CHAPTER 1

DNA MICROARRAYS

"Gene Chips" Permit Genome-wide Profiling of mRNA Expression and Have Many Other Applications

Background/Overview and General Principles:

The relative concentration of any given mRNA in a population of transcripts can be determined by hybridization to its specific complementary sequence. The relative abundance of each of the mRNAs encoded in a given genome can therefore be assessed by their simultaneous hybridization to the complete set of corresponding complementary sequences. Although RNA-RNA hybrids are more stable (i.e. have a higher melting temperature) than RNA-DNA hybrids, which are, in turn, more stable than DNA-DNA hybrids (*1*), DNA is an intrinsically less reactive and more durable polymer than RNA, as well as easier to synthesize, amplify, and manipulate. Hence, each mRNA in a sample is converted to a complementary DNA (cDNA) using reverse transcriptase, with concomitant or post-synthesis labeling with fluorescent tags to permit visualization. Likewise, sequences complementary to the cDNAs for every transcript are prepared in the form of the corresponding single-stranded DNA, either a synthetic oligonucleotide or a denatured product of the polymerase chain reaction (PCR). The most convenient way to display the DNAs that correspond to every transcript is by depositing them as a series of spots at high density on a solid support (usually a glass slide) (see FIGURE 1). Such DNA microarrays ("gene chips") can be produced in different ways. For example, on the commercial chips sold by Affymetrix, Inc., Santa Clara, CA, each gene is represented by many (typically ≥10) different, short (typically 25-base) oligonucleotides that span the coding sequence, which are synthesized directly on the chip and attached covalently at a concentration of ~10^7 copies per spot (*2,3*). In other microarrays, the genes are represented by longer oligonucleotides (typically 70 bases) (*4*) or by full, open-reading-frame (ORF)-length, DNA fragments (or amplified cDNAs) generated by PCR (*5*), which are "printed" post-synthesis as spots on the slide using a robot (*6,7*). Selecting sequences to synthesize oligonucleotides specific for each gene, or selecting primers for PCR amplification of each ORF, is relatively straighforward for organisms for which the complete genome sequence has been determined. Even for organisms for which the genome has not been completely sequenced, useful microarrays can be prepared from amplified cDNAs or from fragments of the genome isolated by random ("shotgun") cloning (*8*). The fluorescently-labeled cDNAs are then hybridized to the "chip" and, after thorough washing, the amount of fluorescence remaining bound to each spot is assessed using an automated scanner. Images are stored digitally and processed by computer because of the very large amount of information generated.

Practical Considerations:

Terminology—Gene transcription generates a single-stranded RNA identical in sequence (except for bearing U instead of T) to one of the strands of the duplex (chromosomal) DNA. The strand of DNA that directs synthesis of the transcript ("messenger RNA" or mRNA) is called the template or anti-sense strand. The other strand of DNA possesses the same sequence as the mRNA (except for bearing T instead of U) and is called the coding strand or sense strand (because mRNA is translated to produce a protein product). Hence, when cDNA copies of mRNA populations are made for microarray analysis, these represent the template (anti-sense) strand. Thus, the complementary DNAs spotted on the slide must contain the coding (sense) strand.

Depositing and Affixing DNA to the Chip—When a robot is used to deliver nanoliter amounts of solutions of oligonucleotides or PCR products (100-500 pg/nl) as spots in a predetermined high-density array (typically, thousands of spots on an area of less than 2 cm^2), the surface of the glass slide must be properly coated so that it is able to bind DNA avidly. The two most commonly used coatings are poly-L-lysine and aldehydes.

Poly-L-lysine adsorbs to the surface of the slide non-covalently; in turn, DNA binds to the poly-L-lysine non-covalently via electrostatic attraction between the polyanionic polynucleotide and the polycationic polymer. Glass slides can given a covalently-bonded positively-charged surface by reaction of the glass (which contains silicon dioxide) with compounds like amino-silane and amino-propyl silane (silanes are silicon hydrides). The use of poly-L-lysine-coated slides to promote the adherence of anionic materials, e.g. chromosome spreads or whole cells (whose surfaces are net negatively-charged because they are covered with glycoproteins, glycolipids and glycosaminoaglycans in which the polysaccharides contain sialic acid, or uronic acids, or sulfate or phosphate groups), predates the advent of DNA microarrays. Irradiation with UV light is often used after spotting to cross-link the DNA to the surface (a double bond in thymidine seems particularly susceptible to forming an adduct with the ϵ-amino groups of the lysine side chains in poly-L-lysine). However, long DNA molecules, like PCR products, seem to stick well to the surface even without this treatment. The slides can be boiled (which is necessary to generate single-stranded sequences from PCR products) and the DNA still adheres to the surface. There is a length dependence for the adsorption of DNA, however, as

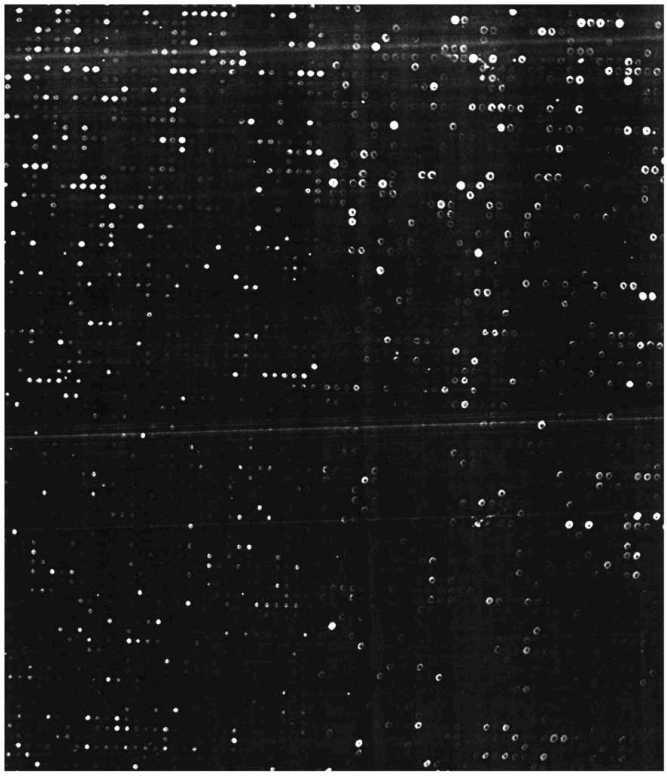

FIGURE 1 **Image of a DNA microarray**
The spots of the microarray each contain DNA from one of the 6,200 genes present in the chromosomes of the yeast *Saccharomyces cerevisiae,* with every yeast gene represented in the array. The microarray has been probed with fluorescently labeled nucleic acid derived from the mRNAs present in the cell when it is (1) growing normally in culture and (2) five hours after the cells have begun to form spores. The spots glowing green represent genes that are expressed at relatively higher levels in cells growing normally, while those glowing red represent genes expressed at higher levels during sporulation. The spots that are yellow represent genes that do not change their levels of expression during sporulation. This image is enlarged; the microarray actually measures only 1.8 by 1.8 cm. Microarrays like this provide a snapshot of global gene expression in a cell.

(Excerpted from Nelson and Cox, Lehninger's Biochemistry)

---

oligonucleotides less than 60 bases do not adhere well. After printing, it is common to cap all the available free amino groups on the slide by reacting them with succinic anhydride. This covalent modification step converts the positively-charged surface to a negatively-charged one and therefore helps block non-specific binding of the cDNA probes during hybridization.

As mentioned above, DNAs shorter than 60 nucleotides do not adhere sufficiently strongly to poly-L-lysine-coated slides. Therefore, covalent attachment is typically used when arrays are printed using short oligonucleotides. However, even long DNA molecules can be tethered to the slide covalently. For this purpose, the glass surface can be modified, as also mentioned above, by silane chemistry to contain any of several different kinds of reactive substituents, including amino groups, aldehydes, carboxyl groups, thiols, and epoxides, which can covalently react with nucleotides (or modified nucleotides included during synthesis) (*9,10*). Similar approaches can be used to attach proteins covalently to glass surfaces (*11,12*). Nonetheless, there are claims that covalent attachment is not obligatory, even for oligonucleotide arrays (*13*). In any event, a common surface for covalent attachment is one derivatized with aldehyde groups (*14,15*). Oligonucleotides synthesized with a nucleotide derivative containing a free amino group, or PCR products in which one of the primers contains such a derivative, react with the aldehyde, forming a Schiff base. Formation of a Schiff base is, however, a reversible reaction. Hence, the double bond in the Schiff base, as well as the free unreacted aldehyde groups on the slide, are reduced by treatment with sodium borohydride. Even if one uses a positively-charged (amino group-containing) surface to initially bind the amino-derivatized oligonucleotides or the PCR products, one can then use a homobifunctional cross-linking reagent containing groups highly reactive with amines (e.g. imidoesters or N-hydroxysuccinimidyl esters) to act as a molecular bridge to covalently link the DNA to the surface of the slide.

Given the pros and cons of all these procedures, spotting PCR products onto poly-L-lysine-coated slides is easy, straightforward and inexpensive. This simplicity makes it a reasonably robust process. The special amine-modified nucleotides (or other modifications) can be quite costly. Also, despite their somewhat uncharacterized interaction with the surface of poly-L-lysine-coated glass, long DNA molecules make good hybridization targets. ORF-length DNAs are sensitive detectors of the presence of complementary sequences under fairly standard hybridization conditions, so that one usually gets data for most genes in a typical experiment. The main disadvantage of spotting long DNA molecules is the increased likelihood of interference during hybridization from related, but non-identical, sequences. Avoiding this problem is a major advantage of oligonucleotide-based arrays.

Oligonucleotides can be designed that are highly specific for any given gene and, hence, can be used to distinguish between different, but highly-related, genes (*16*). Oligonucleotides can also be used to distinguish between splice variants, or for sensitive detection of single-nucleotide polymorphisms (SNPs) in the same gene. Also, oligonucleotides can be synthesized with an additional non-nucleotide linker or spacer, which allows them to be covalently bound, yet be held at a distance from the surface itself, increasing accessibility during hybridization. A 40-base oligonucleotide stuck covalently to a surface in this way is almost as sensitive for detecting its complement as an array spot made with a long DNA, but much more specific.

As mentioned above, there are methods for on-chip synthesis using photolithography and light-activated nucleotide chemistry, pioneered by Affymetrix, Inc. (*2,3,17*). Recently, methods for on-chip assembly of up to 50-base oligonucleotides that exploit Ink Jet-printer technology to deliver the phosphoramidite nucleotides for synthesis have been developed by a number of firms, including Agilent Technologies Laboratories, Inc., in Palo Alto, CA, and Rosetta Inpharmatics, Inc., in Seattle, WA (now a subsidiary of Merck) (*18-20*).

Whether depositing PCR products or long (70-base) oligonucleotides, another important limiting factor that affects the quality of the microarrays is the operation of the robot. The tips or

prongs on the arraying device must be scrupulously clean, properly aligned and as uniform as possible. It helps if there is enough moisture in the air (humidity) and/or a little anionic detergent (e.g. sodium sarkosyl) in the DNA solution to ensure that the spots wet and spread properly on the slide. The microarrays should be stored dry (in a dessicator) and be used as soon as possible. Chips older than 6 months show a noticeable deterioration in performance for reasons no one understands.

Methods for Fluorescent Labeling of the cDNA—When cDNA samples are going to be hybridized to microarrays comprised of complementary DNAs spotted on the chip (see FIGURE 2), four methods are the most widely used to fluorescently label the cDNAs:

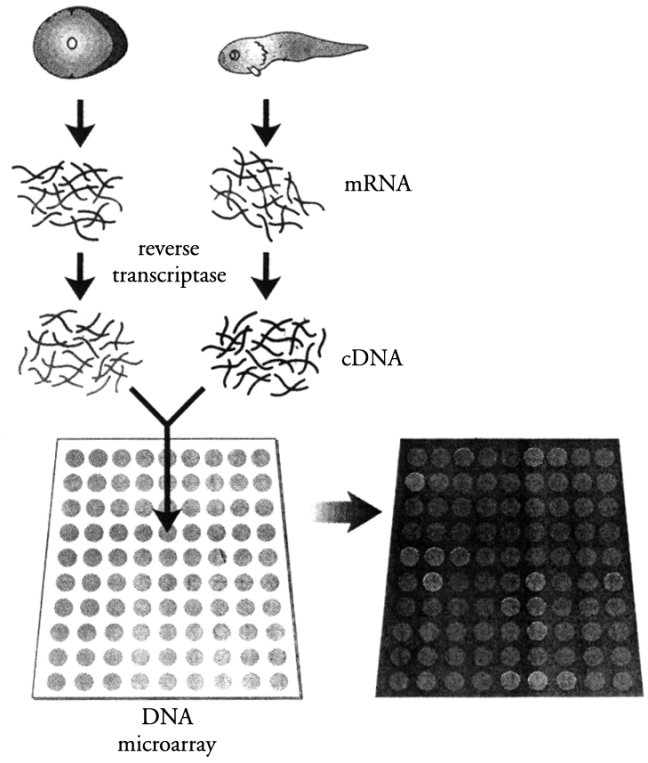

FIGURE 2 Constructing a DNA microarray
Any known DNA sequence, from any source, can be used in a microarray. The DNA can be generated by chemical synthesis or PCR. The DNA is positioned on a solid surface (usually specially treated glass slides) with the aid of robotic device capable of deposisting very small drops (nanoliters) in precise patterns. UV light is then used to cross-link the DNA to the glass slides. Once the DNA is attached to the surface, the microarray can be probed with other fluorescently labeled nucleic acids. For example, the mRNA isolated from a cell (representing all the genes being expressed in that cell) can be converted to cDNA probes by reverse transcriptase, using fluorescently labeled dNTPs. The fluorescent cDNAs anneal to complementary sequences on the microarray. After the removal of unhybridized probe, each spot that fluoresces represents a gene that was being expressed in the sample. Here, mRNA samples are collected from cells at two different stages in the development of a frog. The cDNA probes for each sample are made with nucleotides that fluoresce in different colors; a mixture of the cDNAs is used to probe the microarray. Spots that fluoresce green represent mRNAs more abundant at the single-cell stage, whereas spots that fluoresce red represent sequences more abundant later in development.

(Excerpted from Nelson and Cox, Lehninger's Biochemistry)

(i) Direct incorporation of fluorescently-labeled nucleotide derivatives, such as a Cy3-dNTP or a Cy5-dNTP, during cDNA synthesis. When purified mRNA (polyA+ RNA) is used as the template for a retroviral reverse transcriptase, oligo-dT is used as the primer and either Cy3-dUTP or Cy5-dUTP are included in the reaction (*21*). When total RNA is used, more elaborate methods for enriching for and then labeling mRNA sequences can be applied (*22*) (see also iii below). Of course, for prokaryotes (and for some applications in eukaryotes, e.g. when using an array that is not 3'-biased), random oligonucleotide primers must be used in place of oligo-dT. The main drawback of this method is that reverse transcriptase has a difficult time incorporating the bulky, dye-labeled nucleotides. In addition, reverse transcriptase does not incorporate the tagged nucleotides with equal efficiency. Some studies have estimated that the incorporation of Cy3-dUTP is about 50 times more efficient than the incorporation of Cy5-dUTP.

Cy3 and Cy5 are brighter, more photostable, and give less background than most other fluorophores. Cy3 can be maximally excited near 550 nm with peak fluorescence near 570 nm. The excitation and emission spectra of Cy3 are nearly identical to that of standard tetramethyl-rhodamine. Hence, the same filter sets can be used, Cy3 can be excited to about 50% of its maximum with the 514 nm or 528 nm lines of an argon ion laser, or to about 75% of maximum with a helium-neon laser (543 nm line) or standard mercury lamp (546 nm line). Cy5 is excited maximally near 650 nm and fluoresces maximally near 670 nm. It can be excited with a krypton/argon ion laser (98% of maximum with the 647 nm line) or a helium/neon laser (63% of maximum with the 633 nm line). Cy5 is used with other fluorophores when multiple simultaneous labeling is required because of the wide separation of its emission from that of shorter-wavelength-emitting fluorophores, like Cy3. Another advantage of both dyes is that their spectral properties avoid, respectively, the background autofluorescence of cells due to flavins (emission maximum 510-520 nm) and to heme (emission maximum 565-605). However, because its emission maximum is at 670 nm, Cy5 cannot be seen well by eye and cannot be viewed with a conventional epifluorescence microscope. Therefore, when Cy5 labeling is paired with Cy3 labeling, samples must be analyzed using a scanning confocal microscope equipped with a krypton/argon ion laser and a far-red detector [e.g. a "charged-coupled device" (CCD) camera or, less expensively, a diode array detector]. False-color imaging is then used to give Cy3 fluorescence a green hue, and Cy5 fluorescence a red hue; red and green together give an orange-yellow color.

Using a mixture of two probes to measure the relative (rather than the absolute) abundance of the sequences present corrects for variation in the amounts of DNA originally deposited in each spot on the grid and other sources of error that might cause spot-to-spot variation in a microarray. This ratiometric approach is the preferred mode for analysis. Often, the experiment is performed twice. In one, the labeled probes are prepared from the samples to be compared, say, Cy3-labeled for the control, and Cy5-labeled for the new condition. In the second, to correct for differences in the efficiency of labeling with the two dyes, the same RNA samples are used for probe preparation, but the labels are reversed. On the other hand, given the extensive internal controls available on the

commerical Affymetrix chips, analyses have been done in the absolute mode as well (see ii in the Applications section below).

(ii) Reverse transcription in the presence of aminoallyl-derivatized nucleotide. The resulting cDNA products are then reacted with fluorescent tags. The aminoallyl group is not nearly as bulky as a fluorescent dye molecule, so aminoallyl-dNTPs are much more readily incorporated by reverse transcriptase. For the structure of aminoallyl-dUTP, see FIGURE 3 (taken from the website for Molecular Probes, Inc. - need permission?). In addition, in principle, use of an aminoallyl-dNTP should not introduce any inherent sample bias in the efficiency of cDNA production because the same nucleotide is used for the preparation of all samples with reverse transcriptase. The aminoallyl group is reactive with succinimidyl esters or isothiocyanates. Such amino-reactive derivatives of the Cy series of dyes (see above) marketed by Jackson ImmunoResearch Laboratories, Inc. and for many other good fluorescent dyes (like the Alexa Fluor set marketed by Molecular Probes, Inc.) are readily available (because they are also used in this form for labeling antibodies). Moreover, because aminoallyl-dNTP is incorporated rather efficiently into the cDNA, this higher incorporation will be reflected in a greater degree of fluorescent labeling than can be achieved via the direct incorporation method (i).

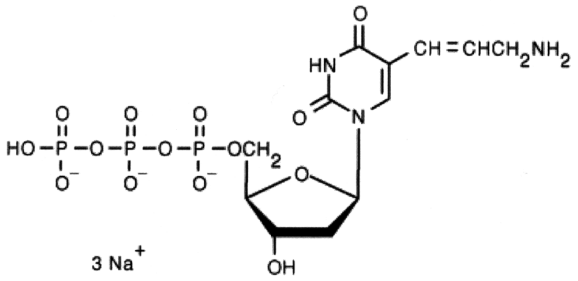

FIGURE 3 The structure of aminoallyl-dUTP
(Excerpted from Molecular Probes Web site)

(iii) As mentioned above, when the total amount of material available is small (e.g. transcripts from a human tissue biopsy versus transcripts from a yeast culture) or the sample is not very enriched in transcripts (e.g. total RNA, which is mostly rRNA and tRNA, instead of purified polyA+ RNA), some sort of amplification can be applied. This can take the form of linear amplification. One trick involves using oligo-dT primers that also contain a promoter for T7 RNA polymerase. The resulting cDNA can be used to make a cRNA copy, which can then be re-converted into cDNA using reverse transcriptase in the presence of dye labeled- or aminoallyl-nucleotide. Alternatively, when one wants to evaluate samples composed of only one or a few cells, or a very small piece of tissue, some form of PCR is performed, typically with a set of essentially random 13-base primers, so as to avoid bias in the species amplified. However, the problem with these amplification methods is that assessing whether the amplification has actually introduced some bias is difficult.

(iv) There is a method available for signal amplification when the amount of mRNA or total RNA is very small that does not involve amplifying the probe. This approach takes advantage of the use of so-called branched DNA (bDNA) or dendrimeric DNA (dDNA), a method first pioneered for diagnostic purposes at Chiron Corporation, Emeryville, CA (*23*). In this technique, cDNA is prepared using an oligo-dT primer that has an extra sequence attached to it. So, sequences that get converted into cDNA will have this tag, and sequences that don't, won't. After hybridization of the cDNA population to the array, the chip is incubated with a DNA dendrimer that contains in the neighborhood of 250 fluors (*24*). The dDNA resembles a giant, tangled ball of DNA "spaghetti" or a "Christmas tree" with the fluors attached to some of the branches and with some of the branches possessing single-stranded regions that are complementary to the tag on the oligo-dT that was used to generate the cDNA. For a schematic depiction of dendrimeric DNA, see FIGURE 4 (taken from the website for Genisphere, Inc. - need permisson?). When the fluorescent dDNA is added to cDNA bound to the array, the tag sequences at the end of the cDNAs hybridize to the complementary capture sequences on the free ends of the dDNA. In this way, any spot that has a cDNA bound to it becomes highly fluorescent. This method has not yet found widespread use because it's rather expensive, and the conditions for efficient dDNA hybridization to the array are a bit finicky, but it may be catching on. Theoretically, it offers unmatched sensitivity, and some people have had good success with it. So, it may be worth it to pursue this approach more persistently.

Overall, for ease and simplicity, method (ii) is currently the one in greatest use. It is an improvement over direct incorporation because of its lack of bias and higher incorporation rate; and, hence, it yields better and brighter probes. It also happens to be the least expensive method too. Its only disadvantage, perhaps, is that it adds an extra step to an already long process.

Specific Examples of Applications:

The uses of DNA microarrays are limited only by one's creativity. Here are some examples of how such arrays have been used.

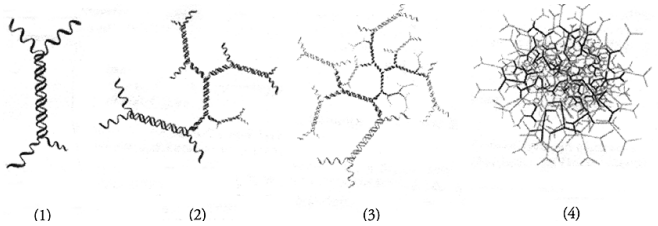

FIGURE 4 Dendrimeric DNA
Each 3DNA monomer is composed of two DNA strands that share a region of sequence complementarity located in the central portion of each strand. When the two strands anneal to form the monomer the resulting structure has a central double-stranded "waist" bordered by four single-stranded "arms (1). This "waist" plus "arms" structure comprises the basic 3DNA monomer. The single-stranded "arms" at the ends of each of the five monomer types are designed to interact with one another in precise and specific ways. Base-pairing between the "arms" of complementary monomers allows directed assembly of the dendrimer as a step-wise series of monomer layers (2-4).

(Excerpted from Genisphere Web site)

(i) Measuring steady-state levels of mRNAs under a given physiological condition or in response to some change in the chemical composition of the extracellular environment. For example, a microarray with 6,200 spots containing PCR products representing every gene in the budding yeast (Saccharomyces cerevisiae) genome was hybridized to a mixture of Cy3-dUTP-labeled cDNA made from purified polyA+ RNA extracted from diploid cells during normal vegetative (mitotic) growth in rich glucose-containing medium and Cy5-dUTP-labeled cDNA made from purified polyA+ RNA extracted from diploid cells at various times after they were induced to undergo meiosis and sporulation on nitrogen-limiting acetate-containing medium (*25*). Spots in the array that glowed green represented genes expressed at a higher level in normally growing cells, whereas those that glowed red represented genes expressed at a higher level during the developmental process of sporulation. Spots that glowed yellow represented genes expressed under both conditions, i.e. genes whose expression does not seem to change. Many genes changed their expression (more than two-fold was the cut-off), and these clustered into about seven different temporal patterns of expression. Such approaches can be used, of course, to study the response of any cell to challenge with any agent (*26*) or stimulus (*26a*).

Similarly, an array bearing ~8,600 different human cDNAs was used to monitor changes in transcript expression pattern when quiescent (serum-starved) human fibroblasts were stimulated to divide by supplying them with fresh serum (*27*). Purified polyA+ RNA from the quiescent cells was converted to Cy3-dUTP-labeled cDNA, mixed with Cy5-dUTP-labeled cDNA prepared from purified polyA+ RNA from the stimulated cells, and hybridized to the array. Again, green fluorescence marked those genes expressed only in quiescent cells, red marked genes expressed in the stimulated cells, and yellow marked genes expressed under both conditions. It was found that the levels of more than 500 genes out of those tested increased more than four-fold. Moreover, the pattern of expression was also revealing. Genes for the bZIP transcription factors, c-Fos and JunB, were induced within 15 min, whereas genes for products required to mediate cell cycle progression (e.g. PCNA, cyclins A and B1, Cdc2/Cdk1) were only elevated significantly later (12-16 hrs). Microarrays have been used to type (compare) clinical isolates of human breast tumors (*28*). It is now clear from this analysis that the malignant state of breast tumors can arise from different underlying molecular causes. This kind of insight should allow the design of more effective tumor-specific therapies with fewer side effects.

A cDNA microarray containing nearly 900 gastrula-specific cDNAs from the frog (Xenopus laevis) was probed to examine the transcriptional regulation of the corresponding genes during early embryonic development (*29*). By comparing hybridization of probes derived from maternal RNA to those of RNA extracted from zygotes at various times after fertilization, the temporal regulation of gene expression during zygotic induction of transcription could be studied. Second, by preparing RNA populations from dorsal and ventral mesoderm dissected from early gastrula embryos, the spatial restriction of these changes in gene expression during development could be examined. Third, by comparing RNA populations from naive embryonic explants and those treated with known developmental inducers (e.g. activin, a TGFβ family member), the transcriptional response of these genes to this extracellular stimulus could be assessed. Independent testing of the results obtained by RT-PCR or by in situ hybridization confirmed the accuracy and reproducibility of the findings made using the microarray approach.

Such genome-wide analyses are very valuable because genes that are regulated together often serve roles in the same or closely-related pathways, or carry out ancillary functions in support of a cell adjusting to a global shift in its physiological state. Hence, this kind of approach can provide lots of new clues as to the function of the thousands of genes in every eukaryotic cell (and hundreds of genes in various kinds of bacteria) about which we know nothing. Examination of the nucleotide sequence of the promoter regions of co-regulated genes can provide clues as to the nature of cis-acting sites(s) that are recognized by the trans-acting transcription factor(s) that control their expression. In fact, DNA microarrays can be combined with the so-called chromatin-immunoprecipitation (a "ChIP" of a different sort) approach to identify all of the cellular sequences that bind to a particular transcription factor (*30,31*), or any other DNA-binding protein for that matter.

Such approachs are not limited to examination of eukaryotic cells. Essentially, identical approaches can be used to study the effects of limitation (or sufficiency) for any nutrient on the pattern of gene expression in standard laboratory strains of bacteria (*32*) or in more exotic bacterial pathogens (*33*).

(ii) Assessing the effect of cellular mutations on the status of gene expression. DNA microarrays have been used to assess in vivo the spectrum of genes that under the control of various general transcription factors or different classes of chromatin-remodeling complexes (34-36). For example, Affymetrix chips containing sets of 25-base anti-sense oligonucleotides (~20 each) spanning every yeast coding sequence, and corresponding oligonucleotides carrying one-base mismatches to serve as negative controls (i.e. should not hybridize to any transcript), were probed with cRNA that was generated as follows (34). Purified polyA+ RNA from the mutant or condition of interest was converted to cDNA with an oligo-dT primer also containing a promoter for T7 RNA polymerase. The resulting cDNA was then converted to cRNA by transcription with T7 RNA polymerase in the presence of biotin-tagged CTP and biotin-tagged UTP. The cRNA was fragmented mechanically and hybridized to the chips. Bound cRNA was then detected by incubation with streptavidin that have been derivatived with the highly-fluorescent cyanobacterial pigment, phycoerythrin, which has a bright red emission (37). The chips were scanned with a confocal microscope and the intensity of the fluorescence signal for each gene could be quantitated and represents the absolute abundance of that particular mRNA. Using this method, it could be estimated from knowing the amount of input RNA initially and the number of cells from which it was derived that most genes in yeast (~80%) are expressed at a level of only ~1 copy of the corresponding mRNA per cell. The most abundant mRNA, like those coding for glycolytic enzymes (such as aldolase or glyceraldehyde 3-P dehydrogenase) or certain ribosomal proteins, were present at 70-90 copies per cell. By extracting RNA from mutant cells carrying a temperature-sensitive allele in a critical subunit of RNA polymerase II at various times after shift to restrictive temperature, the decay rate (half-life) of each mRNA

could be determined after its de novo synthesis was blocked. On the basis of this information, the transcription rate necessary to maintain the observed steady-sate level of any given transcript could be calculated. For example, to maintain the level of the transcripts for the glycolytic enzymes or ribosomal proteins, pol II must produce ~200 of these mRNAs per hour. Similarly, by using the appropriate temperature-sensitive mutations, it was found that loss of TFIIE (tfa1-ts) reduced mRNA expression for only 54% of the genes in the yeast genome, whereas loss of TFIIH (kin28-ts) reduced expression of nearly 90% of all genes. Likewise, using mutations that inactivate different types of chromatin-remodeling complexes or different families of histone acetylase complexes, or that release genes from the effects of different types of global repressors or various classes of histone deacetylase complexes, it is clear that different groups of genes are under the control of different subsets of regulatory factors.

Of course, shifting a temperature-sensitive mutant to a higher (non-permissive) temperature subjects the cells to heat stress. Just how much the transcriptional profile of even a normal cell shifts in response to such a physical change, and other physical insults (e.g. irradiation with ultraviolet light), has also been examined in in a variety of organisms, including yeast (*38,39*) and bacteria (*40*).

(iii) Practically anything else you can think of! Microarrays have been used to analyze the progress of replication forks during bacterial DNA replication and the role of different topoisomerases in this process (*41*), to identify all of the transcripts that are targeted to a particular subcellular destination by virtue of their association with certain RNA-binding proteins (*42*), to examine human physiological response to administration of a drug (*43*), and other applications too numerous to mentioned here, but which are discussed in the general reviews cited already (*3,6,18,20*) and in the additional reviews listed at the end of this document.

Evaluation and Manipulation of the Data Obtained:

Microarray analysis is a very powerful technology. A tidal wave of data has already been generated and more gigantic tsunamis of data are on the way. Hence, it is imperative for the modern molecular and cellular biologist to master these methods. However, the advent of global analysis of this sort has created a problem of unprecedented proportions with regard to how best to evaluate the meaning of the data obtained (*44,45*). Bioinformatics is the name given to the discipline at the interface between computer science and biology that is trying to grapple with these problems. While some of the hardware required to evaluate the results of hybridization to microarrays and store the information are fairly well developed, other aspects of the infrastructure are still rather rudimentary. Massively parallel analysis of the sort that spews forth from microarray studies generates massive amounts of data. How does one deal with data sets consisting of 500,000 individual points or values or relative ratios. Even asking the right questions about the ~10,000-point data set derived from a single chip can be daunting. Software for manipulating the data are pretty primitive. Generally accepted standards and conventions for describing experiments in a common way are still under development [see, for example, the XML ("Extensible Markup Language")-based approach for describing gene expression data at http://www.mged.org]. Tools for integrating microarray data with other genomic observations, or for rapidly querying many different kinds of data and the cumulative literature about a gene and its product are just now being implemented [see, for example, the GO ("Gene Ontology") approach at http://www.geneontology.org]. These methods for interpreting data and thinking about gene expression networks are still in their infancy. In these respects, computational biologists should have a lot to keep themselves busy. Finally, straightforward ways to rapidly follow up on, and verify, conclusions gleaned from analysis of microarray results are not quite in place. One promising method for measuring absolute amounts of any given mRNA is the quantitative RT-PCR technique, dubbed "TaqMan", developed at PE Applied Biosystems, Inc., Foster City, CA (*46,47*). In this method, the 5'-exonuclease activity of Taq DNA polymerase is used to attack a fluorogenic probe comprising a PCR primer oligonucleotide that is labeled with both a fluorescent reporter dye and a quencher dye at the 5'-end. Amplification of the probe-specific product permits exonucleolytic removal of the quencher, generating an increase in fluorescence of the reporter dye. Supposedly, there is a company out there that plans to make affordable TaqMan probes for every known gene in every known organism!

Summary:

In any event, even in the absence of fancy tools, microarrays still provide a good method for rapidly testing hypotheses, for generating ideas, or for screening for the unsuspected involvement of novel genes in any given process, under any physiological condition, or in response to any challenge. However, for microarrays to achieve the same status and ease of application as doing a mini-prep and running an agarose gel, the technology must be more accessible, more freely available, and the resulting data more easily digestible. Most labs have not reached that level of comfort with microarrays, yet.

Acknowledgements: I am grateful to the following people, who provided information and advice during the preparation of this document including Mike Eisen, Alan Sachs and especially Chris Seidel.

Literature Cited:

1. Parish JH (1972), Chapt. 12 ("Hybridization"), pp. 355-379, In *Principles and Practice of Experiments with Nucleic Acids*. John Wiley & Sons, Inc., New York, NY.
2. Jacobs JW, Fodor SP (1994) Combinatorial chemistry—applications of light-directed chemical synthesis. *Trends Biotechnol.* 12: 19-26.
3. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ (1999) High density synthetic oligonucleotide arrays. Nature Genet. 21(Suppl.): 20-24.
4. ten Bosch J, Seidel C, Batra S, Lam H, Tuason N, Saljoughi S, Saul R (2000) Validation of sequence-optimized 70-base oligonucleotides for use on DNA microarrays. Poster presented at TIGR Genome Sequencing and Analysis Conference, September 12-15, 2000, Miami Beach, FL.
5. Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO, Davis RW (1997) Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc. Natl. Acad. Sci. USA 94: 13057-13062.

6. Eisen MB, Brown PO (1999) DNA arrays for analysis of gene expression. Methods Enzymol. 303: 179-205.
7. Alizadeh A, Eisen M, Davis RE, Ma C, Sabet H, Tran T, Powell JI, Yang L, Marti GE, Moore DT, et al. (1999) The lymphochip: a specialized cDNA microarray for the genomic-scale analysis of gene expression in normal and malignant lymphocytes. Cold Spring Harb. Symp. Quant. Biol. 64: 71-78.
8. Hayward RE, Derisi JL, Alfadhli S, Kaslow DC, Brown PO, Rathod PK (2000) Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum* malaria.
Mol. Microbiol. 35: 6-14.
9. Rogers YH, Jiang-Baucom P, Huang ZJ, Bogdanov V, Anderson S, Boyce-Jacino MT (1999) Immobilization of oligonucleotides onto a glass support via disulfide bonds: A method for preparation of DNA microarrays. *Analyt. Biochem.* 266: 23-30.
10. Lindroos K, Liljedahl U, Raitio M, Syvanen AC (2001) Minisequencing on oligonucleotide microarrays: comparison of immobilisation chemistries. *Nucleic Acids Res.* 29: E69-E69.
11. Weetall HH (1993) Preparation of immobilized proteins covalently coupled through silane coupling agents to inorganic supports. *Appl. Biochem. Biotechnol.* 41: 157-188.
12. MacBeath G, Schreiber SL (2000) Printing proteins as microarrays for high-throughput function determination. *Science* 289: 1760-1763.
13. Call DR, Chandler DP, Brockman F (2001) Fabrication of DNA microarrays using unmodified oligonucleotide probes. *Biotechniques* 30: 368-376.
14. Zammatteo N, Jeanmart L, Hamels S, Courtois S, Louette P, Hevesi L, Remacle J (2000) Comparison between different strategies of covalent attachment of DNA to glass surfaces to build DNA microarrays. *Analyt. Biochem.* 280: 143-150.
15. Proudnikov D, Timofeev E, Mirzabekov A (1998) Immobilization of DNA in polyacrylamide gel for the manufacture of DNA and DNA-oligonucleotide microchips. *Analyt. Biochem.* 259: 34-41.
16. Ben-Dor A, Karp R, Schwikowski B, Yakhini Z (2000) Universal DNA tag systems: a combinatorial design scheme. *J. Comput. Biol.* 7: 503-19.
17. Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP (1994) Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl. Acad. Sci.* USA 91: 5022-5026.
18. Burge CB (2001) Chipping away at the transcriptome. *Nature Genet.* 27: 232-234.
19. Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, et al. (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nature Biotechnol.* 19: 342-347.
20. Jain KK (2001) Third Annual Cambridge Healthtech Conference on Lab-on-a-Chip and Microarrays. Pharmacogenomics 2: 73-77.
21. Vernon SD, Unger ER, Rajeevan M, Dimulescu IM, Nisenbaum R, Campbell CE (2000) Reproducibility of alternative probe synthesis approaches for gene expression profiling with arrays. *J. Mol. Diagn.* 2: 124-127.
22. Zhao R, Gish K, Murphy M, Yin Y, Notterman D, Hoffman WH, Tom E, Mack DH, Levine AJ (2000) Analysis of p53-regulated gene expression patterns using oligonucleotide arrays. *Genes Dev.* 14: 981-993.
23. Horn T, Urdea MS (1989) Forks and combs and DNA: the synthesis of branched oligodeoxyribonucleotides. *Nucleic Acids Res.* 17: 6959-6967.
24. Stears RL, Getts RC, Gullans SR (2000) A novel, sensitive detection system for high-density microarrays using dendrimer technology. *Physiol. Genomics* 3: 93-99.
25. Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, Brown PO, Herskowitz I (1998) The transcriptional program of sporulation in budding yeast. *Science* 282: 699-705.
26. Gross C, Kelleher M, Iyer VR, Brown PO, Winge DR (2000) Identification of the copper regulon in Saccharomyces cerevisiae by DNA microarrays. *J. Biol. Chem.* 275: 32310-32316.
26a. Roberts CJ, Nelson B, Marton MJ, Stoughton R, Meyer MR, Bennett HA, He YD, Dai H, Walker WL, Hughes TR, Tyers M, Boone C, Friend SH (2000) Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles *Science* 287: 873-880.
27. Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JC, Trent JM, Staudt LM, Hudson J Jr, Boguski MS et al. (1999) The transcriptional program in the response of human fibroblasts to serum. *Science* 283: 83-87.
28. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA et al. (2000) Molecular portraits of human breast tumours. *Nature* 406: 747-752.
29. Altmann CR, Bell E, Sczyrba A, Pun J, Bekiranov S, Gaasterland T, Brivanlou AH (2001) Microarray-based analysis of early development in Xenopus laevis. *Dev. Biol.* 236: 64-75.
30. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA (2000) Genome-wide location and function of DNA binding proteins. *Science* 290: 2306-2309.
31. Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF *Nature* 409: 533-538.
32. Khodursky AB, Peter BJ, Cozzarelli NR, Botstein D, Brown PO, Yanofsky C (2000) DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in Escherichia coli. *Proc. Natl. Acad. Sci.* USA 97: 12170-12175.
33. Paustian ML, May BJ, Kapur V (2001) Pasteurella multocida gene expression in response to iron limitation. *Infect. Immun.* 69: 4109-4115.
34. Holstege FC, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, Green MR, Golub TR, Lander ES, Young RA (1998) Dissecting the regulatory circuitry of a eukaryotic genome. Cell 95: 717-728.
35. Lee TI, Causton HC, Holstege FC, Shen WC, Hannett N, Jennings EG, Winston F, Green MR, Young RA (2000) Redundant roles for the TFIID and SAGA complexes in global transcription. *Nature* 405: 701-704.
36. Holmes MC, Tjian R (2000) Promoter-selective properties of the TBP-related factor TRF1. Science 288: 867-870.
37. Glazer AN (1989) Light guides. Directional energy transfer in a photosynthetic antenna. *J. Biol. Chem.* 264: 1-4.
38. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* 11: 4241-4257.
39. Causton HC, Ren B, Koh SS, Harbison CT, Kanin E, Jennings EG, Lee TI, True HL, Lander ES, Young RA. (2001) Remodeling of yeast genome expression in response to environmental changes. *Mol. Biol. Cell.* 12: 323-337.

40. Courcelle J, Khodursky A, Peter B, Brown PO, Hanawalt PC (2001) Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*. *Genetics* 158: 41-64.
41. Khodursky AB, Peter BJ, Schmid MB, DeRisi J, Botstein D, Brown PO, Cozzarelli NR (2000) Analysis of topoisomerase function in bacterial replication fork movement: use of DNA microarrays. *Proc. Natl. Acad. Sci* USA 97: 9419-9424.
42. Takizawa PA, DeRisi JL, Wilhelm JE, Vale RD (2000) Plasma membrane in yeast by messenger RNA transport and a septin diffusion barrier. *Science* 290: 341-344.
43 Loguinov AV, Anderson LM, Crosby GJ, Yukhananov RY (2001) Gene expression following acute morphine administration. *Physiol Genomics* 6: 169-181.
44. Bassett DE Jr, Eisen MB, Boguski MS (1999) Gene expression informatics--it's all in your mine. Nature Genet. 21 (Suppl): 51-55.
45. Chiang DY, Brown PO, Eisen MB (2001) Visualizing associations between genome sequences and gene expression data using genome-mean expression profiles. *Bioinformatics 17* (Suppl.): S49-S55.
46. Heid CA, Stevens J, Livak KJ, Williams PM (1996) Real time quantitative PCR. *Genome Res.* 6: 986-994.
47. Kuimelis RG, Livak KJ, Mullah B, Andrus A(1997) Structural analogues of TaqMan probes for real-time quantitative PCR. *Nucleic Acids Symp Ser.* 37: 255-256.

Additional Reading and Other Source Materials:

Baldi P, Hatfield GW (2002) DNA Microarrays and Gene Expression: From Experiments to Data Analysis and Modeling, Cambridge University Press, New York, NY.

Bowtell D, Sambrook D (2002) DNA Microarrays—A Molecular Cloning Manual, Cold Spring Harbor Press, Cold Spring Harbor, NY.

Brown PO, Botstein D (1999) Exploring the new world of the genome with DNA microarrays. Nature Genet. 21: 33-37.

Burgess JK (2001) Gene expression studies using microarrays. Clin. Exp. Pharmacol. Physiol. 28: 321-328.

Cho RJ (2000) Deriving meaning from genomic information. Biotechnol. Genet. Eng. Rev. 17: 91-107.

DeRisi JL, Iyer VR (1999) Genomics and array technology. Curr. Opin. Oncol. 11: 76-79.

Devaux F, Marc P, Jacq C (2001) Transcriptomes, transcription activators and microarrays. FEBS Lett. 498: 140-144.

Deyholos MK, Galbraith DW (2001) High-density microarrays for gene expression analysis. Cytometry 43: 229-238. R

Ferea TL, Brown PO (1999) Observing the living genome. Curr. Opin. Genet. Dev. 9: 715-722.

Jordan B (Ed.) DNA Microarrays: Gene Expression Applications, Springer-Verlag, Inc., New York, NY, 2001, 145pp.

Kerr MK, Churchill GA (2001) Statistical design and the analysis of gene expression microarray data. Genet Res. 77: 123-128.

Kumar A, Snyder M (2001) Emerging technologies in yeast genomics. Nature Rev. Genet. 2: 302-312.

Lockhart DJ, Winzeler EA (2000) Genomics, gene expression and DNA arrays. Nature 405: 827-836.

One-stop shopping: Supplement to Vol. 21 of Nature Genetics (Jan., 1999), entitled, "The Chipping News" (a pun on the wonderful novel, "The Shipping News" by E. Annie Proulx).

Planet PJ, DeSalle R, Siddall M, Bael T, Sarkar IN, Stanley SE (2001) Systematic analysis of DNA microarray data: ordering and interpreting patterns of gene expression. Genome Res. 11: 1149-1155.

Quackenbush J (2001) Computational analysis of microarray data. Nature Rev. Genet. 2: 418-427.

Rampal JB (Ed.) DNA Arrays: Methods and Protocols (Methods in Molecular Biology, Volume 170), Humana Press, Inc., 2001, 312pp.

Shalon D, Smith SJ, Brown PO (1996) A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res. 6: 639-645.

Steinmetz LM, Davis RW (2000) High-density arrays and insights into genome function. Biotechnol. Genet. Eng. Rev. 17: 109-146.

Werner T (2001) Cluster analysis and promoter modelling as bioinformatics tools for the identification of target genes from expression array data. Pharmacogenomics. 2: 25-36.

Young RA (2000) Biomedical discovery with DNA arrays. Cell 102: 9-15.

Practice Problems:

To be done - Watch this space!