disease: Expression Profiling Human SMN-Deficient Muscle
Cinzia Brandoli, Serenella Servidei, and Eric Hoffman

P-07
Gene Expression Profiling in Dysferlinopathies Using a Dedicated Muscle Microarray
Stefano Campanaro, Marina Fanin, Beniamina Pacchioni, Silvia Trevisan, Barbara Celegato, Elena Pegoraro, Yukiko K. Hayashi, Giorgio Valle, Corrado Angelini and Gerolamo Lanfranchi

P-08
Progress to Construct the RIKEN Mouse Full-Length cDNA Encyclopedia
P. Carninci, K. Shibata, M. Itoh, T. Arakawa, Y. Ishii, H. Konno, K. Sato, N. Hayatsu, T. Shiraki, T. Hirozane, K. Aizawa, H. Bono, S. Kondo, K. Waki, J. Kawai, A. Yoshiki, Y. Okazaki and Y. Hayashizaki

P-09
Characterizing Transcriptional Determinants of the Innate Immune Response to Infection Through Gene Expression and Literature Profiling
Damien Chaussabel, Roshanak Semnani, Mary Ann McDowell, David L Sacks, Alan Sher, Thomas B Nutman

P-10
PGA-CPR (Chip Profiling Resource): A Web Queried Database of Affymetrix Profiles with Cross-Profile Analysis Tools

Josephine Chen, Dietrich Stephan, and Eric Hoffman

P-11
Using Multivariate Regression Analysis to Detect Differential Co-Regulation of Transcriptions Between Subgroups of Medulloblastomas
Chun Cheng, Andrew R. Hallahan, James M.Olson, and Lue Ping Zhao

P-12
Definition of the Transcriptional and Translational Response to a Physiological Stimulus: A Cross-Species Time Series Muscle Study in Human Volunteers and Rats
Yi-Wen Chen, Eric P Hoffman, Gustavo Nader, Keith R. Baar, Karyn A. Esser, Monica J. Hubal, Paul D. Thompson, Priscilla M. Clarkson

P-13
Transcription Profiling of Cell Death in Drosophila Melanogaster
S. Chittaranjan, E. Garland, D. Freeman, S. Jones, M. Marra, and S. Gorski

P-14
From Human Tissue Samples to Transcriptome Analysis : How to Obtain High Quality RNA
Virginie Copois, Frédéric Bibeau, Patrick Chalbos, Sandrine Imbeaud, Carmella Cappellano, Marc Ychou, Charles Auffray, Emmanuel Conseiller, Bernard Pau, Maguy Del Rio

P-15
Pathogenic Pathways in Spinal Cord Injury Identified by Gene Expression Profiling
Simone Di Giovanni, Susan Knoblach, Cinzia Brandoli, Alan Faden, Eric Hoffman

P-16
Preparation of cDNA Hybridization Targets by Phi29 DNA Polymerase
Rohini Dhulipala, Mubasher Dar, R. Scott Duthie and John Nelson

P-17
Production and Use of DNA Microarrays for Studies of Aging
Mark W. Eshoo and Krysta Felkey

P-18
Intelligent Bioinformatics Systems for Functional Genome Analysis
Daniel Gerlich, Martin Granzow, Julian Mattes, Wolfgang Tvarusko, and Roland Eils, Dr. Wolfgang Tvaruskoi

P-19
The aRNA Amplification Procedure in mRNA-Poor Samples Array Applications
Esther Graudens, Xavier Barlet, Virginie el Marhomy, Charles Auffray & Sandrine Imbeaud

P-20
New Cancer Genome Anatomy Project (CGAP) Internet Resource for the Research Community
Lynette H. Grouse, Carl F. Schaefer, Susan F. Greenhut, Kenneth H. Buetow and Robert L. Strausberg

P-21
Comparative Analysis of TBX1 Genes and Their Promoters from Different Species
Ilya Ioshikhes, Jun Liao and Bernice Morrow

P-22
Comparative Bioinformatics Analysis of Rat and Human ALB-AFP Gene Locus
Ilya Ioshikhes and Joseph Locker

P-23
Expression of Cytokines and Growth Factors in Aging and hTERT-transfected Human Fibroblasts
Yukari Kanzaki, Toshinori Ide

P-24
Full-length cDNA Library Construction Using Lambda Recombination (GatewayTM Technology)
Larissa G. Karnaoukhova, Mark R. Smith, Chris E. Gruber, and Martin Gleeson

P-25
LNA (Locked Nucleic Acid) Microarrays - A New Platform for Gene Expression Profiling
Sakari Kauppinen, Peter Stein Nielsen, Helle Ohlsson, Dan Jeffares, Tobias Mourierand Peter Arctander,

P-26
Clinical Utility of Quantitative Real-Time PCR for Monitoring Leukemia Patients Using RNA Stabilizer In Blood Samples
Ryuji Kawaguchi, Hideyuki Kudoh, Yasuhumi Matsuda, Thoru Murayama and Ralf Wyrich

P-27
In Silico Detection and Characterisation of Cancer-Specific Transcript Isoforms
Janet Kelso

P-28
Protein Interaction Verification and Functional Annotation by Integrated Analysis of Genome-Scale Data
Patrick P.C.W. Kemmeren, Nynke L. van Berkum, Jaak Vilo, Theo Bijma, Rogier Donders, Alvis Brazma & Frank C.P. Holstege

P-29
Application of Statistical Modeling to Microarray Gene Expression Data in Human Lung Cancer
Najma Khalid, Lue Ping Zhao

P-30
The Immediate Early Gene Response of Human Keratinocytes to Gamma-Rays Analyzed by DNA Micro-Arrays
Jérôme Lamartine, Gilles Waksman, Xavier Gidrol, Michèle Martin

P-31
Specific Subtraction of Abundant mRNAs in Skeletal Muscle
Paolo Laveder, Cristiano De Pittà, Stefano Toppo, Giorgio Valle and Gerolamo Lanfranchi

P-32
Systematic Screening of Tissue Specific Cis-regulatory Elements in *Ciona intestinalis* Embryo
Byung-in Lee, David Keys, Mei Wang, Orsalem J. Kahsai, Sylvia Ahn, Qing Zhang, David K. Engle, Irma Rapier, Jason Olivas, Chris J. Detter, Sharon Doyle, Naoe Harafuji, Anna Di Gregorio, Trevor Hawkins, Dan S. Rokhsar, Michael Levine, Paul M. Richardson

P-33
Molecular Genetic Profiling of Gleason Grade 4/5 Prostate Cancers Versus Benign Prostatic Hyperplasia
Mamatha Mahadevappa, Zhaomei Zhang, Mitchell C. Caldwell, Zuxiong Chen, Zhenbin Fan, John E. McNeal, Rosalie Nolley, Thomas A. Stamey, Janet A. Warrington

P-34
A Novel Approach to Eliminate Vector Background and Increase Sequencing Efficiency of cDNAs
Uday Matrubutham, Jody Mirchandani, Jian Liu, Martin A. Gleeson, K. MacDonald, J. Asano, Y. Butterfield, N. Girn, S. Lee, T. Olson, P. Pandoh, U. Skalska, D. Smailus, L. Spence, Jeff Stott, G. Yang, J. Schein and M. Marra

P-35
Identification of Putative Genes Involved in Cell Transformation
Andrei M. Mikheev, Svetlana A. Mikheeva, Vivian Li, Jeffrey Delrow and Helmut Zarbl

P-36
Quality Control of Microarray Samples Using Lab-on-a-Chip Technology

Odilo Müller, Monika Dittmann, Samar Lightfoot, Kathi Williams, and Robert Dubrow

P-37
Prepare a Set of Human cDNA Clones Encoding Large Proteins in an Expression-Ready Form
Takahiro Nagase, Noriko Okazaki, Daisuke Nakajima, Hisashi Yamakawa, Reiko Kikuno, and Osamu Ohara

P-38
Encoding Biological Prior Information Into Gene Expression Analysis
Michael F. Ochs, Frank J. Manion, Ghislain Bidaut, Thomas Moloshok, Jeffrey D. Grant

P-39
Whole Genome Functional Analysis and Metabolic Pathway of *Zymomonas mobilis*
Sang-Ho Park, Ki-Nam Heo, Hyun-Seok Parkv, Jeong-Sun Seo

P-40
Gene Expression Profiling on Radiation Responses
Woong-Yang Park, Chang-Il Hwang, Chang-Nim Im, Min-Ji Kang, Jang-Hee Woo, Ju-Hoon Kim, Kyoung-A Kim, Hyoung-Jin Yoo, Yun-Sil Lee, Su-jae Lee, Jeong-Sun Seo

P-41
Expression Profiling and TPE Analysis Generates Candidate Tumor Markers for Papillary Thyroid Carcinoma
Erwin Pauws, Truus JM Veenboer, Antoine HC van Kampen, Jan JM de Vijlder, Carrie Ris-Stalpers

P-42
From Transcriptome to Proteome: Genome Scale Collection and Expression of Low Abundant Human Full-Length Transcripts
Yijun Ruan, Chia-Lin Wei, Stephan Dillard-Telm, Erika Peterson, Gina Truesdell, Helen Chan, Gary Wolfe, and Dave Mannion

P-43
The Use of Representative Gene Fragments (Rgss) for Less Populated, Highly Replicated Microarrays
Helena B. Samaia, André C. Zaiats, Otávia O. Caballero, Ricardo R Brentani, Sandro J. de Souza and Andrew J.G. Simpson

P-44
Differential Gene Expression in Leishmania Major Friedlin
A. Saxena, S.Yan, L. Worthey, A. Leland, PJ Myler, K. Stuart

P-45
Identification of Genes in Biosynthesis of Fatty Acids and Cholesterol Pathways Using Microarrays
Nila A. Shah, Young-Ah Moon, Suchismita Mohapatra, Jay D. Horton, and Janet A. Warrington

P-46
Bioinformatics for High-Throughput Mu Transposon Sequencing
Ursula Skalska, Yaron Butterfield, Ran Guin, Martin Krzywinski, Kim MacDonald, Duane Smailus, Jeff Stott, George Yang, Scott Zuyderduyn, Jacqueline Schein, Steven Jones and Marco Marra

P-47
A Robust DNA Based Amplification Technique for Tissue-Limiting Experiments on cDNA Microarrays
L. Smith, P. Underhill, C. Pritchard, Z. Tymowska-Lalanne, S. Abdul-Husein, H. Hilton, L. Winchester, D. Williams, T. Freeman, and A. Greenfield

P-48
Affinity-Based Immobilization Tools for Functional Genomics
Amy L. Springer, Anna S. Gall, Karin A. Hughes, Robert J. Kaiser, Guisheng Li, Deborah D. Lucas, Kevin P. Lund, Doug A. Spicer, Jean P. Wiley
Poster PDF file

P-49
Full-Length cDNA and SAGE Sequencing at the British Columbia Cancer Agency Genome Sequence Centre
Jeff M. Stott, J. Asano, Y. Butterfield, R. Guin, M. Krzywinski, K. MacDonald, S. Lee, T. Olson, P. Pandoh, U. Skalska, D. Smailus, L. Spence, K. Teague, G. Yang, S. Zuyderduyn, J. Schein, S. Jones and M. Marra

P-50
Large-Scale Identification of the Promoter Regions of Human Genes Using Full-length Enriched and 5'-end Enriched cDNA Libraries Constructed by the Oligo-Capping Method
Yutaka Suzuki, Riu Yamashita, Sawako Suzuki, Aya Miura, Munetomo Hida, Junko Mizushima-Sugano, Hata Hiroko, Takao Isogai, Kenta Nakai, Sumio Sugano

P-51
Precision Gene Expression Analysis and Bioinformatics Can Track the Effect of Orally Administered, Anti-inflammatory Therapeutic Agents in Human Subjects
Victor V. Tryon, J. Emmick, N. Nicholls, J. Zuzack, D. Trollinger, J. C. Cheronis, M. P. Bevilacqua

P-52
Assessment of Candidate Genes for Fetal Alcohol Syndrome
Mshengu E. Tshabalala, Michele Ramsay and Denis Viljoen

P-53
Microarray Hybridization by Shear-Driven Flows, a New Approach to Enhance Both Sensitivity and Efficiency of Microarray Analysis
Paul Van Hummelen, Kris Pappaert, Johan Vanderhoeven, Danny Clicq, Gino V Baron and Gert Desmet

P-54
Gene Expression Studies of In Vitro HIV-1 Infection:HIV-1 Activates Transcription of the Sterol Biosynthesis Pathway
Angelique B. van 't Wout, Ginger K. Lehrman, Svetlana A. Mikheeva, Gemma O'Keeffe, Gary K. Geiss, Roger E. Bumgarner and James I. Mullins

P-55
SAGEdb: A Computational Platform for Investigations Using Serial Analysis of Gene Expression
Richard Varhol, Scott Zuyderduyn, Mehrdad Oveisi-Fordoei, Erin Garland, Martin Krzywinski, Marco Marra and Steven Jones

P-56
ORESTES and the International Database of Cancer Gene Expression
André L. Vettore, Anamaria A. Camargo, Sandro J. de Souza, Gregory J. Riggins, Robert L. Strausberg, Andrew J. G. Simpson

P-57
Generating and Sequencing Full-length cDNAs of Novel Human Genes Within the German cDNA Consortium
Ruth E. Wellenreuther, Stefan Wiemann, Daniela Heiss, Nina Claudino, Annemarie Poustka

P-58
The Hypoxic Macrophage Transcriptome Profile: A Strategy to Increase the Discovery of Biologically Relevant and Novel Targets for Angiogenesis and Inflammation
Jonathan R. White, Robert A. Harris, Marie H. Craigon, Georgina L. Beard, Sheena R. Lee, Stuart M. Naylor, Chris R. Mundy

P-59
A Regression-based Method to Identify Differentially Expressed Genes in Microarray Time Course Studies and its Application in an Inducible Huntington's Disease Transgenic Model
Xie L. Xu, James M. Olson, Lue Ping Zhao

P-60
Genetic Hardware and Genetic Software after Human Genome Project
Shuwei Yang

P-61
High Throughput Detection of Single Nucleotide Polymorphisms (SNPs) by OCP Microarrays
Zailin Yu, Allan Peng, Amy Cernetich and Qianjin Hu

P-62
Characterization of the Upstream Regions of 836 Orthologous Human/Mouse Genes
Mihaela Zavolan, Terry Gaasterland, and Nikolaus Rajewsky

P-63
Downstream of MyoD: An in vivo 27-time Point Temporal Series, Defining Novel Gene Transcriptional Pathways in Muscle
Po Zhao, Simona Iezzi, Ethan Carver, Devin Dressman, Thomas Gridley, Vittorio Sartorelli, and Eric P Hoffman

P-64
Defining Disease Pathogenesis by Comparative Cross-Species Profiling in Human and Mouse Knock-Outs: the Downstream Consequences of Dystrophin Deficiency
Eric P. Hoffman, Livia Pasquali, Po Zhao, Ksenija Gorni, Frank W. Booth, Brian Tseng, Yi-Wen Chen

P-01

Definition of Sources of Variability and Noise Factors in Affymetrix Expression Profiling

Marina Bakay[1], Yi-Wen Chen[1], Rehannah Borup[1], Po Zhao[1], Kanneboyina Nagaraju[2], and Eric P Hoffman[1], [1]Research Center for Genetic Medicine, Children's National Medical Center, Washington, DC, [2]Division of Rheumatology, Johns Hopkins School of Medicine, Baltimore, MD

We present a systematic study of the sources of variability in expression profiling data using 56 RNAs (28 subjects) isolated from human muscle biopsies (34 Affymetrix MuscleChip arrays), and 36 RNAs from murine cell cultures and tissues (42 Affymetrix U74A v2 arrays). Variables studied included tissue heterogeneity, cRNA probe production, patient diagnosis, and GeneChip hybridizations. We found that the greatest source of variability was often different regions of the same patient muscle biopsy, reflecting variation in cell type content even in a relatively homogeneous tissue such as muscle. Inter-patient variation was also very high (SNP noise). Experimental variation (RNA, cDNA, cRNA, or GeneChip) was minor. Pre-profile mixing of patient cRNA samples effectively normalized both intra- and inter-patient sources of variation, while retaining a high degree of specificity of the individual profiles (86% of statistically significant differences detected by absolute analysis; and 85% by a 4-pairwise comparison survival method).In conclusion, our data using unsupervised cluster analysis and correlation coefficients of 90 RNA samples on 76 oligonucleotide microarrays showed that experimental error was not a significant source of unwanted variability in expression profiling experiments. Major sources of variability were instead from tissue heterogeneity and inter-individual variations.

P-02

Image Analysis and Signal Extraction for cDNA Microarray Images

Bergemann, T.L., Laws, R.J., and Zhao L.P., Fred Hutchinson Cancer Research Center, Seattle WA

Microarray technologies are now routinely used to monitor genome-wide expression profiles. Steps in analyzing microarray data are to visualize the images, inspect intensity values and their variations, and then extract signal data for down-stream analyses. Due to recent advances, several commercial software packages, such as GenePix, have been developed, and are used routinely in practice. While these technologies are shown to be functional, there remain many concerns about assumptions required for their use as well as lack of statistical rigor in analyzing noisy signal data.In this talk, I will present methods to overcome hurdles in the image analysis bottleneck. These methods are incorporated into a software application that is available in the public domain. The goal of our software is to provide researchers with a tool to visually explore image data and assess their confidence in numerical information generated by the experiment. Meanwhile, this application is useful for statisticians as a means to explore variability in image data. Further, I will discuss models to quantify spot signals and their variability. These models speak to the reliability of array data in a given experiment, accounting for sources of bias and variation.

P-03

Poster PDF file

A Superparamagnetic Particle-Based Kit for the Purification of Fluorescent Dye-Labeled cDNA

Lisa R. Booth, Dwayne T. Campogan, Karin A. Hughes, Douglas A. Spicer, Amy L. Springer, Prolinx, Inc., Bothell, WA

Nucleic acid microarrays allow for the profiling of thousands of genes in a single experiment and are an increasingly valuable tool for analyzing gene expression. A typical microarray is composed of a 3 X 1" slide containing an array of nucleic acid fragments to which fluorescently-labeled cDNAs are hybridized. cDNA is the product of reverse transcription reactions where mRNA is converted to cDNA. The removal of excess dyes from these reverse transcription reactions is vital to the accuracy and efficiency of experimental results. Prolinx,Inc., has developed a new method using superparamagentic beads for the removal of excess Cy3- and Cy5-dye labeled dNTP precursors from these reactions. The protocol affords minimal loss of product, making it ideal when sample size is limited. High cDNA recoveries reduce the potential for introduction of variability, and concomitant loss of data reliability, by enabling replicates using the same cDNA sample and increase the possibility of improved detection of rare message. Studies presented will compare and contrast methods of cDNA purification in expression analysis microarrays. Effect of cDNA recovery on signal intensity and assay sensitivity will be included.

P-04

Development and Production of a Redundant Oligonucleotide Musclechip: Use for Verification and Prioritization of ESTs

Rehannah HA Borup[1], Stefano Troppo[2], Yi-Wen Chen[1], Tanya M Teslovich[1], Gerolamo Lanfranchi[2], Giorgio Valle[2], and Eric P Hoffman[1], [1]Research Center for Genetic Medicine, Children's National Medical Center, Washington, DC, [2]CRIBI Institute, University of Padova, Padova, ITALY

We describe the development, validation, and use of a highly redundant oligonucleotide microarray (MuscleChip) containing 4,601 muscle probe sets representing 1,150 known genes expressed in muscle and 2,075 EST clusters from a non-normalized subtracted muscle sequencing project (28,074 EST sequences). This set included 452 novel EST clusters showing no match to previously characterized genes in any database. Each probe set was designed to contain 20-32 25mer oligonucleotides, with each probe evaluated for hybridization kinetics and similarity to other sequences. Approximately 120,000 oligonucleotides were synthesiz-ed. Hybridization of normal muscle cRNA to this MuscleChip showed a correlation between EST cluster member number (n-EST), and determination of a "present" call based on hybridization patterns to probe sets: 30% of singletons, 46% of duplex, and 66% of triplex EST clusters were identified as "present" in normal muscle. Limiting the analysis to the 452 novel EST clusters, and to _at probe sets (288 unique probe sets fulfilling design rules), we found 181/452 (40%) of the novel clusters to be expressed in normal muscle, thereby verifying these ESTs as expressed genes. A hyperlink table was developed to map each of these novel "present" calls confirmed hypothetical proteins to either the RefSeq database or genomic data.

P-05

Paracel TranscriptAssembler: Identification of Splice Variants for Gene Discovery and Gene Expression Profiling

Cecilie Boysen, Stephanie S. Pao, Jun Qian, Charles P. Smith, Lingyan Zhu, and Joseph A. Borkowski, Paracel, Inc., Pasadena, CA

Alternative splice forms play a major role in modifying physiological processes. Although the human genome may have fewer genes than originally thought, many of these give rise to multiple different transcripts. These alternative splice forms can be specific to certain environments, tissues, and drug treatments. A thorough investigation of genes involved in biological functions requires analysis of the corresponding splice variants. Many splice variants can be identified by comparing EST and mRNA sequences originating from the same gene. This can be done even when the genomic sequence is not known, which makes it widely applicable to plant and animal studies. Paracel TranscriptAssembler has been used to recognize splice variants in the transcriptome of various species. Capable of clustering and assembling millions of ESTs and mRNAs, Paracel Transcript-Assembler addresses technical and biological problems in this process, such as chimeric sequences, low quality sequences, repeats, and splice variants. Following the resolution of splice variants into separate contigs, Paracel TranscriptAssembler performs an alignment of the resulting consensus sequences and displays the relationship between splice forms. Specific gene segments are generated for each variant and can be used directly in splice variant specific oligo design for gene expression assays.

P-06

Downstream of an RNA Regulatory Protein Causing a Common Human Disease: Expression Profiling Human SMN-Deficient Muscle

Cinzia Brandoli[1], Serenella Servidei[2], and Eric Hoffman[1], [1]Children's National Medical Center, Washington, DC, [2]Neurology, Catholic University, Rome, ITALY

Spinal Muscular Atrophy (SMA type 1) is an autosomal recessive neuromuscular disease occurring 1 in 10,000 newborns. While the primary pathological feature of SMA has long been considered to be the loss of the cell bodies of alpha-motor neurons in the anterior horns of the spinal cord, increasing evidence points to important defects of the end target of the motor neurons, namely skeletal muscle. SMA is caused by reduction of the expression of the survival of motor neuron (SMN) gene, which is thought to regulate a series of downstream mRNAs, and the protein is highly concentrated at the neuromuscular junction. Downstream targets have not yet been defined. We describe the use of gene expression profiling of muscle biopsies from human SMA type I infants and age/sex-matched controls to identify genes that were potential downstream targets of SMN mRNA regulation. We found specific transcription factors to be highly down-regulated relative to normal age-matched controls, and disease controls (neonatal Duchenne muscular dystrophy patients). Most striking was a 24-fold under expression of a specific Zinc finger transcription factor of otherwise unknown function. These genes become potential down-stream targets of SMN regulation in muscle.

P-07

Gene Expression Profiling in Dysferlin-opathies Using a Dedicated Muscle Microarray

Stefano Campanaro[1], Marina Fanin[2], Beniamina Pacchioni[1], Silvia Trevisan[1], Barbara Celegato[1], Elena Pegoraro[2], Yukiko K. Hayashi[3], Giorgio Valle[1], Corrado Angelini[1] and Gerolamo Lanfranchi[1], [1]CRIBI Biotechnology Center, University of Padova, ITALY, [2]Department of Neurological and Psychiatric Sciences, University of Padova, ITALY, [3]Department of Neuromuscular Research, National Institute of Neuroscience, Tokyo, JAPAN

We have used gene expression profiling to define transcriptional patterns involved in a muscular dystrophy with known primary biochemical defect (the dysferlin deficiency or limb-girdle dystrophy type 2A) and to correlate them with histopathological changes. We have employed a dedicated muscle-specific microarray composed by 2,700 sequences corresponding to the 400-500 most 3'-terminal region of muscle transcripts. Muscle biopsies from 10 homogeneous patients were used as sources of RNA complex target for array hybridizations. Parallel studies for mutation analysis of the dysferlin gene and immunohistochemical characterization of muscle biopsies were undertaken. We have compared the transcription profiles of patient samples and normal samples for the same type of muscles. The transcription pattern variability among patients was also assessed. We found evidence for incomplete differentiation of myofibers and upregulation of telethonin, FATZ/Myozenin/calsarcin, titin and alpha actin genes. There was also 11-times upregulation of MHC class 1. We also found a metabolic alteration in muscle mitochondria affecting AcylCoA-dehydrogenase and fatty acid binding protein, two genes involved in fatty acid metabolism. Finally, our expression data show that primary dysferlin deficiency induces a corresponding reduction of caveolin 3 and acetylcholinesterase, which are localized in the cell membrane and are involved in intracellular signaling.

P-08

Progress to Construct the RIKEN Mouse Full-Length cDNA Encyclopedia

P. Carninci, K. Shibata, M. Itoh, T. Arakawa, Y. Ishii, H. Konno, K. Sato, N. Hayatsu, T. Shiraki, T. Hirozane, K. Aizawa, H. Bono, S. Kondo, K. Waki, J. Kawai, A. Yoshiki, Y. Okazaki and Y. Hayashizaki, Genome Science Laboratory, RIKEN, Saitama, JAPAN and Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama, JAPAN

We have prepared long insert, full-length, subtracted/normalized cDNA libraries with cap-trapper from >250 tissues/stages using thermoactivated RT and cloning in lambda-FLC vectors, amenable to transfer in functional vectors. We have produced >1,445,000 3’ end ESTs, >252,000 5’ ESTs and fully sequenced and functionally annotated >21,000 fully sequenced cDNAs (FANTOM). Clustering gives >179,000 groups, which cover most genes. Remaining clusters are being fully sequenced. Discrepancy with current gene number prediction (35-40,000 only), derives from: (1) sequencing errors, (2) clustering strategies, (3) polymorphisms of transcription initiation/termination sites; but also (4) non-protein-coding RNAs and (5) genes not found by genome annotation. To select new, rare/very long cDNAs, we are making (1) lambda cDNA libraries subtraction, (2) size selection (~7Kb), (3) stabilization of long cDNAs, (4) improved plasmid preparation and (5) special interest libraries (preimplantation embryos, cancer, tissues of immunological and neurobiological interest). To produce functional proteins, we use cytoplasmic RNA to avoid residual unspliced introns. http://genome.gsc.riken.go.jp

P-09

Characterizing Transcriptional Determinants of the Innate Immune Response to Infection Through Gene Expression and Literature Profiling

Damien Chaussabel, Roshanak Semnani, Mary Ann McDowell, David L Sacks, Alan Sher, Thomas B Nutman, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD

New perspectives for the study of host-pathogen interactions have emerged from the recent development of microarray technologies that allow genome wide measures of gene expression. We used high-density oligonucleotide arrays to monitor changes in gene expression that occur following exposure of human Macrophages and dendritic cells to a wide diversity of infectious agents. An approach was developed for the extraction of meaningful knowledge from the large volume of information generated by this type of experiment in which the transcriptional events taking place during the early response to infection were characterized by sequentially mining the information contained in this expression dataset and the literature. Clustering analysis of expression data revealed complex profiles and allowed to identify groups of tightly co-regulated genes. An original literature profiling technique based on the clustering analysis of term occurrences in abstracts was then used to scan scientific publications for the existence of functional relationships among genes sharing similar expression profiles. This approach allowed us to identify groups of genes associated from both functional and transcriptional standpoint. Interestingly pathogen-specific responses were found to be constituted by unique combinations of these transcriptional determinants, with dramatic differences being observed between the two types of professional antigen presenting cells studied.

P-10

PGA-CPR (Chip Profiling Resource): A Web Queried Database of Affymetrix Profiles with Cross-Profile Analysis Tools

Josephine Chen, Dietrich Stephan, and Eric Hoffman, Research Center for Genetic Medicine, Children's National Medical Center, Washington, DC

The centralized public resource of genomic DNA sequence, and SNPs that have facilitated research world-wide, while also minimizing redundancy in data generation by different laboratories. Development of a similar centralized database of mRNA information (expression profiles) has been hampered by issues such as standardization of data generation, reporting, and storage. Here, we present standard operating procedures, quality control measures, and uniform data reporting enabling a centralized Oracle database of expression profiles (1,300 to date). We have built an SQL web interface to this data warehouse, allowing on-line queries of all profiles by any number of experimental variables (tissue, species, chip type, etc). The database permits access and downloads of all forms of data associated with any particular profile, including raw image files (.dat), processed image files (.cel), and interpretation files. We have also built a function in LIMS that simultaneously exports data to both the SQL web server, and the NCBI GEO database; the GEO files retain links back to the LIMS files. Finally, we have coded a gene-based cross-profile search function, that returns Avg Diff values for all profiles in the data warehouse, enabling rapid ascertainment of transcript levels in dozens of tissues and pathological conditions in mouse, rat, and human. PGA-CPR is available through http://microarray.cnmcresearch.org/pga.htm

P-11

Using Multivariate Regression Analysis to Detect Differential Co-Regulation of Transcriptions Between Subgroups of Medulloblastomas

Chun Cheng, Andrew R. Hallahan, James M.Olson, and Lue Ping Zhao, Divisions of Public Health Sciences and Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, WA

Transcription profiles have been utilized to study the co-regulation of genes under different experimental conditions or from different patients. Here we present a regression-based method that quantitatively measures the strength of association between the genes and evaluates its statistical significance. Furthermore, this method can help identify the genes co-regulated in different fashions between two groups of samples. Using this method, we studied the transcription profiles of 60 medulloblastoma samples published recently. We focused on the co-regulation of molecules that have been shown to correlate with prognosis. These genes include the neurotrophin-3 receptor (TrkC), NeuroD3 and N-Myc. Discovery of genes associated with these molecules may shed light on the pathways important in determining prognosis of medulloblastoma. Using the transcription level of these molecules as explanatory variable respectively, we performed linear regression analysis on the transcription levels of all other genes. We identified co-regulated genes at a desired significance level. Furthermore, to study how the co-regulation of this group of genes differs in tumors with different prognostic outcomes or histological types, we used an indicator variable to denote the identity of a sample and performed multivariate regression analysis. The findings of these studies will be presented here.

P-12

Definition of the Transcriptional and Translational Response to a Physiological Stimulus: A Cross-Species Time Series Muscle Study in Human Volunteers and Rats

Yi-Wen Chen[1], Eric P Hoffman[1], Gustavo Nader[2], Keith R. Baar[2], Karyn A. Esser[2], Monica J. Hubal[3], Paul D. Thompson[4], Priscilla M. Clarkson[3], [1]Children's National Medical Center, Washington, DC; [2]University of Illinois, Chicago, IL; [3]University of Massachusetts, Amherst, MA; [4]Hartford Hospital, Hartford, CT

Muscle is a highly adaptable tissue specifically responds to change in the amount and type of activity. Here, we compare transcriptional cascades resulting from a single bout of muscle activity, in both human volunteers and an experimental rat model system. Expression profiling was done on human or grouped rat time series muscle samples, using Human U95Av2 microarrays (~12,000 genes), and rat U34A (~8,000 genes/ESTs). Six human arrays from three volunteers, and 48 rat arrays were done. Human data were analyzed to detect consistent intra-patient expression changes. Both rat total and polyribosomal RNAs at two different time points were studied to investigate transcriptional and translational regulations. In the human study, 34 genes were differentially regulated in common among three subjects (exercised vs. control). Of these 34, 24 were on the rat arrays, and half (11) showed similar dysregulation in the rat model. Most genes were involved in cell growth regulation and stress response pathways. About 10% of the differentially regulated genes in the rat showed regulation that was primarily translational. Muscle is post-mitotic, and must show cell growth without activation of mitotic pathways in myofibers. The expression profiles show evidence of this balance in growth promoting and growth inhibiting gene regulation.

P-13

Transcription Profiling of Cell Death in Drosophila Melanogaster

S. Chittaranjan, E. Garland, D. Freeman, S. Jones, M. Marra, and S. Gorski, Genome Sequence Centre BC Cancer Agency, Vancouver, CANADA

Programmed Cell death - PCD - is a highly conserved and genetically controlled event that plays important roles in animal development, homeostasis and disease. We are employing a genomics approach, using EST and SAGE, to identify new PCD genes that are transciptionally regulated in the salivary glands - SGs - of Drosophila. We constructed a SG-specific cDNA library and SG-miniSAGE libraries from three consecutive developmental stages leading up to PCD.5461 high quality 3' ESTs from our SG cDNA library were used to perform BLAST analysis against Drosophila predicted genes and genomic sequence from the Drosophila Genome Project. The majority of the ESTs matched known or predicted genes in Drosophila including a number of ecdysone-induced genes and known PCD genes. We also identified a significant number of putative new genes not determined by gene prediction programs. SAGE tags derived from the SAGE libraries were matched to our Drosophila constructed transcripts database that incorporates our SG EST data; E.Garland et al. Comparison of the first 40,000 SAGE tags from each miniSAGE library indicated differential expression for some known PCD genes and for many uncharacterized genes. Expression profiling data and verification data using real time quantitative RT-PCR will be presented.

P-14

From Human Tissue Samples to Transcriptome Analysis : How to Obtain High Quality RNA

Virginie Copois[1], Frédéric Bibeau[2], Patrick Chalbos[2], Sandrine Imbeaud[3], Carmella Cappellano[4], Marc Ychou[2], Charles Auffray[3], Emmanuel Conseiller[4], Bernard Pau[1], Maguy Del Rio[1], [1]Institut de Biotechnologie et Pharmacologie, Montpellier, FRANCE; [2]CRLCC Val D'Aurelle, Centre Paul Lamarque, Montpellier, FRANCE; [3]Genexpress, FRANCE; [4]Aventis Pharma, Oncology Genomics, Vitry sur Seine, FRANCE

Chemotherapeutic drug resistance is a major clinical problem and a frequent cause of treatment failure. The aim of our work is to identify cellular targets responsible for the resistance to anti-cancer drugs used in colorectal cancer treatment. For the analysis of gene expression profiles of patients suffering from colorectal carcinoma by microarray technology, we had to obtain high quality RNA from the tumor tissue samples. Indeed, the quality of data from microarrays is strongly related to the quality of the extracted RNA. We analyzed the importance of several factors susceptible to influence the integrity of RNA within the tissues. Among those, ischemia appeared to be of critical importance. We recorded the time between surgical extirpation of human tissues and storage of 84 samples from colorectal adenocarcinoma or liver metastasis. Total RNA was extracted from those samples and controlled on RNA 6000 LabChips® demonstrating that RNA was damaged after 30 min of ischemia. Furthermore, we showed that other factors like necrosis might be involved in the degradation of RNA pointing out that macroscopic and microscopic controls of the lesion by a pathologist is essential. A strictly protocoled monitoring of samples allowed us to obtain high quality RNA for transcriptome exploration as confirmed by different analyses.

P-15

Pathogenic Pathways in Spinal Cord Injury Identified by Gene Expression Profiling

Simone Di Giovanni[1,2]; Susan Knoblach[2], Cinzia Brandoli[1], Alan Faden[2], Eric Hoffman[1], [1]Childrens National Medical Center, Center for Genetic Medicine, Washington, DC, [2]Georgetown University Medical Center Laboratory for the Study of Central Nervous System Injury Department of Neuroscience Washington, DC

Spinal cord injury is a major cause of disability, and it is known that much of the functional deficit results from delayed cellular consequences of injury repair mechanisms. To define the temporal series of gene expression changes following a spinal cord injury, rats were subjected to a controlled mild injury by weight drop. Rats were sacrificed at four time points (30 min, 4h, 24h and 7 days), with 4 to 6 individual rat spinal cords expression profiled at each time point (total 26 U34A profiles). Genes showing forty or more “present” calls in 26 profiles by Affymetrix analyses were retained for further analysis (data scrubbing), and p values and fold changes correlated, with temporal and functional clustering. Specific RNAs were verified by QMF-RT-PCR using infrared primers, and by immunocytochemistry. We found rapid induction of immediate early genes (30 min), followed by induction of DNA damage-inducible genes and cell cycle related genes. The profile of the cell cycle genes suggested favoring the G1 to S transition (4 and 24h). Inflammatory and oxidative stress genes were upregulated in all profiles. We conclude that DNA damage and cell cycle related genes seem to be important in both cellular apoptosis (glia and neurons) and in proliferation (glia cells) as shown by their localization in both degenerating and proliferating glia and damaged neurons.

P-16

Preparation of cDNA Hybridization Targets by Phi29 DNA Polymerase

Rohini Dhulipala, Mubasher Dar, R. Scott Duthie and John Nelson, Amersham Biosciences, Piscataway, NJ

We have investigated a method of cDNA amplification using phi29 DNA polymerase. The unique strand displacement property of this polymerase allows repeated replication of input target cDNA in an isothermal exponential amplification. Our novel method overcomes two issues we originally identified using this polymerase. Amplification of linear DNA, such as cDNA, is more efficient in the center portion of the molecules so ends are under represented in the amplified product. Additionally, larger linear fragments are preferentially amplified over smaller fragments. We learned that ligation to form relatively large concatemers prior to amplification eliminated these issues. Double stranded cDNA was prepared from either total RNA or mRNA . The resulting fragments were then ligated to form a mix of both concatemers and circles. This material was then used as input for random primed amplification using phi29 DNA polymerase. Following a subsequent labeling step of the amplified cDNA, gene expression analysis was performed by microarray hybridization. Data will be presented to show results were consistent with those obtained using standard techniques requiring much larger amounts of RNA. We are investigating this novel amplification methodology to achieve gene expression analysis using smaller samples including mRNA amounts expected from a single cell.

P-17

Production and Use of DNA Microarrays for Studies of Aging

Mark W. Eshoo and Krysta Felkey, Buck Institute for Age Research, Novato, CA

The Buck Institute for Age Research is a non-profit, independent, research institute located 20 miles north of San Francisco in Marin County California. Research at the Buck Institute is focused upon aging and diseases of aging such as Cancer, Parkinson's and Alzheimer's disease. To support researchers in the field of Aging the Buck Institute has set up a high throughput Genomics Core facility that offers researcher investigators both here at the Buck Institute and at other institutions an integrated gene expression resource center. The Genomics Core performs all the steps necessary for microarray analysis of gene expression. The Genomics Core also assists the researchers in the experimental design of their gene expression studies and provides all the resources necessary for the performance and of microarray gene expression studies from samples as small as a few hundred cells.

One of the key strengths of the Buck Institute's Genomics Core is our capacity to efficiently produce the tens of thousands of DNA samples required to produce our DNA microarrays using the 384 well microtiter plate. This high throughput format combined with our ability to robotically select individual genes from our collection of 85,000, unique and sequence verified cDNA clones enables the Genomics Core to quickly produce microarrays that can be custom designed to meet the individual researcher's needs.

The Genomics Core has developed a number of human, mouse, Drosophila and C. elegans DNA microarrays that are specifically designed for studies of aging and age related diseases. Data will be presented on microarray methods development and examples of data generated by the Genomics Core from studies of aging.

P-18

Intelligent Bioinformatics Systems for Functional Genome Analysis

Daniel Gerlich, Martin Granzow, Julian Mattes, Wolfgang Tvarusko, and Roland Eils, Dr. Wolfgang Tvaruskoi (Intelligent Bioinformatics Systems) German Cancer Research Institute, Heidelberg, GERMANY

The BioFuture-Group "Intelligent Bioinformatics Systems" works on the development of bioinformatics systems for the interpretation of complex data generated by analytic processes in genetic diagnostics and cell biology research. In the recent years an increasing number of high troughput screening systems have been developed in molecular biology. While in the past decade most such techniques were devoted to sequencing the genome of humans and other organisms, more recently such techniques emerged to relating genomic structure to function. One example is the DNA chip technology, which allows the screening of the expression level of all genes in a given organism under different conditions in one experiment. With this technology the study of genetic effects of certain drugs has become possible thus paving the way for optimised design of new drugs and therapies. High throughput screening techniques generate a huge amount of data, which is difficult or even impossible to analyse without computer assistance. The development of fully automated computerized systems for the analysis of complex data in molecular biology is at the core of the group established in 2000 at the German Cancer Research Center (DKFZ Heidelberg). The research areas of major interest include:- Knowledge based data mining of molecular biological and clinical data for deciphering pathomechanisms in genetic diseaeses- Analysis and modelling/simulation of complex processes in cellular biology- Improving molecular cytogenetic dignostics by fully automated and highly sensitive image analysis systems.

P-19

The aRNA Amplification Procedure in mRNA-Poor Samples Array Applications

Esther Graudens, Xavier Barlet, Virginie el Marhomy, Charles Auffray & Sandrine Imbeaud, Array s/IMAGE, Genexpress team, Villejuif Cedex, FRANCE

For several years, the Genexpress has investigated transcriptomes through high-throughput gene expression profiling developing cDNA array approaches. One important limitation is the requirement for large amounts of RNA for array hybridization, typically 10-20µg of total RNA. This is especially a problem when working with limited samples such as laser capture microdissection. Our team is focusing high-attention on the optimization of the array procedures in term of reduced sample consumption keeping data precision. The aRNA amplification is a procedure we used in mRNA-poor samples programs. It consists of reverse transcription with a T7 promoter-oligo(dT) and in vitro transcription of the resulting cDNA to generate antisense RNA copies of each mRNA in a sample. In our hand, a single reaction, starting with 1µg total RNA (around 10 ng mRNA) yields a minimum of 10µg aRNA (e.g. 1000x amplification). We show the aRNA synthesis being tightly dependent on the efficiency and fidelity of such complex enzymatic reactions. We identify the primer sequence itself as a critical factor and report the analysis of three different ones, testing the influence of either the dT size or T7 sequence. Characterization of the cDNA and aRNA has been done using UV measurement, Agilent LabChip and array hybridization.

P-20

New Cancer Genome Anatomy Project (CGAP) Internet Resource for the Research Community

Lynette H. Grouse, Carl F. Schaefer, Susan F. Greenhut, Kenneth H. Buetow and Robert L. Strausberg, National Cancer Institute, National Institutes of Health, Bethesda, MD

The Cancer Genome Anatomy Project is a collaborative project launched by the National Cancer Institute in 1997. This Project is an interdisciplinary initiative to build a comprehensive information infrastructure for profiling molecular changes in cells associated with cancer. Critical to deciphering the molecular profiles of cancer cells is the availability of both a robust molecular database and analysis tools. The goal of such a database is to facilitate the seamless integration of molecular and clinical data. The new CGAP interface (http://cgap.nci.nih.gov/) supports a variety of molecular analysis tools that allow flexible retrieval and analysis of data. Data is accessed through new entry points that provide analysis of the data from specific molecular or biological viewpoints, such as Genes, Chromosomes, Tissues, or Pathways. Data for both human and mouse tissues is included in the database. In summary, the new CGAP Internet resource provides a comprehensive database and informatics tools that will assist both basic and clinical investigators in defining and characterizing the molecular profiles of cancer cells, with anticipated applications in all areas of cancer research.

P-21

Comparative Analysis of TBX1 Genes and Their Promoters from Different Species

Ilya Ioshikhes, Jun Liao and Bernice Morrow, Department of Molecular Genetics, Albert Einstein College of Medicine, Bronx, NY

TBX1 gene plays a very important role in regulation of VCF syndrome and other diseases. In our work we present results of comparative bioinformatics sequence analysis of human TBX1 gene locus, in particular its promoter region, with syntenic mouse sequences. Conserved regions among the two species and their possible functional role are described. Results of sequence comparison of the human and mouse TBX1 genes to a closely related Drosophila T-box gene are also represented.

P-22

Comparative Bioinformatics Analysis of Rat and Human ALB-AFP Gene Locus

Ilya Ioshikhes[1] and Joseph Locker[2], Department of Molecular Genetics[1] and Pathology [2], Albert Einstein College of Medicine, Bronx, NY

Alpha-fetoprotein (AFP) is an important developmental marker, expressed in fetal liver and liver cancer, but silent in adult liver. AFP is part of a two gene locus with its sister protein Serum Albumin (Alb). We present comparative bioinformatics sequence analysis of an extended region comprising the Alb-AFP locus from different species. In particular, regions conserved between the rat (a 70 kb region proprietarily sequenced at the Albert Einstein College of Medicine) and human loci were determined and related to known functional elements, coding and noncoding. This demonstrated conserved promoter and enhancer regions and evolutionary divergence through insertion of repetitive DNA elements. Additional analysis predicted TF-binding sites and MARs, correlating these with known regulatory elements. These analyses were extended to more limited sequences available from the mouse genome.

P-23

Expression of Cytokines and Growth Factors in Aging and hTERT-transfected Human Fibroblasts

Yukari Kanzaki, Toshinori Ide, Hiroshima University School of Medicine, Hiroshima, JAPAN

Telomerase is the unique reverse transcriptase that synthesizes telomeric DNA onto chromosomal ends using its RNA component as a template. Telomerase expressed in the most cancer cells, but not somatic cells except germline cells and stem cells. We surveyed, using membrane array hybridization, the expression of some growth factors and cytokines in human fibroblasts transfected with human telomerase reverse transcriptase subunit (hTERT) cDNA. Twelve genes out of 268 were induced and 7 out of 268 were repressed in hTERT-transfected fibroblasts. Telomerase are specialized structure at the end of eukaryotic chromosomes and concerned with chromosomal stability. Telomere DNA shortens by 50-200 b.p. per cell division and telomere length has been proposed as a counting mechanism of cell division number that controls cellular senescence. We have examined the changes in gene expression in accordance with aging *in vitro*. Moreover we examined whether telomere elongation rescues expressional changes, which occur during aging.

P-24

Full-length cDNA Library Construction Using Lambda Recombination (GatewayTM Technology)

Larissa G. Karnaoukhova, Mark R. Smith, Chris E. Gruber, and Martin Gleeson, Research and Development, Invitrogen Corporation, Carlsbad, CA

The Invitrogen GatewayTM Cloning Technology is a versatile and efficient system for cloning and subcloning DNA. It is based on a lambda phage recombinase that can be applied in vitro to stimulate site specific (att site) recombination. One application of this technology is the construction of directional cDNA libraries as Entry Clones. Once in Entry Clone format, the library can be recombined with any of the Destination Vectors creating an expression ready library. To construct Entry Clones, the cDNA is synthesized so that attB sites flank it. The attB-adapted cDNA is then recombined with a Donor Vector (attP sites) to create the attL Entry Clone library. Libraries constructed using this technology show that such recombination cloning of cDNA produces a higher quality library exceeding the standards seen using traditional cDNA library construction methods. Unlike standard cDNA cloning protocols, this approach requires no restriction enzyme digestion of the cDNA to create unique ends suitable for directional cloning. The recombination cloning reaction produces on average 1.3 million cfu/ng cDNA with average insert size of 2.3 kb and clone size ranging from 0.6-12 kb. Recombination cloning combined with the GeneRacerTM technology has been applied to clone cap-selected cDNA and create full-length enriched cDNA libraries.

P-25

LNA (Locked Nucleic Acid) Microarrays - A New Platform for Gene Expression Profiling

Sakari Kauppinen[1], Peter Stein Nielsen[1], Helle Ohlsson[1], Dan Jeffares[2], Tobias Mourier[2] and Peter Arctander[2], [1]Department of Expression Microarrays, Exiqon, Vedbaek, DENMARK, [2]Department of Evolutionary Biology, University of Copenhagen, Copenhagen, DENMARK

LNAs (locked nucleic acids) constitute a novel class of bicyclic DNA analogues that have an exceptionally high affinity toward complementary DNA and RNA due to increased thermal stability of the heteroduplexes. We have developed a novel technology platform for gene expression profiling by combining the design of highly specific LNA/DNA mixmer oligonucleotides with photoactivated immobilization of the oligonucleotides onto polymer microarray slides via a 5'-anthraquinone moiety. We have chosen a subset of genes for the use in LNA microarrays to monitor heat-shock responses in the yeast *Saccharomyces cerevisiae*. In addition, we have analysed alternative splicing as well as heavy metal-induced toxicity responses in the nematode *Caenorhabditis elegans* using highly specific LNA oligoarrays. Technical aspects of the LNA microarray platform as well as data obtained with the yeast and *C. elegans* model systems will be presented.

P-26

Clinical Utility of Quantitative Real-Time PCR for Monitoring Leukemia Patients Using RNA Stabilizer In Blood Samples

Ryuji Kawaguchi[1], Hideyuki Kudoh[1], Yasuhumi Matsuda[1], Thoru Murayama[2] and Ralf Wyrich[3], [1]Genetics Research Institute, SRL, Inc., Tokyo, JAPAN, [2]Hyogo Medical Centre for Adults, Hyogo, JAPAN, [3]QIAGEN GmbH, Hilden, GERMANY

Chimeric mRNA detection is used for minimal residual disease (MRD) monitoring in patients with leukemia. However, degradation of mRNA in patient samples may impede accurate data analysis. We have evaluated a new system for sample collection, RNA stabilization and purification (PAXgene[tm] Blood RNA System, PreAnalytiX GmbH). For this evaluation we used real-time PCR methods, that were optimised to determine BCR-ABL (Major bcr and minor bcr) in CML patients with t(9;22), AML1-MTG8 in AML patients with t(8;21) and PML-RARA in APL patients with t(15;17). Real-time quantitative PCR is done using the LightCycler[tm] (Roche) with hybridisation probes for BCR-ABL, AML1-MTG8, PML-RARA and G6PDH. Each quantification assay is optimised to achieve sensitivity between 10 and 100 copies/[mu]g of total RNA, with dynamic ranges between 1 to 10^6 dilutions of a positive cell line in a negative cell line. The mRNA quantifications were carried out using blood from normal donors that was spiked with cells from cell lines carrying the genetic markers described above. Expression levels of the house keeping genes, G6PDH, were used as controls. Blood samples from identical donors were collected and stored in PAXgene Blood RNA tubes and EDTA tubes for extended periods of time (day0 - day21). Our data shows significantly higher mRNA detection levels in PAXgene-stabilized samples as compared EDTA blood samples.

P-27

In Silico Detection and Characterisation of Cancer-Specific Transcript Isoforms

Janet Kelso, The JambORESTES Consortium, Winston Hidel, South African National Bioinformatics Institute, University of the Western Cape, Bellville, SOUTH AFRICA

Large-scale mining of expressed sequence tags (ESTs) and comparison with available human genome sequence has allowed us to detect variations in the exon composition of the mature transcripts of genes commonly associated with cancer. The occurrence of exon skipping, the most common form of alternative splicing, has been linked to various disease phenotypes including cancer. Disease-specific transcript isoforms may prove to be useful diagnostic markers or therapeutic targets. We have developed and implemented a controlled vocabulary, which partitions expression information extracted from cDNA library annotation into four categories: anatomical site, cell type, developmental stage and pathological state, to determine the specificity of the expression state of both skip and constitutive transcript isoforms. We have detected and characterised 323 exon skipping events in 241 genes, some of which are uniquely associated with cancer.

P-28

Protein Interaction Verification and Functional Annotation by Integrated Analysis of Genome-Scale Data

Patrick P.C.W. Kemmeren[1], Nynke L. van Berkum[1], Jaak Vilo[2], Theo Bijma[1], Rogier Donders[3], Alvis Brazma[2] & Frank C.P. Holstege[1], [1]Genomics Laboratory, Department of Physiological Chemistry and Department of Medical Genetics, University Medical Center Utrecht, THE NETHERLANDS, [2]Microarray Informatics group, EMBL-EBI, Hinxton, ENGLAND, [3]Centre for Biostatistics, University Utrecht, THE NETHERLANDS

Whole genome sequences and assays capable of determining the properties of thousands of genes in parallel present challenges with regard to accurate data-processing and functional annotation. Here, large collections of DNA microarray expression data are used to verify high-throughput protein interactions and provide functional annotation. This is done by determining which interacting pairs also exhibit mRNA coexpression. Several strategies for determining coexpression are assessed and the results indicate that more than half of all the genuine interactions contained within the high-throughput data sets can be verified in this way. Confidence in 973 out of 5342 recently reported putative *S. cerevisiae* interactions is increased. Tenative function is assigned to over 300 previously uncharacterized genes. The robustness of these approaches are demonstrated by experiments that test the *in silico* predictions made. The study shows how the utility of different types of data can be improved, and how well their combination can contribute to functional annotation.

P-29

Application of Statistical Modeling to Microarray Gene Expression Data in Human Lung Cancer

Najma Khalid, Lue Ping Zhao, Fred Hutchinson Cancer Research Center Public Health Sciences, Seattle, WA

Microarray technology allows investigators to genotype thousands of transcripts in experimental studies. Cluster analysis techniques, typically used to classify expression profiles, cannot utilize clinical information. We applied a more robust statistical approach to the analysis of expression data collected in a study to subclassify lung adenocarcinomas. (Bhattacharjee et al, PNAS 2001). Expression data on 12,600 genes were available for 17 normal lung samples and 139 adenocarcinoma tumors (125 samples associated with clinical data). Using an estimating equation technique that adjusts for heterogeneity across microarray chips, and for multiple comparisons, we compared normal samples with tumor samples and identified nearly 700 genes that were differentially expressed between the two groups ($p < 0.001$). We had similar results by including clinical variables (tumor stage, size, vital status recurrence, survival, age at resection, smoking history and sex), as covariates in the model. Finally for adenocarcinomas only, we compared groups by vital status and by recurrence. No differentially expressed genes were identified with either outcome. Further adjustment for age at resection and survival also failed to identify any differentially expressed genes. We conclude that microarray data of samples taken at resection yield diagnostic markers but do not predict prognosis.

P-30

The Immediate Early Gene Response of Human Keratinocytes to Gamma-Rays Analyzed by DNA Micro-Arrays

Jérôme Lamartine, Gilles Waksman, Xavier Gidrol, Michèle Martin, CEA, Service de Génomique Fonctionnelle, Evry, FRANCE

Epidermis keratinocyte is the first target cell of external exposures to genotoxic stress. Thus radiotherapy can induce early inflammation and desquamation of the epidermis, as well as keratosis or carcinoma as late effects. However, the response of keratinocytes to ionizing radiation (IR) is poorly characterised. We studied the early gene response of differentiated human keratinocytes to gamma-rays using cDNA microarrays. After 10 days in culture, HaCaT cells were irradiated with doses of 0.5, 2 and 15 Gy and total RNA was isolated at 3 and 24 hours. We used glass slides where 7600 human cDNAs were spotted. We found that 3800 genes were expressed in keratinocytes (50%). The number of genes that were at least 2-fold differentially expressed in irradiated cells versus control cells was dose dependent, varying at 3 hours from 4% of the genes after 0.5 Gy to 10 % after 15 Gy. The genes modulated by a dose of 0.5 Gy were not significantly different from those modulated by higher doses. Most of the response was lost at 24 hours. The present results revealed many potential regulators of IR response, such as transcription factors, and uncovered a large number of new markers of exposure among differentiation and metabolism genes.

P-31

Specific Subtraction of Abundant mRNAs in Skeletal Muscle

Paolo Laveder, Cristiano De Pittà, Stefano Toppo, Giorgio Valle and Gerolamo Lanfranchi, CRIBI Biotechnology Center, University of Padova, ITALY

Through systematic sequencing of cDNA clones restricted to the 3'-end of each transcript (more than 35,000 clones analyzed), our group has built a catalogue of the most expressed genes in human skeletal muscle (http://muscle.cribi.unipd.it). A few exceptionally abundant (super-prevalent) mRNAs encoding myofibrillar and mitochondrial proteins accounted for about 40% of the total clones. Based on these data we have developed a new two-step strategy for producing subtracted cDNA libraries. The first subtractive step is limited only to the super-prevalent mRNAs and involves a novel use of oligonucleotide-directed RNase H digestion: DNA-oligonucleotides are selected informatically to digest the 3'UTRs of the target transcripts; mRNAs are obtained by in vitro transcription from 3'-end specific cDNA libraries. In the second step the hundred most abundant mRNA species are removed through subtractive hybridization. RNA drivers produced from 3'-end cDNA clones of our collection assured the maximum specificity of the hybridization reactions. Removal of mitochondrial transcripts required particular care, since they often are prematurely terminated. We showed that this technology allows a highly specific subtraction and that the resulting cDNA libraries are effectively enriched for genes expressed at low levels. In general, our subtractive approach would be appropriate for connective and epithelial tissues.

P-32

Systematic Screening of Tissue Specific Cis-regulatory Elements in *Ciona intestinalis* Embryo

Byung-in Lee[1], David Keys[2], Mei Wang[1], Orsalem J. Kahsai[1], Sylvia Ahn[1], Qing Zhang[1], David K. Engle[1], Irma Rapier[1], Jason Olivas[1], Chris J. Detter[1], Sharon Doyle[1], Naoe Harafuji[1], Anna Di Gregorio[1], Trevor Hawkins[1], Dan S. Rokhsar[1], Michael Levine[1,2], Paul M. Richardson[1], [1]DOE Joint Genome Institute, Walnut Creek, CA, [2]Department of Molecular and Cellular Biology, University of California at Berkeley, Berkeley, CA

Regulatory DNA elements such as promoters and enhancers work by serving as docking sites for specific protein complexes. These complexes are comprised of cooperative groups of transcription factor proteins that recognize the target DNA sequences quite specifically and their presence or absence governs the off or on status of the target. Therefore understanding DNA regulatory elements is key to understanding the composition and function of the biochemical networks and pathways that carry out the essential processes of living organisms. To characterize gene regulatory networks, we used electroporation assays to screen genomic DNA fragments for tissue specific regulatory activities in *Ciona intestinalis*. The *Ciona* genome is one of the smallest and most compact of all chordate genomes and the Ciona tadpole represents the most simplified chordate body plan. Since, exogenous DNA can be introduced into the synchronously developing embryos via simple electroporation, we used this method with a vector containing the *lacZ* reporter to determine cis regulatory DNA modules that lead to the specification of key chordate tissues. We screened ~300kb of *Ciona* genomic DNA containing *Hox* genes for tissue specific enhancer elements using the shotgun approach, and found 30 clones (80kb) that were positive for *cis*-regulatory activity. Among 30 positives, 20 unique putative enhancer elements identified which indicate about one putative enhancer elements every 18kb with expression patterns in epidermis, tail muscles, cerebral vesicle, notochord, neural tube. Among 20 enhancers, 8 expression patterns may be associated with Hox genes. These are being confirmed by whole mount *in-situ* hybridization.

P-33

Molecular Genetic Profiling of Gleason Grade 4/5 Prostate Cancers Versus Benign Prostatic Hyperplasia

Mamatha Mahadevappa[1], Zhaomei Zhang[1], Mitchell C. Caldwell[2], Zuxiong Chen[2], Zhenbin Fan[2], John E. McNeal[2], Rosalie Nolley[2], Thomas A. Stamey[2], Janet A. Warrington[1], [1]Health Management Research, Affymetrix Inc., Santa Clara, CA, [2]Department of Urology, Stanford University School of Medicine, Stanford, CA

Prostate cancer (PC) is one of the most common causes of cancer death among American men. Serum prostate-specific antigen (PSA) between 2-10 ng/ml has been widely used in the U.S. as a marker for PC. Serum PSA is largely related to benign prostatic hyperplasia (BPH), but correlates poorly with PC curative outcome at PSA (2-10 ng/ml). Gene expression characterization of grade 4/5 cancers could aid in the development of new PC serum markers. We compared the results of large-scale gene expression monitoring from 9 Gleason grade 4/5 cancers to 8 BPH. Hierarchical clustering separated two groups. 86 candidates were identified with $p<0.0005$. The candidates included genes previously associated with PC, e.g. hepsin, prostate-specific membrane antigen and prostate differentiation factor, as well as genes associated with oncogenesis, tumor suppression, transcription, signal transduction and apoptosis. A number of new candidates are presented as possibilities for further study.

P-34

A Novel Approach to Eliminate Vector Background and Increase Sequencing Efficiency of cDNAs

Uday Matrubutham[1], Jody Mirchandani[1], Jian Liu[1], Martin A. Gleeson[1], K. MacDonald[2], J. Asano[2], Y. Butterfield[2], N. Girn[2], S. Lee[2], T. Olson[2], P. Pandoh[2], U. Skalska[2], D. Smailus[2], L. Spence[2], Jeff Stott[2], G. Yang[2], J. Schein[2] and M. Marra[2], [1]Invitrogen Corporation, Carlsbad, CA, [2]Genome Sequence Centre, BC Cancer Agency, Vancouver, CANADA

We, at Invitrogen Corporation, have developed a novel and highly efficient transposon-mediated method for generating cDNA sequences with low quantities of vector background. The cDNA sequencing templates are generated in two consecutive in vitro reactions utilizing Mu transposition (GeneJumperTM) and att/clonase recombination (GatewayTM, Invitrogen Corporation). First, the GeneJumperTM transposon is randomly inserted into the cDNA clone. Second, the cDNA insert is transferred to a compatible GatewayTM vector by recombination. Transposon-containing cDNA recombinants are identified using antibiotic selection and sequenced bi-directionally with transposon-specific primers. This process has been applied to cDNAs cloned into pCMVSport and GatewayTM clones that have att recombination sites flanking the cloned insert DNA. This methodology was successfully evaluated for high-throughput process adaptation as part of the Mammalian Gene Collection initiative. The approach reduces the number of reads required to complete cDNA sequences through reduction of reads initiated from within the cDNA vector. In addition, we have adapted the approach to the pooled clone strategy currently in use at the BC Cancer Agency Genome Sequence Centre. We anticipate that the increased efficiency provided by the method will reduce dramatically the cost of sequencing cDNA clones or other similarly sized clones.

P-35

Identification of Putative Genes Involved in Cell Transformation

Andrei M. Mikheev, Svetlana A. Mikheeva, Vivian Li, Jeffrey Delrow and Helmut Zarbl, Fred Hutchinson Cancer Research Center, Seattle, WA

We previously described the isolation of two independent, non-tumorigenic variants of the HeLa cervical carcinoma cell line following exposure to the mutagen ethylmetansulfonate. Unlike the parental HeLa cells, the HA and HF revertants were unable to form colonies in semi-solid medium and failed to induce subcutaneous tumors upon injection into nude mice. The levels of E6 and E7 DNA, RNA and protein in the revertants were unaltered compared to HeLa cells. The levels of p53 were increased in HA and HF cells due to protein stabilisation. Using cDNA microarray hybridization, we identified genes differentially expressed between the revertants and Hela cells. Among identified genes, we found insulin growth factor binding receptor 3 (IGFBP-3) and Dickkopf-1 (DKK-1) upregulated in both revertants. Ectopic expression of either gene in Hela cells inhibited soft agar growth. Although human papillomavirus ectopic over-expression of E6/E7oncogene decreased IGFBP-3 protein levels in the revertants, IGFBP-3 and DKK -1 mRNA levels were unresponsive to E6/E7 expression. These results suggested that p53 independent activation of the DKK-1 and IGFBP-3 genes contributed to phenotypic reversion of Hela cell transformation. The HeLa revertant model is useful for identification of putative suppressors of cell transformation.

P-36

Quality Control of Microarray Samples Using Lab-on-a-Chip Technology

Odilo Müller[1], Monika Dittmann[1], Samar Lightfoot[2], Kathi Williams[3], and Robert Dubrow[3], [1]Agilent Technologies, Waldbronn, GERMANY, [2]Agilent Technologies, Palo Alto, CA, [3]Caliper Technologies Corp. Mountain View, CA

Microarray experiments allow the fast and efficient assessment of the expression of thousands of genes in one single experiment. The quality of the array data is largely dependent on the quality of the array itself, as well as on the quality of the labeled probes that are hybridized to the array. The Agilent 2100 bioanalyzer – the first commercially available Lab-on-a-Chip system – is the tool of choice for the quality assessment of RNA samples at different steps within the microarray workflow. A set of new RNA assays has been developed to better suit the needs of microarray users. Sample alignment using an internal marker allows better comparison of different samples, also providing rough size estimation. Quantitative measurement of RNA samples allows in many cases to forgo UV measurements. For the analysis of RNA samples from micro-dissected tissue, a new assay is being developed, which allows the detection of RNA down to 100 pg of sample.

P-37

Prepare a Set of Human cDNA Clones Encoding Large Proteins in an Expression-Ready Form

Takahiro Nagase, Noriko Okazaki, Daisuke Nakajima, Hisashi Yamakawa, Reiko Kikuno, and Osamu Ohara, Kazusa DNA Research Institute/Department of Human Gene Research, Kisarazu, Chiba, JAPAN

We have focused our sequencing efforts on long cDNAs (>4 kb) encoding relatively large proteins (>500 a.a.) from brain. The total number of cDNAs entirely sequenced (KIAA plus some of FLJ cDNAs) has exceeded 2000. Besides these cDNA clones, we have also accumulated more than 1500 cDNA clones coding for large proteins first identified by other groups. However, the coding sequences of some cloned cDNAs were predicted to be truncated due to incompleteness of the cDNAs or spurious interruption by intron(s) and/or reverse transcription error(s). Although we have continuously updated our cDNA sequences through experimentally revising these spurious coding interruption as far as possible, our cDNA clones cannot always produce the authentic proteins because these revisions have been made only in sequence information. Thus, preparation of a set of expression-ready cDNA clones encoding large proteins becomes our urgent mission because this must play a critical role for comprehensive functional analysis of human genes. We have conducted manual curation of our cDNA clones one by one to yield cDNA clones which can produce authentic KIAA proteins by combination of currently available methods. Because large cDNAs are generally difficult targets to obtain in an expression-ready form, our efforts will provide an invaluable human gene resource to the research community.

P-38

Encoding Biological Prior Information Into Gene Expression Analysis

Michael F. Ochs, Frank J. Manion, Ghislain Bidaut, Thomas Moloshok, Jeffrey D. Grant Fox Chase Cancer Center, Philadelphia, PA

The fundamental goal of gene expression data analysis is to recover biological information in the form of coexpression groups and signaling or metabolic pathway information. However, present methods assign each gene to only one coexpression group, which limits their ability to recover fundamental biological behavior, since genes often lie in multiple coexpression groups, which are activated in response to different external stimuli. In addition, published algorithms are generally based on standard statistical tools and data mining methods and do not encode any biological information in their structure. We present a new system designed to overcome these limitations. The algorithm is based on our previous work in identifying spectral signatures from mixtures through the incorporation of prior knowledge. We show preliminary results from the analysis of public domain data, which demonstrate the ability of the system to identify genes that are multiply expressed in response to different stimuli. In addition we discuss the importance of inclusion of biological information in the form of the underlying model encoded within the analysis software. Finally, future additions to the algorithm, which should increase its power to recover biologically significant features in the data will be discussed.

P-39

Whole Genome Functional Analysis and Metabolic Pathway of *Zymomonas mobilis*

Sang-Ho Park[1], Ki-Nam Heo[1], Hyun-Seok Parkv, Jeong-Sun Seo[1, 2], [1]Macrogen Co., Seoul, South Korea, [2]Seoul National University, Seoul, SOUTH KOREA

Recent advances in genomics, transcriptomics, metabolomics and bioinformatics have led to a growing interest in using whole genome sequence data to reconstruct precise metabolic pathways in microbials for industry and pharmaceutics. Zymomonas mobilis was chosen as a model organism for the study of microbials, and the entire genome has been sequenced using a random shotgun approach. This bacterium is a facultative anaerobic Gram-negative rod with a high level of ethanol fermentation through the Entner-Doudoroff pathway. We have completed whole genome assembly and initial prediction of protein coding regions and functions by utilizing several different tactics. To confirm gene expression and functions, we are developing an integrated functional genomics tool using the initial annotation, DNA Microarray data, protein profiles, and metabolite profiles. We are also developing algorithms for promoter prediction in the whole genome. Our goal is to develop effective tools for metabolic pathway mapping and its application in microbial chemical production.

P-40

Gene Expression Profiling on Radiation Responses

Woong-Yang Park, Chang-Il Hwang, Chang-Nim Im, Min-Ji Kang, Jang-Hee Woo, Ju-Hoon Kim, Kyoung-A Kim[1], Hyoung-Jin Yoo[1], Yun-Sil Lee[2], Su-jae Lee[2], Jeong-Sun Seo, Department of Biochemistry and Molecular Biology and Ilchun Molecular Medicine Institute, Seoul National University College of Medicine, [1]Macrogen Inc., [2]Korea Cancer Center Hospital, Seoul, KOREA

Frequent mutations in tumor cells prevent the expression of p53, which is the key molecule to determine the responses to ionizing radiation. Gene expression profiling in p53-deficient cells will unveil the p53-independent mechanisms of checkpoint control, repair and cell death following gamma-irradiation. Jurkat T cells with nonsense mutation on p53 and PTEN were exposed to 4Gy to 16Gy gamma-irradiation. Using 2,400 human cDNAs, the gamma-irradiation-responsive genes were examined by microarray. We could find the induction of phospholipases after the gamma-irradiation. Other radical scavenger was also induced, which might explain the interpersonal difference in radio-resistance. However, we could not find any significant changes in cell cycle-related genes. Using 384 genes related to gamma-irradiation, different tumor cell lines and peripheral blood mononuclear cell (PBMC) were found to be similar in expression pattern. From these results, we could list the novel radiation-related genes and also suggest the possibility of usage of these biomarkers in DNA chip for checking radiation exposure.

P-41

Expression Profiling and TPE Analysis Generates Candidate Tumor Markers for Papillary Thyroid Carcinoma

Erwin Pauws[1], Truus JM Veenboer[1], Antoine HC van Kampen[2], Jan JM de Vijlder[1], Carrie Ris-Stalpers[1], [1]Pediatric Endocrinology, Amsterdam, THE NETHERS-LANDS, [2]Bioinformatics Laboratory, Academic Medical Center, Amsterdam, THE NETHERLANDS

Using Serial Expression of Gene Expression (SAGE) the expression profiles of normal thyroid and papillary thyroid carcinoma are compared. Each profile consists of 10,000 tags corresponding to 6,000 unique transcripts. Substraction generates 22 upregulated transcripts and 121 downregulated transcripts ($p<0.05$). An algorithm is developed to calculate a Tissue Preferential Expression (TPE) value of these differentially expressed transcripts, comparing the expression of each tag to the expression of that tag in a cohort of SAGE libraries from normal and neoplastic tissues publicly available at NCBI/CGAP. A high TPE value (range 0-100) indicates specific expression in the primary library compared to the cohort. Ordering the differentially expressed transcripts according tot their TPE value shows that transcripts like XPC and ECM1 are specifically upregulated in thyroid carcinoma while transcripts like VEGF and CD9 are more generally expressed in tumor tissues. Confirmed downregulated thyroid specific transcripts like TG and TPO show high TPE values as expected. The role of the resulting shortlist of transcripts in the initiation and/or progression of thyroid cancer is now under investigation. SAGE analysis and subsequent determination of TPE value fascilitates the rapid distinction of genes specifically expressed in cancer tissues in general, or in a specific tumor tissue.

P-42

From Transcriptome to Proteome: Genome Scale Collection and Expression of Low Abundant Human Full-Length Transcripts

Yijun Ruan, Chia-Lin Wei, Stephan Dillard-Telm, Erika Peterson, Gina Truesdell, Helen Chan, Gary Wolfe, and Dave Mannion, Large Scale Biology Corporation, Vacaville, CA

A fundamental aspect of understanding human genome is to clone, sequence, and analyze all of full-length transcripts including their alternative splicing variants. A complete collection of full-length cDNA clones and their sequences will not only refine the global view of a genome and detail the structures of genes, but also fuel the whole range of proteomics through high throughput recombinant protein expressions. Recently we developed an efficient process to select low abundant cDNA clones based on their abundance level in transcriptome for sequencing analyses. With this approach we have achieved the highest discovery efficiency of full-length transcript cloning and sequencing. So far we cloned and characterized tens of thousands unique human full-length transcripts from hundreds of human tissues, cell lines, and biopsy samples in expression ready vectors, and identified large number of novel transcripts that have not been described, as well as many new alternative slicing variants of known genes. We are in process of generating thousands of human recombinant proteins using GENEWARE® (a plant viral expression system). These high quality collections of low abundant human full-length genes and recombinant proteins will be invaluable resources for gene function and proteomics studies. Detailed methodology and full-length transcripts content will be discussed.

P-43

The Use of Representative Gene Fragments (Rgss) for Less Populated, Highly Replicated Microarrays

Helena B. Samaia, André C. Zaiats, Otávia O. Caballero, Ricardo R Brentani, Sandro J. de Souza and Andrew J.G. Simpson, Ludwig Institute of Cancer Research, São Paulo, BRAZIL

The possibility of extracting the expression profile of thousand of genes in a single experiment makes the use of microarray technology very promissing. The extraction of reliable information from the large data sets generated is critical. Processes including probe selection, spotting, hybridization, image processing and normalization are the major causes of data variability and have been estensively discussed in the literature. However, the potential for hybridization cross-reactivity, the length of the immobilized fragments and the use of replicate spots have been less often reported in these studies. We developed an array using transcript fragments that have been carefully designed so as not to exhibit cross reactivity and that have similar lengths, GC content and position within the transcript (RGSs). They were produced by RT-PCR using primers designed by an in-house computer algorithm. Our current aproach to the statistical analysis of our data also involves working with the variation between replicate spots on the array. We show evidence that hybridization intensity is dependent on the length and the position within the transcript from wich the RGS is derived, that the use of RGS avoids cross-hybridization and that the use of multiple spotting is essential for a reliable interpretation of the data.

P-44

Differential Gene Expression in Leishmania Major Friedlin

A. Saxena, S.Yan, L. Worthey, A. Leland, PJ Myler, K. Stuart, Seattle Biomedical Research Institute, WA

Leishmania, a flagellated protozoan is responsible for a wide spectrum of human diseases. The sequencing of Leishmania genome is currently underway. A large number of genes have been identified and a small proportion of these have been ascribed functions. However, it is essential to carry out functional studies in order to utilize the mass of data generated by the sequencing project. Using DNA microarrays it is possible to study the expression of thousands of genes at the same time. Microarrays, containing PCR amplified DNA from a random amplified genomic library of L. major Friedlin (Ref: Akopyants et al, MBP 113 (2001)), were hybridized with fluorescent probes made from procyclic and metacyclic RNA. The fluorescent signal showed an excellent level of signal to noise ratio. The data was normalised for background and probe intensity. The relative abundance of RNA for each spot was calculated. We observed statistically significant increase in signal intensity (1 to 5-folds, $p<0.01$) in 5% of DNAs with the metacyclics probe. Meanwhile in procyclics, 1.5 % of DNAs showed 1 to 3 folds increase in signal intensity. We also carried out flip-dye experiments to take into account the inherent ability of some species of RNA to bind to a particular dye. Northern blots were used to corroborate our results. We have identified several of genes up-regulated in both procyclics and metacyclics.

P-45

Identification of Genes in Biosynthesis of Fatty Acids and Cholesterol Pathways Using Microarrays

Nila A. Shah[1], Young-Ah Moon[2], Suchismita Mohapatra[2], Jay D. Horton[2], and Janet A. Warrington[1], [1]Health Management Research, Affymetrix Inc., Santa Clara, CA, [2]Dept of Molecular Genetics, Internal Medicine and Biochemistry, University of Texas South Western Medical Center, Dallas, TX

Sterol regulatory binding proteins (SREBP) are a family of basic helix-loop-helix leucine zipper transcription factors that regulate biosynthesis of fatty acids and cholesterol. To identify candidate genes in the fatty acid synthesis pathway, transgenic mice over expressing SREBP-1a and SREBP-2 isofroms in liver were used. Labeled cRNA was prepared from liver and hybridized to mouse expression microarrays interrogating ~30,000 genes/ESTs. Some of the known genes, ACC, FAS, and ATP citrate lyase regulated by over expression of SREBP were identified by microarray analysis. A novel gene, long chain fatty acyl elongase (LCE) regulated by the SREBP1a and 2 was also identified. LCE was over expressed with a fold change of 25 and 9 in SREBP-1a and SREBP-2 mice respectively. The confirmation of LCE was done by northern blot analysis, sequence homology searches and biochemical tests. This enzyme specifically catalyzes the elongation of 12, 14 and 16 carbons but does not elongate the C-18 fatty acids or very long chain fatty acids.

P-46

Bioinformatics for High-Throughput Mu Transposon Sequencing

Ursula Skalska, Yaron Butterfield, Ran Guin, Martin Krzywinski, Kim MacDonald, Duane Smailus, Jeff Stott, George Yang, Scott Zuyderduyn, Jacqueline Schein, Steven Jones and Marco Marra, Genome Sequence Centre BC Cancer Agency, Vancouver, BC, CANADA

As participants in the Mammalian Gene Collection full-length cDNA sequencing initiative, we have developed an efficient, high-throughput method for accurate sequencing of entire cDNA clone inserts. Sequencing is accomplished through the insertion of Mu transposon into cDNAs, followed by sequencing reactions primed with Mu-specific sequencing primers. Our approach uses a clone pooling strategy that eliminates the need for a transposon insertion library to be constructed for each clone. This process includes a number of key bioinformatic stages. Prior to pooling, analysis of Expressed Sequence Tags confirms clone identity. Accurate clone insert size and DNA quantitation data are used to ensure proportional representation of each cDNA clone in the pool. Sequences are assembled using Phred, Phrap, and Consed. Clones remaining incomplete are either re-pooled or finished using directed primer reads. Finished sequences are inspected for frameshifts and chimeric clones using BLAST and BLAT searches. Additionally, we have developed a number of programs to complement these technologies and to expedite sequencing of clones. Using these techniques, we have to date generated over 7.3 Mbp of accurate sequence from 3,891 cDNA clones. A detailed description of our strategy and the associated bioinformatics tools will be presented.

P-47

A Robust DNA Based Amplification Technique for Tissue-Limiting Experiments on cDNA Microarrays

L. Smith[1], P. Underhill[1], C. Pritchard[1], Z. Tymowska-Lalanne[1], S. Abdul-Husein[1], H. Hilton[1], L. Winchester[1], D. Williams[2], T. Freeman[2], and A. Greenfield[1], [1]Medical Research Council, Mammalian Genetics Unit, Harwell, Oxfordshire, UNITED KINGDOM, [2]Medical research Council, UK Human Genome Mapping Project-Resource Centre, Hinxton, Cambridge, UNITED KINGDOM

AbstractThe potential of cDNA microarray analyses to address complex problems in a wide variety of fields is steadily being realised. The limiting factor in such analysis is often the amount of RNA required usually microgram quantities. To address this problem, researchers have turned to methods of improving detection sensitivity, either through increasing fluorescent signal output per molecule or increasing the amount of RNA through amplification. The key features of any amplification technique are that it is 1, reproducible, 2, reliable, 3, unbiased, 4, easy to use, and 5, cost effective. In vitro transcription of cDNA template has to date been the method of choice, however, this method has potential problems associated with RNA instability and degradation. Here we present a novel alternative method of amplification that increases the stability of target by using a DNA rather than RNA amplification procedure. The reproducibility, bias, and sensitivity of this technique has been examined using a 6000 EST microarray set and gives comparable results with as little as 30 ng total-RNA starting material.

P-48

Poster PDF file

Affinity-Based Immobilization Tools for Functional Genomics

Amy L. Springer, Anna S. Gall, Karin A. Hughes, Robert J. Kaiser, Guisheng Li, Deborah D. Lucas, Kevin P. Lund, Doug A. Spicer, Jean P. Wiley, Prolinx, Inc., Bothell, WA

Studies of gene expression and subsequent interactions of gene products on a systems scale require methods that are robust, simple to implement and yield data that can be analyzed efficiently. Such methods often require immobilization of biomolecules for capture and detection of analytes, impurities, secondary products or metabolites. Immobilization methods such as direct conjugation to surfaces (e.g., glutaraldehyde coupling) or biologically-based affinity systems (e.g., (strept) avidin/biotin), can be limited by poor reproducibility, low surface capacities, essential purification steps and significant non-specific binding. Prolinx®, Inc. has developed a small molecule affinity system suitable for immobilization of nucleic acids and proteins on a variety of surfaces. This technology is based on the reversible complexation of phenyl(di)boronic acid (P(D)BA) with salicylhydroxamic acid (SHA). Surfaces can be reproducibly modified with SHA resulting in high capacities for P(D)BA-modified biomolecules and excellent assay sensitivities. P(D)BA-modification is performed in solution, independent of immobilization, and PDBA-conjugates can be directly immobilized on an SHA-modified solid support without purification; any excess reagent is removed by washing. As a result, multiple conjugations may be performed in an automated format suitable for high-throughput applications such as protein microarrays. The advantages demonstrated using this system make P(D)BA-SHA technology a convenient platform for systems-scale research.

P-49

Full-Length cDNA and SAGE Sequencing at the British Columbia Cancer Agency Genome Sequence Centre

Jeff M. Stott, J. Asano, Y. Butterfield, R. Guin, M. Krzywinski, K. MacDonald, S. Lee, T. Olson, P. Pandoh, U. Skalska, D. Smailus, L. Spence, K. Teague, G. Yang, S. Zuyderduyn, J. Schein, S. Jones and M. Marra, Genome Sequence Centre BC Cancer Agency, Vancouver, CANADA

A large component of the DNA sequencing activities at the BCCA Genome Sequence Centre is directed towards the study of expressed genes. cDNA sequencing activities are primarily focused on the NCI-funded, multi-centre Mammalian Gene Collection initiative identifying and sequencing full-length cDNA clones for human and mouse genes. Sequencing of cDNA clones is accomplished through a combination of 5' and 3' EST generation, transposon-mediated sequencing and directed primer reads. To date we have generated more than 7.3 Mb of accurate sequence from 3,891 candidate full length cDNAs. Automated bioinformatics tools have been developed to coordinate library construction and sequencing and streamline sequence assembly and analysis.We are also involved in a number of projects employing the SAGE method for gene expression surveys in the study of various cancers, programmed cell death and ageing. SAGE sequences are generated by single-pass vector-primed reads. To date we have generated over 150,000 sequencing reads from 62 libraries. Among these are 26 brain libraries analyzed in collaboration with Greg Riggins. Software for assessment of the SAGE data, tag to gene mapping and related biological interrogation are also being developed.

P-50

Large-Scale Identification of the Promoter Regions of Human Genes Using Full-length Enriched and 5'-end Enriched cDNA Libraries Constructed by the Oligo-Capping Method

Yutaka Suzuki[1], Riu Yamashita[1], Sawako Suzuki[1], Aya Miura[1], Munetomo Hida[1], Junko Mizushima-Sugano[1], Hata Hiroko[1], Takao Isogai[2], Kenta Nakai[1], Sumio Sugano[1], [1]The Institute of Medical Science, The University of Tokyo, Tokyo, JAPAN, [2]The Helix Research Institute, Chiba, JAPAN

In order to identify the promoters from the large volumes of genomic sequences, we utilized mRNA start sites determined by a large scale sequencing of the cDNA libraries constructed by the 'oligo-capping' method. So far, we have accumulated 271,304 5' end sequences through random sequencing of more than 130 kinds of goligo-capped (full-length enriched and 5 f-end enriched) human cDNA libraries. Among them, 149,950 have been corresponded to cDNA sequences of known genes (8,959 genes). Sequence comparison between our entries and those of a reference sequence database, RefSeq, revealed that 4270 (48%) of RefSeq sequences should be extended towards the 5' ends. We, then, computationally mapped these 5 f-end sequences onto the human draft genomic sequence obtained from Golden Path. The results of the mapping provided us with not only the precise positional information of the transcriptional start sites and the adjacent promoters of the corresponding genes but also with the detailed information on distribution patterns of transcriptional start sites, which should reflect the dynamic nature of the transcription initiation events in vivo. All of these data are presented in our new database, DataBase of Transcriptional Start Sites (DBTSS; http://elmo.ims.u-tokyo.ac.jp/dbtss/).

P-51

Precision Gene Expression Analysis and Bioinformatics Can Track the Effect of Orally Administered, Anti-inflammatory Therapeutic Agents in Human Subjects

Victor V. Tryon, J. Emmick, N. Nicholls, J. Zuzack, D. Trollinger, J. C. Cheronis, M. P. Bevilacqua, Source Precision Medicine Inc., Boulder, CO

Differential gene expression can be used to precisely and reproducibly track the molecular responses of an individual to orally administered, anti-inflammatory therapeutic agents. Oral treatment with steroids or NSAIDS led to profound modulation in the expression of selected genes associated with the inflammatory response. A three day treatment with prednisone resulted in greater than 70% suppression of expression of IL1A and IL1B and a greater than 10 fold increase in expression of IL10. mRNA from 0.2 ml of whole, heparin anti-coagulated blood was extracted immediately post-blood draw using an aqueous method followed by cDNA synthesis. Quantitative PCR was performed with the ABI Prism 7700 instrument from Applied Biosystems. High dose ibuprofen (800 mg TID) also demonstrated an anti-inflammatory pattern, reducing the expression of selected pro-inflammatory genes including TNF. This pattern was distinguishable from that induced by treatment with steroids (e.g., ibuprofen decreased IL10 expression). Control subjects were examined over time and demonstrated a stable and reproducible pattern of gene expression using a 24 gene Precision Profile^{TM}for inflammation. These observations support the clinical utility of gene expression analysis in tracking the response of an individual to anti-inflammatory therapy when such assays are conducted with controlled and reproducible levels of precision.

P-52

Assessment of Candidate Genes for Fetal Alcohol Syndrome

Mshengu E. Tshabalala, Michele Ramsay and Denis Viljoen, Department of Human Genetics, The South African Institute for Medical Research and School of Pathology, University of Witwatersrand, SOUTH AFRICA

Assessment of Candidate Genes for Fetal Alcohol syndrome Considerable variation in offspring outcome is observed after intrauterine alcohol expo-sure. The underlying mechanisms may include genetic diversity in the enzymes responsible for alcohol metabolism. Several genetic polymorphism differences have been ob-served in enzymes involved in alcohol metabolism including the alcohol dehydrogenase-2 (ADH2); alcohol dehydrogenase-3 (ADH3); and aldehyde dehydrogenase-2 (ALDH2) genes. The variant alleles code for enzymes that are >30-fold different in their kinetic properties and metabolic effects. Previous studies from the Western Cape Province have shown that the ADH2*2 allele has a protective effect against FAS among individuals of mixed ancestry. The aim of this study is to ascertain families with fetal alcohol syn-drome (FAS) and to study the potential involvement of candidate genes including the ADH2, ADH3, ALDH2 and CYP2E1 loci in FAS by determining and comparing frequencies of the alleles in subjects and controls. Genotyping for ADH2, ADH3, ALDH2 and CYP2E1 loci will be done in 50 mother-FAS child pairs as well as in 50 black, 50-mixed ancestry and 50 white control subjects. Several methods will be used including the PCR followed by restriction-enzyme digestion. The allele frequencies in the mixed ancestry population will give an indication of the contributions from the black and white population to the mixed ancestry gene pool and may provide insight into the origins of susceptible and protective alleles.

P-53

Microarray Hybridization by Shear-Driven Flows, a New Approach to Enhance Both Sensitivity and Efficiency of Microarray Analysis

Paul Van Hummelen, Kris Pappaert, Johan Vanderhoeven, Danny Clicq, Gino V Baron and Gert Desmet, MicroArray Facility, Flanders Interuniversity Institute (VIB), Leuven, BELGIUM and Department of Chemical Engineering, Brussel, BELGIUM

Traditional diffusion driven hybridization takes at least 15 hours with an efficiency of less than 1%. Therefore, it is highly unlikely that labeled molecules from low expressed genes will reach the corresponding spot to hybridize. To elevate diffusion limitation we used Shear-Driven flows to generate a rapid lateral convective transport across the microarray surface and compared this to traditional overnight hybridizations. Shear-Driven flows relies on the dragging action exerted by a moving surface on an adjacent fluid layer and can transport extremely thin fluid layers at high velocities. We built two prototype modules, one excerting linear and one excerting rotating driving flows onto regular microarray slides. Hybridization tests were done with fluorescent cDNA pools, labeled by nick-translation. We could demonstrate that in 1 hour fluorescent intensities were achieved comparable to conventional overnight hybridization. Using the rotating module, the hybridization rate was increased by increasing the fluid velocity. However, using the linear module opposite observations were made. This unexpected result of the linear module may be caused by the striking lack of flatness of the microarray slides used. Future experiments will focus on studying hybridizations of low copy number genes, once a second generation shear-driven flow chambers is build.

P-54

Gene Expression Studies of In Vitro HIV-1 Infection:HIV-1 Activates Transcription of the Sterol Biosynthesis Pathway

Angelique B. van 't Wout, Ginger K. Lehrman, Svetlana A. Mikheeva, Gemma O'Keeffe, Gary K. Geiss, Roger E. Bumgarner and James I. Mullins. University of Washington, Seattle, WA

HIV requires host cell factors to survive and replicate and needs to interfere with cellular pathways aimed at destroying viral replication. We used cDNA microarrays to assess the interplay between the various biochemical processes involved in HIV infection of T cells. The expression levels for ~4600 human cDNAs were assessed in mock versus HIV infected T cells at different times after infection. Using human T cell lines and the lab strain HIV BRU we simultaneously infected millions of cells and studied subsequent changes in gene expression levels. Some changes induced by HIV BRU were found only in certain cell types. A subset of the regulated genes was verified by RT-PCR or FACS analysis. Genes involved in transcription, translations, splicing and protein processing, including MYC, were inhibited by HIV BRU infection. Several genes involved in sterol synthesis were stimulated in infected cells. Previous work by other laboratories has shown that cholesterol is essential for HIV infectivity. Our studies suggest that HIV actively ensures sufficient levels of cholesterol in the cell membrane for release of infectious viral particles. We are currently verifying these observations using metabolic labeling to measure functional changes in the sterol synthesis pathway.

P-55

SAGEdb: A Computational Platform for Investigations Using Serial Analysis of Gene Expression

Richard Varhol, Scott Zuyderduyn, Mehrdad Oveisi-Fordoei, Erin Garland, Martin Krzywinski, Marco Marra and Steven Jones, Genome Sequence Centre BC Cancer Agency, Vancouver, CANADA

Serial Analysis of Gene Expression, SAGE, a comprehensive gene expression profiling technique, generates acurate profiles of mRNA populations, making biological inferences from such expression data remains challenging. Our goal is to provide a platform facilitating the analysis and characterization of SAGE data. An important characteristic of our effort is the exploitation of existing connections between disparate data sources. This has been accomplished by developing a relational database containing large amounts of both public and internally generated biological information. These data include raw nucleotide and protein sequences, biological pathways, protein models, disease information, molecular interactions and tissue specificity. Where no connection exists, similarity is inferred computationally using BLAST & HMM. Emphasis is currently placed on designing effective methods to add new data sources and update existing data sources with new information. A complementary software package, SAGEspace, provides statistical analysis of raw data, allowing rapid formation of biological questions and enabling investigators to visualize and interpret results. An example of the software's capabilities involves modeling a pair-wise SAGE library comparison in 3D space. Gene expression levels are represented on the x-y axes; the degeneracy of a particular data point is represented on the z-axis; and results of statistical analyses and biological questions are denoted by colour, shape and texture.

P-56

ORESTES and the International Database of Cancer Gene Expression

André L. Vettore[1], Anamaria A. Camargo[1], Sandro J. de Souza[1], Gregory J. Riggins[2], Robert L. Strausberg[3], Andrew J. G. Simpson[1], [1]Ludwig Institute for Cancer Research, São Paulo, BRAZIL, [2]Duke University, Duham, NC, [3]National Cancer Institute, Bethesda, MD.

Open reading frame expressed sequences tags (ORESTES) differ from conventional ESTs by providing sequence data from the central, protein-coding portion of transcripts. Up to now, over 740,000 ORESTES sequences from human normal and tumor tissues have been deposited in the GenBank database. We estimate that ORESTES sampled over 80% of all highly and moderately expressed, and between 40% and 50% of rarely expressed human genes in the tumors studied. We thus believe that the capacity of the ORESTES strategy for gene discovery exceeds that of conventional ESTs. The distribution of ORESTES is such that many human transcripts are now represented by a scaffold of partial sequences distributed along the length of each gene. The experimental joining of the scaffold components, by reverse transcription-PCR, represents a direct route to transcript finishing that is a useful alternative to full-length cDNA cloning.Using ESTs (expressed sequence tags), SAGE (serial analysis of gene expression) and ORESTES sequences, an International Database of Cancer Gene Expression has been assembled comprising six million gene tags (ORESTES, CGAP-ESTs and CGAP-SAGE), which reflect the expression gene profiles of a wide variety of cancerous and normal tissues. A suite of informatics tools has designed to facilitate analysis of these datasets and is available through the NCI Cancer Genome Anatomy Project web site (http://cgap.nci.nih.gov/).

P-57

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

Ruth E. Wellenreuther, Stefan Wiemann, Daniela Heiss, Nina Claudino, Annemarie Poustka, Department of Molecular Genome Analysis, German Cancer Research Center, Im Neuenheimer Heidelberg, GERMANY

We generate human cDNA libraries that are enriched in full length clones i.e. from the translation start to the poly A tail. These libraries are used for a) systematic sequencing within the cDNA consortium of the Genome Project aiming at the identification and analysis of as many new genes as possible and b) for screening to isolate full length clones of partial genes. Libraries are created by directional cloning of cDNA's into plasmid vectors. Full-length enrichment is achieved via Clontech's SMART technology. This method is PCR-based, and in our modified strategy, we amplify and clone selective size windows of the cDNA fraction above 3 kb.Clones from the libraries generated within this project are the major source for the cDNA sequencing effort of the German Human Genome Project that is carried out by a consortium of nine laboratories. Over 900 full-length cDNAs of novel human genes could be identified by the consortium within the last 4 years. The clones generated in this project are available through the Resource Center of the German Human Genome Project (www.rzpd.de).

P-58

The Hypoxic Macrophage Transcriptome Profile: A Strategy to Increase the Discovery of Biologically Relevant and Novel Targets for Angiogenesis and Inflammation

Jonathan R. White, Robert A. Harris, Marie H. Craigon, Georgina L. Beard, Sheena R. Lee, Stuart M. Naylor, Chris R. Mundy, Oxford BioMedica (U.K.) Ltd, Medawar Centre, Oxford Science Park, Oxford, UNITED KINGDOM

Macrophage infiltration at sites of hypoxia is found in several human diseases, including cancer, rheumatoid arthritis and atherosclerosis. The macrophage has been shown to play roles in disease initiation, progression and treatment. Since hypoxia modulates important macrophage functions, we have investigated the transcriptional response to hypoxia. We have used a strategy to increase the sensitivity and biological relevance of conventional gene array approaches. To increase sensitivity, the key transcription factors involved in the hypoxia response were over-expressed in primary human macrophages, using viral vectors, together with the exposure of cells to hypoxia. RNA samples were subjected to gene array analysis, using non-biased arrays representing the majority of the human transcriptome. From this data, a focused secondary array was produced, also including a panel of indicator genes of macrophage function. The secondary array was used to increase the biological meaning of the original findings. Firstly an extensive panel of samples examining other stimuli frequently occurring with hypoxia in specific diseases were tested. Cells treated with a panel of cytokines were also used to help elucidate inflammatory functions of the arrayed hypoxia regulated genes. As final validation, clinical material was applied to the array. Following in depth analysis of the whole data set, potential novel disease targets are being selected for validation using primary human cells.

P-59

A Regression-based Method to Identify Differentially Expressed Genes in Microarray Time Course Studies and its Application in an Inducible Huntington's Disease Transgenic Model

Xie L. Xu, James M. Olson, Lue Ping Zhao, Fred Hutchinson Cancer Research Center, Seattle, WA

Time course studies with microarray technologies provide enormous potential for exploring underlying mechanisms of biological phenomena in many areas of biomedical research, but the large amount of gene expression data generated by such studies also present great challenges to data analysis. Here we introduce a regression-based statistical modeling approach that identifies differentially expressed genes in microarray time course studies. To illustrate this method, we applied it to data generated from an inducible Huntington's disease transgenic model. The regression method accounts for the induction process, incorporates relevant experimental information, and includes parameters that specifically address the research interest: the temporal differences in gene expression profiles between the mutant and control mice over the time course, in addition to heterogeneities that commonly exist in microarray data. Least squares and estimating equation techniques were used to estimate parameters and variances, and inferences were made based on efficient and robust Z statistics under a set of well-defined assumptions. A permutation test was also used to estimate the number of false positives, providing an alternative measurement of statistical significance useful for investigators to make decisions on follow-up studies.

P-60

Genetic Hardware and Genetic Software after Human Genome Project

Shuwei Yang, GeneCopoeia, Inc., Frederick, MD

With the progress of Human Genome Project, other animal genomes and pathogenic microorganism genome projects, many full-length genes were discovered and more and more will be added into the list. I proposed the two sections of the post-genome project, Genetic Software and Genetic Hardware. The opportunities and challenges of these two sections will be discussed. The contents of the Genetic Hardware Section are high speed computer, high accuracy genetic data, high accuracy of full-length cDNA clones, high efficiency gene transfer vector and host, and high specific and sensitive molecular monitoring system in vitro and in vivo. The study of the interaction among proteins encoding by human and pathogenic microorganism genes is the next wave of the life science. The ORFs of these genes have to be transferred from one vector to another very frequently in order to produce the gene product (protein) in vitro and in vivo. Classic subcloning methods can not meet the requirement of rapidly growing of the studying of function of many discovered genes. New technologies of semi-automation we have invented to accelerate re-cloning of the ORFs of full-length genes with higher fidelity into suitable vector(s) for specific host(s) and the QC/AC will be discussed. The problems such as the difference in the sequence analysis of the ORFs from public cDNA database and genomic annotation will be explored and discussed.

P-61

High Throughput Detection of Single Nucleotide Polymorphisms (SNPs) by OCP Microarrays

Zailin Yu, Allan Peng, Amy Cernetich and Qianjin Hu, Mergen LTD., San Leandro, CA

Mergen's patented method (OCP: On-Chip PCR) enables us to quickly screen thousands of SNPs by combining the advantages of the high throughput and detection from DNA microarray with the sensitivity and specificity from in situ PCR.The key advantages of the method are high throughput, specificity and low price The targeted end user price is $0.35 per SNP score and the technique has a high success rate. Mergen's technique allows up to 40,000 SNPs (for example, all four alleles and both strands scored among 10,000 loci) to be detected per microarray pair. In a "proof of principle", experiment Mergen screened a hundred human SNPs that were randomly chosen from a public accessible SNP database - The SNP Consortium Ltd. (Cold Spring Harbor Laboratory, http://snp.cshl.org/). Up to 85% of SNPs tested were detected and the results matched the data generated from a standard sequence method. (These failures are caused by poor SNP and PCR primers design and, if needed, the rate of failures could be reduced by subsequent redesign of the oligos.) In another experiment, a mutation at 273C/T on a human tumor suppressor gene (p53) was identified from normal and tumor DNA samples.

P-62

Characterization of the Upstream Regions of 836 Orthologous Human/Mouse Genes

Mihaela Zavolan[1], Terry Gaasterland[1], and Nikolaus Rajewsky[2], [1]Laboratory for Computational Genomics, The Rockefeller University, New York, NY, [2]Center for Studies in Physics and Biology, The Rockefeller University, New York, NY

In recent years, it has been become evident that cross species comparisons are a very powerful tool for discovering functional regions in genomic DNA [1-4]. For example, regulatory elements such as binding sites for transcription factors are frequently found in conserved parts of upstream sequences. In order to study the conservation of upstream sequences on a genome--wide level, we have developed a fully automated computational pipeline to identify pairs of orthologous genes, to unambiguously map them to the corresponding genomic sequences, to isolate their upstream regions, and to identify conserved sequence elements. Starting with the RefSeqdatabase [5] of mouse and human protein sequences, this pipeline extracted data for 836 pairs of orthologous human/mouse genes. We present the analysis of the sequence composition, degree of conservation, and presence of regulatory motifs in the upstream region of these genes.

[1] R. C. Hardison, J. Oeltjen, and W. Miller, Long Human—Mouse Sequence Alignments Reveal Novel Regulatory Elements: A Reason to Sequence the Mouse Genome, Genome Research 7:959-966 (1997).

[2] N. Jareborg, E. Birney, and R. Durbin, Comparative Analysis of Noncoding Regions of 77 Orthologous Mouse and Human Gene Pairs, Genome Research 9:815-824 (1999).

[3] W.W. Wasserman, M.Palumbo, W. Thompson, J.W. Fickett, C.E. Lawrence, Human-mouse genome comparisons to locate regulatory sites, Nature Genetics 26:225-228 (2000).

[4] G.G.Loots, R.M. Locksley, C.M. Blankespoor, Z.E. Wang, W. Miller, E.M. Rubin, K.A.Frazer, Identification of a Coordinate Regulator of Interleukin 4, 13, and 5 by Cross--Species Sequence Comparisons, Science 288:136-140 (2000).

[5] K.D. Pruitt and D.R. Maglott, RefSeq and LocusLink: NCBI gene-centered resources Nucleic Acids Research 29:137-140 (2001).

P-63

Downstream of MyoD: An in vivo 27-time Point Temporal Series, Defining Novel Gene Transcriptional Pathways in Muscle

Po Zhao[1], Simona Iezzi[2], Ethan Carver[3], Devin Dressman[1], Thomas Gridley[3], Vittorio Sartorelli[2], and Eric P Hoffman[1], [1]Children's National Medical Center, George Washington University, Washington DC, [2]NIAMS, National Institutes of Health, Bethesda, MD, [3]The Jackson Laboratory, Bar Harbor, ME

Temporal expression profiling was utilized to define transcriptional regulatory pathways in vivo in a mouse muscle regeneration model. Potential downstream targets of MyoD were identified by temporal expression in a 6 time point cardiotoxin (CTX) degeneration/regeneration series in mouse muscle in vivo. Nucleated clusters with known downstream targets were further queried by promoter database mining, and gel shift and supershift assays; Slug, calpain6, Peg3, Sox11 and nectin3 were all identified as novel MyoD targets. Slug, a member of the snail/slug family of zinc finger transcriptional repressors critical for mesoderm/ectoderm development, was further shown as a downstream target by promoter/reporter constructs, and demonstration of defective muscle regeneration in Slug null mice. The E box consensus identified in genes regulated by MyoD in vivo was found to be considerably more stringent (CACAG[G/C]TGT) than the in vitro consensus (CAnnTG). We then continued the temporal expression profiling approach with expression profiling of 27 time points to more finely define temporally regulated gene clusters. In addition, we defined potential downstream targets of Slug by expression profiling muscle regeneration in Slug null mice. The results reported here demonstrate that transcriptional pathways can be defined in vivo in vertebrates, not only for genes downstream of well characterized transcription factors (e.g. MyoD), but also for less well-characterized transcription factors (e.g. Slug).

P-64

Defining Disease Pathogenesis by Comparative Cross-Species Profiling in Human and Mouse Knock-Outs: the Downstream Consequences of Dystrophin Deficiency

Eric P. Hoffman[1], Livia Pasquali[1], Po Zhao[1], Ksenija Gorni[1], Frank W. Booth[2], Brian Tseng[2], Yi-Wen Chen[1,]
[1]Research Center for Genetic Medicine, Children's National Medical Center, Washington, DC, [2]Department of Veterinary Biomedical Sciences, University of Missouri at Columbia, Columbia, MO

The absence of dystrophin protein leads to different phenotypes in different species, despite complete loss of dystrophin in all muscles. Human DMD patients show chronic degeneration/regeneration with progressive muscle wasting and an early death. The mdx mouse initially shows normal muscle, with widespread skeletal muscle necrosis at the age of 3-4 weeks, followed by “successful” muscle fiber regeneration. We hypothesized that comparison of 60,000-gene profiles in human DMD muscle (U95 GeneChip series and custom Affymetrix MuscleChip), and 36,000-gene profiles in mdx and experimental muscle regeneration (U74 GeneChip series) would identify differentially expressed genes as potential candidates for conferring protection to murine dystrophin deficient muscle. These same genes would be targets for modulation of the progressive pathology of the human disease. We used as a starting point a “late muscle regeneration cluster” gleaned from a 27 time-point murine regeneration series. We then compared the expression of these genes in dystrophin-deficient mdx mouse muscle and human DMD muscle. We then generated genelists of those genes upregulated in both mouse models, but not upregulated in DMD. Our true genome-wide cross-species comparative candidate gene analysis showed inhibitors of negative regulators of skeletal muscle mass, specific cell division and differentiation genes, and specific connective tissue modulatory genes as candidates for the species-specific response to dystrophin-deficiency.