E1–TRANSCRIPTOMICS

Organised by Karin van de Sande and Sarah Blackford, Co-sponsored by BBSRC

E1.1–Microarray expression analysis in an academic setting

T. Freeman, MRC HGMP-Resource Centre, Hinxton, Cambs., UK

The application of microarrays for gene expression profiling has been demonstrated to be one of the most powerful and direct ways of using the sequence data for functional studies. As the use of microarray technology grows, bottlenecks in array fabrication, sample preparation, hybridisation and analysis need to be overcome. The Microarray Programme at the HGMP-RC has a remit to develop and distribute the technology, specifically microarrays of human and mouse gene probes, to a UK academic user base. I will describe some of the measures we are taking to ensure an effective pipeline for the generation of microarray data. This will include the use of oligonucleotide probes for microarray fabrication, methods of RNA labelling and hybridisation to ensure high sensitivity and use of samples where the amount of available bioclogical material is limited. I will also our efforts currently being put in to developing bioinformatics support for microarray data storage and analysis.

E1.2–Hybridisation-array technology in the integrated analysis of the yeast genome and its activities

S.G. Oliver, M.E. Gent, D. Delneri and A. Hayes, School of Biological Sciences, University of Manchester, Manchester, UK

Hybridisation-array technology is the most facile method currently available for the global analysis of gene expression. It represents an approach that is both comprehensive in its scope and high-throughput in its application. Hybridisation arrays may be applied, not just to transcriptome analysis, but also to the comparative analysis of phylogenetic relationships between closely related species using genomic hybridisations. The technology may also be exploited in the high-throughput phenotyping of bar-coded deletion mutants.

In transcriptomics, the comprehensive nature of hybridisation-array analysis is both its greatest advantage and its greatest difficulty. It is a difficulty because traditional methods of expression analysis (e.g. northern hybridisations) that examine only a few genes at a time (e.g. northern hybridisations) led us to design experiments in which only genes whose expression was analysed were those known, or suspected, to be relevant to the biological problem under study. With hybridisation-array analysis, expression data on *all* genes is obtained, and the problem is to ensure that only *relevant* information is acquired. The consequences of this for experimental design and the normalisation and processing of array data will be discussed and illustrated with examples from yeast. Finally, the advantages of integrating transcriptome analyses with other kinds of functional genomic data will be explored.

E1.3–Using arrays to dissect the process of anther and pollen development in *Arabidopsis*

Z.A. Wilson, G. Vizcay Barrena and C-Y. Yang. Plant Sciences Division, University of Nottingham, Nottingham, UK

We have been using the *Arabidopsis* Affymetrix arrays to analyse the process of anther and pollen development in the *male sterility1* (*ms1*) mutant. In the *ms1* mutant the process of pollen development begins normally, with meiosis and tetrad formation progressing as in the wild type. However, immediately after microspore release from the callose wall the immature pollen begins to breakdown. Around this stage the tapetal tissue also appears abnormally vacuolated. Degradation of the locule contents continues, resulting in empty anthers with no functional pollen. The *MS1* gene is therefore critical for the production of viable pollen. Female fertility is however unaffected and the mutant can be rescued by fertilization with wild type pollen. The *MS1* gene has homology to PHD-finger motif transcription factors (Wilson et al, Plant J. 28:27–39). *MS1* gene expression is highly regulated, with low-level expression seen briefly during the final stages of microspore maturation and tapetal degeneration.

We are interested in determining the genes that *MS1* may regulate and one approach we are using is that of transcriptomic analysis. We have analysed bud development in Ler (wt) and the *ms1* mutant to ensure comparable developmental stages were selected. RNA extraction was initially conducted on a range of floral tissues from both *Arabidopsis* and *Brassica* flowers to assess the extraction procedures. However, due to the difficulties

in sampling and staging, whilst maintaining the quality of tissue samples, RNA from intact buds was selected for the microarray analysis. Extractions were made from Ler (wt) and the *ms1* mutant buds at developmental stages during and post- *MS1* expression. These have been used to screen the *Arabidopsis* Affymetrix gene array. Results from these analyses will be presented.

E1.4–Microarrays for microbial pathogens

J.A. Mangan, Dept. Medical Microbiology, SGHMS, London, UK

The wealth of pathogen genome sequence information available has increased the demand for tools to gain a greater understanding of the general biology of the microbe and complex host–pathogen interactions. To address this need our group (The Bacterial Microarray Group at St. George's Hospital Medical School; B?G@S) has been funded by The Wellcome Trust to generate whole genome DNA microarrays for 12 bacterial pathogens in two years and to make these available to a multi-collaborative network of research groups for numerous and widespread applications for a further three years. We have completed whole-genome arrays for the bacterial pathogens *Campylobacter jejuni, Mycobacterium tuberculosis, Haemophilus influenzae, Yersinia pestis* and *Streptococcus pneumoniae* as well as an array of the *Yersinia pestis* and *Salmonella typhi* plasmids. Whole-genome arrays will also be generated for the following pathogens: *Neisseria meningitidis*, *Listeria monocytogenes, Helicobacter pylori, Bordetella pertussis, Clostridium difficile, Staphylococcus aureus, Chlamydia spp, Mycobacterium spp.*
Some of the unique problems associated with bacterial pathogens will be discussed within the context of microarrays.

E1.5–The UK's Public Wheat Transcriptomics Project

I. Wilson, D. Edwards, G. Barker, S. Shepherd, R. Beswick, J. Coghill, P. Owen, K.J. Edwards, Biological Sciences, University of Bristol; M. Holdsworth, P. Shewry and J. Lenton, Rothamsted Research

Grain development and to a lesser extent plant development under abiotic stresses are areas of biology that are of considerable interest to the UK cereal community. Within the BBSRC Investigating Gene Function (IGF) program we (http://www.cerealsdb.uk.net) are producing the resources required to investigate alterations in the transcriptome of hexaploid, winter wheat during these developmental processes. To this end we have single pass sequenced cDNAs of 1000 randomly-picked clones from each of 35 cDNA libraries representing various stages of grain and plant development. Annotated sequencing results are stored in a publicly accessible, online database at the above URL or via GenBank. Our database currently holds ~25 000 EST sequences which may be accessed *via* similarity (BLAST searches) to a user-submitted sequence or by the submission of keyword text queries through a user-friendly web interface. We have also generated high-density microarrays of a 10 000-wheat unigene set, which is available to the wheat community at cost price. Interested cereal researchers can also view online the expression profiles of selected ESTs as determined by microarray experiments carried out in our laboratory. These resources are an obvious aid to selecting physiologically-relevant ESTs for further investigation. Our current experiments are focusing on grain development and responses to drought, waterlogging, salt-stress, reduced nitrogen stress and lowered temperatures.

E1.6–Transcriptome Analysis of Haem-Associated Proteins

A.G. Smith[*], H.H. Ottenhof[*], I. Wilson[**] and L. Dennis[**]. Plant Sciences, University of Cambridge, Cambridge, UK[*], CSIRO Plant Industry, Canberra, Australia[**]

We are using a gene-driven approach to ask the question 'what is the transcript profile of haem-associated protein genes under several different treatments?' By this means we aim to analyse all the enzymes of a single pathway, and determine if there is coordination of cofactor and apoprotein production. There are about 500 genes for haem-related proteins in the *Arabidopsis* genome, equivalent to 2% of the genome. These include about 30 proteins involved in the biosynthesis of haem and the related tetrapyrroles chlorophyll, sirohaem and phytochromobilin, and a further 450 encoding proteins that have haem as cofactor, such as cytochromes, catalases, peroxidases and several oxidases. Many of these have been identified solely by their haem-binding motif, and are of unknown function, including the majority of the 292 cytochrome P450 genes. We have used spotted arrays that contain almost all the biosynthetic genes, and representatives of the apoprotein genes. We have studied their transcript profiles in seedlings grown under different conditions, and in mutants impaired in plastid–nuclear signalling. Cluster analysis reveals that there is little if any overall coordination of the biosynthetic genes, but enzymes of the chlorophyll branch are regulated simi-

larly to those for chlorophyll apoproteins. A comparison of our data with those for published microarray experiments using GeneSpring is currently being undertaken to establish other patterns of expression, in the hope of identifying the function of uncharacterised haemoproteins.

E1.7–Differential gene expression in *Drosophila*

S. Russell, Department of Genetics, University of Cambridge, Cambridge, UK

I will discuss the provision for high-throughput gene expression analysis in *Drosophila*, including Affymetrix and cDNA arrays. I will present results from an analysis of strain variation and examples of an exploration of sex-specific gene expression. Examples of our recent work on approaches to dealing with issues of data acquisition and analysis will also be presented.

E1.8–NASC Affymetrix service and Data services

S. May, Nottingham Arabidopsis Stock Centre, University of Nottingham, Sutton Bonnington Campus, Loughborough, UK

The NASC/GARNet Affymetrix service has been running since February 2002. We have processed a large number of chips on behalf of the community. The website about the Affymetrix GeneChip service includes:
– The logistics of handling and processing requests at NASC;
– How users can access their own and other people's chip data;

We also cover frequently asked Questions (FAQs) about costs, timing, preparation of materials, analysis of data and the perennial issue of replicates (technical and biological).
NASC stores large amounts of data for Arabidopsis, obtained from three sources: 1) NASC's in-house services (e.g. our transcriptomic service or our catalogue). 2) our other GARNet service collaborators (chiefly the proteomics service at Cambridge). 3) large public data banks (such as EMBL). This data is put into our four major databases. Our genomics database AGR contains sequence data from Arabidopis. Among other things, it allows easy searching for inserts. NASC also has a web-accessible transcriptomics database. NASC's current seed catalogue has undergone renovation to improve it's status as a data resource with emphasis on ease of use and interoperability with external databases. Finally NASC is building a proteomics database to store data from Cambridge. With these new and existing databases, as well as web interfaces, we intend to add a facility to make them 'web-services' enabled. This will allow bioinformaticians to run automated queries to link the power of these databases together and integrate them with other databases elsewhere.

E1.9–Transcriptome analysis of metabolic and developmental transitions in *Streptomyces* bacteria

C.P. Smith, G. Hotchkiss, V. Mersinias and G. Bucca, Biomolecular Sciences, UMIST, Manchester, UK

We are applying DNA microarray analysis to study global patterns of gene expression in *Streptomyces*. This organism presents a number of challenges for this technology; it has a very high G+C composition and a relatively high level of apparent gene redundancy. We have focussed largely on producing PCR-based DNA microarrays, although we have recently completed favourable side-by-side trials with long oligonucleotide arrays. For generating PCR products we designed an automated primer selection programme taking full advantage of the genome sequence. Similarity searches on each candidate probe predicted potential cross-hybridisations and allowed selection of 'unique' PCR products. The PCR products were generated by a two-stage procedure, firstly using gene-specific primers with universal tags and subsequently, using universal primers, one of them being 5′-amidated. Arrays comprising the majority of open reading frames (93%) have been produced and protocols for RNA isolation, cDNA synthesis and hybridisation have been optimised (www.arrays.umist.ac.uk). We are currently investigating patterns of gene expression in time course experiments to investigate changes that correlate with developmental and metabolic transitions. Our analysis has focussed mainly on 'surface-grown' cultures and has revealed dramatic changes in gene expression at the 'decision' phase prior to the onset of aerial mycelium and secondary metabolite formation. The design and production of the array will be reviewed as will recent developments with oligonucleotide arrays. Most importantly, we will discuss recent biological insights into developmental regulation that were revealed through the microarray-based analyses.

E1.10–transcriptomics using oligoarrays to discover genes involved in seed germination and seedling establishment in *Arabidopsis*

K. Madagan, Department of Biology, University of York, York, UK

The availability of the full genome sequence of *Arabidopsis* means that the expression levels of its 26 000 genes can be monitored on a global scale using DNA microarrays. During germination and early post-germinative growth, *Arabidopsis* seeds break down oil storage reserves to provide sucrose for growth until the seedlings become photosynthetically competent. We have performed a series of hybridisations using oligo microarrays to investigate the transcriptome of *Arabidopsis* during germination, from dry seed until the seedlings become established. The arrays consist of 26 092 70-mer oligos on a single slide. Each oligo represents one predicted open reading frame in the *Arabidopsis* genome. Alien and constitutive controls are also included. The aims of the project are (1) to identify genes co-ordinately regulated during germination and early post-germinative growth (2) to evaluate data generated from Affymetrix and oligo arrays and compare the sensitivity and accuracy of the two.

E1.11–Ensuring data integrity and identifying differential expression without normalisation on a chloroplast genome microarray

D. MacLean, C. A. Jerome and J. C. Gray. Department of Plant Sciences, University of Cambridge, Cambridge, UK

A DNA microarray containing reporters for all the genes (approx. 100) in the 156 kbp chloroplast genome of *Nicotiana tabacum* was used to examine patterns of gene expression in tobacco seedlings. The problems raised by analysing a comparatively small number of genes have required different approaches to those used on larger arrays. Such a system requires that all array reporters give signal in each experiment, requiring modification of priming regimes and labelling conditions. Standard normalisation techniques cannot be applied and both internal and biological replication become essential. Changes in transcript abundance between experimental conditions can be subtle and approaches to identifying differential expression based on power analyses, Bootstrap t-tests and effect size analyses have been used to identify consistently differentially expressed genes at low ratios.