tion of anonymous sequences. Toward systematic gene function analysis from microarray data, we evaluated several methods of cluster analysis using a publicly available yeast microarray data set, which was effective in grouping functional units by sequence similarity. These methods include the nearest-neighbour method (single linkage clustering), the furthest-neighbour method (complete linkage clustering) and the group average method. Clustered groups were evaluated by looking up functional annotations in the GENES database of the Kyoto Encyclopedia of Genes and Genomes (KEGG). The GENES database contains additional functional annotations compared with the original databases. One example is the EC number of an enzyme, which makes it possible to map the enzyme on the metabolic pathway diagrams in our PATHWAY database. Genes clustered by expression information were analysed in conjunction with sequence similarity and three-dimensional structural classification to study the correlation between these 'expression clusters' and functional units of genes that have been stored in KEGG. We will demonstrate our results to evaluate clustering methods from the biological aspect using KEGG databases.

Bottinger, Erwin

Genetic/genomic analysis of signal transduction pathways using cDNA microarray technology

Yaw-Ching Yang[1], Ester Piek[3], Aldo Massimi[2], Rongguang Yang[2], Joerg Heyer[2], Raju Kucherlapati[2], Anita B. Roberts[3] & Erwin P. Bottinger[1,2]

[1]Departments of Medicine and [2]Molecular Genetics, Albert Einstein College of Medicine, Bronx, New York, USA
[3]Laboratory of Cell Regulation and Carcinogenesis, National Cancer Institute, Bethesda, Maryland, USA

cDNA microarray technology provides a novel approach to identify individual target genes and survey global genetic programs under the control of specific signalling pathways in mammalian systems. Smad2 and Smad3 are highly homologous members of the receptor-regulated subfamily of Smad proteins with a central role in TGF-β signalling and target gene regulation.

The purpose of this study is to identify genes that are regulated by TGF-β through activation of Smad2 or Smad3, or through Smad-independent pathways. Using cDNA microarray technology developed at AECOM, we have profiled the expression of 9,000 genes in wild-type mouse embryonic fibroblasts ($MEFs^{+/+}$), $Smad2^{-/-}$ MEFs or $Smad3^{-/-}$ MEFs at baseline and following exposure to TGF-β1 or activin for 1 hour, 4 hours and 10 hours, respectively. Cells are cultured in exponential growth conditions. Total RNA from $MEFs^{+/+}$ and matched test samples ($Smad2^{-/-}$ or $Smad3^{-/-}$ MEFs) is fluorescently labelled with either Cye3- or Cye5-dUTP. Probes are then co-hybridized to genes gridded on the array and single-channel fluorescence intensities are measured. Results are expressed as a normalized ratio (Cye3/Cye5) in which deviations from 1 indicate increased (>1), or decreased (<1) levels of gene expression in $Smad2^{-/-}$ or $Smad3^{-/-}$ MEFs relative to $MEFs^{+/+}$. Using microarray data mining and analysis tools (microarray analysis tools or MAT) developed at AECOM, we have identified a large number of novel genes (expressed sequence tags) that are differentially regulated in $Smad2^{-/-}$ or $Smad3^{-/-}$ compared with $MEFs^{+/+}$, depending on genotype, cytokine treatment and time. MAT is used for global survey (cluster analysis) of gene expression profiles in this genetically defined model system.

In conclusion, using a high-throughput cDNA microarray approach, we have identified gene modules differentially expressed in Smad2- and Smad3-deficient fibroblasts compared with wild-type fibroblasts in response to TGF-β. Stratification of gene activation profiles according to specific signalling mediators should provide new insights into the mechanisms that determine biological specificity of TGF-β.

Brazma, Alvis

A public repository for DNA microarray-based gene expression data

Alvis Brazma, Alan Robinson & Jaak Vilo

European Bioinformatics Institute, EMBL Outstation, Hinxton, UK

The European Bioinformatics Institute has discussed the possibility of establishing a public repository for DNA microarray-based gene expression data with many of the major laboratories developing and using these technologies in Europe and the USA, and established a consensus that the time is right to develop standards for storing these data. There are many reasons for establishing a public repository of these data. First, by storing data obtained from different experiments under varying conditions, the repository will build up progressively more detailed expression profiles for genes and aid in the study of functional genomics. Second, it will facilitate the cross-validation of data obtained by different technologies, characterise various techniques and establish error rates, benchmarks and gold standards. Third, by making these data available publicly, bioinformatics and analysis of gene expression data will be promoted and pushed forward. Fourth, that the experimental data should be public will be consistent with the policy of most journals with regard to the verifiability of published conclusions. As with biological sequence and molecular structure public repositories, this will not prevent submitters from being the first to exploit and publish their data from gene expression studies. Currently the EBI is establishing a pilot database containing the microarray gene expression data either available publicly or provided to the EBI by laboratories generating these data. We would like to encourage laboratories wishing to discuss submitting their data in such a database to contact us.

Budach, Wolfgang

Customized cDNA chips for pharma development

W. Budach, J. Gut, D. Neuschäfer, Ch. Wanke, E. Baer, H. Schürch, O. Grenet, S. Chibout & J. Vonderscher

Novartis Pharma AG, 4002 Basel, Switzerland

The development of novel pharmaceutically active compounds requires effective methods in order to increase throughput and focus corporate research and development. The increasing number of potential compounds in pharmaceutical research and the shortened time frame in the drug development have created a need for novel methods in order to accelerate development processes. Among other methods, such as Real-Time RT-PCR, gene expression analysis with customized cDNA chips providing study relevant markers have shown to be a valuable tool for surrogate markers identification and monitoring. The present paper describes the preparation and application of cDNA chips for apoptosis targets genes as well as for calbindin D-28k and cytokines. Validation of the results are performed using the Real-Time RT-PCR technology. In addition, investigations regarding the requirement of replicates as well as the development of tools for chip quality control are discussed.Results of an in-house cDNA mid-density feasibility study are shown. Customized multiple use chips with genes for apoptosis, cytokines, and CYP450 were prepared. The chip preparation steps such as ink-jet printing, incubation and baking were monitored with Differential Interference Contrast microscopy (DIC) which gives information on e.g. missing spots, overlap, and homogeneity and can therefore be used as tool for quality control. The correlation between number of replicates and coefficient of variation (CV) of the fold change values was investigated with specially designed chips and ink-jet and pin spotter principle was compared. In addition, direct determination of the background and background extrapolation was compared and the impact on the