nature genetics
volume 14 no. 4 december 1996

To affinity ... and beyond!

Santa Clara, California. The notion that the next paradigmatic advance in genetic analysis might take place on a silica chip the size of a thumbnail would have been dismissed just a few years ago. But if the buzz at the recent annual gathering of the American Society of Human Genetics (ASHG) is anything to go by, scientists are becoming increasingly enthralled with DNA microchip technology, especially as practiced by a young biotechnology company named Affymetrix, located in an unassuming industrial park complex just south of San Francisco. The company's GeneChip technology has captured the imagination of a number of academic and corporate investigators, spawning applications ranging from DNA sequencing and mutation detection to expression monitoring and tagging of genes. Three of these applications are described in a series of new reports[1–3] — two in this issue of *Nature Genetics* and one in *Nature Biotechnology* — that graphically illustrate the disarming simplicity and potential of the GeneChip technology.

In 1991, the year that Affymetrix was created as a division of Affymax, Steven Fodor (now president of Affymetrix) and colleagues published an award-winning paper adapting photolithography for synthesizing peptides on a solid support[4]. (Among the co-authors were the three other co-inventors of the chip technology patent — Lubert Stryer (author of the best-selling textbook *Biochemistry*), Michael Pirrung and Leighton Read.) However, it quickly became evident that nucleic acids, not peptides, would be the most malleable raw material. Five years ago, Fodor predicted[4]: "Oligonucleotide arrays produced by light-directed synthesis ... would be valuable in gene mapping, fingerprinting, diagnostics, and nucleic acid sequencing." That prophecy is now being realised. The technology itself has been focussed in some earlier publications[5–9]. The approach is an expedient marriage between light-directed synthetic (photolithography) methods routinely used in the semiconductor industry and standard oligonucleotide synthetic chemistry (see Box). The result is that in about a day, a series of microchips can be prepared containing potentially hundreds of thousands of oligonucleotides of predetermined sequence, which are designed to interrogate a DNA or RNA sample for sequence or expression information.

The applications of these microchips generally fall into one of two classes: diagnostics and genomics. The potential in both areas can be gleaned from the elegant recent study from Mark Chee and colleagues on the resequencing of the complete 16.6-kilobase (kb) mitochondrial (mt) genome[10]. The mtDNA chip consists of some 136,000 oligonucleotides (25mers) spaced on the chip at a resolution of 35

Box Chips Ahoy!

The process of synthesizing potentially hundreds of thousands of oligonucleotides on a chip is deceptively simple. In the first step, a mercury lamp is shone through a standard computer-industry photolithographic mask onto the synthesis surface, activating specific areas for chemical coupling with a nucleoside which itself contains a 5' protecting group. Further exposure to light removes this group, leaving a 5'-hydroxy group capable of reacting with another nucleoside in the subsequent cycle. The choice of which nucleotides to activate is thus controlled by the composition of the mask. Successive rounds of deprotection and chemistry can result in an exponential increase in oligonucleotide complexity on a chip for a linear number of steps. For example, it requires only (4 x 15) = 60 cycles to synthesize a set of 15mer oligonucleotides — whether a few thousand, or a complete inventory of every possible 15mer, numbering more than one million[5–7].

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The space occupied by each specific oligonucleotide sequence is termed a 'feature', housing a million or more molecules. A 1.6-cm^2 chip can easily house several hundred thousand features at a resolution of 20 μm, and more than 1 million different probes could be synthesized on a chip with a spacing of 10 μm. The computer industry can achieve resolutions of 0.3 μm, and Rich Rava, Affymetrix's vice president of research and engineering, thinks that resolution will approach the 1-μm level in the next few years.

At the manufacturing plant at Affymetrix, the microchips start life as a 'wafer' or grid of 49–400 chips. Two technicians shuttle the wafers between the lithography machine and the oligonucleotide synthesizer, with each step taking only 15 minutes or so. (Waiting in the wings is a new machine that will perform the chemistry on six wafers in parallel, ensuring that the lithography machine seldom lies idle.) After the DNA synthesis is complete, the individual chips are diced apart, and each one housed in a sealed flow chamber, ready for hybridization. The hybridization reactions take place in a custom-designed 'fluidics station' using fluorescently labelled DNA or RNA as the target sequence. Once the washing procedure is completed, the flow chamber is transferred to a scanner, where the fluorescent signal is detected by scanning confocal microscopy in a process that takes less than 30 minutes. The current scanner is manufactured by Molecular Dynamics, and can read chips containing 65,000 probes at a resolution of 50 μm. In 1997, a new reader will be available from Hewlett Packard that can discriminate some 400,000 probes on a single chip, at 20-μm resolution. (At 40 probes per gene, this means that the complete human genome could probably be represented on a mere ten chips.)

The design of the chips generally follows one of two schemes. If the application is mutation detection or sequence analysis, where every nucleotide position of a gene has to be interrogated, a complete series of four oligonucleotides spanning each position in the sequence will be designed, differing only in the identity of the central base. The relative intensity of hybridization to each quartet of probes reveals the identify of the base at that location. (Note that the key element is the relative intensity at a given location; intensity of hybridization will vary across a long gene (or genome) depending on the characteristics of the sequence.) If a mutation has occurred, then a different probe will light up because of the mismatch. However, the overlapping oligonucleotides which are designed to interrogate the flanking nucleotides will not hybridize as strongly as before, leaving a characteristic 'footprint' or loss of signal[1,10].

For applications monitoring the expression of genes, where details of the precise sequence are not required, a set of approximately 20 to several hundred suitable oligonucleotides spanning the length of a gene are designed following criteria such as uniqueness and hybridization characteristics, and laid out in horizontal rows on the chip. Results so far suggest that the system can monitor differences in expression over three orders of magnitude[3].

The price of the commercially available HIV chip is $100 for two chips, although Affymetrix's Rob Lipshutz says that the price of future chips will vary according to the amount of information they can extract. The complete scanner system is $120,000, comparable to the price of an automated DNA sequencer.

μm. The complete sequence can be read with 99% accuracy and allows unequivocal detection of known mtDNA mutations. The speed of the fluorescent detection steps is such that a complete genome sequence can be obtained roughly once every 12 minutes. Chee and colleagues are presently exploiting the chip to study mtDNA variation patterns between and within populations, and intend to address other issues such as the interaction between the mitochondrial and nuclear genomes, and the possible role of mtDNA variation in neurological disorders and longevity.

More specific diagnostic applications are already in production. Affymetrix already has a commercially available HIV chip which contains 1,040 bases of the protease and reverse transcriptase genes, to correlate mutations with the development of viral resistance. As described earlier this year, a study of HIV-1 viral isolates detected variation in almost 50% of 99 residues of the protease gene[11]. But there is immense interest in applying the technology to screen for cancer, especially in individuals who may be carrying a mutation that places them at risk of the disease. On page 441, Francis Collins, Joseph Hacia and co-workers describe their results[1] in screening individuals for mutations in the breast cancer gene, *BRCA1*. Hacia *et al.*[1] placed almost 100,000 20mer oligonucleotides on a chip, matching the wild-type sequence of exon 11 as well as substitutions and small insertions and deletions. The assay measures the relative hybridization to the oligonucleotide array of wild-type and test RNA targets, labelled with fluorescent green and red tags respectively. Equal hybridization of the two transcripts produces a mixed yellow signal. RNA from a heterozygote carrying a *BRCA1* substitution will hybridize to two oligonucleotides interrogating the specific nucleotide, revealing the presence of the mutant transcript. If a frameshift mutation exists, then a loss of signal will leave a 'footprint', or loss of signal, easily visualized by comparing the intensity ratio of the reference and test targets. In all, 14 of 15 heterozygous *BRCA1* exon 11 mutations were correctly identified. Thus, this study shows that individuals who are heterozygous carriers for a mutation can still be detected despite the presence of the wild-type gene. The system can potentially be adapted to cover any gene of interest — for example, a 'p53 chip' is close to being ready for production.

The next generation: Apart from mutation detection and sequence analysis, the microchips are being exploited for the creation of a third-generation genetic map. Affymetrix is collaborating with Eric Lander's group (Whitehead Institute, Cambridge, Massachusetts) to develop a set of single nucleotide polymorphic (SNP) markers, which can be easily scored on the microchip. Preliminary data presented by Lander at the ASHG meeting suggest that these SNPs occur roughly once every 1,000 bases in the human genome. Because these markers are biallelic, a larger number of SNPs is required to extract the same amount of information as the multi-allelic microsatellite markers — about 2.25 times as many, says Lander. His group has already isolated about 450 SNPs, most of which work robustly, and has its sights set on an initial target of 2,000. Lander foresees a wealth of genomic applications, including linkage and association studies (including haplotype analyses), loss of heterozygosity and DNA fingerprinting[12].

But the uses of the microchip are by no means confined to human genome analysis. As described on page 450, Ron Davis and colleagues have used the chips to detect yeast deletion mutants[2]. Researchers are systematically disrupting the estimated 6,000 genes in *Saccharomyces cerevisiae* to analyse their functional effects. Davis and colleagues[2] have performed a pilot study in which 20mer oligonucleotides serve as unique 'tags' for 11 auxotrophic yeast mutants, facilitating the identification of each deletion strain after selection. This 'molecular barcoding' approach can be scaled up for a full-scale functional assault on the yeast genome.

Another exciting use of the technology lies in the realm of gene expression. As described in this month's *Nature Biotechnology*, David Lockhart and colleagues at Affymetrix (in collaboration with Genetics Institute) have produced the first oligonucleotide arrays to monitor gene expression[3]. Their results show that transcripts can be measured across several orders of magnitude down to levels of just a few copies per cell, allowing detection of increases in gene expression following stimulation of a T-cell line. As few as 20 probe pairs per gene appear to be sufficient for reliable monitoring of gene expression (for moderately expressed transcripts), potentially allowing as many as 10,000 genes to be encoded on a single chip. This system is being adapted by Dave Mack and co-workers to study changes in gene expression in cancerous cells. Mack has compared the expression profile in normal and malignant breast tissue on two different kinds of chips — one encoding 250 cancer-related genes, the other several thousand expressed sequence tags (ESTs). Mack sees significant alterations in expression for several of these genes associated with malignancy. A related but complementary use of DNA chips for studying gene expression in cancer is presented on page 457 by the groups of Patrick Brown and Jeffrey Trent, who have spotted cDNAs, rather than oligonucleotides, onto chips to analyse expression changes in a human melanoma cell line following suppression of the tumorigenic phenotype[13]. Using a similar colour assay to the *BRCA1* work, Brown and Trent compared the expression of nearly 900 genes in two cell lines, finding that 1.7% were downregulated whereas more than 7% (including *WAF1*) increased expression in conjunction with suppression of tumorigenicity. The cDNA approach may prove more popular in the short term, although Affymetrix's director of corporate development, Rob Lipzhutz, believes that the luxury of analysing splice variants as well as mutations using oligonucleotide arrays will eventually prove irresistible.

How widespread will this technology become? Fodor says that sequence analysis on DNA chips is able to link genetic variability with function. He is optimistic that the clinical applications linked to the technology will grow dramatically over the next few years, but is less certain whether the present intense interest in genomics applications can be maintained in the long term. Nevertheless, Affymetrix has already struck deals with Genetics Institute and Incyte to study gene expression; with Glaxo on gene polymorphisms; with Roche Molecular Systems on cystic fibrosis; and with other companies to examine bacterial classification. Despite this corporate interest, company executives know it will not be easy to place their technology into widespread use. There are many potential problems ahead, such as regulatory hurdles, competition from rival technologies, patenting issues and so on. But the technology has clearly reached the point where meaningful questions can be answered experimentally, not just theoretically. Some foresee the chips being used in non-invasive prenatal diagnosis screening procedures; others dream of the unprecedented analysis of the entire complement of human genes on one (or at most a few) high-resolution chip(s), highlighting the genetic interactions at play in common diseases. It will be interesting to watch the progress of this scientifically as well as aesthetically pleasing technology over the next few years … and beyond!

1. Hacia, J.G., Brody, L.C., Chee, M.S., Fodor, S.P.A. & Collins, F.S. *Nature Genet.* **14**, 441–447 (1996).
2. Shoemaker, D.D., Lashkari, D.A., Morris, D., Mittmann, M. & Davis, R.W. *Nature Genet.* **14**, 450–456 (1996).
3. Lockhart, D. *et al. Nature Biotech.* (in the press).
4. Fodor, S.P.A. *et al. Science* **251**, 767–773 (1991).
5. Fodor, S.P.A. *et al. Nature* **364**, 555–556 (1993).
6. Pease, A.C. *et al. Proc. Natl. Acad. Sci. USA* **91**, 5022–5026 (1994).
7. Lipshutz, R.J. *et al. Biotechniques* **19**, 442–447 (1995).
8. Southern, E.M. *Trends Genet.* **12**, 110–115 (1996).
9. O'Donnell-Maloney, M.J., Smith, C.L. & Cantor, C.R. *Trends Biotech.* **14**, 401–407 (1996).
10. Chee, M.S. *et al. Science* **274**, 610–614 (1996).
11. Kozal, M.J. *et al. Nature Med.* **2**, 753–759 (1996).
12. Lander, E. *Science* **274**, 536–539 (1996).
13. DeRisi, J. *et al. Nature Genet.* **14**, 457–460 (1996).