The Plant Cell, Vol. 12, 613, May 2000, www.plantcell.org © 2000 American Society of Plant Physiologists

IN THIS ISSUE

Microarrays: Determining the Balance of Cellular Transcription

The advent of high-throughput DNA sequencing and the ability to investigate global patterns of gene expression are revolutionizing biology, and plant biology is no exception. These new technologies allow biologists to systematically evaluate gene expression patterns on a genomewide scale. It is now possible to examine the expression of many genes at once over multiple developmental stages and in response to environmental cues. Such experiments enable us not only to better assess the behavior of genes previously implicated in a given process, but also to identify novel, similarly regulated genes. In addition, links between physiological processes can be established by looking for common gene expression. The versatility of these technologies can be breathtaking, as is exemplified by two articles presented in this issue of THE PLANT CELL. **On pages 707–719, Reymond et al.** examine a wounding response of Arabidopsis, and **on pages 647–661, Aharoni et al.** explore the developmental process of fruit ripening in strawberry.

Anyone who has bitten into a not-quite-ripe strawberry realizes that there are a number of profound differences between the ripe and unripe fruit. The obvious changes in color, aroma, and weight that occur during ripening are underlain by significant biochemical events. Among these are the accumulation of sugars and the synthesis of aromatic compounds that contribute to the flavor and scent of the ripe fruit. The commercial as well as scientific interests in identifying these differences are considerable. Unfortunately, the high levels of phenolic compounds and polysaccharides in fruit tend to interfere with enzyme isolation and have impeded the biochemical characterization of the ripening processes. Currently, one of the most tractable approaches to these complex processes is to examine changes in gene expression that occur during fruit development.

A number of technologies have been used in the past decade to study gene expression in plant development. Approaches such as subtractive hybridization of cDNA libraries and differential display have been used with some success (Kuno et al., 2000; Manning, 1998; Wilkinson et al., 1995); however, these techniques are laborious and are more suited to the analysis of individual genes. In contrast, methods for high-throughput analysis of broader changes in steady state gene expression have recently been developed. One method is SAGE (for serial analysis of gene expression), which has been used to examine gene expression in a variety of situations, including human cancer cells and yeast cell-cycle regulation (Polyak et al., 1997; Velculescu et al., 1997). One advantage of SAGE is that it gives a quantitative estimate of gene expression levels. However, because this technique relies upon the identification of genes based on short sequence tags (ten to 14 base pairs), it is not appropriate for organisms for which large databases of expressed sequence tags are not available. In addition, rare transcripts are likely to be missed by the SAGE technique.

A popular new approach for the examination of global changes in gene expression is the use of high-density DNA microarrays, which in many ways parallels RNA gel blotting techniques. Specifically, gene sequences are anchored to a solid support, RNA is labeled and applied to the microarray, and individual hybridization signals are used to infer cellular levels of gene expression. Many permutations of this technique exist and can involve either DNA–DNA or DNA–RNA hybridization.

Researchers at Stanford University have developed a microarray method in which a systematized collection of cDNAs (often derived from collections of expressed sequence tags) is deposited on a glass slide by a robot using a high-speed printing process (Brown and Botstein, 1999). cDNAs generated from two experimental cell populations, representing two different experimental conditions, are then labeled with two different fluorescent tags. These tagged fragments are mixed and hybridized to the microarray in a competitive fashion, and a high-resolution scanner then quantifies the amount of both fluorescent labels bound to each cDNA that had been deposited on the glass slide. The comparison of the two fluorescence signals, allowing for assessment of the relative levels of particular mRNAs from the two cell populations, is crucial to meaningful conclusions because spotting reproducible amounts of fluorescently labeled cDNA is prohibitively difficult.

Aharoni et al. have used such an approach to examine changes in gene expression that occur during strawberry fruit ripening, a complex developmental process. The authors isolated 1701 cDNA clones from strawberry and 480 clones from petunia (as a control) and arrayed them in duplicate on glass slides, using a robot to spot out the cDNAs (i.e., the "probes"). They then isolated mRNA from strawberry fruits at various stages of development, as characterized by fruit color, and produced single-stranded cDNA "target" sequences by reverse transcription. cDNA product ("target") generated from one stage of fruit development (i.e., the "test" state) and labeled with the first dye, and cDNA produced from

IN THIS ISSUE

fruit at another stage of development (i.e., the "reference" state) and labeled with a second dye, were then mixed and applied to the microarrayed probe sequences. The competitive hybridization of the two fluorescently labeled target cDNAs onto the microarray revealed the relative levels of transcript corresponding to each probe sequence in the test and reference cell populations. Importantly, this scheme provides an internal control for each experiment in that the relative abundance of gene expression is always based upon a direct comparison of cDNA samples.

The analysis of microarray data presents its own challenges. Paramount is the decision as to what constitutes a significant difference in gene expression between test and reference populations. Aharoni et al. found that 401 of the 1701 randomly chosen strawberry probes were differentially expressed among green, white, and red fruit. They arrived at this number by using statistical analysis-of-variance models to establish minimum thresholds for significant differences in expression. Their analysis assumes the same statistical variability for each gene represented on the array, and therefore disregards the likelihood that rare messages will in fact show greater variability than highly abundant transcripts. Of course, the best way to assess the variability of individual messages is with multiple independent replicates of the same experiment; however, such replicates can be prohibitively costly. (Alternative means of statistical analysis, less sensitive to gene expression levels, have therefore been suggested [Wittes and Friedman, 1999] and might be better suited for establishing the significance of observed changes in gene expression with a limited number of replicate experiments.)

Given the above statistical caveat, it is thus important to confirm the differential expression of genes predicted by microarray analysis. Aharoni et al. therefore use RNA gel blot analysis to confirm that two of the genes predicted to be more abundant in ripe fruit are indeed upregulated in red berries relative to green or white berries. One of these two genes is a novel strawberry alcohol acyltransferase gene (*SAAT*) that may well play an important role in the synthesis of chemicals that give ripe strawberries their wonderful smell and taste. Alcohol acyltransferases (AATs) produce the esters that contribute to the distinctive tastes of many fruits. The microarray data of Aharoni et al. that suggested their novel *SAAT* gene to be much more highly expressed in ripe red fruit than in green or white fruit was confirmed by RNA gel blot analysis.

The predicted SAAT protein exhibits little amino acid sequence similarity to other AATs, but key residues are conserved. The authors expressed the *SAAT* gene in *E. coli* and confirmed that the recombinant protein does indeed possess AAT activity. The substrate specificity of the SAAT protein is largely consistent with AAT activity previously isolated from strawberry and can indeed produce the predominant esters found in ripe strawberry fruit. Significantly, SAAT exhibits relatively little sequence similarity to other AAT enzymes, and so database searches alone might not have identified the functional role of the *SAAT* gene product. The authors' microarray data are thus crucial to establishing the role of SAAT in creating the compounds that make strawberries taste like strawberries.

cDNA microarray technology has a number of strengths that are well illustrated by Aharoni et al. For example, prior knowledge of the genome under study was not necessary in that clones isolated directly from cDNA libraries were arranged into the microarray. An additional advantage to microarray technology is that custom chips can be easily fabricated within a laboratory once a printing robot has been acquired. Such custom arrays can be tailored for organisms, such as strawberry, that are not broadly used as model systems. In addition, genes can be chosen to reflect the interests of the experimenter. Once the system is established in the laboratory, moreover, each new chip printed is only moderately expensive. (These qualities are not currently shared by the competing oligonucleotide-based microarray technology. The oligonucleotide chips are more expensive and their development is dependent upon having a database of gene sequences.)

On the other hand, the work of Aharoni et al. also highlights some of the limitations of the cDNA spotting technology. For example, 30% of the genes on the probe array prove to be redundant, reducing the identified number of differentially expressed genes from 401 to 239. Such redundancy can occur even when ESTs, rather than cDNAs, are used to create the microarray, and the use of large DNA fragments as probes can be similarly problematic due to cross-hybridization of closely related family members to the same probe. Aharoni et al. did in fact see cross-hybridization between closely related petunia and strawberry genes. Such problems might be circumvented with synthetic oligonucleotide-based microarrays, which can discriminate between nucleotide sequences that are up to 93% identical. In addition, oligo-based arrays can provide the added experimental rigor of probes that hybridize to multiple regions of the same RNA (Lipshutz et al., 1999).

In the second paper of this issue that illustrates the power of cDNA microarray technology, Reymond et al. look at genes in Arabidopsis induced by wounding. Results from wild-type and mutant plants suggest that jasmonate signaling is required for the expression of almost half of the genes normally induced by wounding. Surprisingly, the authors find essentially no difference between wild-type and ethylene-insensitive plants. These findings illustrate how microarrays, in combination with defined mutants, can be used to infer

which signaling pathways are associated with an environmental response.

Like the article from Aharoni et al., the work of Reymond et al. demonstrates the utility of DNA microarrays that cover even a small portion of a genome. Nevertheless, microarrays that more extensively represent the genome of a given organism are emerging with powerful capabilities. When combined with genome sequence databases, one such capability is the identification of common promoter regulatory elements shared by coregulated genes, as has been achieved for yeast genes under cell-cycle control (Cho et al., 1998; Spellman et al., 1998).

Despite the powerful uses of DNA microarrays, the assumption that mRNA levels reflect protein abundance has been questioned in certain experimental settings (Gygi et al., 1999). New techniques are therefore being developed to move beyond the analysis of message levels. For example, mRNA can be separated into translationally active and inactive fractions and then subjected to microarray analysis (Zong et al., 1999). In addition, advances in proteomics may eventually make it possible to determine not only the levels of proteins present in a cell, but also their posttranslational modifications, intracellular locations, and intermolecular interactions. We can now only imagine the fundamental insights into cellular regulation that we will eventually gain from such knowledge.

Stacey L. Harmer

Steve A. Kay
The Scripps Research Institute
10550 North Torrey Pines Road
La Jolla, CA 92037
staceyh@scripps.edu
stevek@scripps.edu

REFERENCES

Aharoni, A., et al. (2000). Identification of the *SAAT* gene involved in strawberry flavor biogenesis by using DNA microarrays. Plant Cell **12**, 647–661.

Brown, P.O., and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. Nat. Genet. **21**, 33–37.

Cho, R.J., Campbell, M.J., Winzeler, E.A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T.G., Gabrielian, A.E., Landsman, D., Lockhart, D.J., and Davis, R.W. (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. Mol. Cell **2**, 65–73.

Gygi, S.P., Rochon, Y., Franza, B.R., and Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. Mol. Cell Biol. **19**, 1720–1730.

Kuno, N., Muramatsu, T., Hamazato, F., and Furuya, M. (2000). Identification by large-scale screening of phytochrome-regulated genes in etiolated seedlings of Arabidopsis using a fluorescent differential display technique. Plant Physiol. **122**, 15–24.

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. (1999). High density synthetic oligonucleotide arrays. Nat. Genet. **21**, 20–24.

Manning, K. (1998). Isolation of a set of ripening-related genes from strawberry: Their identification and possible relationship to fruit quality traits. Planta **205**, 622–631.

Polyak, S.W., Forsberg, G., Forbes, B.E., McNeil, K.A., Aplin, S.E., and Wallace, J.C. (1997). Introduction of spacer peptides N-terminal to a cleavage recognition motif in recombinant fusion proteins can improve site-specific cleavage. Protein Eng. **10**, 615–619.

Reymond, P., Weber, H., Damond, M., and Farmer, E.E. (2000). Differential gene expression in respone to mechanical wounding and insect feeding in Arabidopsis. Plant Cell **12**, 707–719.

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol. Biol. Cell **9**, 3273–3297.

Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Jr., Hieter, P., Vogelstein, B., and Kinzler, K.W. (1997). Characterization of the yeast transcriptome. Cell **88**, 243–251.

Wilkinson, J.Q., Lanahan, M.B., Conner, T.W., and Klee, H.J. (1995). Identification of mRNAs with enhanced expression in ripening strawberry fruit using polymerase chain reaction differential display. Plant Mol. Biol. **27**, 1097–1108.

Wittes, J., and Friedman, H.P. (1999). Searching for evidence of altered gene expression: A comment on statistical analysis of microarray data [editorial; comment]. J. Natl. Cancer Inst. **91**, 400–401.

Zong, Q., Schummer, M., Hood, L., and Morris, D.R. (1999). Messenger RNA translation state: The second dimension of high-throughput expression screening. Proc. Natl. Acad. Sci. USA **96**, 10632–10636.