DNA chips, microarrays and genomics

P. K. Gupta*, Joy K. Roy and Manoj Prasad

During the last few years (1995–99), genomics research has witnessed an unprecedented progress due to the availability of microarrays on DNA chips. For an organism these microarrays allow parallel acquisition of massive data for thousands/millions of specific DNA sequences, which can be analysed using automated computer devices. The microarrays can be produced using in situ synthesis of oligonucleotides of known sequences at specific sites or by deposition, at specific sites, of DNA molecules already prepared. Robotics are used to achieve automation in the production and the microarrays thus produced may involve oligonucleotides or cDNAs, that are used as probes for hybridization with labelled genomic DNA or mRNA/cDNA used as the target. These microarrays are being put to a variety of applications including the following: (i) DNA sequencing/resequencing, (ii) single nucleotide polymorphisms (SNPs), (iii) functional genomics, (iv) studies involving reverse genetics, (v) diagnostics and genetic mapping, (vi) genomic mismatch scanning (GMS), (vii) agricultural biotechnology and (viii) proteomics. These applications of DNA chips will be realized and will expand in future for the benefit of mankind, in the fields of health care and agricultural biotechnology.

DURING the last three decades, due to development of newer technologies, significant progress has been made in the study of animal and plant genomes. In the 1960s and the 1970s, using cytospectrophotometry, estimates were made for the genome size (nuclear DNA content) in a large number of animal and plant species. Concurrently, during the 1970s, using the technique of reassociation kinetics, studies were also undertaken on estimation of relative proportions of repetitive and unique DNA sequences and their interspersion within the genomes of many species. However, the major revolution in the study of genomes of eukaryotes, was brought about due to availability of recombinant DNA and PCR technologies that became available during 1970s and 1980s, respectively. These technologies helped in the preparation of molecular maps (both genetic and physical maps) for a variety of animal and plant genomes. However, ideally the objective of genomics research in any species should be to sequence its entire genome (as been done for the human genome) and to decipher functions of all the different coding and non-coding sequences. A study of the structures of all proteins encoded in DNA sequences is also a part of the genomics research, popularly described as ‘structural genomics’. In most eukaryotes, the sequencing of an entire genome may not be possible, except for small genomes represented by organisms like yeast, human, nematode, fruit fly and Arabidopsis thaliana. However, a study of the functions of all the sequences in the above organisms will certainly take time. Since sequencing an entire genome is not possible in most other cases in the foreseeable future, the genomics technology is being used to develop expressed sequence tags (ESTs), which provide a cost effective and rapid approach for describing all the genes of an organism. Developing complete EST database for individual genomes, in itself, is also an enormous task and cannot be easily achieved using cloning/PCR technologies. In view of this, during 1990s, while PCR technology was increasingly used to replace and/or supplement the recombinant DNA technology for a variety of purposes, yet another remarkable tool had emerged for genomics research. This new technology is popularly described as DNA chip technology and, as believed by some, will play in future the same role in the field of genomics, which in the past, the semiconductors played in the field of electronics. A comparison of the work already done using DNA chip technology with that done using other technologies would also suggest that the pace of research in the area of genomics has greatly accelerated due to development of this new technology. The DNA chip technology facilitated parallel acquisition of massive data for thousands/millions of specific DNA sequences at a fast rate through automation, followed by the analysis of this data using computer devices. Thus, during the last few years this technology has revolutionized research in the area of genomics, which has been described either as ‘functional genomics’ or as ‘structural genomics’. While ‘functional genomics’
involves a study of the functions of all specific gene sequences and their expression in time and space in an organism, structural genomics involves the task of solving the structures of all proteins encoded in a fully sequenced genome. Due to increasing demand of DNA chips for genomics research, contracting services have also become available to facilitate the use of DNA chips, by those who do not have the facility for production of microarrays on DNA chips. In this article, a brief overview of this technology is given and its various applications are discussed.

**Characteristic features of microarrays**

The characteristic features of DNA chips/microarrays are: (i) Parallelism: Microarray analysis allows parallel acquisition and analysis of massive data. This greatly increases the speed of experimental work. It allows meaningful comparison between genes or gene products represented in microarrays and may eventually allow the analysis of the entire genome of any organism in a single reaction. Recent gene expression experiments in yeast are important examples of achieving this goal. (ii) Miniaturization: Microarray analysis involves miniaturization of DNA probes and reaction volumes, thus reducing time and reagent consumption. (iii) Speed: Microarray analysis is highly sensitive and allows rapid data acquisition with either confocal scanner or cameras equipped with charged coupled devices (CCD). (iv) Multiplexing: This is a process by which multiple samples are analysed in a single assay. The labelling and detection methods that involve multicolour fluorescence allow comparisons of multiple samples on a single DNA chip. Multiplexing also increases the accuracy of comparative analysis by eliminating complicating factors such as chip to chip variation, discrepancies in reaction conditions and other shortcomings inherent in comparing separate experiments. It has already been used in expression analysis, genotyping and DNA resequencing. (v) Automation: Advanced manufacturing technologies permit the mass production of DNA chips, and the automation led to proliferation of microarray assays by ensuring their quality, availability and affordability. As a result, DNA chips may eventually become like commodity items in the computer industry. (vi) Combinatorial synthesis: Using the combinatorial synthesis strategy, a set of all $4^k$ oligonucleotides of the length $k$ nucleotides ($k$-mers) can be generated in $4k$ synthesis cycles. For example, the set of all 4-mers (256) can be synthesized in 4 rounds, each round having 4 cycles, thus making a total of 16 cycles (Figure 1).

**Types of DNA chips (microarrays) and their production**

Two major types of DNA chips are available for DNA analysis.

**Oligonucleotide-based chips**

This type of DNA chips contains a high density of short oligonucleotide microarrays which are prepared by photolithography. Such arrays contain 100,000 to 400,000 oligonucleotides immobilized within an area of 1.6 cm$^2$ (ref. 2). This allows the use of targetted regions of genomic DNA for sequencing or for a large-scale analysis of single nucleotide polymorphisms (SNPs).

**DNA-based chips**

This type of DNA chips contains a high density of DNA microarrays, most often derived from cDNA (ref. 4) (hence they are also called cDNA microarrays). These chips are currently made by robotically spotting a large number of PCR-amplified DNA fragments onto glass or nylon surfaces. The hybridization is carried out with fluorescently labelled mRNA or its corresponding cDNA and the hybridized duplexes are identified by colour fluorescence detection methods. These DNA chips, thus, can be used for studying gene expression patterns in time and space.

The above two types of microarrays (DNA chips) can be produced by using two different approaches: synthesis and deposition. In the synthesis approach, microarrays are prepared in a stepwise fashion by in situ synthesis of nucleic acids from biochemical building blocks, the nucleotides. With each round of synthesis, individual nucleotides are added to growing chains until the desired length is achieved. In the deposition or delivery approach, on the other hand, separately prepared samples of nucleic acids are deposited exogenously for chip fabrication. Molecules, such as cDNA fragments, are amplified by PCR and purified; small quantities of these fragments are then deposited onto known locations using a variety of delivery technologies. The key parameters for evaluating both the techniques include microarray density and design, biochemical composition, quality, cost and ease of prototyping.

**Production of oligonucleotide microarrays**

Oligonucleotide microarrays are produced by in situ synthesis techniques using phosphoramidites, surface linker chemistry and combinatorial synthesis strategies. There are two most common approaches for in situ synthesis of oligonucleotides on a solid surface.

Light generated oligonucleotide arrays (photolithography deprotection). It utilizes photolithographic deprotection of 5'-OH groups at specific sites on the chip, so that individual specific bases may be added at these deprotected sites during chemical synthesis (Figure 2a).
Initially, synthesis of entirely independent sequences at individual grid positions was not easy. But the use of photolithography in oligonucleotide synthesis has resolved the above shortcoming and allowed miniaturization of the arrays. Furthermore, the sequence of the oligonucleotide at each site is known. However, the synthesis of oligonucleotides longer than 25-mers is difficult with the currently available efficiency. The major disadvantage of photolithography is the need of photomasks, which are expensive and time consuming to design and build. The most popular manufacturers of oligonucleotide microarrays are listed in Table 1. The following main steps are involved in the production of oligonucleotide arrays:

(i) Synthesis support: The synthesis support consists of, for example, a glass substrate. The surface is derivatized with *bis*(2-hydroxyethyl) aminopropyltriethoxy silane. The 5'-O-(α-methyl-6-nitropiperonyloxy carbonyl)-N-2'-deoxynucleoside phosphoramidites are then attached to derivatized substrate through a synthetic linker e.g. 4,4'-dimethoxy triyl (DMT)-hexaethoxy-o-cyanoethyl phosphoramidite. (ii) Masking: Photolithographic masks having apertures are used to cover those regions, where a particular nucleotide is not desired to be added. (iii) Illumination: The surface of the solid support is illuminated, so that specific regions of the support are activated for synthesis, by illumination (wavelength more

---

**Figure 1.** Combinatorial synthesis of all 256 tetranucleotides. In cycle 1 of round 1, mask 1 activates one-fourth of the substrate surface for coupling with the first of four nucleosides (MeNPdC-dT). In cycle 2 of the round 1, mask 2 activates a different quarter of the surface for coupling with the second nucleoside (MeNPdC-dC). The process is continued to produce four regions of four different mononucleotides. The masks of the next three rounds are perpendicular to that of the previous round. By repeating the above activation and coupling processes, 16 di-, 64 tri- and 256 tetranucleotides are produced in round 2, 3 and 4, respectively. (Note: Solid black strips represent activated deprotected areas, they should not be confused with masks.)
than 280 nm) through microapertures (~800 x 12800 μm) of the photolithographic mask. Because the bases have strong π-π transitions in the 280 nm region, the deprotection wavelength should be longer than 280 nm to avoid undesirable nucleoside photochemistry\(^1\). (iv) Coupling reaction: A 3'-O-phosphoramidite activated deoxynucleoside (its 5'-hydroxyl is protected with a photolabile group) is presented to the surface that was illuminated, and coupling occurs at sites that were exposed to light. Following this, the substrate is rinsed and the surface is illuminated through a second mask, to expose additional hydroxyl groups for coupling. A second 5'-protected, 3'-O-phosphoramidite-activated deoxynucleoside is presented to the surface. The selective photodeprotection and coupling cycles are repeated until the desired set of products is obtained\(^1\).\(^5\) (Figure 2a).

Since the synthesis areas are defined by the photolithographic process, high density arrays can be formed using the above steps. The relationship between the size of each synthesis site and the density of synthesis sites (a synthesis site being the region in which a homogeneous set of probe molecules is synthesized) is given in Table 2, and the relationship between probe length, number of chemical synthesis steps and number of synthesis sites is given in Table 3. To support the production of increased number of arrays, instrumentation has been developed to synthesize multiple arrays on a single large substrate (wafer). After synthesis, the wafers are diced to yield

---

**Figure 2.** Three different approaches for microarrays production technologies. a, photolithography; b, mechanical microspotting; c, ink jetting.

**Table 1.** Some firms producing DNA chips

<table>
<thead>
<tr>
<th>Company</th>
<th>Contact information</th>
<th>Key products and services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affymetrix</td>
<td>Santa Clara, CA, USA</td>
<td>GeneChip(^TM) technology, microarray contract service, complete microarray system</td>
</tr>
<tr>
<td>Alphagene</td>
<td>Woburn, MA, USA</td>
<td>AlphaGenomics(^TM), full-length cDNAs, microarray contract services</td>
</tr>
<tr>
<td>Genome Systems</td>
<td>St. Louis, MO, USA</td>
<td>Expressed sequence-tag libraries</td>
</tr>
<tr>
<td>Genomatrix</td>
<td>The Woodlands, TX, USA</td>
<td>Microarray technology platform, contract services</td>
</tr>
<tr>
<td>Hyseq</td>
<td>Sunnyvale, CA, USA</td>
<td>HyChip(^TM) products, genomics plateform, contract services</td>
</tr>
<tr>
<td>Incyte Pharmaceutical</td>
<td>Palo Alto, CA, USA</td>
<td>LifeSeq (^TM) database, Genetec(^TM) and GEM(^TM) technology, microarray contract services</td>
</tr>
<tr>
<td>Intelligent Automation Systems</td>
<td>Cambridge, MA, USA</td>
<td>Custom automation, microarray instrumentation, contract work</td>
</tr>
<tr>
<td>Molecular Dynamics</td>
<td>Sunnyvale, CA, USA</td>
<td>Microarray Technology Access Programme, complete microarray systems</td>
</tr>
<tr>
<td>Nanoegen</td>
<td>San Diego, CA, USA</td>
<td>APEX(^TM), electronic microarray technology, contract services</td>
</tr>
<tr>
<td>Protesyn Laboratories</td>
<td>Palo Alto, CA, USA</td>
<td>Ink jetting technology, microarray contract services</td>
</tr>
<tr>
<td>Synteni</td>
<td>Fremont, CA, USA</td>
<td>GEM(^TM) technology, microarray contract services</td>
</tr>
<tr>
<td>Sequenom</td>
<td>Hamburg, Germany, and San Diego, CA, USA</td>
<td>Integrated DNA chip-based diagnostic system</td>
</tr>
<tr>
<td>Atomic Sciences</td>
<td>Oak Ridge, TN, USA</td>
<td>PNA chip</td>
</tr>
</tbody>
</table>

---

878 CURRENT SCIENCE, VOL. 77, NO. 7, 10 OCTOBER 1999
individual arrays. These arrays are packaged in individual injection-molded flow cells, making them easier to handle.

Ink jet: Ink jet printing technology can also be used for in situ production of oligonucleotide microarrays. It involves direct delivery of tiny droplets of phosphoramidite solutions to precise locations on a surface rather than the use of photolithography masks. Piezoelectric-based delivery of phosphoramidite reagents has recently been used for the manufacture of high density oligonucleotide microarrays.

Production of DNA (e.g. cDNA) microarrays

DNA microarrays are produced by post synthesis attachment of DNA probes. Their attachment on a solid surface is done by delivery/deposition of DNA probe aliquots at specific sites on the surface of the chip. This delivery can be performed manually or robotically. The two delivery technologies used extensively for the production of DNA microarrays are:

Mechanical microspotting: Microspotting, a miniaturized version of earlier DNA spotting techniques, produces an automated microarray by printing small quantities of pre-made cDNA fragments onto a solid surface (Figure 2. b). Printing is accomplished by direct surface contact between the printing substrate and a delivery mechanism that contains an array of tweezers, pins or capillaries that serve to transfer biochemical samples to the surface. The microspotted microarrays currently manufactured at Synteni (Fremont, CA, USA) contain each as many as 10,000 groups of cDNAs in an area of ~3.6 cm². Although microspotting is unlikely to produce the densities achievable by photolithography, improvements are being made in microspotting technologies for automated production of chips containing 100,000 groups of cDNAs in an area of ~6.5 cm².

The following steps are involved in microspotting for production of cDNA microarrays: (i) probe DNA samples are prepared generally through amplification of cloned cDNA fragments using PCR; (ii) a 1 x 3 inch microslide with chemically modified surface is prepared for microspotting; (iii) DNA samples prepared as above, are loaded each into a spotting pin by capillary action and a small volume is transferred to chemically modified solid surface at a specified position; this is achieved by a physical contact between the pin and the solid surface; (iv) after the first spotting cycle, the pin is washed and the second sample is similarly loaded and deposited at an adjacent position. These steps are speeded up by robotic control systems and multiplexed printheads. Common advantages include ease of prototyping, rapid implementation, low cost and versatility. One disadvantage of microspotting, however, is that each sample must be synthesized, purified and stored prior to microarray fabrication.

Ink jets: Ink jets provide another approach to the manufacture of microarrays. The most advanced version of this technology includes piezoelectric and other forms of propulsion to transfer biochemical substances from miniature nozzles to solid surfaces (Figure 2. c). Thus, like microspotting approaches, ink jets also allow high density grading of any biomolecule of interest including cDNAs, genomic DNAs, antibodies, etc.

Hybridization and detection methods

Hybridization of the target DNA to a microarray yields sequence information. The target DNA is labelled and incubated with the array. If the target DNA has regions complementary to the probes on the array, then the target DNA will hybridize with these probes. Under a fixed set of hybridization conditions, e.g. target concentration, temperature, buffer and salt concentration, etc., the fraction of probes bound to targets will vary with the base composition of the probe and the extent of the target-probe match. In general, for a given length, probes with high GC content will hybridize more strongly than those with high AT content. Similarly, probes matching the target will hybridize more strongly than probes with mismatches, insertions and deletions. Various detection methods are currently available for the analysis of hybridization patterns on microarrays of immobilized probes. Some rely on the use of enzymes to enable detection, while others detect hybridization directly.

For the detection of hybridization patterns on DNA chips, the technique of reverse dot-blot, used earlier on

<p>| Table 2. Photolithographic resolution and synthesis site density |
|---------------------------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Resolution (μm)</th>
<th>Synthesis site density (sites/cm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>400</td>
</tr>
<tr>
<td>200</td>
<td>2500</td>
</tr>
<tr>
<td>100</td>
<td>10000</td>
</tr>
<tr>
<td>50</td>
<td>40000</td>
</tr>
<tr>
<td>20</td>
<td>25000</td>
</tr>
<tr>
<td>10</td>
<td>100000</td>
</tr>
</tbody>
</table>

<p>| Table 3. Combinatorial synthesis |
|----------------------------------|------------------|------------------|</p>
<table>
<thead>
<tr>
<th>Probe length (mers)</th>
<th>No. of chemical steps</th>
<th>Number of possible probes</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>16</td>
<td>256</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>65536</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>1048576</td>
</tr>
<tr>
<td>15</td>
<td>60</td>
<td>1073741824</td>
</tr>
</tbody>
</table>
the membranes, is utilized. The technique is so described, because as opposed to dot-blots, where the target DNA is dot blotted on the membrane and the probes are labelled, on DNA chips, the probes are anchored in the form of microarrays and the target DNA is labelled. Once hybridization is completed, the detection of hybridization is achieved either with the help of an enzyme system (enzyme-assisted detection) or directly due to radio-labelling and/or fluorescence.

The target DNA is either nonradioactively labelled (biotin or digoxigenin labelling) or radioactively labelled, the former requiring enzymatic detection and the latter requiring direct detection through autoradiography, gas phase ionization and phosphorimagers. However, there are drawbacks with the detection methods involving radioactivity (low resolution, etc.). In order to circumvent these problems, fluorochromes may be used which will also allow direct detection due to fluorescence. This would also allow multiplexing, where more than one target DNA labelled with different fluorochromes can be used for hybridization of microarray on the DNA chips. The hybridization patterns can be scanned in this case using automatic scanner (e.g. Scan Array 3000). These detection systems are based either on lens-based systems (epifluorescent and confocal microscopes) or on CCD-based systems. The lens-based systems, including confocal microscopy, allow selective detection of the surface bound molecules, as opposed to those in the surrounding fluid medium. However, these are not well suited to the level of miniaturization already achieved in DNA chip technology. Therefore, more recently CCD detection systems have been developed to detect small quantities of array-bound molecules. In this method labelled target DNA is hybridized to an immobilized probe on a silicon wafer; the wafer is then placed on the CCD surface and a signal is generated. A fluorescence microscope fitted with a CCD camera and a computer is used. This will allow sequencing by hybridization with an oligonucleotide matrix (SHOM).

Applications

Microarrays have a large number of applications, that will expand in future. These include:

DNA sequencing by hybridization

The two popular methods of sequencing include the Sanger's dideoxy synthetic method, and the Maxam and Gilbert's degradation method. Sanger's method is even currently used as a routine method for DNA sequencing. However, the efficiency, cost and reliability of the above two methods, were not able to cope with the requirements of large-scale genome sequencing. Therefore, in the late 1980s, a new approach towards DNA sequencing was suggested simultaneously by four groups. The approach was described as sequencing by hybridization or SBH. The method involves manufacturing the sequencing DNA chips that contain a complete set of immobilized oligonucleotides of a particular size (e.g. 8-mers), and hybridization of the target DNA of unknown sequence (whose sequence is to be determined) onto these DNA chips. The hybridization patterns are then recorded using one of the several suitable devices discussed earlier in this article. Identification and analysis of the overlapping oligomers that form perfect duplexes with the DNA of interest permits reconstruction of the target DNA sequence (Figure 3). During the 1980s, it was believed that SBH using microarrays carrying all the possible 65,536 octamer oligonucleotides could possibly be used as an alternative to Sanger's dideoxy and Maxam and Gilbert's methods of sequencing. However, this objective has not been successfully achieved, since uniform hybridization signals are not available for a large number of oligonucleotides in parallel due to sequence-dependent variability in heteroduplex formation. This leads to false positives and false negatives, so that unambiguous determination of an unknown sequence is not always possible. Further complications arise due to repeated sequences. Consequently, the technical barriers of SBH are now obvious, and microarrays which were initially considered to be useful only for SBH, are now used for a variety of other purposes.

Single nucleotide polymorphisms and point mutations

Restriction fragment length polymorphisms (RFLPs) and simple sequence length polymorphisms (SSLPs) were the markers of choice in the past, but these markers had some drawbacks. For instance, they need gel-based assays and are, therefore, time-consuming and expensive. Recently, single nucleotide polymorphisms (SNPs), as biallelic genetic markers, have been extensively used as the markers of choice. Although they have the disadvantage of being biallelic as against SSLPs, which are polyallelic, their abundance (more than 1 per 1000 bp) makes them attractive. Genotyping individuals using SNPs needs only plus/minus assay, permitting easier automation. Further, high density oligonucleotide arrays allow genotyping at a large number of these biallelic loci in parallel. The approach used for this purpose relies on the capacity to distinguish a perfect match from a single base mismatch.

A set of four groups of oligonucleotides of known and related sequences is used, such that corresponding oligomers from the four groups differ only for the central base. For this purpose, a 'tiling strategy' proposed by Affymetrix makes use of a microarray of 40,000 oligomers for resequencing a 10 kb gene. The use of this strategy is shown in Figure 3 and can be used for
detecting SNPs or point mutations. As shown in the
figure, in this strategy, a number of columns are available
and in each column, there are four sites, each having an
oligomer that differs from the other only at the central
position. In each row, the oligos at the adjoining sites
represent sequences that have an overlap of 14 out of the
15 nucleotides. The sequences of 24 oligomers for a chip
containing 6 columns are given in Figure 3.

Use of SNPs offered great promise for rapid and highly
automated genotyping, leading to rapid advancement in
developing human genetic map. Earlier methods for de-
velopment of SNPs were dependent primarily on gel-based
sequencing of DNA from several individuals and have
therefore been relatively slow and expensive. Several
semi- or fully automated techniques including DNA chips
are being used now in the human genome project for
development of a large number of SNP markers. Recently,
the ‘First International Meeting on SNPs and Complex
Genome Analysis’ was held in Sweden (29 August to
1 September 1998), signifying the importance of SNPs. It
was emphasized in this conference that there are also
some problems with this technology, since association of
SNPs with individual traits, can break due to recombina-
tion, thus making it necessary to have many SNPs
associated with a trait. 5

Functional genomics

Microarrays for gene expression analysis provide an
integrated platform for functional genomics. Samples of
mRNA from a variety of cells and tissues are used for

\[
\begin{array}{ccccccc}
1A & G & C & T & A & C & T \\
1C & G & C & T & A & C & T \\
1G & C & T & A & C & T & A \\
1T & G & C & T & A & C & T \\
2A & C & T & A & C & T & A \\
2C & C & T & A & C & C & T \\
2G & C & T & A & C & G & T \\
2T & C & T & A & C & T & A \\
3A & T & A & C & T & A & T \\
3C & T & A & C & T & A & C \\
3G & T & A & C & T & A & C \\
3T & T & A & C & T & A & C \\
4A & A & C & T & A & T & G \\
4C & A & C & T & A & C & T \\
4G & A & C & T & A & C & G \\
4T & A & C & T & A & C & C \\
5A & C & T & A & G & C & T \\
5C & T & C & A & T & G & C \\
5G & T & C & A & C & T & G \\
5T & C & T & A & G & C & T \\
6A & T & C & A & G & C & T \\
6C & T & C & A & G & C & T \\
6G & T & C & A & G & C & T \\
6T & T & C & A & G & C & T \\
\end{array}
\]

(a) microarray

\[
\begin{array}{ccccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
A & 1 & 0 & 0 & 0 & 0 & 0 \\
C & 0 & 1 & 0 & 0 & 0 & 0 \\
G & 0 & 0 & 1 & 0 & 0 & 0 \\
T & 0 & 0 & 0 & 1 & 0 & 0 \\
\end{array}
\]

(b) hybridization pattern in
wild type

\[
\begin{array}{ccccccc}
T & A & C & G & C & A \\
\end{array}
\]

sequence deciphered

\[
\begin{array}{ccccccc}
\begin{array}{ccccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
A & 0 & 0 & 0 & 0 & 0 & 0 \\
C & 0 & 0 & 0 & 0 & 0 & 0 \\
G & 0 & 0 & 0 & 0 & 0 & 0 \\
T & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{array}
\end{array}
\]

(c) hybridization pattern in a
mutant

\[
\begin{array}{ccccccc}
\begin{array}{ccccccc}
T & A & C & T & C & A \\
\end{array}
\end{array}
\]

sequence deciphered

24 nucleotides arranged in 6 columns and
used in a tiling pattern shown on right

Figure 3. Tiling strategy for DNA resequencing with oligonucleotide microarrays to detect point mutation or SNP. In the schematic DNA chip there are 6 columns; in each column there are four
sites, each having an oligomer (15 mer) that differs from the others only at the central position
(underlined). The hybridization between the probe and target DNAs showing complete match is
represented by solid blocks (wild-type). Comparative fluorescence intensity analysis of each column
allows base-by-base resequencing of a wild-type (TACGCA) and mutant (TACTCA) gene. The single
base mutation in the mutant sequence leads to general reduction of signal (dotted blocks) in the
proximity of the point mutation.
microarray analysis and would yield information about specific changes in gene expression patterns. The mRNA samples of interest are labelled and used for hybridization-based microarray analysis, yielding quantitative data on the expression of thousands of cellular genes. Parallel measurement of transcript levels for thousands of genes is one of the most widespread uses of DNA chip technology. Both oligonucleotide and cDNA microarrays are very useful for estimating levels of transcripts. While developing a DNA chip for gene expression studies, the following three strategies can be adopted: (i) A limited number of genes involved in the biological pathway may be selected; this has been done in an analysis of genes for inflammatory diseases, where cDNA microarrays were used\(^{21}\). (ii) Clones from a cDNA library, prior to sequencing may be used as has been done in the human genome for monitoring 1000 genes\(^{20}\); in this strategy only cDNA microarrays can be used; (iii) A complete expressed sequence content of an organism is used, as done in yeast\(^{21}\). Through these genomic chips, novel expressed sequences can be detected\(^{21}\). In plant systems also, expression profile of a large number of genes in parallel has been achieved using cDNA microarrays\(^{3}\). More than 35,000 expressed sequence tags (ESTs) are now available in dbEST of Gene Bank and can be used for the preparation of cDNA micrarrays. In *Arabidopsis* alone, 16,000 genes are represented as ESTs, next to only human and mouse ESTs. In a more recent study more than 1400 *Arabidopsis* EST cDNA clones were used for the preparation of cDNA microarrays. These microarrays were used for hybridization leading to the preparation of expression profile of genes in major organs (root, leaf, flower) of *Arabidopsis* plants. Novel expression profiles were identified for many sequences, leading to the understanding of their possible functions. DNA chips can also be used for distinguishing the transcripts from individual members of multigene families that share extensive sequence homology. For this purpose DNA chips can be prepared synthesizing oligonucleotides representing the regions of nonidentity among members of a multigene family.

Reverse genetics and DNA chips

DNA chips can also be used for characterization of mutant populations exposed to various selection pressures, to collect information about the fitness value of a variety of alleles for each of the large number of genes in a species. Particularly in organisms where complete sequence of the genome is already available, post-genomics research activity may involve introduction of deletions/insertions or substitutions at will followed by analysis of their fitness (such an approach where we start a study with DNA sequence and conclude it with the analysis of phenotype is described as ‘reverse genetics’). This can be achieved, if the mutants are first subjected to a selection pressure and then characterized. This can be illustrated using the example of yeast, where the genome has been completely sequenced and was shown to carry 6000 open reading frames (ORFs). Unique molecular sequences or ‘bar codes’ can be introduced in each of the above 6000 ORFs in the yeast genome. A mixture of yeast strains containing individual bar-codes for all 6000 genes, is then subjected to a selection pressure. Samples of cells are taken and bar-code sequences are labelled using multiplex PCR with fluorescent primers. A pool of fluorescent amplicons is then hybridized to an oligonucleotide microarray containing sequences complementary to each of the amplified bar-codes, and after detection of fluorescent signals, an estimate of fitness of each strain under a given selection pressure can be worked out.

In species, where the genome sequence is not yet fully determined, ESTs can be used to identify mutants. In *Arabidopsis*, a pool of lines carrying insertion elements at random locations were used. Hybridization of PCR amplicons (derived from these lines carrying insertion elements) to microarray of ESTs can be used to identify mutant lines of *Arabidopsis*.

Diagnostics and genetic mapping

DNA chips are also being used for diagnostics. Since some information about the alleles belonging to genes responsible for a number of diseases is available, the search can be focused on a restricted number of polymorphisms, thus reducing the required number of features on a DNA chip. For instance, diagnostic chips have been prepared to detect mutant alleles in *CFTR* (cystic fibrosis), *BRCA 1* (cancer susceptible gene) and *beta globin* genes. For *CFTR*, one microarray containing 428 features was designed to detect mutations in exon 11 of *CFTR* and another microarray containing 1480 features was designed for detection of known deletions, insertions or base substitutions. Hybridization of genomic DNA samples from *CFTR* patients with already characterized mutations to diagnostic chips for *CFTR* gave expected results. Similarly, genotyping of patients with uncharacterized mutations by microarrays could be confirmed by techniques of RFLP and PCR. These results confirmed the utility of micrarrays in diagnostics. Similarly, in *BRCA 1* experiments conducted by Hacia et al.\(^{22}\), microarrays containing 96,000 20-mers were used to identify mutations over the entire 3.45 kb DNA belonging to exon 11. Utilizing samples from patients with known mutations, it was shown that no false mutations were available. Oligonucleotide DNA chips have also been used to detect *beta*-thalassemia mutations in patients by hybridizing PCR-amplified DNA with the DNA chips\(^{23}\). DNA chips containing microarrays of up to 135,000 probes comple-
mentary to the 16.6 kb human mitochondrial genome have also been developed for simultaneous analysis of the entire mitochondrial genome of individuals\textsuperscript{24}. DNA chip technology was also successfully applied to the genotyping of hepatitis C virus in blood samples\textsuperscript{25}. More recently, Sequenom (a German/US gene chip company), launched its fully integrated DNA chip-based diagnostic system. This system is fairly big in size and efforts will be made to assemble smaller, simpler and cheaper systems to be used conveniently by doctors in their clinics to identify as many as 10,000 genotypes per day. Efforts are being made to develop diagnostic systems, that would not need any labelling. In such systems, mass spectrometry will be used for direct electronic detection of DNA hybridization on the chip. Microarrays, composed of peptide nucleic acids (PNAs) will also facilitate the detection without label, so that phosphate ions of hybridized DNA would act as intrinsic labels that can be detected by mass spectrometry\textsuperscript{26}.

**Genomic mismatch scanning\textsuperscript{27}**

Genomic mismatch scanning (GMS) is a hybridization-based method for linkage analysis. Homologous segments are identified by the formation of heteroduplexes that are free of any mismatches. Fragments of chromosomal DNA representing inherited regions are hybridized to a microarray of ordered genomic clones and positive hybridization signals pinpoint regions of identity-by-descent at high resolution. The mapped PCR products could be used to prepare a microarray of physically ordered fragments, for use in detecting meiotic recombination breakpoints. GMS is only one example of the use of gene microarrays to characterize the composition of nucleic acid mixture subjected to in vitro selection. Restriction endonucleases protection, selection and amplification (REPSA) is another example of a selection method that could be adapted to a DNA microarray-based detection\textsuperscript{28}. REPSA makes use of a combination of restriction enzyme cleavage, PCR amplification and filter binding to selectively identify DNA sequences used for binding of DNA-binding proteins.

**DNA chips and agricultural biotechnology**

DNA chips with ESTs can also be used to collect data on expression in an agricultural crop under different conditions. This information can prove to be of practical utility in agricultural biotechnology. For instance, if the expression of genes on hormone application is known, hormone application can be monitored. Transgenic plants can also be rapidly analysed using microarray on DNA chips, and expression patterns under different environmental conditions can be predicted at the gene level. Action of herbicide can be similarly determined and decision be taken on the application of herbicides.

DNA microarray will also be extensively used in future, for a study of DNA polymorphism (e.g. SNPs) to develop molecular markers tagged to specific economic traits. The molecular markers thus developed can be used in diagnostics and for actual molecular marker aided selection in breeding programmes. The major advantage of DNA chips for developing molecular markers is the simultaneous analysis of thousands of polymorphisms in a single experiment. This will of course require a cost-effective microarray technology. The current excitement and activity in this area suggests that the complete microarray system will soon be available at an affordable price\textsuperscript{2}.

**Proteomics**

Like genomics, the term 'proteomics'\textsuperscript{29} relates to the study of protein–protein interactions. DNA chips can also be used for this area of study. Protein linkage maps can also be created using genomic sequence information. Protein–protein interactions can be studied using the yeast two-hybrid system. In this system, two fusion proteins are used for the activation of transcription of a reporter gene in yeast. The first fusion protein contains a DNA-binding domain fused to a protein of interest and the second fusion protein carries an acidic transcription-activation domain fused to a second protein of interest. Specific interaction between two chimeric proteins leads to transcriptional activation of the reporter genes which can be easily scored with colour-based assays. The identity of the two proteins of interest is confirmed by sequence analysis of each clone thus identified. Therefore major sequencing work is involved in the above two-hybrid system.

As an alternative to DNA sequencing needed in two-hybrid analysis as mentioned earlier, DNA chip arrays can be used to identify the genes involved in protein–protein interactions. In cases where the entire genome sequence is available as in *E. coli* or yeast, DNA chips can be used for parallel resequencing, so that clones involved in the two-hybrid system can be identified through single hybridization to genomic chips. Phage presentation libraries can also be used for DNA chip-based detection systems. This involves the use of fusion proteins encoded by chimeric sequences of phage viral coat protein and genes of interest.

**Conclusion**

DNA chip technology provides a variety of advantages, so that the currently used many gel-based and filter-based assays will be replaced in future by DNA chip-based microarray analysis. The availability of contract services through involvement of private sector will also make this technology cost-effective in due course.
parallel gene analysis, recently described in Ref. 13, will help in the rapid proliferation of this technology at an affordable cost. Microarray feature size will further decrease, thus increasing the density of these features on DNA chips, which will further increase their utility.


Received 16 January 1999; revised accepted 21 July 1999