

Gene chips: Array of hope for understanding gene regulation

Mark Johnston

High density arrays of DNA fragments on a solid surface allow the expression of thousands of genes to be assessed in a single experiment. The development of this 'gene chip' technique heralds a new era of studies that promises to provide an integrated view of the expression of all genes of an organism.

Address: Department of Genetics Box 8232, Washington University Medical School, 4566 Scott Avenue, St. Louis, Missouri 63110, USA.

Current Biology 1998, 8:R171–R174
<http://biomednet.com/elecref/09609822008R0171>

© Current Biology Ltd ISSN 0960-9822

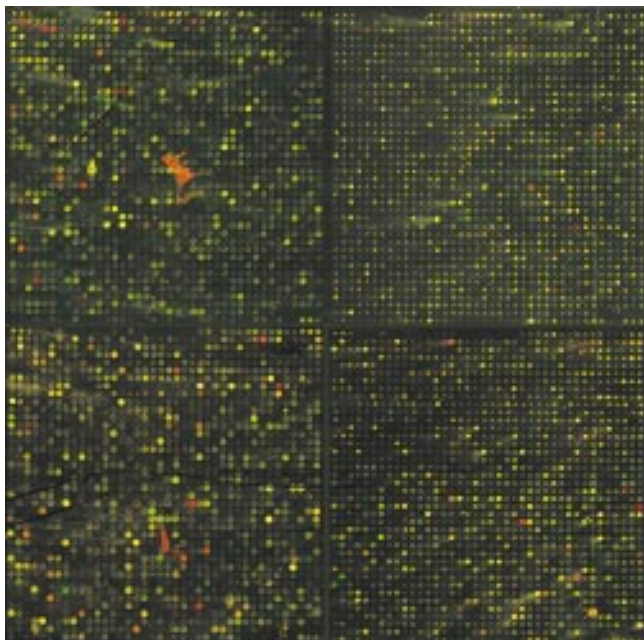
The realization almost a half century ago that genes that are co-regulated often encode proteins of related function fueled a remarkable period of discovery about mechanisms of gene regulation. The paradigms provided by this work form a cornerstone of molecular biology [1]. These paradigms were verified and extended by painstaking dissection of regulatory mechanisms, operon by operon (for eukaryotes, regulon by regulon). While 'global' regulatory mechanisms that act upon large sets of genes were recognized early on [2], their analysis has been, for the most part, limited to a small number of representative genes subject to global control. Our present knowledge of how genes are regulated thus stems from analysis of a limited number of genes. A revolutionary new technology for measuring expression of all genes of an organism in a single experiment has now been devised [3–5]. This may herald a new era of investigation of gene regulation that promises to provide a much deeper understanding of how cells coordinate expression of thousands of genes.

This advance was made possible by the development of technology that allows DNA fragments to be arrayed at high density on a solid support for use in hybridization experiments [6,7]. Thousands of DNA fragments can be arrayed on a surface no larger than a fingernail and used to probe the mRNA content of cells. Thus, whole genomes can be assessed for their pattern of gene expression, enabling us, for the first time, to view gene regulation in the context of all the complex networks of pathways that operate in cells. We can now identify all the genes of an organism that change expression under a given condition, and hope to make sense of the cell's response to that condition. Such information can provide key clues to the function of individual proteins. Moreover, the ability to acquire data of this kind is a big step toward achieving the ultimate goal of molecular biology: a complete understanding of cellular function.

Two different methods for arraying large numbers of DNA molecules in a very small space have been developed. In one, cDNA-sized fragments — usually produced by the polymerase chain reaction (PCR) — are spotted onto polylysine-coated glass slides [6]. In the other, short (~25 nucleotide) oligonucleotides are synthesized on a glass surface [7]. The arrays produced by both methods have been called 'chips', but this moniker fits the oligonucleotide arrays better, because they are made using photolithographic masks similar to those used for fabricating computer chips. Both methods pack thousands of DNA fragments into a very small area: the current oligonucleotide chips display all 6000 yeast genes on four 1.28 × 1.28 cm chips; the DNA fragment microarrays fit the same information onto a single 1.8 × 1.8 cm glass slide (see Figure 1). On the oligonucleotide chips, each gene must be represented by several (typically 20) different oligonucleotides, because of the differences in hybridization properties and reduced hybridization specificity inherent in such short probes. In addition, each oligonucleotide on the chips has a partner adjacent to it that differs at just one central base, which serves as an internal control for hybridization specificity. Each gene thus encompasses about 40 'features' — a feature being an area of the glass surface occupied by DNA molecules of one sequence — on an oligonucleotide chip, whereas it takes up only one feature on a DNA fragment microarray.

The DNA arrays are used to interrogate complex mixtures of nucleic acids, and thus are similar to the 'dot blots' that have been in use for a long time [8,9]. They differ from dot blots in the nature of the labelled species that serves as the probe — in dot blots, the complex mixture of mRNAs is fixed to the solid surface and probed with a single labelled DNA fragment; in the DNA microarrays, individual unlabelled DNA fragments are fixed on the solid support and probed with a complex mixture of labelled cDNAs or mRNAs. The major advance of the arrays over the older technology is a significant increase in sensitivity, primarily as a result of two factors. Because the labeled probe is usually the limiting component in nucleic acid hybridization, probably the more important factor is the small area occupied by the arrays, which significantly reduces the volume of the hybridization solution — from milliliters to microliters — and thereby greatly increases the concentration of the probe. Because of the small area they occupy, sophisticated lasers and sensitive detection systems are required to measure the hybridization signals. The second factor is that the glass surface of an array generates a smaller background hybridization signal than the porous membranes used for dot blots. Both kinds of

Figure 1



A DNA fragment array of all ~ 6000 yeast genes probed with labeled cDNA made from galactose- and glucose-grown cells. Each spot (element) on the array contains a cDNA-sized DNA fragment representing one yeast coding sequence. mRNA from galactose-grown cells was converted to red-labeled cDNA (using dUTP labeled with the fluorescent dye Cy3); mRNA from glucose-grown cells was converted to green-labeled cDNA (with the dye Cy5). These two preparations of labeled cDNA were mixed and used to probe the array. Red spots bind only galactose-grown cDNA, and thus represent genes expressed only in galactose-grown cells; green spots bind only cDNA from glucose-grown cells, and therefore represent genes expressed only in glucose-grown cells. Spots containing genes expressed under both conditions hybridize to both cDNAs, and thus appear yellow. The intensity of the color of each spot (from red to green) reveals the relative expression level of genes under the two conditions. (Figure courtesy of Joe DeRisi, Vishy Iyer, and Pat Brown; for more of these images, see [17].)

microarray thus permit very sensitive detection of gene expression: currently, an mRNA present at a level less than one molecule in 100,000 can be detected, equivalent to a transcript present at only one copy per 20 yeast cells!

The DNA fragment microarrays can be produced by anybody with the ability and modest means required to assemble the equipment to print the arrays [10]. Production of the DNA fragments to be arrayed does, however, require a large number of oligonucleotides for the PCR, which can be prohibitively expensive, and generation of the PCR products is labor intensive. (For yeast, much of this work has already been done [11].) A limitation of the oligonucleotide chips is that knowledge of the DNA sequences to be studied is necessary to produce them, whereas random cDNA clones can be used in the DNA fragment microarrays. Also, dependence on commercial

sources for the oligonucleotide chips may present limitations of availability and affordability. Both methods require fairly sophisticated microscopy and software for detecting, measuring and identifying hybridization signals from the arrays. This technology currently seems out of the reach of the average lab, but commercial services are sprouting to provide the microarrays and equipment necessary to make this technology widely accessible. In the meantime, the whole genome dot blots that have recently become available, at least for yeast, may fulfil the needs of most labs that want to perform these kinds of experiments [12].

The utility of the two kinds of microarray for measuring expression of a large number of genes was established previously [6,13–15], but was spectacularly demonstrated recently by two groups who used them to measure expression of all 6000 genes of the bakers' yeast, *Saccharomyces cerevisiae*, grown under a few different conditions [3,4]. Wodicka *et al.* [4] compared gene expression in yeast cells grown on rich and minimal media. They isolated polyA⁺ RNA from cells grown under the two conditions, converted it into cDNA flanked by a promoter for T7 RNA polymerase, and copied it into antisense, biotin-labeled RNA by transcription *in vitro*. This final step amplifies the mRNA probe, apparently without introducing significant bias. Labeled RNA made in this way from cells grown under the two conditions was used to probe the oligonucleotide chip, and the bound RNA was detected and quantified using streptavidin conjugated to a fluorescent dye, yielding highly reproducible results. More than 87% of yeast mRNAs were detected, with a dynamic range of about three orders of magnitude.

Similar results were obtained by DeRisi *et al.* [3], who used the DNA fragment microarrays to measure gene expression in yeast cells as they run out of glucose. They isolated polyA⁺ RNA from a culture of cells at several different times after inoculation into glucose media, fluorescently labeled it by reverse transcription, and used the labelled product to probe DNA fragment microarrays. The two types of array seem roughly comparable in their sensitivity, range and reproducibility. The oligonucleotide chips may be better at measuring relative expression differences, because they easily revealed more than 50-fold differences in expression, whereas the maximum expression difference measured with the DNA fragment microarrays was 20-fold (although it is difficult to compare the results of the two experiments, as they employed very different growth conditions).

The results presented by DeRisi *et al.* [3] and Wodicka *et al.* [4] mostly serve to validate the experimental approach, but in a very satisfying way, as many of the changes in gene expression that were observed were expected. DeRisi *et al.* [3], for example, rediscovered the fact that, when yeast cells run out of glucose, the expression of

genes for oxidative metabolism and gluconeogenesis increases, and the expression of genes for fermentation and protein synthesis decreases. That these results conform almost perfectly to what is known about regulation of these well-studied genes lends great confidence to the technique.

Similarly, many of the genes expected to have higher levels of expression in cells grown on minimal media than on rich media — such as those involved in nitrogen acquisition or amino acid synthesis — were identified with the oligonucleotide chips, as were many genes that have the converse expression pattern, such as those involved in amino-acid transport. The technique is not perfect, however, as the DNA fragment microarrays missed several genes whose expression is known to be regulated by glucose — for example, *HXT1*, which is induced about 300-fold by glucose [16], and *GAL4*, which is about 75% repressed by glucose [17]. (These omissions could be easily uncovered, because all the results are publicly available in a terrific, searchable database [18].) Nevertheless, the microarrays work better than most of us imagined they would, and provide a wonderful tool that greatly expands our horizons.

What have these experiments taught us about cellular function? They revealed that almost 90% of yeast genes are expressed, most at very low levels (69% with one or fewer mRNAs per cell) [4], but this has long been known from the classic work of Hereford and Rosbash [19]. Similarly, many of the genes DeRisi *et al.* [3] found to be regulated by glucose have long been known to be subject to such regulation. A substantial number of genes, however, were found for the first time to be regulated in these two studies, and nothing is known about a significant proportion of these. The regulatory patterns of these proteins thus provide a first clue to their function. These results also allow genes to be grouped by their expression pattern, as was done insightfully by DeRisi *et al.* [3]. The function of at least some of the genes in a group is usually known, allowing inferences to be made about the possible function of the other genes in the same group. Clearly, this technology will speed the pace of discovery of protein and cellular function.

One of the most promising applications of DNA microarrays is the identification of all the genes whose expression changes when a gene is inactivated. This is a boon for those interested in transcription factors, as this information should help reveal their role in cellular physiology, and might even speak to their mechanism of action. DeRisi *et al.* [3] identified all yeast genes whose expression changes when the Tup1 transcription factor is inactivated by mutation. The expression of many genes increased significantly as a result of deletion of *TUP1*, which would probably lead one to conclude correctly that

Tup1 is a general repressor. Interestingly, expression of a few genes decreased significantly in a *tup1* mutant, suggesting that Tup1 may also activate transcription in certain cases. In a separate set of experiments, genes whose expression changes when the Yap1 transcription factor is overexpressed were identified. This revealed a set of genes whose expression increased, indicating that Yap1 is a transcriptional activator. Again, expression of a few genes decreased significantly upon Yap1 overexpression, suggesting that Yap1 may also be a repressor.

A major problem with interpretation of these results is the difficulty in ascribing them to direct action of the transcription factor that is inactivated. In fact, it seems a good bet that indirect effects account for the unexpected responses to Tup1 absence and Yap1 overexpression. Nevertheless, the wealth of data provided by the microarrays allows the formulation of hypotheses that can be tested with other, more conventional experiments. The practical uses of this technology to identify candidate compounds for drug development are obvious. Furthermore, the microarrays are sure soon to be in wide clinical use, where they will undoubtedly aid in disease diagnosis and treatment.

Now that the DNA microarrays are clearly working well for the analysis of gene expression, the major challenge is to handle and interpret the massive amounts of data that will quickly accrue. Just from the two reports of DeRisi *et al.* [3] and Wodicka *et al.* [4], there is a rich vein of information waiting to be mined that is sure to grow as this technology becomes widely available. But the problem we are faced with is a pleasant one: we are not limited by the amount of data we can collect, but by our ability to interpret it. If we are able to do so successfully, great insight into cellular function is promised. It is unlikely to change our paradigms, but it will take us one large step closer to the goal of a complete understanding of how cells work.

Acknowledgements

I thank Stan Fields for insightful comments on the manuscript, and Joe DeRisi, Vishy Iyer and Pat Brown for generously providing the image of the yeast genome DNA fragment array.

References

1. Jacob F, Monod J: Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 1961, 3:318-356.
2. Magasanik B: Catabolite repression. *Cold Spring Harbor Symp Quant Biol* 1961, 26:249-256.
3. DeRisi JL, Iyer VR, Brown PO: Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 1997, 278:680-686.
4. Wodicka L, Dong H, Mittmann M, Ho M-H, Lockhart DJ: Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat Genet* 1997, 15:1359-1367.
5. Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO, Davis RW: Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc Natl Acad Sci USA* 1997, 94:13057-13062.
6. Schena M, Shalon D, Davis RW, Brown PO: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995, 270:467-470.

7. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D: **Light-directed, spatially addressable parallel chemical synthesis.** *Science* 1991, **251**:767-773.
8. Kafatos FC, Jones CW, Efstratiadis A: **Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure.** *Nucleic Acids Res* 1979, **7**:1541-1552.
9. Chuang SE, Daniels DL, Blattner FR: **Global regulation of gene expression in *Escherichia coli*.** *J Bacteriol* 1993, **175**:2026-2036.
10. **The microarray home page.**
<http://cmgm.stanford.edu/pbrown/array.html>
11. Hudson JR Jr, Dawson EP, Rushing KL, Jackson CH, Lockshon D, Conover D, Lanciault C, Harris JR, Simmons SJ, Rothstein R, Fields S: **The complete set of predicted genes from *Saccharomyces cerevisiae* in a readily usable form.** *Genome Res* 1997, **7**:1169-1173.
12. **Gene filters.**
http://www.resgen.com/online_catalog/cat_view_group.html?group=151009&item_id_no=-1
13. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL: **Expression monitoring by hybridization to high-density oligonucleotide arrays.** *Nat Biotechnol* 1996, **14**:1675-1680.
14. Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW: **Parallel human genome analysis: microarray-based expression monitoring of 1000 genes.** *Proc Natl Acad Sci USA* 1996, **93**:10614-10619.
15. DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM: **Use of a cDNA microarray to analyse gene expression patterns in human cancer.** *Nat Genet* 1996, **14**:457-460.
16. Ozcan S, Johnston M: **Three different regulatory mechanisms enable yeast hexose transporter (*HXT*) genes to be induced by different levels of glucose.** *Mol Cell Biol* 1995, **15**:1564-1572.
17. Griggs DW, Johnston M: **Regulated expression of the GAL4 activator gene in yeast provides a sensitive genetic switch for glucose repression.** *Proc Natl Acad Sci USA* 1991, **88**:8597-8601.
18. **Exploring the metabolic and genetic control of gene expression on a genomic scale.**
<http://cmgm.stanford.edu/pbrown/explore/index.html>
19. Hereford LM, Rosbash M: **Number and distribution of polyadenylated RNA sequences in yeast.** *Cell* 1977, **10**:453-462.

If you found this dispatch interesting, you might also want to read the February 1998 issue of

Current Opinion in Biotechnology

which includes the following reviews, edited by Michael J Gait and Stephen H Hughes, on **Analytical biotechnology**:

Advances in DNA diagnostics

Joel H Graber, Maryanne J O'Donnell, Casandra L Smith and Charles R Cantor

Advances in fluorescent *in situ* hybridisation

Rosemary Ekong and Jonathan Wolfe

Applications of mass spectroscopy to the characterization of oligonucleotides and nucleic acids

Pamela F Crain and James A McCloskey

The polymerase chain reaction: from functional genomics to high school practical classes

Graham R Taylor and P Robinson

Advances in quantitative PCR technology: 5' nuclease assays

Yolanda S Lie and Christos J Petropoulos

New uses for old DNA

Alan Cooper and Robert Wayne

Recent developments in biological sequence databases

Patricia G Baker and Andy Brass

Modern methods for probing RNA structure

Jørgen Kjems and Jan Egebjerg

RNA as a drug target: chemical, modelling and evolutionary tools

Thomas Hermann and Eric Westhof

Recent advances in glycoconjugate analysis and glycobiology

Thomas W Rademacher

Hydrogen exchange studies of protein structure

Tanya M Raschke and Susan Marqusee

Advances in transient state-kinetics

Kenneth A Johnson

Macromolecular matchmaking: advances in two-hybrid and related technologies

Robert M Frederickson

BIAcore for macromolecular interaction

Matthew Fivash, Eric M Towler and Robert J Fisher

Strategies for selection of antibodies by phage display

Andrew D Griffiths and Alexander R Duncan

The full text of *Current Opinion in Biotechnology* is in the BioMedNet library at

<http://BioMedNet.com/cbiology/bio>