ically organize them in biologically meaningful ways. There are many options for the implementation of feature extraction and gene clustering programs to help accomplish these goals. We have developed Matlab prototypes that emphasize flexibility, accuracy, transparency and the systematic incorporation of statistical analysis. These prototypes facilitate the development and refinement of algorithms as we converge on fully automatic analysis.

There are many possible approaches to feature extraction, and the choices are influenced by factors including: feature morphology and uniformity, and positioning errors, array deposition methodologies and tolerances, labelling methods and scanner specifications. Our feature extraction software includes several methodologies and a number of adjustable parameters that can be modified to suit the application. For example, image file formats, image alignment procedures for two-colour images, irregular modifiable grids, feature locating, grid fitting, extraction, pixel-level outlier rejection, feature-level outlier rejection, background-subtraction and dye-normalization processes have all been incorporated. With this feature extraction prototype, we are can explore different methods of image analysis, outlier rejection, background subtraction and normalization while optimizing sensitivity and specificity. Relative expression levels are reported along with their respective statistical information.

To analyse data from multiple array experiments we developed a package built around a clustering method based on a graph theoretic approach. This approach requires no assumptions as to the number of clusters expected and is flexible with regard to the selection of a similarity function. In particular, similarity functions that take into account statistical characteristics, such as confidence intervals, are supported. Results from applying both analysis tools to raw images and to processed gene expression data from humans, C. elegans and other organisms will be presented.

Scott, James

Use of biochip microarrays to discover genes for dyslipidaemia, insulin resistance and hypertension in rats and humans

James Scott, Rossi Naoumova, Carol C. Shoulders & Timothy J. Aitman

Hammersmith Hospital, Du-Cane Road, London W12 ONN, UK

The metabolic overlap syndromes hyperlipidaemia, diabetes mellitus, insulin resistance, obesity and hypertension together constitute the major risk factors for coronary heart disease. The genetic predisposition to these conditions is strong but complex, and many of the genes involved have not been identified. The identification of the genes or metabolic pathways involved in these disorders would lead to new diagnostic and therapeutic targets with wide applicability. The spontaneously hypertensive rat (SHR) is a model of essential hypertension which also displays abnormalities of lipid metabolism and insulin resistance similar to those found in the human metabolic syndromes. These include raised blood triglycerides and fatty acids, defective catecholamine-mediated lipolysis and excessive growth of intra-abdominal adipocytes. Similar abnormalities have been identified in combined hyperlipidaemia, maturity-onset diabetes mellitus and obesity in humans. We are using the SHR as a model to identify genes and pathways that are likely to be in common with those causing the human diseases. Quantitative trait loci (QTL) for glucose and fatty acid metabolism, hypertriglyceridaemia and hypertension map to a locus on rat chromosome 4. We used cDNA microarrays, congenic strains and radiation hybrid mapping to identify a defective gene, *Cd36* (encoding fatty acid translocase), at the peak of linkage. *Cd36* cDNA contains multiple sequence variants and the encoded protein product is undetectable in SHR adipocyte plasma membrane. Transgenic mice overexpressing Cd36 have reduced blood lipids. We conclude Cd36 deficiency underlies insulin resistance, defective fatty acid metabolism and hypertriglyceridaemia in SHR (ref. 1). Human CD36 is being examined for mutation in subjects with insulin resistance from different ethnic backgrounds. The role of other genes showing primary or secondary alteration in expression and their role in metabolic pathway tracking ('pathwayomics') and QTL identification will be discussed.

1. Aitman, T.J. *et al. Nature Genet.* **21**, 76–83 (1999).

Shams, Soheil

Information processing tools for microarray technology

S. Shams, P. Kalocsai, J.D. Park, J. Chen & Y. Zhou

BioDiscovery Inc., 11150 W. Olympic Blvd., Suite 805, Los Angeles, California 90064, USA

With the maturation of the technology for fabricating and scanning high-density gene expression microarrays, information processing issues loom over as the fundamental challenge to the widespread use of this technology. These issues can be broadly organised into three interrelated categories: data tracking and management, image analysis and expression level quantification, and data mining and analysis. Data tracking and management requires tools for automated tracking of samples and generation of mapping tables to specify what has been spotted at each location on a microarray slide. The image processing issues are one of the more important and difficult issues at this time. Issues with proper image segmentation, separating signal from background and contamination, must be effectively dealt with in an automated fashion. A number of sophisticated approaches have been developed and will be presented. The data analysis and mining issues require a uniform and flexible method of data representation and normalisation to allow for sharing of data among different "data suppliers". Once the data is made available a number of clustering and dimensionality reduction approaches can be used to analyse, visualise and compare various experiments together. We will describe software tools for performing these operations.

Shannon, Karen

Differentiation of homologous sequences using oligonucleotide microarrays

Karen W. Shannon[1], Paul K. Wolber[1], Robert Kincaid[1], Julja Burchard[2] & Alan Blanchard[2]

[1]Hewlett-Packard Company, M/S 25U-5, 3500 Deer Creek Road, Palo Alto, California 94304, USA
[2]Rosetta Inpharmatics, 12040 115th Ave NE, Kirkland, Washington 98034, USA

Microarrays of surface-linked oligonucleotide probes are an emerging technology for understanding the genome-wide relationships among gene sequences and cellular functions. Microarrays can potentially measure the inter-relationships among the expression levels of all an organism's genes in parallel, enhancing our ability to use sequences to understand disease processes. Chemically synthesized oligonucleotide probes can be designed to differentiate between closely related homologues and alternatively spliced mRNAs; arrays based on probes derived from cDNA libraries cannot perform such measurements. Estimates of the degree of cross-hybridization (specificity) of that probe to other biological targets represented in the sample are essential requirements for the extraction of meaningful data and results. To provide such estimates, we have characterized the performance of *in situ*-synthesized oligonucleotide microarrays fabricated by a novel, highly parallel printing process. The microarrays contained 25-mer oligonucleotide probes for yeast b-actin (*ACT1*) and 4 homologous actin-related genes (*ARP1*, *ARP2*, *ARP3* and *ARP5*). Cy3-labelled actin-specific targets were