Microarray technology
An array of opportunities

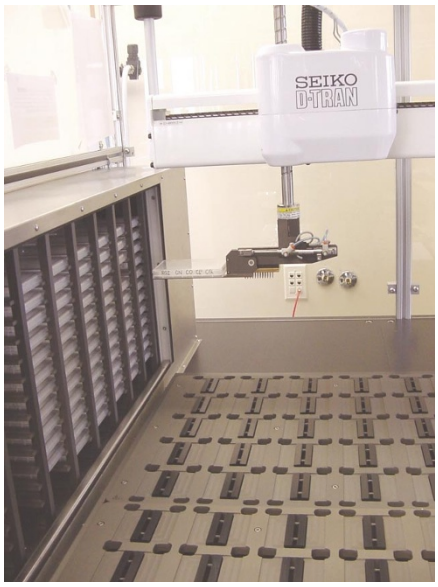
The mountain of information that is the draft sequence of the human genome may be impressive, but without interpretation that is all it remains — a mass of data. Gene function is one of the key elements researchers want to extract from the sequence, and the DNA microarray is one of the most important tools at their disposal.

The past few years have seen rapid growth within the microarray field, with the falling price of technology allowing biologists to abandon their home-made equipment in favour of one of an expanding range of commercial instruments now on the market.

"Medicine is going molecular in a major way, and microarrays are being used to profile everything from autism and schizophrenia to Alzheimer's and Parkinson's," says Mark Schena, visiting scholar at TeleChem International/ arrayit.com in Sunnyvale, California, which sells microarray reagents and parts for arrayer robots.

Microarrays exploit the preferential binding of complementary single-stranded nucleic-acid sequences. The underlying principle is the same for all microarrays, no matter how they are made — the unknown sample is hybridized to an ordered array of immobilized DNA molecules whose sequence is known. Each array features thousands of different DNA probe sequences arranged in a defined matrix on a glass or silicon support. Unlike conventional nucleic-acid hybridization methods, microarrays can identify thousands of genes simultaneously, which means that genetic analysis can be done on a huge scale.

This has revolutionized the way in which researchers analyse gene expression in cells and tissues. Microarrays — also referred to as DNA arrays, DNA chips, biochips and GeneChips — allow researchers to determine which genes are being expressed in a given cell type at a particular time and under particular conditions. They can be used to compare the gene expression in two different cell types or tissue samples, for example, healthy versus diseased tissue, and to examine changes in gene expression at different stages in the cell cycle or during embryonic development.



JEREMY HASSEMAN

Custom-built robot arrayer in action at TIGR can spot 100 slides at a time.

Microarrays are also being used in comparative genomic hybridization studies, a molecular cytogenetic approach for genome-wide detection of chromosomal deletions and amplifications, as well as for genotyping individuals for genetic differences, such as single-nucleotide polymorphisms (SNPs),

DIY OR OFF-THE-SHELF?

When microarray technology came onto the market in the late-1990s, the high price tag pushed it beyond the reach of most academic labs, so researchers were forced to use their initiative. At the time, Stanford University's *MGuide* on how to build your own arrayer from scratch proved invaluable. Today, there are many more options.

Although it is possible to build an arrayer for about US$50,000, a basic instrument can now be bought for about the same price from a number of suppliers — such as BioRobotics of Cambridge, UK, Genetix of New Milton, UK, Cartesian Technologies of Irvine, California, and GeneMachines of San Carlos, California — although prices for arrayers vary widely depending on their speed and capacity.

At the same time, the GeneChips made by Affymetrix are now more affordable; and ink-jet systems are starting to trickle onto the market, offering greater speed and more uniform spot morphology over contact printing, but these still come with a fairly hefty price tag.

The demand for the technology is so great at some of the major research institutions that they have established core facilities to produce inexpensive microarrays and so make the technology more broadly available. Some facilities with spare production capacity are also selling arrays at cost to investigators from outside institutions. Stanford University's Stanford Functional Genomics Facility, for example, offers human microarrays containing 49,000 cDNAs, of which 15,000 or more are unique human genes. The KIChip core facility at the Karolinska Institute in Stockholm offers a range of services to external researchers on a fee basis, including the production of custom-spotted microarrays.

When Vivian Cheung, of the departments of neurology and pediatric oncology at the University of Pennsylvania, Philadelphia, first thought about using microarrays in late 1996, there were no commercial arrayers and GeneChips were out of her price range. Her only option was to build a DNA arrayer in-house. Cheung's SPOT DNA arrayer has churned out arrays for her lab for several years, where one of the main aspects of research is narrowing down the location of genes responsible for genetic diseases. She still makes and reads her own arrays, but has switched to a commercial instrument from Affymetrix. "The price is coming down to a point where it's worth our while to buy the instruments and have someone else take care of them," she says.

On the other hand, Michael Miles, of the department of pharmacology and toxicology at the Virginia Commonwealth University in Richmond, admits to being "sort of biased towards the commercially available arrays", which he has been fortunate enough to be able to afford. Miles is studying the

that might be associated with disease.

At a fundamental level, microarrays are also being used in attempts to assign probable functions to newly discovered genes by comparison with the expression patterns of known genes, to identify key players in signalling pathways and to uncover new categories of genes.

But their use is not restricted to basic biology. They are also finding applications in the identification of new targets for therapeutic drugs, in disease diagnosis, and in toxicogenomics, the study of the genetic basis of an individual's response to environmental factors such as drugs and pollutants.

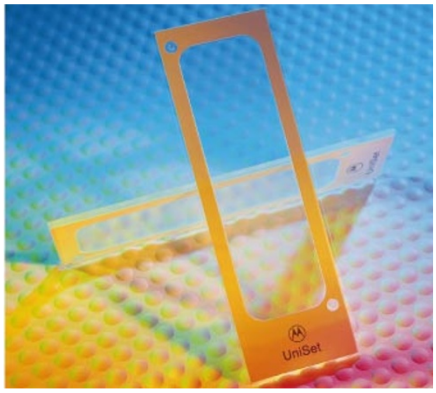
Spot specs

The most commonly used substrate for microarrays is glass — although they can be made of other materials, such as silicon — onto which thousands of spots of single-stranded DNA probes, in the form of cDNAs or oligonucleotides, are placed by a robot arrayer using contact or non-contact printing methods.

Alternatively, oligonucleotides can be synthesized *in situ*, building up each element of the array nucleotide by nucleotide and using ink-jet printing or photolithographic methods similar to those used in the semiconductor industry.

The spots are typically less than 200 μm in diameter and need to be read by specialized imaging equipment — confocal laser scanners. The spot sizes on 'macroarrays', by contrast, are about 300 μm or more and can be imaged using conventional gel and blot scanners. Contact printing and ink-jetting methods typically give spots of 100 μm in diameter, whereas those produced by photolithography are about 20 μm. This produces microarray densities of 10,000 and 250,000 spots per cm^2, respectively.

AP PHOTO/RICH PEDRONCELLI



MOTOROLA

Affymetrix (top) leads the high-density chip market; new kid on the block is Motorola's CodeLink chip (right).

Industry landscape

Affymetrix of Santa Clara, California, was one of the first commercial microarray companies and still has command over the high-density microarray market. The company uses 25-mer oligonucleotides synthesized *in situ* using its proprietary process, which combines solid-phase chemical synthesis with photolithography. Its GeneChip — an Affymetrix trademark — Human Genome U133 set of two microarrays contains over 1 million different oligonucleotides, representing more than 33,000 of the best-characterized human genes.

The price of GeneChips has come down by about half, bringing them within the reach of at least some academic researchers.

molecular plasticity of drug abuse and says DNA array studies provide a genomic-level, non-biased approach. He buys commercially produced chips but processes them on his lab's Affymetrix scanner, still a fairly expensive item at just under $200,000.

Whether it makes sense to buy off-the-shelf or make your own arrays also depends on how many you need, and whether commercial arrays contain the genes you are interested in. But doing it yourself is not always easy. Jan Vijg, of the department of physiology at the University of Texas Health Science Center, San Antonio, says it took endless telephone calls, numerous lab visits and more than a year to develop a workable system. His research interests centre on the molecular basis of ageing and cancer. In 1999 he looked into buying commercial arrays but realized that to do large-scale experiments of 100–200 arrays at a time would mean making his own arrays in-house. "Everything we did is really based in one way or another on information in the public domain," says Vijg. He bought a BioRobotics arrayer and an Axon Instruments scanner, and adapted the Stanford protocols, initially printing 2,000 genes per slide in duplicate. He has since been asked to turn his facility into an institutional core and has bought a second arrayer, this time from GeneMachines. This comes with a price tag of $120,000–130,000 but has better throughput and capacity. "I think eventually we'll be able to make 20,000-gene slides in duplicate available for less than $100," says Vijg.

ONTARIO CANCER INSTITUTE

Data analysis at the Ontario Cancer Institute's Microarray Centre.

When Jim Woodgett began to dabble in microarray technology three years ago, he never expected to end up running a core facility that supplies high-density microarrays and technical support to academic researchers across the globe. The Microarray Centre at the Ontario Cancer Institute in Toronto, which he directs, was established through a partnership between the institute, the government and industry to ensure that Canadian scientists had access to affordable high-quality microarrays.

Woodgett contracted with Toronto-based Engineering Services, now Virtek Vision International, to design a contact arrayer that uses a split-pin configuration. The company now sells a third-generation version of the original. Woodgett's centre runs four machines in parallel, printing from 48 genes at a time. The 30 staff generate the probes, produce the arrays and carry out quality control, and include technicians, researchers and bioinformaticians. Last year they produced 16,000 off-the-shelf arrays, including human and mouse arrays — most of which were high density, and 40% of which went to academic labs outside Ontario, many to the United States. **D.G.**

The company shipped over 280,000 GeneChips in 2001 and reported revenues of US$194.9 million, up 12% on 2000.

But recent years have seen a shake-up in the industry. Last year, Incyte Genomics of Palo Alto, California, a leading supplier of microarrays, quit the chip-making business, deciding instead to refocus its efforts on its core information business. By forging strategic collaborations with microarray manufacturers, which get access to the company's extensive database and patent portfolio, Incyte hopes to benefit from microarray sales without having to make them. Incyte may be gone, but some heavy hitters — most notably Agilent Technologies in Palo Alto and Motorola of Northbrook, Illinois — have recently entered the market-place.

It is perhaps not surprising that Motorola is making a play in this area. The company has a keen nose for business opportunities in emerging markets and the deep financial pockets needed to secure some market share. It also has core expertise in manufacturing, microfluidics, miniaturization, software engineering and systems integration.

Its subsidiary, Motorola Life Sciences, launched its first microarray product last summer. The CodeLink bioarray system for gene-expression profiling and SNP genotyping includes off-the-shelf arrays, optimized reagents and software to capture the images and carry out a first-level analysis of the array. Labs can use their own scanners. The company offers human and rat arrays, each representing 10,000 full-length gene sequences, and expects to launch a mouse array next month. Its genotyping array contains 72 SNPs from the P450 cytochrome family. Motorola's agreement with Incyte Genomics allows it to develop microarrays based on Incyte's comprehensive gene databases.

Motorola synthesizes 30-mer oligonucleotides 'off-line' and spots them onto slides coated with a three-dimensional, branched polymeric substrate gel surface, using Hewlett-Packard's non-contact, piezo-dispense technology. The company also produces custom arrays to order and sells 'activated' non-spotted slides for researchers to make their own arrays.

Agilent Technologies, on the other hand, uses proprietary SurePrint ink-jet technology and offers human, mouse and and rat cDNA arrays and custom oligonucleotide arrays. In the latter case the oligonucleotides (either 25- or 60-mer) are synthesized *in situ* and built up a base at a time on standard 1 × 3-inch glass slides to give arrays of either 8,400 or 22,000 features. Doug Amorese, R&D section manager responsible for chemistry and molecular biology in Agilent's DNA Microarray Program, says the cDNA type of microarray is useful when large numbers of identical arrays are needed, whereas the *in situ* system provides the flexibility to tailor designs to suit individual needs.

As a subsidiary of Hewlett-Packard, Agilent has access to considerable expertise in ink-jet printing methods and high-end analytical instrumentation — principally high-performance liquid chromatography and mass spectrometry. So the microarray area "seemed like a very good fit" for the company, says Amorese. Hewlett-Packard had been looking for a way into molecular biology, and microarrays "seemed like an area that was going to grow", he says.

The cross-licensing agreement Agilent signed in 1999 with Oxford Gene Technology (OGT) of Oxford, UK, is seen by the company as key to making this happen. OGT was set up by Edwin Southern and the University of Oxford in 1995 to commercialize Southern's DNA microarray patents. Agilent's other main collaborators are Rosetta Inpharmatics of Kirkland, Washington, and Incyte Genomics.

David and Goliath

As well as the big guns, several smaller companies are seeking to carve out a niche. One example is febit, a young biotechnology company employing some 70 people in Mannheim, Germany. It has developed a prototype DNA analysis device that fully automates and integrates all the steps in the analysis process. Its machine, Geniom one, is designed for both gene-expression analysis

DEALING WITH THE DATA DELUGE

The massive amount of microarray data collected so far has been generated on multiple platforms and is stored in a host of different formats, levels of detail and locations. This makes it difficult for any group to re-analyse or verify the data, or compare the results with their own. "It's apples to oranges," says Steven Gullans of the department of medicine at Brigham and Women's Hospital/Harvard Institutes of Medicine in Boston, Massachusetts.

Moreover, there are no uniform standards for reporting microarray data in journal articles, and there is no requirement for authors to deposit their data — and any supporting information — in the public domain. "I think the journals have to force it," says Gullans, "just like they forced us to put sequence data in the public databases, and they are a little at a loss how to do that."

Although most researchers agree that public databases for microarray data are a good idea, many are hesitant about depositing their own data in the public repositories now being developed. These include the Gene Expression Omnibus (GEO), operated by the US National Center for Biotechnology Information (NCBI); ArrayExpress, run by the European Bioinformatics Institute (EBI) in the UK; and CIBEX, the gene-expression database being developed by the DNA Data Bank of Japan.

"I think everyone realizes that the value of [microarray] data is not in looking at them in isolation but really trying to look at them in a broader context," says John Quackenbush, head of the whole-genome functional analysis group at The Institute for Genomic Research in Rockville, Maryland.

The problem is that expression data are much richer than sequence data, and many factors can affect how genes are expressed. You need to capture more information, says Quackenbush, including details of the experimental design, array design, samples, controls and experimental conditions, and the data manipulation and analysis methods used.

The Microarray Gene Expression Data (MGED) group was established in 1999 to develop a framework for describing information about a DNA microarray experiment, as well as a standard format for data exchange. The first version of its MIAME (minimum information about a microarray experiment) was proposed last year (see *Nature Genet.* **29**, 365–371; 2001 and *Nature* **415**, 946; 2002). The MAGE-ML (Microarray Gene Expression Markup Language) data-exchange format, which the MGED is developing along with the

JEREMY HASSEMAN

Data, data everywhere

and genotyping. It offers "a plug-and-play solution", says Peer Stähler, febit's vice-president and chief scientific officer, and one of the company's founders. "You don't have to become an expert in surface chemistry, you don't have to optimize the processes. All you need is data," he says.

At the heart of Geniom one is the programmable DNA processor — a special reaction carrier with a three-dimensional microchannel structure. Both the synthesis of the oligonucleotide probes — which uses a light-dependent technique that does not rely on physical masks — and the hybridization of the labelled samples takes place in the channels. "You insert the reaction carrier and never touch it again until you throw it away," says Stähler. "If you're efficient you can do two runs a day."

The current design can produce microarrays containing up to 64,000 different oligonucleotides — it runs eight arrays in parallel, each with 8,000 spots per array. With between one and four spots covering a gene, each array can cover a few thousand genes. This is not as dense a coverage as Affymetrix's GeneChips, but Stähler expects future versions of Geniom to have 10 times as many spots per array.

The prototype is being tested by Jörg Hoheisel and his team at the German Cancer Research Centre in Heidelberg. Stähler expects Geniom one, which has a price tag of a few hundred thousand dollars, to hit the market by the end of the year.

Room for improvement

There is still a lot of room for improvement in microarray technology, say players in the field. TeleChem International/arrayit.com, for example, is exploring the use of reflective substrates. Although still in the development phase, Schena says it seems that printing microarrays on mirrors rather than glass improves the signal-to-noise ratio by as much as 1,000%.

Several companies are pursuing the development of 'active' hybridization technologies. Advalytix, a recent spin-off from the Center for NanoScience at the Ludwig-Maximilians University of Munich, will begin shipping a hybridization device this month, which has no moving parts and is designed to speed up hybridization reactions, as well as to produce more homogeneous reaction conditions than with 'passive' hybridization, eliminating so-called edge effects. The mixer chip uses surface acoustic waves to control the motion of reagents. It is used in a 'sandwich' arrangement, with a conventional DNA microarray slide on the bottom, the mixer chip on top and the hybridization solution in between.

"Microarrays will get better over time and a lot of that will be in content as we better understand which genes are important and, specifically, perhaps which splice variants are most important," says Amorese. In addition to improvements in the probes themselves, he expects advances in labelling technologies for the sample nucleic acid, allowing researchers to use less starting material. As for chips in the clinic, Schena believes they will be there within five years, and probably a lot sooner on the genetic screening side. ■

Peer Stähler (left) and board members of febit with Geniom one.

FEBIT

Diane Gershon is Assistant Editor, New Technology at *Nature Medicine*.

- **www.microarray.org**
- **www.gene-chips.com**
- **www.lab-on-a-chip.com**
- **cmgm.stanfordedu/pbrown/mguide/index.html**
- **www.mged.org/**

Life Sciences Research Task Force of the Object Management Group (OMG), a software standards organization, moved a step closer to implementation after a recent vote within the OMG.

"It all boils down to whether we want to continue in the life sciences with a tradition that the supporting data should be available, or not," says Alvis Brazma, team leader for microarray informatics at the EBI. Brazma is responsible for spearheading efforts to adopt minimum standards for microarray data and a standard data-exchange format.

The MGED has sought the input of the microarray community, including software and hardware companies. Rosetta Inpharmatics, for example, was working on its own standard, but has since joined forces with the MGED. "Our goal was to have a standard that everyone would use and that was at risk if we had a lot of smart folks working on two different applications," says Doug Bassett, vice president and general manager of Rosetta Biosoftware, the recently formed software arm of the company. Bassett expects the company's software products, which include the Rosetta Resolver gene-expression data analysis system, to be among the first to offer full support for MAGE-ML.

EBI's ArrayExpress currently houses only three data sets, but it now accepts data in the MAGE-ML format. The EBI is beta-testing the web-based data submission capabilities for ArrayExpress, and Brazma expects this phase to last another 2–3 months.

The GEO, launched by the NCBI last July, has been operational for longer, contains more data, and both accepts data submissions and supports data queries. But some researchers find it difficult to work with. "GEO has the disadvantage that all of the data are stored basically as a big tab-delimited file inside the database. That makes it very difficult to query," says Quackenbush. The NCBI is developing a set of tools on top of the GEO to try to extract the information and make it more accessible. Yoshio Tateno, of the Center for Information Biology, part of the National Institute of Genetics in Mishima, Japan, expects CIBEX to be publicly accessible and support MAGE-ML some time this summer.

JEREMY HASSEMAN

Quackenbush: supports data standards

Some private databases are also working towards supporting MAGE-ML and being MIAME-compliant. Gavin Sherlock, director of Microarray Informatics at the Stanford Microarray Database, hopes the database will be MIAME-compliant by the end of this year. "One of the things that makes it hard for us is the quantity of data we already have," he says, which amounts to information from some 22,000 arrays.

The MGED is also about to come up with a checklist for authors, editors and reviewers of what information should be given in microarray-based papers and what supporting information should be revealed electronically — details of which will be posted on its website. Brazma hopes it will serve as a useful guide that "will put everything on a more level playing field". **D.G.**