Arabidopsis Microarray Service Facilities[1]

Ellen Wisman* and John Ohlrogge

Michigan State University, East Lansing, Michigan 48824

Now that we know the complete sequence of the Arabidopsis genome, how can plant biologists most effectively use this 130 million-bp database to move toward fuller understanding of plants? Microarray technology is one of a collection of tools that can accelerate our transition from raw data toward broader understanding. It is clear that knowing a gene sequence often does not tell us its function. In fact, more than 30% of the approximately 25,000 genes of Arabidopsis show no homology to genes of known or hypothesized function. Thousands of additional genes are only identified as members of classes such as protein kinases or transcription factors, but no information is available about their specific roles. The knowledge that there are 6,000 to 10,000 plant genes with functions still to be discovered helps to define our level of ignorance and challenges us to find new tools for investigating gene function. It is an exciting challenge because uncovering the function of these poorly understood genes is likely to lead to a very rich harvest of new discoveries. To help meet this challenge, the Arabidopsis Functional Genomics Consortium (AFGC) was established to coordinate the study of gene function by two synergistic methods: microarray gene expression profiling and gene knockout mutagenesis. AFGC is comprised of a team of investigators from four institutions. Michigan State University and the Carnegie Institute of Washington at Stanford University have established microarray facilities. The University of Wisconsin and Yale University coordinate gene knock-out efforts as described in the accompanying article by Sandra Austin-Philips. A website explains in more detail the objectives, procedures, and services of AFGC (http://afgc.stanford.edu).

PRINCIPLE OF MICROARRAYS

Microarrays are conceptually very simple, but their production and analysis can be technically demanding. Another name for such technology might be "reverse northern-dot blots." DNA representing thousands of genes is deposited on a solid surface at high density (1,000–10,000 "spots"/cm^2). These DNA samples are then hybridized with labeled probes derived from the mRNA population present in plant sample(s). When two mRNA samples are compared (for example, from control and treated plants) the intensity of the signal from label bound to each spot reflects the relative mRNA abundance for each gene represented on the array. Therefore, information on gene expression can be obtained simultaneously for thousands of genes (Schaffer et al., 2000). The probes are usually labeled with fluorescent nucleotide derivatives and the arrays scanned by confocal microscopy. mRNA species present at very low levels (a few copies per cell) can be detected and the dynamic range over which expression can be monitored is several orders of magnitude.

Microarrays can also be produced using oligonucleotides deposited by a photolithographic process (Fodor et al., 1993; Lipshutz et al., 1999) and such Arabidopsis arrays representing about 8,000 genes are commercially available from Affymetrix (Table I). Although cDNA-based and oligonucleotide-based arrays are well-proven technologies that provide reliable data on expression patterns of thousands of genes, each type of array has distinct advantages. For example, oligonucleotide arrays can in many cases more easily distinguish between closely related members of gene families. The arrays offered by AFGC, which are based on spotting PCR products of cDNA or genomic DNA, also have several advantages, which include lower costs, the ability to provide data on different Arabidopsis ecotypes or closely related species such as *Brassica*, and the ability to cohybridize probes from two or more mRNA samples simultaneously on the same array.

WORLD-WIDE ARABIDOPSIS MICROARRAY SERVICES

An increasing number of laboratories have generated Arabidopsis-specific microarrays (for review see, Richmond and Somerville, 2000). The first papers reporting the use of Arabidopsis microarrays applied to scientific problems have recently appeared (for example, Schena et al. 1995; Ruan et al., 1998; Kehoe et al., 1999; Girke et al., 2000; Reymond et al. 2000; Wang et al., 2000). However, producing well-designed, comprehensive microarrays and the controls needed to validate their output requires substantial investments in microarray technology that are not feasible for many laboratories or small institutions. Therefore, centralized service facilities can allow many more scientists to implement this technology using a con-

[1] The Arabidopsis Functional Genomics Consortium is funded by the National Science Foundation (grant no. DBI–9872638 to Pam Green).

* Corresponding author; e-mail wismanp@pilot.msu.edu; fax 517–353–9168.

Table I. *Some URLs with microarray resources for Arabidopsis*

Organization	Web Site
AFGC	http://afgc.stanford.edu
GARNet	http://www.york.ac.uk/res/garnet/may.htm
Stanford Microarray Database	http://genome-www4.Stanford.EDU/MicroArray/SMD/
Developing Seeds	http://www.bpp.msu.edu/Seed/SeedArray.htm
The Arabidopsis Information Resource	http://www.arabidopsis.org/
The Institute for Genome Research	http://www.tigr.org/tdb/at
Monsanto	http://www.monsanto.com/
Affymetrix	http://www.affymetrix.com/
Synteny/InCyte	http://gem.incyte.com/gem/index.shtml

sistent array design. To date there are at least two publicly funded programs that provide world-wide microarray services; the AFGC in the USA and GARNet in the UK. The AFGC, funded by the National Science Foundation, began their services in early 2000, whereas GARNet's is scheduled to start from the first half of 2001. Both microarray services will be available to the community with no proprietary rights with respect to intellectual property. Two investigators coordinate the microarray service of the AFGC: Dr. Ellen Wisman at Michigan State University and Dr. Shauna Somerville who leads the AFGC Microarray group at the Carnegie Institute of Washington at Stanford University. The group of Dr. Mike Cherry at Stanford University provides bioinformatic support for data analysis and tracking. Alternative Arabidopsis array services are available from Monsanto and Incyte (Table I).

STANFORD MICROARRAY DATABASE

Microarrays can, in one experiment, provide information on expression patterns for thousands of genes and this information often can supply new insights about already described genes, as well as useful clues about the function of uncharacterized genes (Fig. 1). Knowing in which tissue(s) a gene is expressed and how expression changes under different environmental conditions or is altered in mutants or transgenics can often build a "modus operandi," which leads toward testable hypotheses about the gene's function. Examining even a single microarray from, for example, a pathogen-infected plant can provide many previously unattainable and unexpected insights. However, even greater power of microarray technology emerges when a large number of experiments can be compared. Increasingly sophisticated bioinformatics tools are being developed to "cluster" or organize millions of datapoints and these tools sometimes reveal completely unexpected relationships that can shift our thinking in exciting new directions. To fully realize these opportunities requires that many experiments be collected into a common database. In this regard, all experimental data generated by AFGC are deposited in the Stanford Microarray Database (SMD; http://genome www4. stanford.edu/MicroArray/SMD/). The expression profiles will be accompanied by a thorough description of the experimental conditions for each RNA sample, which provides other users of the SMD a meaningful interpretation of the observed changes in gene expression levels. A user of the SMD can now search for the "expression history" of their favorite gene and receive microarray expression data for over 100 different experiments. Just as the huge sequence databases have provided fertile territory for in silico discovery, the SMD offers users throughout the world an opportunity to mine this large data set on gene expression. Given the capacity of AFGC of several hundred slides per year, its operation should yield a wealth of gene expression data available to all plant scientists. This exchange of information in the scientific community will generate an important and extensive resource helpful in understanding the role of genes in their respective biological pathways.

DESIGN OF THE AFGC MICROARRAY

Microarrays currently offered by AFGC are produced by spotting PCR products of expressed sequence tag (EST) clones onto glass slides. The Year 2000 array produced by AFGC contains DNAs representing 11,000 clones from the Michigan State University collection of ESTs (Newman et al., 1994). The DNA preparations used as PCR templates were from the same tubes used for the original EST sequencing performed by Dr. Tom Newman. This unique resource, developed with National Science Foundation funds and kindly made available to the AFGC by Dr. Tom Newman, has allowed the AFGC to avoid handling errors often associated with large gene sets. For The Year 2001 array a new collection of ESTs has been selected that represent a significant increase in information content and reduction of redundancy. The new set consists of many ESTs from the original collection, a few thousand new EST clones, and thousands of gene fragments amplified from genomic DNA with gene-specific primers. The gene-specific fragments will be produced in collaboration with the Nottingham Arabidopsis Stock Centre (NASC) in the UK and Australia's Commonwealth Scientific and Industrial Research Organization. It is anticipated that the 2001 arrays will represent at least 12,000 Arabidopsis genes.

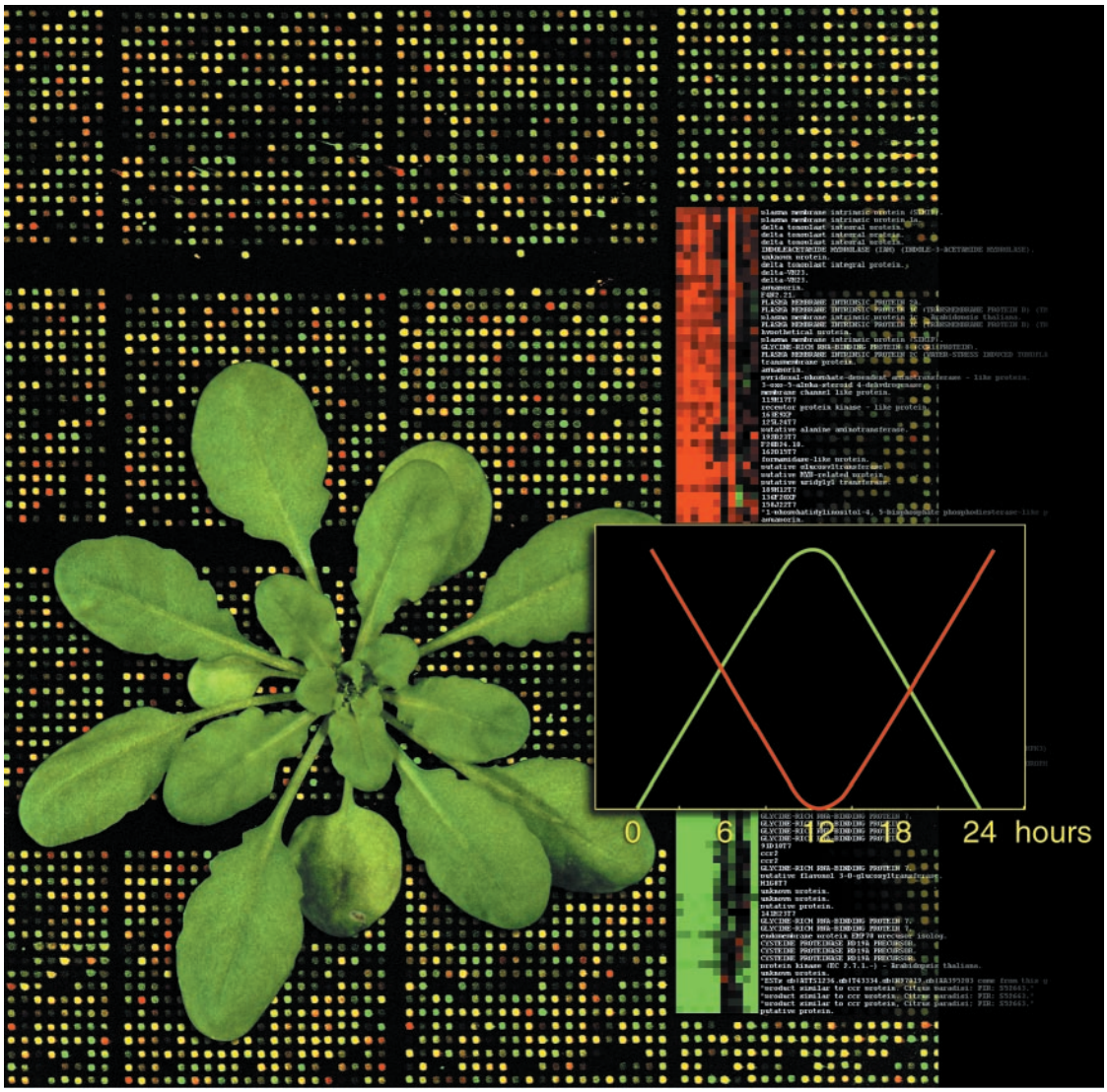

Figure 1. Arabidopsis microarray identifies novel circadian genes. Image design by Marlene Cameron.

The AFGC service accepts RNA samples from the users and performs labeling, hybridization, and data collection. Protocols for preparing samples, a list of clones on the arrays, and the guidelines for application are available from the Internet at http://afgc.n. The microarray service facility will have approximately 400 arrays (slides) available in year 2001. Most experiments include two arrays to be used for biological or technical replicates of the same experiment. Academic laboratories may request up to six experiments per year. Users of the service can choose for immediate data release or for a 3-month delay. If microarray services are oversubscribed, a science review committee independent of the AFGC will rank the applications. Although the National Science Foundation provides the major support for establishing AFGC arrays, the price of $1,200 per experiment ($300 per mRNA sample hybridized) recovers the costs for labor and materials.

Members of the AFGC microarray team also are performing surveys of gene expression patterns in different developmental stages and different genotypes of Arabidopsis and much of this data is already publicly available via SMD. The tissue survey including roots, leaves, siliques, flowers, and stems has now provided a base data set for the community and will enhance other expression studies by allowing researchers to identify tissue specificity of their differentially expressed genes. In addition to the tissue survey, AFGC will record the response to a number of biological conditions such as circadian rhythms (Fig. 1).

The microarray service facility of GARNet is part of the Biotechnology and Biological Sciences Research Counsel Investigating Gene Function initiative and aims to create national facilities of genomics resources for Arabidopsis and other plant research. The GARNet consortium includes a public array service

with the intention to generate an array containing the complete transcriptome of Arabidopsis by 2003. This service is the responsibility of the NASC in collaboration with the John Innes Center, Norwich and Horticulture Research International (HRI) Wellesbourne. NASC's goal is to provide arrays in micro- (slide), and macro- (filter) format to the community and to provide a complete microarray service (labeling, hybridization, analysis, and data mining) from the first half of 2001. GARNet policy determines that emergent data will be made public within 3 to 6 months. It will integrate prevailing public standards on array content and data distribution, particularly those of the AFGC, The Arabidopsis Information Resource, the National Center for Genome Resources, and the European Bioinformatics Institute.

CONCLUSION

In conclusion, AFGC and GARNet will provide the ability for hundreds of laboratories to carry out microarray experiments using carefully designed and well-controlled microarrays processed under consistent conditions. Collection of millions of datapoints from these experiments in common public databases will allow each researcher and the world-wide community to develop a much richer understanding of gene expression patterns and how these patterns change.

Received September 7, 2000; accepted September 25, 2000.

LITERATURE CITED

Fodor SP, Rava RP, Huang XC, Pease AC, Holmes CP, Adams CL (1993) Multiplexed biochemical assays with biological chips. Nature **364:** 555–556

Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J (2000) Microarray analysis of developing Arabidopsis seeds. Plant Physiol **124:** 1570–1581

Kehoe DM, Villand P, Somerville S (1999). DNA microarrays for studies of higher plants and other photosynthetic organisms. Trends Plant Sci **4:** 38–41

Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ (1999) High density synthetic oligonucleotide arrays. Nat Genet **21:** 20–24

Newman T, Debruijn FJ, Green P, Keegstra K, Kende H, McIntosh L, Ohlrogge J, Raikhel N, Somerville S, Thomashaw M, Retzel LE, Somerville C (1994) Genes galore: a summary of methods for accessing results from partial sequencing of anonymous Arabidopsis cDNA clones. Plant Physiol **106:** 1241–1255

Reymond P, Weber H, Damond M, Farmer EE (2000) Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis. Plant Cell **12:** 707–720

Richmond T, Somerville S (2000) Chasing the dream plant EST microarrays. Curr Opin Plant Biol **3:** 108–116

Ruan Y, Gilmore J, Conner T (1998) Towards *Arabidopsis* genome analysis: monitoring expression profiles of 1,400 genes using cDNA microarrays. Plant J **15:** 821–833.

Schaffer R, Landgraf J, Perez-Amador M, Wisman E (2000). Monitoring genome wide expression in plants. Curr Opin Biotechnol **11:** 162–167

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270:** 467–470

Wang R, Guegler K, LaBrie ST, Crawford NM (2000) Genomic analysis of a nutrient response in *Arabidopsis* reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. Plant Cell **12:** 1491–1510