
Emerging Genomic Technology: DNA Microarrays

by Ryan M. Roper '01 & Gregory P. Fournier '01

Abstract

This paper provides a brief overview of microarrays and their application, and proposes several improvements to existing technologies. The continued application of traditional genetic analysis techniques, combined with the recent availability of vast amounts of raw genomic information has prompted the development of DNA microarrays, massively parallel technologies for the assessment of multiple gene expression on a genomic scale. Exploring the current applications of this technology, we also suggest several improvements to the design, analysis, and implementation of DNA microarrays in both research and clinical settings.

BACKGROUND

Genetic research has long depended on procedures for the identification of specific genes, and the correlation of these genes to a protein fulfilling a specific function within an organism. Until recently, this was done through random generation of mutants, followed by phenotypic screening and restriction mapping. A long, often inaccurate process was required for the analysis of each individual gene of interest (Lander 1996). While much useful information about the structure and function of many organisms has been (and continues to be) gained through these techniques, they are limited in the quality of information they provide, and often produce an incomplete picture of the genetics of an organism.

With the advent of high-speed gene sequencing in the late 1990's, the DNA sequences of tens of thousands of individual genes became available, replacing cumbersome restriction and genetic maps, and providing an enormous reservoir of raw information (Lander 1999). With this new information came a shift in paradigm: genes were now identified before their function was investigated. Furthermore, it became possible to speculate on the function of any given gene through sequence comparison to genes of known function.

DNA microarrays (notably, Affymetrix's *GeneChip*TM) are a technology designed to take advantage of this wealth of genetic data, by allowing massively parallel hybridization experiments to be performed, using mRNA contained in the nuclei of target cells.

Microarray technology stems from the blot technology that has been used by biologists for decades. In a blot, mRNA or digested DNA is applied to a base membrane such as nitrocellulose or fiberglass and covalently linked by UV irradiation. A short piece of DNA labeled either fluorescently or radioactively, is then applied as a probe (Lewen 1997). Once the excess probe is washed off, the spots it hybridized to can be identified. These techniques were used for comparing a number of different cell types for an expression or presence of single gene. These procedures were very large scale, and required lots of mRNA and lots of probe.

The next technological advancement was attaching a number of known probes to a base membrane filter and applying labeled cDNA made from cellular mRNA to it. This technique, known as an array, is useful for identifying multiple gene expression in a single cell type. For example, it is possible to purchase an array of probes for 198 genes expressed during apoptosis (programmed cell death), or an array with genes for most known cytokines (immune system secreted proteins).

In this way, every transcriptionally active gene in a cell can be identified by binding to a known genetic sequence in the array. Applications of microarray technology are numerous, and their use is promising in such fields as oncology, developmental and biochemical genetics. Furthermore, enhancements to the existing DNA microarray technology may make even more applications possible, including real-time detection of binding, identification of point mutations, and analysis of increasingly smaller

volumes of genetic material. Microarrays also significantly improve on array technology through miniaturization. Using glass instead of base membrane and much smaller probes than an array does allows a microarray to contain thousands instead of dozens of genes on a chip a few inches in size.

Microarray Analysis

Microarrays are generally used to probe cells for presence or expression of certain genes. Activated genes in cells are transcribed in the nucleus to produce mRNA fragments, which are then exported into the cytoplasm and used by ribosomes to produce corresponding proteins. Consequently, the existence of particular mRNA in a cellular extract is an excellent gauge of gene expression. An assembled DNA microarray is used to test for presence of a given sequence in an mRNA library. After extracting mRNA from a sample tissue, a purified library is first bound to fluorescent labels and then applied to a microarray. Over a sufficient exposure time, complementary sequences bind, resulting in RNA/DNA hybrids. Washing unbound mRNA away leaves fluorescence only where hybridization occurred; an image of the microarray under UV light can quickly be analyzed to determine which sequences were present in the original cellular mRNA. A freeware software packaged called Dapple can perform just such an analysis.

Microarray Synthesis

Before a DNA microarray is synthesized, the

manufacturer must be provided with a list of oligonucleotides (small sequences of single-stranded DNA) to synthesize. A DNA sequence twelve bases long allows for over 16,000,000 possible combinations, so some twelve-base oligonucleotide should be unique to every gene in a genome. Genomic data is readily available online, thanks to the completion of the Human Genome Project, so the task of generating appropriate oligonucleotides is trivial.

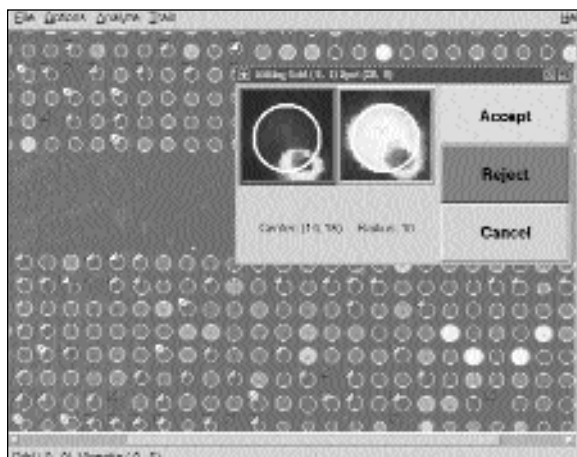
There are currently two general models for synthesizing microarrays: sequential or parallel. Sequential synthesis focuses manufacturing resources into creating a single probe at a time. The most common method for sequential synthesis relies a tiny sprayer, much like that found in an inkjet printer, to deposit pre-synthesized oligonucleotides onto a chip. This approach is being used by a number of biotech companies, including Combion, Rosetta, ProtoGene, and Affymetrix (Henke 2001). Manufacturers spray a chemical solution contain-

ing the gene probes in a pattern onto the chip substrate, in a manner similar to traditional blot arrays. Another sequential

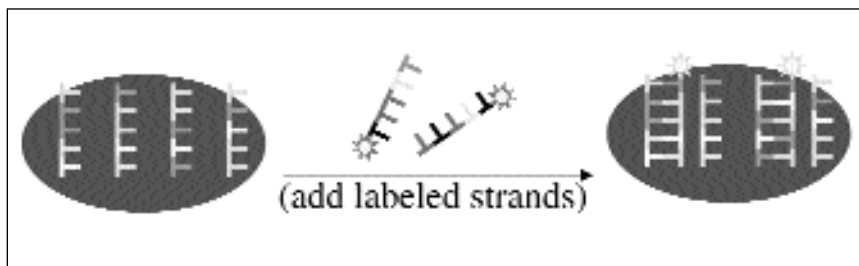
method is currently being developed by Nanogen, and uses microscopic robots to deposit the gene probes onto the substrate. Electrophoresis can be used to speed up hybridization.

Parallel synthesis generates and deposits large numbers of oligonucleotides *simultaneously*. One technique for accomplishing this is called photolithography. It is analogous to computer chip synthesis and relies on light to synthesize oligonucleotides *directly* on the substrate chip.

In a process patented by Affymetrix,



Dapple. Freeware software electronically detects levels of red and green fluorescence emitted by microarray.



Single stranded probes. Fixed to microarray surface, then exposed to labeled mRNAs. Complementing strands bind and remain on the gene chip after washing.

nucleotides are first tagged with photoreactive chemical caps and washed across the substrate chip. The chip is then exposed to a specific pattern of light, determined by a photolithographic mask, which removes selected caps. Another set of tagged nucleotides is then washed across the chip – binding only where previous caps were removed. The process is repeated, one nucleotide layer at a time, until specific, predetermined sequences are synthesized at each array site (Affymetrix 2001). Scientists at Argonne National Laboratory are presently developing a similar process, using a series of gel blots to deposit nucleotide layers. Since the synthesis time for these models is independent of the number of sequences placed on a chip, the result is a greater economic scalability and a much lower cost for microarrays.

APPLICATIONS

Developmental Genetics

While the clinical applications of DNA microarrays involve the direct analysis of a human genome, these same technologies can be applied to the genome of any species.

Drosophila, *C. elegans*, and yeast genomes have already been sequenced, their genomic data available for use in microarrays. *Drosophila* and *C. elegans* (nematode worm) are intensely studied as models for the cell-cell interactions that lead to the differentiation and organization of tissue and biological structures in animal development (Reinke 2000). DNA microarrays provide a powerful tool in this analysis, as they can be used for the direct comparison of gene expression in particular cells (or the same cell at particular stages of development), and elucidate the genetic regulation behind specific cellular differentiation events (Jiang 2001).

Applied globally, this information is useful because of the pervasive conservation of genes and function seen between most animals, including humans. Debilitating mutations in a particular gene of an experimental system animal often have counterparts in humans, which allow us to pinpoint the specific cause of many genetic diseases. Also, as cancer is primarily a failure of cell

differentiation/proliferation regulation, a thorough, genomic understanding of the genetic regulation of development can lead to the better understanding of the mechanisms in humans which cause this cell proliferation/regulation mechanism to go awry. As well as the understanding of disease, developmental genomic analysis allows for the better creation of stem-cell lines for a wide range of therapies.

Biochemical Genetics

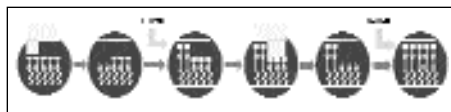
Just as *Drosophila* and *C. elegans* are used to analyze development, yeast is the choice organism for studies in biochemical genetics, due to its simple eukaryotic structure and well-understood metabolic pathways. As these pathways are generally regulated through feedback that regulates the transcription of particular genes (and hence generating a distinctive mRNA library), biochemical pathway analysis is an excellent problem for

DNA microarrays. For example, DNA Microarrays have been used to characterize the Gal pathway in yeast (DeRisi 1997). Normally, yeast make use of glucose as a food substrate, and run a normal metabolism. However, if their

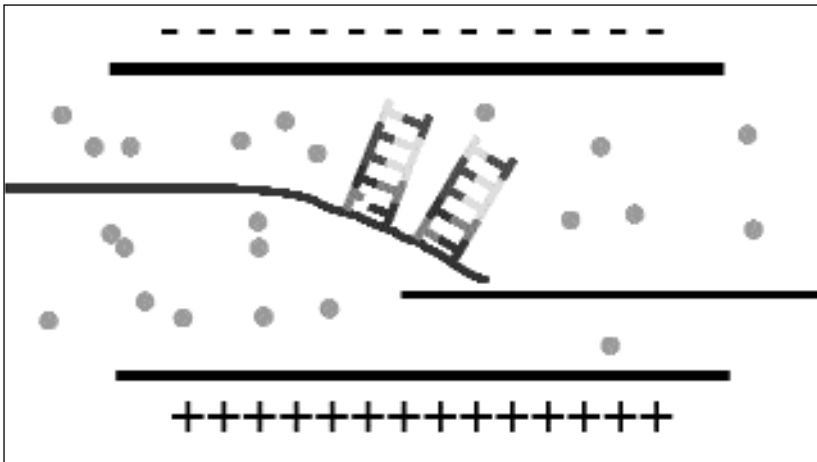
substrate is switched to galactose (a type of sugar requiring a different set of enzymes for metabolic processing) the presence of the new sugar triggers a systemic change of gene expression, resulting in a yeast cell competent to metabolize galactose, and thus allowing the cell to thrive in its new environment. A DNA microarray, using the mRNA from a Gal-competent yeast cell, compared with a typical yeast cell shows the genes involved in each metabolic pathway, including the regulatory components common to each. Once again, this information is relevant to human disease in its application to metabolic disorders such as diabetes, and in a more general sense, the processes by which cells shift expression patterns when presented with insults such as injury or viral infection.

Oncology

Perhaps the most immediate clinical application for DNA microarrays is in the diagnosis of



The Affymetrix Process. Photo-sensitive caps terminate growing oligos are terminated and prevent elongation. These can be selectively removed by a light mask.



Deflection. When microarray surface is placed in a constant magnetic field, negatively-charged DNA will experience an electromotive force perpendicular to the field, bending the lever. This force will double on hybridization, increasing lever deflection.

cancer. Cancer is an unregulated monoclonal expansion of a mutated cell line, resulting from the accumulated genetic damage of a single progenitor cell. Therefore, every cancer cell is genetically identical in a patient, and an analysis of the particular genes responsible for a given case of cancer is an excellent way to identify and characterize an individual's illness. While there are several stages to a cell becoming cancerous, all involve the mutation of oncogenes and/or tumor suppressor genes. As many of these genes have now been identified in humans and their sequences known, it is possible for a DNA microarray to be constructed which can assay mutations/deletions of these genes. Furthermore, assaying the mRNA libraries of tumor cells will show the absence/presence of gene expression abnormal for that particular tissue type. Some examples of this are the ectopic expression of embryonic proteins, and the absence of typical cell-surface proteins, such as CD4 (Lydyard 2000). Studies using DNA microarrays in breast cancer research have shown that genomic profiles of cancer types strongly correlate with prognosis and treatment efficacy; such data will surely continue to be used in the early detection and targeted therapy of cancer (Perou 2000). Genomic data about specific cancer types also allows for the generation of cancer-specific immunotherapies, such as drug-carrying or radioactive antibodies targeted specifically against Tumor Associated Antigens (TAA) or Tumor Specific Antigens (TSA) (Lydyard).

TECHNOLOGY ASSESSMENT / PROPOSALS

Current microarrays require a fair amount of RNA exposure time (roughly 24 hours) in order to ensure complete hybridization. Since unbound fluorescent mRNA cannot be washed before exposure is complete, hybridization cannot be monitored in real-time. A true real-time approach must use a non-fluorescent mechanism (most likely electronic) for detecting hybridization.

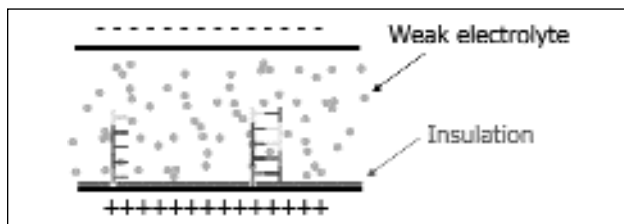
There are many potential advantages in real-time hybrid detection.

First, it allows approximation of protein concentrations. mRNA's with a higher concentration will statistically reach their target probes in less time. Accordingly, an ordering of when probes hybridize serves as an approximation of how abundant each mRNA is (and roughly how common its encoded proteins are). Electronic detection is also subject to less error than fluorescent detection. First, fluorescent tags are sequence specific, not necessarily binding every strand of mRNA. Second, electronic detection isn't dependent on a third generation signal (reflected light, converted into an electrical signal by a charged-coupled device), so it's subject to much less noise interference.

There are also downsides to electronic/real-time detection, however. Due to the expense and delicate nature of microarrays (which can be used a limited number of times before probes begin to degrade), two mRNA libraries are often analyzed simultaneously with green and red fluorescent labels (Korbel 2001). This also allows for multiple experiments to be conducted under identical conditions. Electronic detection methods can't provide either of these advantages – only one library at a time can be tested.

Weighed Cantilever

One approach to the problem of real-time detection relies on a microscopic cantilever that bends under the weight of a hybrid DNA/RNA. This technique was successfully developed and applied by IBM's Zurich Laboratory. Using pho-



Microarray surface is placed in electrolytic solution and voltage is applied, resulting in a steady-state current "through" DNA molecules. Hybridization increases current through molecules and can be detected.

tolithography, IBM constructed a 1 mm-thick metallic lever and plated it with gold. A layer of lysine was then applied, thereby providing an adhesion layer for DNA, and oligonucleotides were bound to the cantilever (Fritz 2000). As expected, the tiny levers bent on hybridization.

Unfortunately, the actual deflection could only be measured by laser reflection, subjecting the gene analysis to the same CCD noise common in fluorescence-based detection. Several extensions, however, could both resolve these issues and improve on the IBM model. First, the bending cantilever could be used to close an electric circuit, absolving the need for light-based analysis. Second, the negative charge inherent to the DNA backbone could be used to generate a force on the cantilever under a voltage differential – a hybrid DNA/RNA would have twice the charge as its single-stranded counterparts, and therefore have twice the electronic "weight" across a voltage potential. This approach could amplify cantilever movement, thereby improving data quality.

Here we propose a series of novel approaches to the real-time hybridization problem:

Molecular Detection

Another possible way of detecting an mRNA hybridization event would be through a change in confirmation of the oligonucleotide sugar-phosphate backbone. If a certain hypothetical nucleotide analog were used during the oligo synthesis, binding of an mRNA molecule may cause it to change orientation/confirmation in a detectable way (by perhaps a cis-trans transition, bond-energy change, or release of an attached "reporter" region). This may also make it possible to more accurately detect point mutations, as the absence of this molecular alteration at any

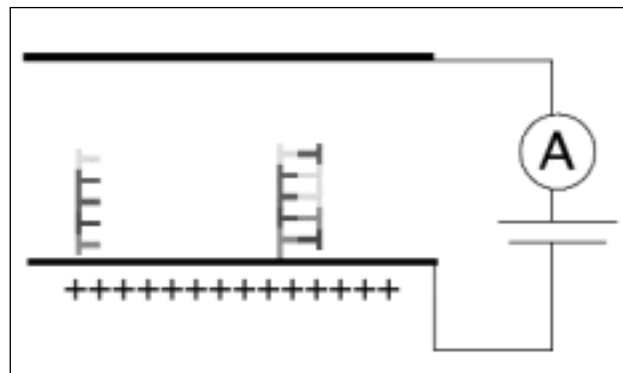
length in the polymer could be detectable, either directly or through a conformation shift carried down the length of the molecule.

Electrical Resistance

Another potential electronic means of detecting mRNA hybridization events in real-time is through a measurement of electrical resistance in the DNA molecule. Since DNA has mild conducting properties, it is possible (using an insulated chip) to keep a continuous low-level charge running through the oligo, out into a conducting solution. If an mRNA hybridizes to this oligo, there will be an instantaneous shift of resistance, which can be measured at the DNA strand/chip interface, generating a signal for VSLI input. This approach avoids the problem of charge inhibiting hybridization or causing promiscuous hybridization, due to the extremely low level of electron flow that is necessary in order to have a measurable differential signal.

Capacitance

A final mechanism for potential real-time detection involves a change in DNA's electronic capacitance during a hybridization event. Since electrons in the DNA/RNA backbone are largely delocalized (surrounding phosphates and nitrogenous bases in clouds of negative charge), a loss of charge can easily be distributed across multiple nuclei. If an oligonucleotide is bound to the positive face of a capacitor, an applied voltage will expel electrons from the backbone until electro-attractive forces from the nuclei overwhelm the electromotive force (i.e. *capacitance*).



Surface of microarray represents one plate in a capacitor. Hybridization increases the DNA's capacity for excess charge (under a constant voltage differential), resulting in a brief current in the circuit.

A binding event would increase the molecule's capacitance by providing more "storage" orbitals for unbalanced charge, and would thus be associated with a momentary flow of electrons away the DNA backbone. This microcurrent could be easily detected.

CONCLUSION

Over the past decade there has been an exponential increase in genomic information available for analysis. This trend will surely continue, as sequencing becomes faster and less expensive, and an understanding of genomic biology propagates through the medical community. Industry and clinicians both agree on the tremendous opportunity presented by functional genomics for the development of the next generation of medical treatment. However, in order for this new information to be applied we first must have the technology to understand it. DNA microarrays provide the ability to understand the genetics of disease on a complete systemic scale never before possible. As this technology advances, it will surely occupy a central role in the characterization, testing, and treatment of human disease. ■

REFERENCES

Affymetrix (2001). *GeneChip® Probe Array Synthesis*. Retrieved May 15, 2001 from the World Wide Web:
http://www.affymetrix.com/technology/tech_probe.html.

DeRisi, J., Iyer, V., Brown, P. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, 278(5338), 680-686.

Fritz, J. (2000). Translating biomolecular recognition into nanomechanics. *Science*, 14;288(5464), 316-318.

Henke, C. (2001). DNA-chip technologies I: Research fundamentals and industry catalysts. *IVD Technology*.

Jiang, M., Ryu, J., Kiraly, M., Duke, K., Reinke V., and Kim, S. (2001). Genome-wide analysis of developmental and sex-regulated gene expression profiles in *C. elegans*. *PNAS*, 98, 218-223.

Korbel, G., et. al (2001). Reaction Microarrays: A Method for Rapidly Determining the Enantiomeric Excess of Thousands of Samples. *Journal of the American Chemical Society*, 123(2), 361-362.

Lander, E. (1999). Array of Hope. *Nature*, 21 (Supplement), 3-4.

Lander, E. (1996). The new genomics: global views of biology. *Science*, 274, 536-539.

Lewen, B. (1997). *Genes VI* (pp. 631-633). New York, NY: Oxford University Press.

Lydyard, P., Whelan, A., Fanger, M. (2000). *Instant Notes in Immunology*. New York, NY: BIOS Scientific Publishers Limited.

Perou, C., Sorlie, T., Eisen, M. (2000). Molecular Portraits of Human Breast Tumors. *Nature*, 406, 747-752.

Reinke, V., Smith, H., Nance, J., Wang, J., Van Doren, C., Begley, R., Jones, S., Davis, E., Scherer, S., Ward, S., and Kim, S. (2000). A global profile of germ line gene expression in *C. elegans*. *Molecular Cell*, 6, 605-616.

ABOUT THE AUTHORS

Ryan M. Roper '01 is a Computer Science major and Molecular Biology minor. He plans to attend medical school following graduation.

Gregory P. Fournier '01 is a Genetics, Cell and Developmental Biology major, with an interest in genomics, computational genetics and bioinformatics. Upon graduation he plans to pursue a career in healthcare technology consulting.