better software integration and the industrialization of the necessary biology.

References
16 Sanchez, J. C. et al. (1997) Electrophoresis 18, 324–327
18 Cid, J. E. et al. (1995) Electrophoresis 12, 2177–2204
37 Anderson, L. and Bitlmen, J. (1997) Electrophoresis 18, 533–537

Powerful tools for genetic analysis come of age

David J. Graves

Microarrays of oligonucleotides or cDNA containing up to several hundred thousand different sequences are starting to influence methodologies and paths to discovery in genomics. Gene polymorphisms and mutations can be found and gene expression measured with unprecedented speed and parallelism. The principles of this modern technology and some of the problems awaiting further study are discussed.

Microarrays of DNA and oligonucleotides are beginning to have the same impact on the biological sciences that integrated circuits have already produced on the physical sciences, and for similar reasons: they can do many things in parallel, with very little material and with a modest investment of labor. It is too early to say whether complex microfabricated chemical systems will result in anything as revolutionary as the personal computers that grew out of the development of electronic devices on a chip, but arrays and other microfabricated devices are already beginning to generate sufficient interest to make them look very promising.

These arrays consist of many microscopic spots, each of which contains identical single-stranded polymeric molecules of deoxyribonucleotide (typically oligonucleotides or cDNAs) attached to a solid support such as glass or a polymer. Each spot contains many copies

D. J. Graves (genetics.upenn.edu) is at the Department of Chemical Engineering, 111 N. Grace Building, University of Pennsylvania, Philadelphia, PA 19104-6393, USA.
of a particular sequence, which can range in length from ten or twenty bases up to one or two thousand. The utility of these spots arises from the tendency of their component bases to pair up or hybridize with a second strand containing a complementary sequence. As there are four bases, the probability \( P \) of a single base matching a given molecule \( X \) bases long is \( \frac{1}{4} \); the probability of a two-base sequence matching is \( \frac{1}{4^2} \times \frac{1}{4} \); the probability for a three-base sequence is only \( \frac{1}{4^3} \times \frac{1}{4} \); and, for an \( n \) base sequence, it becomes \( \frac{1}{4^n} \times \frac{1}{4} \). If \( X \) represents a very large piece of DNA, for example, 1000 bases, and \( n \) is a modest number such as 20, the probability that the entire sequence of 20 bases will match some location in the chain by chance alone is only \( \frac{1}{4^{20}} \), or less than one in a million. Thus, even a sequence as short as 20 bases or so can uniquely bind to and thus identify a given gene, provided that the sequence is carefully chosen so it does not appear frequently in several genes (e.g. in a promoter or microsatellite).

With this background, it is easy to see how an array of different sequences can be used to identify one or more pieces of DNA or RNA in a solution. These unknown molecules are all tagged by, for example, attaching a fluorescent dye to them, and then exposed to an array containing hundreds or thousands of different sequences, each in a known location; when the array is examined, one can tell which molecules are present in the solution by determining which spots fluoresce.

Although this brief description captures the essence of the method, many details have been omitted. Just as in Southern blotting, from which the method originated, one must wash the array following hybridization to remove loosely bound molecules that do not match the test sequence exactly. Also, the techniques differ significantly depending on whether one is dealing with genomic DNA or mRNA and whether one is searching for the presence of a particular sequence (to detect mutations, for example) or attempting to quantify the amounts of many sequences likely to be present (as in gene-expression patterns). Figure 1 highlights some of the major differences, but this article will not attempt to catalog the various techniques as these have been well and recently summarized by others. Instead, I will concentrate on describing some of the important potential applications and unsolved problems in microarray analysis.

Applications of microarrays

Broadly speaking, the applications of microarrays fall into two categories: studies of genomic structure and studies of active-gene expression. The idea for DNA arrays arose during the early years of gene sequencing; when it was felt that the traditional Sanger–Coulson and Maxam–Gilbert methods for sequencing would never provide the speed necessary to tackle the massive workload resulting from the Human Genome Project (Fig. 2). For example, an array of \( 4^4 = 16.384 \) different spots can be constructed and a fairly short piece of DNA (say, 1000 bases) then tagged and applied to the array; the spots that glow (statistically, about 1 in 10) can then be located. Each sevenmer sequence is decoded from its location and, from an analysis of the overlapping short sequences, the original long sequence can be reconstructed. Researchers quickly realized, however, that a given short sequence appeared more than once in the long sequence undergoing analysis, one would have a branch point in the reconstruction diagram and could not assign an overall sequence without ambiguity. This application has thus largely fallen out of favor, although ‘resequencing’, to look for mutations or polymorphisms in a region whose ‘normal’ sequence is already known, is still considered to be a worthwhile goal.

Many of the potential applications of microarrays are in the health-related sciences and medicine. Recently, F. Collins (Director, National Human Genome Research Institute, Bethesda, MD, USA) stated that, although it had initially been assumed that only a few diseases involved genetics, ‘we now believe that the only disease not having some genetic component is trauma.’ One might even go further than this and include trauma, because the way an individual responds to and recovers from trauma is likely to be determined genetically, to some degree.

Applications of arrays to genomics

The applications of arrays to genomic studies, including resequencing, primarily involve the search for single-nucleotide polymorphisms (SNPs), which may have considerable importance regardless of whether they cause an overt disease. From these, one can use analytical techniques such as genetic-linkage mapping or association analysis to discover genetic predispositions to disease, and to classify diseases according to defect and the best treatment option. If one were not sure of the general location of the altered bases within the genome, such applications would require large numbers of relatively inexpensive chips and powerful analytical methods to be widely available; but the technology is young enough that satisfactory economics are eventually likely to be achieved. When the locations of nearly all of the specific defects have been determined (expected to occur within the next ten to twenty years), the job will become quite a bit easier, because a specific subset of probes can be constructed for a given purpose. One can imagine that, on a routine basis, all newborns might undergo 10 000 or more simultaneous genetic tests, using only a single chip and a few drops of blood.

Another interesting potential genomic application suggested recently has been called ‘pharmacogenomics’. Because each individual has a slightly different genetic makeup, each will have a unique set of polymorphic sites; although these polymorphisms might not be sufficiently aberrant to cause disease, some of them would determine how each individual responds to a particular drug. Thus, with the proper screening, one might know ahead of time that he or she would have an adverse reaction to a particular drug or that it would be ineffective for the disease it was directed against. Such testing might resuscitate drugs that had been discarded in the past because of low effectiveness or bad side effects in the entire population, but significant subpopulations could find these drugs to be highly effective and free of deleterious side effects.

Other genomic applications are numerous, including the identification of criminals or of blood
Applications to gene-expression studies

There are numerous areas in which gene-expression studies will be equally useful. Arrays are already being used to study how cells respond to environmental changes and stress through changing mRNA patterns. It has been suggested that this technology might be extended to studies of environmental toxicity caused, for example, by dioxin or mercury, by looking for subtle changes in gene expression. Eventually, such studies will be equally useful. Arrays are already being used to evaluate the many products resulting from combinatorial chemistry. Such products might not cause obvious changes in cellular appearance or behavior but could cause subtle metabolic changes that would show up when the mRNA was interrogated by an array; this has obvious implications in drug development. Such applications are not confined to human cells, of course. One day, it might be possible to control the life cycle of a plant much more precisely (stimulating early blooming, for example) or to find a "magic bullet" insecticide that does not affect other species, if the complex metabolic-control systems of cells were more fully understood.

Problems with microarrays

These are only a few of the possible uses for microarrays. Many more are mentioned in recent reviews, and still others cannot be foreseen at this time or may develop as a result of combining microarrays with other microfabricated devices to create entire analysis systems on a chip. In the remainder of this article, some of the most significant remaining problems lying between our present level of expertise and such intriguing possibilities will be discussed. Many of these problems are well understood and are being addressed, others are not as obvious or have been ignored in the rush towards some immediate practical application.

Probe selection and target preparation

Before the molecular interactions that take place during hybridization are discussed, it is necessary to define two terms used throughout this paper: the names given to the oligonucleotide immobilized on the surface and the portions of gene sequences for many different genes (expression analysis) or relative mRNA concentrations (mutation analysis).

Figure 1

Similar methods are used to apply microarrays to genomic analysis and gene-expression studies. This figure illustrates both techniques, where the steps differ for the two types of analysis. Finding mutations and other polymorphisms in genomic DNA involves: transcription, hybridizing and washing to create fluorescent hybrids; analysing spots for fluorescent products, in particular looking at the relative degrees of fluorescence for the four probe spots (A,C,G and T) at a given point mutation (d). There may also be additional spots for a given mutation consisting of a deletion or containing different lengths of nucleotide sequence spanning the mutation point. Studying gene-expression patterns, however, involves: isolating mRNA from the sample undergoing analysis and preparing cDNA containing fluorescent tag 1 (a); hybridizing and washing to create fluorescent tags (b); hybridizing and washing to create fluorescent hybrids (c); analysing spots for fluorescent products, in particular looking at the relative degrees of fluorescence for the four probe spots (A,C,G and T) at a given point mutation (d).

Figure 2

The original concept for sequencing by hybridization is shown here. A large array of, for example, all possible heptanucleotides is prepared. The unknown will form fluorescent hybrids (shown here as black squares) with those sequences that are complementary to the ones it contains. From the series of overlapping short sequences found in this way, the overall sequence of a long piece of DNA (top row, right) can be deduced.
the one in solution that hybridizes to it. The conven-
tion used here, and by a number of groups18, is to call
the former a probe and the latter a target; this is logi-
cal because the immobilized molecule is usually the
one with a known sequence, and it is used to study the tar-
get molecule in solution. Unfortunately, others14 use
exactly the opposite convention, which derives from
the original Southern-blotting application, in which the
targets on a membrane were interrogated by solu-
tion-phase probes. A stated convention, even if it is
contrary to one's expectation, is necessary to avoid
ambiguity. There are several issues to consider in choos-
ing probes and targets. These include the absolute and
relative size of the probe, the target, and the hybridizing
region.

Issues involving probe size
Early work tended to use short probes of
8–19 bases15,16 because they were less expensive to
synthesize or to grow in place on the surface. Unfor-
tunately, these have several drawbacks, including
less than ideal specificity and the need to carry out the
hybridization at or below room temperature, because of
the well-known relationship between size and hybrid
melting temperature. The technology and costs have
changed enough that, today, probes less than 20 bases
in length are seldom employed. Sizes between 20 and
30 bases are now common, and some probes generated
by PCR reactions are 1000 or more bases in length. In
this case, the exact sequence may not even be known
initially, although the added length is said to enhance
the specificity of hybridization. For these longer probes,
several thousand different clones containing genomic
inserts are prepared and the sequences immobilized as
probes, those that prove useful at identifying genes of
interest can be sequenced later, and those that show no
utility are simply discarded without the expense of
sequencing. Even after sequencing has revealed the
identity of the gene being probed, however, the exact
region or regions that interact with the target when
these probes are used may not be known.

Target size
Target size is a parameter for which the ideal value
has not been well established. Full length mRNAs or
cDNAs of substantial length have been, and in some
cases are still being, used. However, it is now widely
recognized that long targets are not desirable: they dif-
fuse from the bulk of solution to the surface relatively
slowly, and their bulkiness may interfere with effective
hybridization by causing collisions with the surface
(steric effects). A theoretical analysis that provides some
general target-size predictive capability has been carried
out17, but it has also been found experimentally that
fragmenting the target molecules is helpful. A general
rule to follow is that the target should be no larger than
100–200 bases in length. A target size of 50 is consid-
ered to be better, and fragments as short as 20 or 30 are
being used by some.

One fragmentation method in current use depends
upon the use of magnesium ions at approximately 94°C
to cause random cutting18. A second method involves
the addition of deoxyuridine triphosphate (dUTP) to a
PCR mixture also containing thymidine triphosphate
(TTP). T (or dLU) should appear by chance about once
in four bases and so, if the mixture is 1/5 dUTP and
4/5 TTP a dU should be inserted about once every
20 bases (1/5 X 1/4). The enzyme uracil-N-glycosylase
is then added to the PCR product19, which cleaves
where uridine is present and gives targets that are on
average roughly 20 bases long. In practice, these theo-
retical numbers need to be modified to account for
unequal incorporation rates. With this technique, one
has excellent control over the length of the target mol-
ecule. The only remaining reasons for using long tar-
gets are: (1) to have a higher density of fluorescent
(or other) tags; (2) to keep these tags farther from the
hybridizing region, so that they interfere less with
probe–target interactions.

Selecting the probe and target sequences
By far the largest unsolved puzzle in probe and tar-
get design and construction, however, is which sequence
to choose from the many that could uniquely identify
the target. Many authors have reported that certain
sequences hybridize very well while others do so
poorly, if at all20. A similar problem exists when trying
to find appropriate antisense oligonucleotides21,22.

Good hybridization potential does not appear to correlate well with any of the obvious parameters, such as
G–C content or melting temperature. Several investi-
gators have tried to correlate usefulness with the sec-
condary structure of the single-stranded target molecule
in solution. To this end they have modified computer
programs such as mfold, which is used to predict RNA-
folding patterns, to predict where loops and stems
should exist and determine whether probes can
match on or near loop regions performed better than
those matching stem regions. Others have tried to look
at the kinetics rather than the thermodynamics of
hybridization23. However, none of these attempts has
yet proved very successful. Moreover, it is not clear
that secondary structures would be particularly important
for short target molecules, and unexpected behavior is
still observed for short targets.

About the best that has been achieved is a set of
heuristic rules advanced by Lockhart et al24 including
the following criteria, which result in both good
hybridization to the correct target and a lack of promis-
cuous incorrect hybridization, for 20-base probes. (1)
The total number of As should be fewer than ten (also,
less than ten T s) and fewer than nine Cs (also, fewer
than nine Gs). (2) In a window of any eight consecu-
tive bases, fewer than seven As or Ts and fewer than six
Cs or Gs. (3) No more than five consecutive Cs or Gs
and no more than six consecutive As or Ts. (4) A palin-
drome score (a measure of self-complementarity) of
less than seven; also, no more than 16 positions should
match an incorrect target, with a mismatched loop of
up to three bases being bypassed in the probe sequence.

The importance of this problem is hard to overstate.
It means that one cannot predict a priori which probes
will work well and that one cannot use the absolute
fluorescent intensity of any given spot intelligently.
Instead, one is limited to comparing the relative
response of matched and mismatched probes or to
examining a mixture containing the unknown target
molecules (with tag 1) plus a known set of target mol-
ecules (with tag 2) to normalize the results. Fluorescent
intensity varies as much with the probe chosen as it
does with the fidelity of the probe–target match. The relationship between sequence and strength of hybridization should constitute a solvable problem but, to date, it has not been attacked rigorously.

**Probes immobilization**

**Building up sequences on the surface**

Probe–covered chips are prepared in three different ways: (1) growing oligonucleotides on the surface, base by base; (2) linking presynthesized oligonucleotides or PCR products to a surface; and (3) attaching such materials within a small 3-dimensional spot of gel. The first method has pitfalls which result in what is called a GeneChip™, has both advantages and disadvantages.

The main advantage is that, it is heavily dependent on the photolithographic techniques used for semiconductor fabrication. It produces dense arrays with a great many different probes in a small area. Currently, Affymetrix (Santa Clara, CA, USA) packs over 400,000 different square probe regions into an area of about 1 cm². This permits considerable probe redundancy: a probe can be present in which a particular base under scrutiny is present in all four forms (A, C, G and T). The location of the mutant base within a probe can also be shifted systematically by moving the two end points of the probe relative to the defect (either a longitudinal shift or a change in length). This provides considerably more information than a single spot would and an increased degree of certainty in identifying a particular base (base calling). Often, several dozen probes will be involved in determining the identity of a particular base with certainty.

The major disadvantage is also related to this chip production method. Although 400,000 appears to be a very large number, it is in fact a very small fraction of all the possible 20-mers (10¹²). Thus, anyone wishing to use 20-mer probes will have to use an existing GeneChip or have one custom made – a very costly alternative. Also, it would not be feasible for researchers to construct their own chips by this complex technique, even if the process were not patented. When the genes under study are of interest to many, as is the case with HIV, it is likely that GeneChips are already or soon will be available; for less-common diseases or new applications, the lack of an appropriate GeneChip could present a severe problem.

Another problem arises from the process itself. A photocleavage step is used in every cycle of the sequential base-addition process. The efficiency of this step is said to lie between 80 and 95% which translates into the number of hybrids formed typically ranges from 0.820 to 0.9520 and so the number of hybrids formed typically ranges from 10¹¹–10¹² per cm². This means that a little over 50% of the probes are likely to be packed on the surface at a higher molar density than longer ones.

We find that approximately 10¹¹–10¹² 20-mer probes can be attached per cm² of glass surface, a result that is in line with reported values. Often, not all of the molecules that become attached are fully anchored at many points as opposed to an end-attached brush-like structure for oligonucleotides attached through the previous method, using a single amino group and linker arm per molecule. However, we would anticipate that this would result in a thin layer of probe anchored at many points as opposed to an end-attached brush-like structure for oligonucleotides attached through the previous method, using a single amino group and linker arm per molecule. However, we would anticipate that the density of probes will be lower than the density of probes used in the previous method.

**Deposition and attachment of pre-synthesized probes**

Using pre-synthesized probes may avoid some problems but cause still others. The oligonucleotides or PCR product can be purified before attachment, but attachment may require the addition of special linker arms and/or reactive groups such as amine. This can increase their cost by a factor of five to ten relative to that of plain oligonucleotides of the same length. A relatively simple technique that avoids the need to add modifying groups but requires using long probes rather than oligonucleotides has been described (see below). In either case, these probe molecules are deposited in microdroplets, typically from a few picoliters to a few nanoliters each. The robotic systems used to plate for droplets cannot pack them as tightly as photolithography can, so the current level of droplet packing density that has been achieved is approximately 16,000 per cm² on average. Furthermore, the minimum volume of probe solution that can be picked up is of the order of one microliter so, if 100 microscope slides are prepared with 1 nl droplets, approximately 95% of the valuable probe solution will be wasted when the deposition tip is washed out for the next sample. The big advantage of any of these attachment techniques, however, is that almost any small laboratory can make such arrays and that they can be custom designed for any new studies that are devised.

The methods used for attachment itself deserve some mention. A number of groups use silanes, primarily 3-aminopropyltrimethoxysilane or a silane epoxide, as modifying groups but require using long probes rather than oligonucleotides. This can take place at all (it is hard to see how a double-helix brush could form if the probe is firmly bound to the surface along its entire length), and suggests that shorter probes are likely to be packed on the surface at a higher molar density than longer ones.

We find that approximately 10¹¹–10¹² 20-mer probes can be attached per cm² of glass surface, a result that is in line with reported values. Often, not all of the molecules that become attached are fully anchored at many points as opposed to an end-attached brush-like structure for oligonucleotides attached through the previous method, using a single amino group and linker arm per molecule. However, we would anticipate that this would result in a thin layer of probe anchored at many points as opposed to an end-attached brush-like structure for oligonucleotides attached through the previous method, using a single amino group and linker arm per molecule. However, we would anticipate that the density of probes will be lower than the density of probes used in the previous method.

**Deposition and attachment of pre-synthesized probes**

Using pre-synthesized probes may avoid some problems but cause still others. The oligonucleotides or PCR product can be purified before attachment, but attachment may require the addition of special linker arms and/or reactive groups such as amine. This can increase their cost by a factor of five to ten relative to that of plain oligonucleotides of the same length. A relatively simple technique that avoids the need to add modifying groups but requires using long probes rather than oligonucleotides has been described (see below). In either case, these probe molecules are deposited in microdroplets, typically from a few picoliters to a few nanoliters each. The robotic systems used to plate for droplets cannot pack them as tightly as photolithography can, so the current level of droplet packing density that has been achieved is approximately 16,000 per cm² on average. Furthermore, the minimum volume of probe solution that can be picked up is of the order of one microliter so, if 100 microscope slides are prepared with 1 nl droplets, approximately 95% of the valuable probe solution will be wasted when the deposition tip is washed out for the next sample. The big advantage of any of these attachment techniques, however, is that almost any small laboratory can make such arrays and that they can be custom designed for any new studies that are devised.

The methods used for attachment itself deserve some mention. A number of groups use silanes, primarily 3-aminopropyltrimethoxysilane or a silane epoxide, as a preliminary treatment step for glass surfaces. Amino-covered glass and amino-linked probes can then be coupled using a bifunctional molecule such as a dithiobis or a disothiocyanate; epoxy-covered surfaces use somewhat different chemistry.

In another method, a polycation such as polylysine is laid down and permits the polyamionic probe molecule to adsorb on this surface. If the probe molecule is long enough (typically several hundred bases), the interactions of multiple positive and negative groups is strong enough to hold it firmly in place, but one can also use photocovalinkage with an ultraviolet source to add covalent linkages to the amino ones. One might anticipate that this would result in a flat layer of probe anchored at many points as opposed to an end-attached brush-like structure for oligonucleotides attached through the previous method, using a single amino group and linker arm per molecule. However, it would appear that even the latter probes are normally also plastered down flat on the surface (see below). This limits the surface concentration that can be achieved and raises questions about how exactly hybridization can take place at all (it is hard to see how a double-helix hybrid could form if the probe is firmly bound to the surface along its entire length), and suggests that shorter probes are likely to be packed on the surface at a higher molar density than longer ones.

We find that approximately 10¹¹–10¹² 20-mer probes can be attached per cm² of glass surface, a result that is in line with reported values. Often, not all of the molecules that become attached are fully anchored at many points as opposed to an end-attached brush-like structure for oligonucleotides attached through the previous method, using a single amino group and linker arm per molecule. However, it would appear that even the latter probes are normally also plastered down flat on the surface (see below). This limits the surface concentration that can be achieved and raises questions about how exactly hybridization can take place at all (it is hard to see how a double-helix hybrid could form if the probe is firmly bound to the surface along its entire length), and suggests that shorter probes are likely to be packed on the surface at a higher molar density than longer ones.

We find that approximately 10¹¹–10¹² 20-mer probes can be attached per cm² of glass surface, a result that is in line with reported values. Often, not all of the molecules that become attached are fully anchored at many points as opposed to an end-attached brush-like structure for oligonucleotides attached through the previous method, using a single amino group and linker arm per molecule. However, it would appear that even the latter probes are normally also plastered down flat on the surface (see below). This limits the surface concentration that can be achieved and raises questions about how exactly hybridization can take place at all (it is hard to see how a double-helix hybrid could form if the probe is firmly bound to the surface along its entire length), and suggests that shorter probes are likely to be packed on the surface at a higher molar density than longer ones.

We find that approximately 10¹¹–10¹² 20-mer probes can be attached per cm² of glass surface, a result that is in line with reported values. Often, not all of the molecules that become attached are fully anchored at many points as opposed to an end-attached brush-like structure for oligonucleotides attached through the previous method, using a single amino group and linker arm per molecule. However, it would appear that even the latter probes are normally also plastered down flat on the surface (see below). This limits the surface concentration that can be achieved and raises questions about how exactly hybridization can take place at all (it is hard to see how a double-helix hybrid could form if the probe is firmly bound to the surface along its entire length), and suggests that shorter probes are likely to be packed on the surface at a higher molar density than longer ones.

We find that approximately 10¹¹–10¹² 20-mer probes can be attached per cm² of glass surface, a result that is in line with reported values. Often, not all of the molecules that become attached are fully anchored at many points as opposed to an end-attached brush-like structure for oligonucleotides attached through the previous method, using a single amino group and linker arm per molecule. However, it would appear that even the latter probes are normally also plastered down flat on the surface (see below). This limits the surface concentration that can be achieved and raises questions about how exactly hybridization can take place at all (it is hard to see how a double-helix hybrid could form if the probe is firmly bound to the surface along its entire length), and suggests that shorter probes are likely to be packed on the surface at a higher molar density than longer ones.
was provided by studies of the adsorption of oligonucleotides to cationic latex particles. For a series of DNA fragments of different sizes, a constant weight was found to be adsorbed per unit area. However, it was found that the actual number of molecules decreased as the size of the fragments increased. An important implication of these results for those preparing probe-covered surfaces is that the larger the probe is, the lower the hybridization signal is likely to be, because fewer probes are present to hybridize with. The calculated coverage for short chains is only 100–10 000 molecules μm⁻². A small number when one hopes to establish detection limits that are perhaps 1% of these maximum values. However, better results have been obtained when the number of probe molecules per unit area was artificially decreased below its maximal value, suggesting that there may be some interference between adjacent molecules if they are packed too closely together. The fluorescence of tagged target molecules can also be quenched, apparently by fluorescence-resonance-energy transfer between adjacent molecules, if their concentration on the surface is too high. The question of linker-arm length is not totally clear at present, although it does appear that a spacer between the surface and the hybridizing region is desirable. In studies that vary the spacing between the oligonucleotide probe and the surface linker systematically from 0 to 34 carbon atoms in length, a two- to three-fold enhancement in the hybridization efficiency has been seen. However, up to 150-fold enhancement has been reported with a 40-carbon-atom spacer. The reasons for such a disparity are not clear.

The effects and ambiguous data cited above strongly indicate that additional fundamental studies on how immobilization affects hybridization could be valuable. Often, the differences between the solution-phase and immobilized reactions are not obvious, and such studies might have a significant impact on the efficiency and utility of microarrays.

Three-dimensional probe spots

The third method of making arrays, which has so far been adopted by only one group of researchers, involves attachment within small three-dimensional spots of gel that are attached to a surface. This method provides increased sensitivity because molecules are attached within a volume element instead of on a surface, but at the price of a significant increase in time for the nucleotide chains to penetrate into the gel and interact with the probe. Because adequate sensitivity can be achieved with surface attachment, the added diffusional problems do not appear to be a reasonable tradeoff for improved sensitivity. It is, however, a technology that bears watching. One interesting approach that attempted to gain three-dimensional sensitivity while retaining two-dimensional speed was a porous-three-dimensional silicon chip; but no information has appeared on its performance since the initial report.

The equipment used to make and analyze microarrays

The preparation and analysis of microarrays is a technically sophisticated process, but arrays can be prepared rather easily. For example, by hand. Submicroliter droplets are deposited on a chemically prepared slide, using an underlying patterned template (L. A. Sanguedolce, PhD thesis, University of Pennsylvania, PA, USA, 1997) or simple jigs to produce patterns of up to a few dozen spots. A robotic system suitable for depositing 100–10 000 or more spots per slide can be constructed for US$15 000–US$20 000 (P. O. Brown, http://cmgm.stanford.edu/pbrown/). Alternatively, prepared arrays can be purchased from commercial sources. Hybridizing and washing are simple processes, but analysing the arrays after hybridization requires a specialized piece of equipment to scan the arrays. Figure 3 shows why scanning is the method of choice by comparing the characteristics of several microscope objective lenses: a low-power lens can capture a large area at once (i.e. a whole array), but only a very small fraction of the fluorescent light emitted from each spot reaches the detector; alternatively, with a high-power lens (with a high numerical aperture) can capture a large fraction of the emitted light but can image only one spot or a very few at a time. This explains why the most commonly adopted design involves using a high-power lens and moving the sample, the lens and/or the light beam so that the whole array is analyzed in a scanning mode. An electronic image is then built up and displayed on a computer. This is virtually the only way to achieve good light efficiency (necessary because of the low surface concentration) and good area coverage simultaneously.

Problems in informatics

As the probe densities increase and the throughputs of analytical instruments rise, the volume of information generated by microarrays will eventually require much better ways of analyzing and cataloging information. Even now, examining individual arrays of 4 000–10 000 spots and locating significant changes is taxing the abilities and patience of investigators to their limits. When one wishes to compare 10 or 100 such slides, it is likely that important correlations will be missed or nonexistent correlations will be incorrectly identified as real ones. Solving these problems will require active collaborations between biologists, personnel and informatics specialists. There is also a
need for a completely automated sample handling and tracking system, especially for arrays formed from droplets from the wells of microwell plates. When one must rely on humans to organize, transfer and archive information manually, order a stack of microwell plates or pipette a large number of samples by hand, the likelihood of error is increased. These are challenges that must be met if the technology is to live up to its promise.

Future developments likely to affect microarray technology

One of the most desirable improvements in microarray technology would be an analytical method that produces a stronger signal without the drawbacks of three-dimensional gel. With strong signals, one could use a fixed camera rather than a scanning system to reduce complexity and cost while speeding up the analysis process (currently a typical array scan takes 10–20 min). In many cases, it has been observed that the accuracy of the recovered data is not limited by the absolute sensitivity but by the difference between the signal and background. Any technique that could reduce background would therefore be especially valuable. Total internal reflection (TIR) techniques have been used to help separate the excitation and emission light beams in some studies, and TIR plus the addition of a metal layer between the glass and the attached monolayer has also been used to create a strong surface-plasmon-resonance signal. Other techniques proposed in the past have included surface-enhanced Raman probes and a light-addressable potentiometric sensor. None of these has shown special promise, but new methods are constantly being evaluated as sensitive DNA-detection techniques.

The strength and specificity of hybridization can be improved considerably by preventing repulsion between negative charges on the two hybridizing nucleotide chains. One effective way to do this is to use peptide-nucleic-acid probes, which are reported to have an increased duplex melting point of 1.5°C per base pair.

Another completely different technology that may eventually compete with microarrays for genomic and genetic-expression analysis is matrix-assisted laser desorption and ionization mass spectrometry (MALDI MS). A laser beam is used to vaporize and ionize a small amount of material (such as the DNA in a microarray spot), which is then analysed in a mass spectrometer. One advantage is that each probe can be labeled with a unique mass marker – theoretically, hundreds of different markers could be used simultaneously. In this technique, the target is immobilized as a spot and simultaneously probed with a solution of mass-labeled oligonucleotides. However, the equipment is much more expensive than the ‘standard’ fluorescent microarray system, so very high throughputs or other major advantages would be needed in order to make such a method attractive. Even then, it is likely that only large service operations would be able to purchase such an instrument.

Future prospects

Microarray technology is already having a major impact on our understanding of the effects of genomic polymorphisms and analyses of gene expression. The pace of change will only accelerate as the Human Genome Project is completed and more economical ways are developed to make and analyse arrays. Practical applications are already running well ahead of the scientific understanding of the basic phenomena, and it is clear that microarray chips can profit from additional fundamental studies on how sequence influences hybrid stability, on how probes are best attached, with hybrid stability, on how long a spacer and at what intermolecular spacing, and on how the kinetics of hybridization can be improved. Some proprietary studies of this type have undoubtedly been carried out by commercial microarray producers. However, individual researchers attempting to design their own arrays, either for contracted or in-house production, would benefit greatly from an improved understanding of the basic biophysical and biochemical phenomena involved.

Acknowledgments

I am indebted to the University of Pennsylvania for granting me a sabbatical leave to learn more about arrays and to try some new things with them. My colleagues at the Penn Medical School and Children’s Hospital of Philadelphia, Thomas Jefferson Medical School and Dupont Hospital for Children, particularly S. Mckenzie, S. Surrey and P. Fortina, have provided

<table>
<thead>
<tr>
<th>Lens</th>
<th>Numerical aperture</th>
<th>Light efficiency</th>
<th>Relative area</th>
</tr>
</thead>
<tbody>
<tr>
<td>5x dry</td>
<td>0.14</td>
<td>1%</td>
<td>376.4</td>
</tr>
<tr>
<td>10x dry</td>
<td>0.25</td>
<td>3.2%</td>
<td>94.1</td>
</tr>
<tr>
<td>43x dry</td>
<td>0.65</td>
<td>24%</td>
<td>5.1</td>
</tr>
<tr>
<td>97x oil</td>
<td>1.25</td>
<td>45%</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 3

The inverse relationship between light-collection efficiency and imaged area illustrated by the data in this figure for several typical microscope objective lenses explains why scanning has become a standard technique for analysing microarrays. As the magnification and numerical aperture of a lens increase, the fraction of light emanating from a given region that is captured by the lens (the light efficiency) increases, but the relative area that can be seen at any time decreases. A high power objective lens moved relative to the array in two dimensions provides both high efficiency and a large effective field of view. Confocal pinholes for the light source and light detector ensure high resolution of individual spots and the elimination of stray light.
valuable discussions and insights. My former students, V. Chan and L. A. Sanglard, gathered some of the data that has been mentioned. Their research has been supported by grants from the University of Pennsylvania Research Foundation and Cancer Center pilot projects program, the Forder Fund for Excellence from the Childrens Hospital of Philadelphia, NIH grants RO1 DK 16691 and P60-HL3832, the Nemours Foundation and a Whittaker Foundation special opportunities grant. Finally, I wish to thank P. Brown and D. Botstein and their graduate students and postdoc-
torial fellows at Stanford University for hosting me dur-
ing this period and adding to my understanding of this 
rapidly developing field.

References
41 Hauer, I. B. et al. (1997) Molecular Biology (translation of Molecular Biology) 31, 133–140

Organizing a meeting?

Are you organizing a meeting that is of interest to biotechnologists? If so, please supply Trends in Biotechnology with the details (date, title, location) and a contact name, and we will include your meeting in our Diary page, without charge. Contact the Editor of Trends in Biotechnology at Elsevier Trends Journals, 68 Hills Road, Cambridge, UK CB2 1LA.

Please ensure that we receive details of your meeting at least ten weeks prior to the start of the meeting.