High-throughput gene expression analysis for drug discovery

Gregory G. Lennon

The ability to rapidly survey and compare gene expression levels between reference and test samples is moving the drug discovery process towards a more genomic orientation. The success of the Human Genome Project and related private genomics initiatives, combined with new technologies to probe, image and access expression data, are responsible for this transformation. This article reviews the history, status and future direction of high-throughput gene expression analysis. It describes classical approaches, explains the development of methods such as differential display for discovering novel genes, and discusses how microarray technology is exploiting collections of known sequences to pinpoint drug targets.

Methods for assaying gene expression can be classified into two major types: open methods, which do not require prior knowledge of the genes being measured, and closed methods, which measure expression levels of already collected clones or sequences. Some expression analysis techniques can only measure on a gene-by-gene basis while others can assay multiple genes simultaneously. Finally, some methods can directly measure differential expression between two samples and some examine expression levels from one sample at a time, followed by computation-based comparisons. Understanding the differences between these methods is essential for choosing the best technology for a given drug discovery application. Regardless of the method chosen, researchers must identify or access (through databases) vast quantities of expression information to find effective drug targets.

The history of gene expression

The history of gene expression analysis began when laboratory methods were developed to examine expression of individual known genes. The northern blot technique, introduced in 1977, hybridizes labeled DNA or RNA probes of known genes to RNA blots[1]. The resulting expression patterns of mRNA transcripts can then be read. This technique is still widely used to confirm the results of other types of gene expression studies. In 1977, another method was published that protects a DNA-labeled probe against degradation by the single-stranded nuclease S1 if the probe is annealed to an RNA (Ref. 2). Ten years later, RNase protection assays were developed to detect the expression of specific, previously characterized RNAs and to compare their levels of expression[3]. With this technique, a specific labeled cDNA forms a hybrid with its corresponding mRNA. When exposed to a single-strand-specific nuclease, the hybrids resist degradation and can be detected using gel electrophoresis. A later approach, differential plaque-filter hybridization, can detect differences in the expression of cloned cDNAs between two samples[4].

In 1993, subtractive hybridization techniques became available for constructing subtractive cDNA libraries. This methodology hybridizes cDNA from one pool to mRNA from the other[5]. Then, cDNA libraries are constructed from the transcripts that are not hybridized, these being used to

Gregory G. Lennon, Gene Logic, 708 Quince Orchard Road, Gaithersburg, MD 20878, USA. tel: +1 301 987 1700, fax: +1 301 987 1701, e-mail: glennon@genelogic.com

 1359-6446/00/$ – see front matter PII: S1359-6446(99)01448-8

identify specific mRNAs (Refs 6,7). A modification of this technique, representational difference analysis (RDA), also uses preferential amplification of non-subtracted fragments. In RDA, 'representations' or simplified versions of the genomes being studied (amplicons) are created using restriction digestion. This method was first developed to examine the differences between genomes, but has proven useful for cloning differentially expressed genes. From this method, suppressive subtractive hybridization (SSH) was derived, which enables further suppression amplification of non-subtracted fragments[8]. SSH combines normalization (equalizing the abundance of cDNAs within the target population) and subtraction (excluding the common sequences between the target and driver populations) in a single procedure. Results from both RDA and SSH should be validated using other methods.

High-throughput parallel methods

Early gene expression methods, such as those already mentioned, are relatively small-scale techniques. They either focus on measuring mRNA expression levels for individual, well-characterized genes, or use *in vitro* nuclear 'run-on' transcription assays to determine the transcriptional profiles of several active genes simultaneously[9–11]. They are therefore inadequate for conducting large-scale screening and developing expression profile patterns for tissues or cells (the basic requirements for efficient pharmaceutical research). Thus, several newer methods for high-throughput screening (HTS) have been developed over the past decade, including differential display, expressed sequence tag (EST) methodology and many array techniques. Collectively, they have made it possible to identify the expression levels of novel genes and characterize them, correlate mRNA expression patterns in many tissue types with disease states, identify side effects of current and experimental treatments, and determine the effects of compounds on non-target tissues.

Differential display

Differential display of eukaryotic mRNA, first reported in 1992, was a major advance in the comparison of gene expression differences between cells or tissues[12]. Encompassing the use of either arbitarily or specifically primed PCR, it is perhaps the most widely used method involving gel electrophoresis for comparing gene expression. Both methods amplify partial cDNAs from subsets of mRNA samples by using reverse transcription and PCR. These short cDNA fragments are then typically displayed on polyacrylamide gels. Differential display can simultaneously measure both up- and down-regulation across tens of samples.

Originally, this method used an oligo(dT) primer with an anchor of one or two bases at the 3′ terminal. Reverse transcription and denaturation were followed by arbitrary priming on the resulting first strand of cDNA. A series of products were then derived from the 3′ end of the mRNAs by using PCR with the original primer (a radiolabeled nucleotide) and a set of short, random decamer primers. Each random primer annealed to the mRNA at a different position relative to the anchor primer. Products showing significant differential expression were sequenced after size fractionation of the PCR sample using denaturing gel electrophoresis, generally after overnight autoradiographic exposure.

READS™

In 1996, Prashar and Weissman significantly modified this standard differential display technique, using selective PCR amplification and display of 3′-end restriction fragments of double-stranded cDNAs. This methodology, named Restriction Enzyme Analysis of Differentially Expressed Sequences (READS), can be used to identify novel drug targets. As with another method that uses restriction endonucleases in differential display, READS is now a well-validated fingerprinting technique[13]. READS was developed to overcome limitations of standard differential display methods, such as a lack of quantitative correlation with mRNA fold changes, a significant number of false-positive signals, inconsistency in reproducing display patterns and failure to fully represent the redundancy of mRNA signals[14–16]. More importantly, the band size produced by arbitrarily primed differential display was not always predictable from the mRNA sequence under study[17]. The classical approach to differential gene expression is also limited because cDNAs are amplified at the low primer-annealing temperature of 40°C, which is a nonstringent PCR condition.

The READS approach (Fig. 1) involves attaching a restriction enzyme-digested double-stranded cDNA to an adapter that mediates selective PCR amplification of only the 3′-end fragments of cDNAs under high-stringency PCR conditions. Approximately 100 combinations of restriction enzymes and oligo(dT) primers (with specific heel sequences) are usually used to selectively measure the expression of 100–200 transcripts per combination. This technique offers several advantages:

- Consistent and reproducible patterns (such as accurate measurement of up- and down-regulation)
- High coverage of the mRNAs in a sample
- Less under-representation or redundant representation of mRNAs

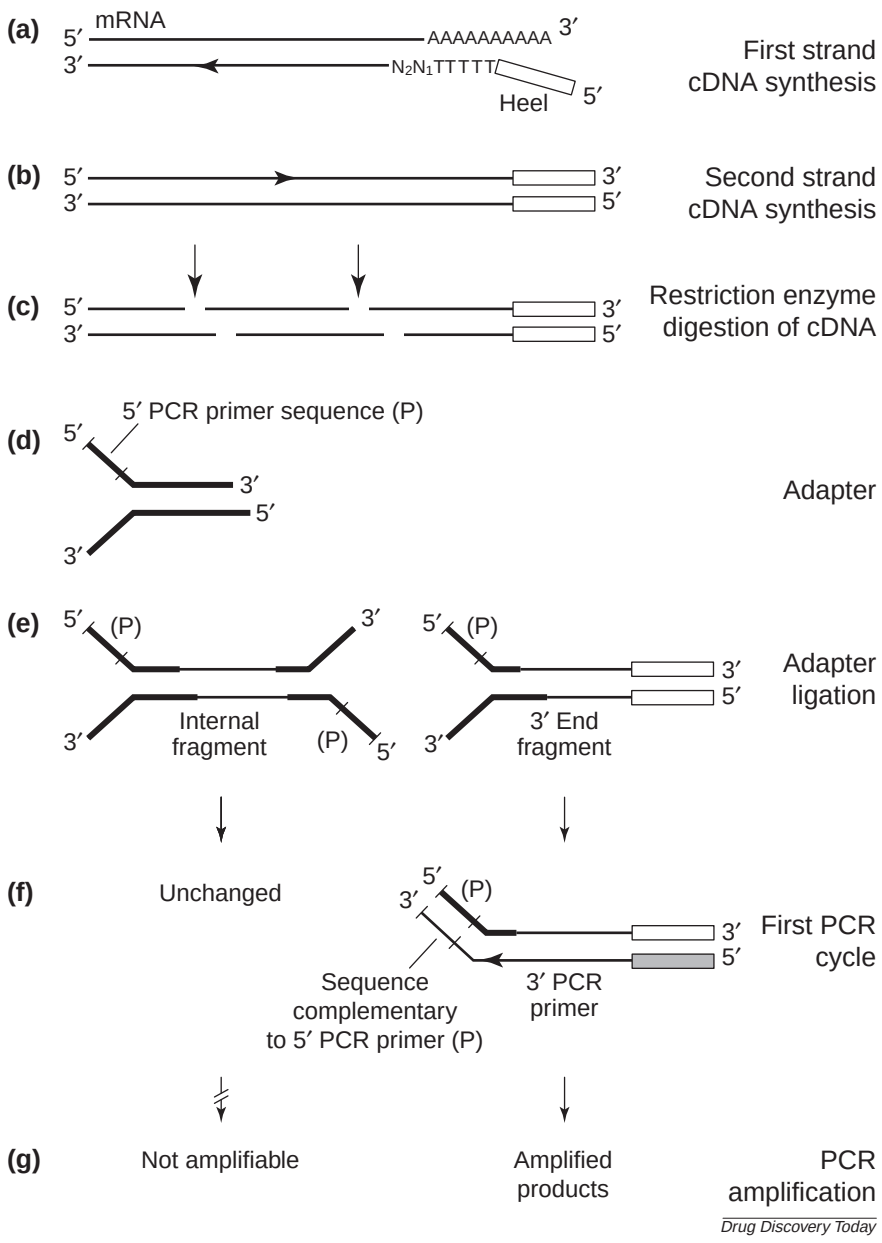

Figure 1. *READS™ measures gene expression by selective PCR amplification and display of 3′-end restriction fragments of double-stranded cDNAs (Ref. 17). (a) An oligo(dT) primer, with one or two additional bases at one end (N1, N2) and a 'heel' sequence at the other, is used for synthesis of the first strands of cDNAs. N_1 can be any base other than thymine, while N_2 can be any base. (b) The second strands of the cDNAs are synthesized using standard methods. All cDNA molecules have a common 3′ heel sequence. (c) cDNAs are digested with a restriction enzyme. (d) A Y-shaped adapter with an overhang on its 3′ end for ligation to the cDNA fragments. The 5′ end contains a noncomplementary sequence, giving rise to its Y shape. The 5′ PCR primer, positioned on this Y region, cannot anneal to the adapter. (e) All fragments are ligated to Y-shaped adapters. Internal fragments have adapters at both ends. The 3′ end fragment has the Y-shaped adapter at one end and the 3′ heel at the other. (f) The first PCR cycle is started, using a 3′ PCR primer that anneals to the heel of the 3′ end fragment. A complementary sequence is made of this entire fragment, including the 5′ PCR primer. The 5′ PCR primer can now anneal to this complementary sequence. Other fragments are unchanged, as they lack functional PCR primers. (g) The two PCR primers can now selectively amplify the 3′ end fragments, yielding quantities similar to the sample's mRNA.*

- Adequate information for quantifying levels of gene expression
- The capacity to distinguish highly homologous and alternatively polyadenylated or spliced transcripts.

Various differential display methods have proven helpful in detecting genes that correlate to certain disease states, genes responsible for growth factors and developmentally regulated genes. With a sufficient number of primer combinations, differential display can detect virtually all expressed mRNAs. As only small quantities of total RNA (a few micrograms) are required for this technique, differential display is especially useful when the RNA supply is limited. Thus, clinical and pharmaceutical researchers can test small biopsies as well as more homogenous cell types, such as those isolated using laser-capture microdissection or flow-sorting.

Expressed sequence tags

EST methodology can determine the expression profile of an entire cell or tissue under analysis. During the 1990s, EST

methodology played the largest role in increasing the catalog of known genes. Using this approach, cDNA clones are randomly picked and a single pass of sequencing is performed from one or both ends of each clone. Subsequent comparison with existing sequence databases immediately identifies novel sequences. Measuring how often a given sequence appears in a (representational) library enables the estimation of expression levels for each gene.

Although this method can accurately identify the presence of a proportion of genes, relatively low sampling (typically 5,000–10,000 sequences are generated from a tissue containing >20,000 distinct transcript types) makes it difficult to measure abundance of expression or to identify differentially expressed genes except where genes are highly up- or down-regulated.

Serial analysis of gene expression

Serial analysis of gene expression (SAGE) can potentially tag and analyze all transcripts in a given cell population or tissue. It has been used to successfully compare expression profiles between normal and cancerous cells, and detect p53 levels prior to apoptosis[18,19]. In theory, SAGE is an 'open' system. However, in practice, the short length of the tags means that it is most useful for expression profiling of fully sequenced genes. Thus, the value of this technique might increase as the Human Genome Project progresses.

This method uses two samples that are ligated and tagged with separate primers and then amplified. Subsequently, the primers are removed, revealing sticky ends that form concatemers. The concatemers are both cloned into a vector, with sequence information for the two different cDNA tags contained between anchoring sites. This cloning and sequencing process is time-consuming, as it must be performed for each sample and followed by extensive computational analysis.

Filter arrays

The public EST efforts, spearheaded by sequencing work at Washington University (St Louis, MO, USA)[20] and the arraying efforts of the IMAGE Consortium (founded by researchers at the Lawrence Livermore National Laboratory, Columbia University, National Institutes of Health and Centre National de la Recherche Scientifique)[21], have made sequences and clones for more than one million cDNA clones publicly available. A network of five distributors (see http://www–bio.llnl.gov/bbrp/image/image.html) across the globe supplies researchers with clones and related research services, such as sets of sequence-verified cDNA clones spotted onto nylon membranes. As standard laboratory protocols can be used and the filters are commercially available at a relatively modest cost, they are a popular forerunner to microarrays. Hybridization of radioactively labeled complex RNA (Ref. 22) to these membranes yields signals for moderately and abundantly expressed genes and, depending on several factors, some of the less abundant transcripts. Thus, differential expression is best measured using genes that are moderately expressed in at least one of the two (or more) states under study.

DNA microarrays

DNA microarrays measure expression by using templates containing hundreds or thousands of probes that are exposed simultaneously to a target sample. They make it possible to systematically survey DNA and RNA variation for the first time and are becoming a standard tool for drug discovery and evaluation. Microarray techniques are so powerful that their uses are often limited largely by the challenge of managing and analyzing the data they generate.

DNA microarray technology evolved from a paper published in 1975 by E.M. Southern (the originator of the Southern blot), who showed how a solid support could be used to examine nucleic acids[23]. This was advanced by the development of non-porous solid supports, leading to miniaturization and the use of fluorescence-based detection methods. The two main types of templates are long DNA fragments[24] (over 100 base pairs) and oligonucleotides (generally 18–25 mers).

Microarrays are expensive, although efficiencies should improve and costs should drop dramatically in the next couple of years, enabling these tools to become accessible to most research laboratories. Besides cost, microarrays are limited by the fact that they can only probe genes for which clones or sequences are already available. Furthermore, their accuracy can be limited by the purity of the RNA and the quantity of RNA for each hybridization.

The projected impact of DNA microarrays on drug discovery and development

By understanding gene expression patterns, researchers can gain information that can link sites of expression, biochemical pathways, and normal or pathological functions in organs and whole organisms. Because of their speed and breadth, microarrays should impact drug discovery in several ways:

- Accelerate the understanding of the molecular basis of disease

- Improve knowledge of model systems
- Explore pathogens and pathogenic reactions in terms of gene expression
- Pinpoint new drug targets
- Examine efficacy and toxicity responses to new drugs.

The remainder of this review will explore a few examples of each type of contribution.

Microarrays have already determined how several important genes are abnormally regulated in disease. For example, a microarray of approximately 100 genes that have a role in inflammation was used to examine rheumatoid tissue. This revealed upregulation of the genes encoding interleukin-6 and several matrix metalloproteinases[25]. In another instance, a novel gene involved in promoting tumors was discovered by using a 1000-element microarray of unknown cDNAs to examine how treatment with phorbol testers affects expression levels[26].

Microarrays should provide more detailed knowledge about pathogens by systematically examining every gene in a microbe to uncover the overall expression pattern. In addition, microarrays will continue to contribute to the understanding of responses to drug treatments. For example, a recent study used microarrays to measure the effects of kinase inhibitors on the entire yeast genome by measuring changes in mRNA levels before and after treatment[27]. In another example, microarray studies of yeast cells showed that the immunosuppressive drug FK506 had the same effect on gene expression level patterns as ablation of the gene that FK506 suppresses. Furthermore, this study showed that, in the absence of this gene, FK506 affected expression levels in other ways[28]. This suggests that the drug might have more than one target. Microarrays are also proving useful in the determination of drug toxicity[29].

Summary of microarray techniques

Expression profiling using cDNA microarrays begins by arraying many gene-specific amplicons derived from the cDNA clones onto a single matrix. Using two-color hybridization, cDNA representations of total RNA pools are created from test and reference cells, fluorescently tagged with two different colors, then mixed together before being hybridized to the matrix. For each transcript, the resulting fluorescence signals reflect the difference in abundance between the two samples. Two-color hybridizations provide rapid comparisons between the two samples, but they do not measure the absolute levels of gene expression for either sample. By contrast, one-color hybridization is slightly slower, as hybridizations of the two samples must be performed separately to reach meaningful comparisons. However, each one-color hybridization measures absolute levels of gene expression rather than comparative levels. After these actual levels are recorded in databases, they can be compared with levels from other samples without the need to perform comparative experiments. Although performing 1000 two-color hybridizations results in 1000 pair-wise comparisons, conducting 1000 one-color hybridizations yields almost half-a-million pair-wise comparisons, as the absolute values of one-color hybridizations can be evaluated against each other.

Using either the one- or two-color methods, microarray experiments must be performed repeatedly to ensure accuracy of the data. However, computational averaging of the signals of one-color hybridizations from multiple independent samples is more straightforward. The choice between using one-color versus two-color methods depends on several factors, including the number of transcripts under examination, the need for speedy results and cost differences. Hence, one-color hybridizations are often more useful for surveying a large number of genes, while two-color hybridizations can be preferable for more restricted sampling.

Sample collection issues

Developing useful databases for microarray sample analysis begins with finding and analyzing a wide range of normal and diseased tissues and cells. Perhaps the 'cleanest' and most accurate sample source is tissue removed at surgery, followed by extensive collation of the clinical and demographic data as well as molecular and cellular data. Several companies have established relationships allowing them to collect such tissues.

Training of personnel in surgical centers to follow the highest possible standards for sample collection is also important, as RNA is notoriously labile. Tissue samples must be characterized by pathological examination, microdissected (when appropriate) and documented with a digitized image for future reference. In addition, extensive clinical information must be gathered from each patient to define the biological state of the sample, including patient background and laboratory data, pathological descriptions and medicinal consumption and dosages.

As databases maintained by both the National Center for Biotechnology Information (NCBI) and European Bioinformatics Institute (EBML) make DNA microarray-based gene expression data repositories publicly available, questions regarding sample collection will become increasingly important. These data will be obtained from many different labs that use varying collection conditions and a variety of technologies. It will be important to establish firm

standards for sample collection and processing if these important public resources are to be maximally useful. Single-source databases can overcome this challenge by ensuring that consistent standards and platforms are used internally for all of their experiments.

Creating cDNA microarrays

The PCR products used in cDNA arrays usually come from purified templates to prevent contamination by cellular products. The amplified samples are partially purified to eliminate components of the PCR cocktail. A robot typically 'prints' an aliquot of each product onto a matrix using a method that resembles dot-matrix printing[30]. Meanwhile, techniques similar to ink-jet printing, which do not make direct contact with the matrices, are also being used. Membranes are often made of nitrocellulose or charged nylon. The glass-based arrays are usually manufactured on microscope slides, which have the advantage of low (background) fluorescence. The results of microarray probing can be detected by using digital image readouts, after local sampling determines the threshold for accurate signal detection. Sophisticated algorithms help to read even the weakest signals. However, because of overall questions of reliability and variance in measurement, the results of microarray studies should be validated by other methods (generally of a gene-by-gene nature, such as quantitative RT-PCR or northern blots).

Types of microarray products

The main committed competitors in commercializing microarray analysis are Affymetrix (Santa Clara, CA, USA), which produces GeneChip® probe arrays and Incyte (Palo Alto, CA, USA), which produces GEM™ slide arrays. Other companies that currently, or will shortly, provide various slide array products or services include Molecular Dynamics (Sunnyvale, CA, USA), Genometrix (The Woodlands, TX, USA) and Hyseq (Sunnyvale, CA, USA).

Gene Logic currently incorporates the Affymetrix GeneChip probe array in its closed gene expression studies, while using the open READS system for detecting novel genes. In addition, the company is developing the Flow-thru Chip™ probe array as a more flexible platform for smaller scale projects that might be suitable for both protein and nucleic acid arrays. The following discussion is based on the author's experience with these three approaches.

Almost all DNA array technologies use mechanical or lithographic means to bind elements on a two-dimensional substrate, which limits performance in two ways. Firstly, it limits the number of recognition elements per spot as, whenever spot density increases, the spot size decreases accordingly. Secondly, increasing density (and therefore decreasing spot size) lowers sensitivity because it limits the surface area where hybridization of the target can occur. As the upper limit for detecting targets is proportional to the number of potential binding sites on each spot, a technology that increases the number of binding sites per spot will improve sensitivity.

One way to improve on the standard process is to maximize binding sites, using a three-dimensional substrate

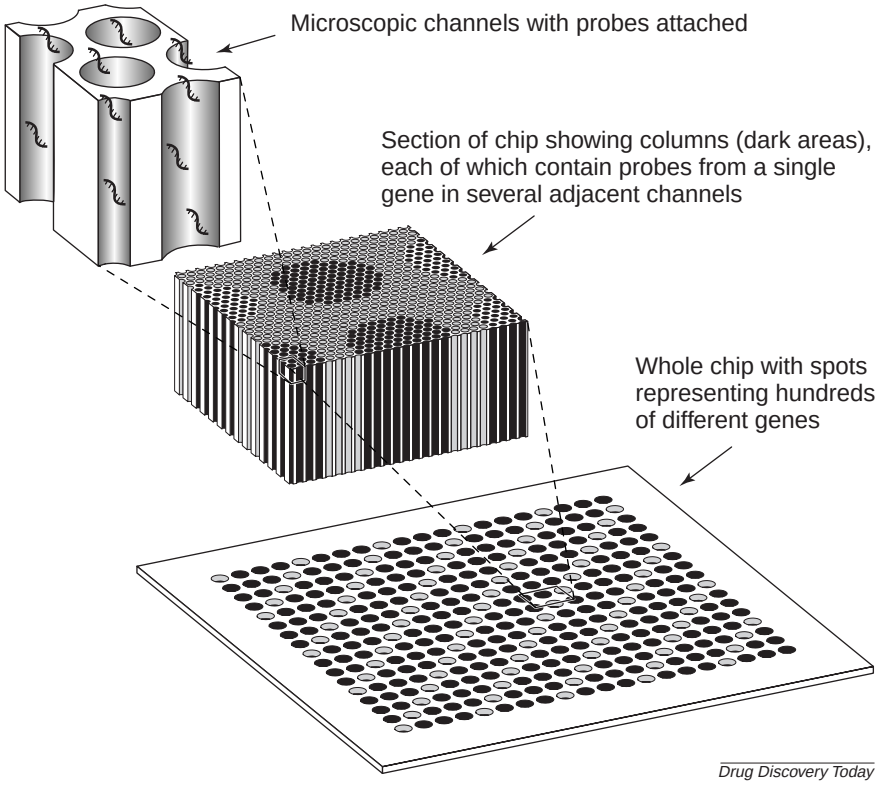

Figure 2. *Each Flow-thru Chip™ probe array has hundreds of thousands of microscopic channels that pass though it. The array is spotted with hundreds of different probes. The sample flows through the channels, coming into close proximity with the probes. This design results in hybridization speeds greater than flat arrays.*

that increases the surface area available for binding probes (Fig. 2). The Flow-thru Chip utilizes this method by using microchannels that connect the upper and lower levels of the chip so that target substances can flow through it. The probes are mobilized on the walls of the microchannels. Experiments have suggested several advantages of this type of chip over two-dimensional chips, including:

- Enhanced reaction times based on increased molecular 'collision' rates in the narrow channels
- The ability to deposit probes more uniformly and at higher densities because of the wetting properties of microporous materials
- The potential for using smaller sample and reagent volumes because the reaction volume is reduced.

Technical details of the GeneChip probe array

The Affymetrix GeneChip probe array, the first and most widely used array on the market, owes its construction to two fundamental technological developments: the ability to make hundreds of thousands of oligonucleotides in specific locations with high spatial resolution, and the capacity to accurately measure molecular binding events using laser, confocal fluorescence scanning. This probe array is based on light-directed synthesis using both photolithography and solid-phase DNA synthesis (Fig. 3). After attaching synthetic linkers with photochemically removable protecting groups to the glass substrate, light is directed through a photolithographic mask to achieve selective, local photodeprotection. Hydroxyl-protected deoxynucleosides are then exposed to the surface, resulting in binding to the deprotected sites. The information capacity of arrays using these techniques is limited only by the physical size of the array and the lithographic resolution. The recent use of semiconductor-like photoresist techniques has enabled the synthesis of arrays with features as small as 2 μm (Refs 31,32) that can work routinely with 1–2 μg mRNA or less. Array sets are currently available covering 42,000 human, 19,000 mouse and 34,000 rat genes.

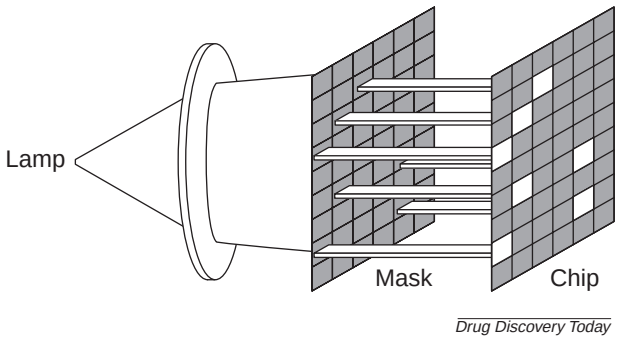

Figure 3. *Schematic representation of the light-directed synthesis that Affymetrix uses for construction of its high-density GeneChip®probe arrays[32]. Light is directed through a mask to protect and activate selected sites on the array. Protected nucleotides couple to the activated sites.*

Applications of gene expression analysis – the use of databases

How can a pharmaceutical company use microarrays to prioritize drug targets and lead compounds? One approach is to compare samples against large gene expression databases. Without presenting an exhaustive review on this subject, some illustrations will be discussed from experience gained in developing GeneExpress™ databases. The databases include quantitative information on gene expression for the relevant genes in the sample tissues, based on results from GeneChip array expression studies. These types of databases can identify disease, toxicity and drug response pathways by comparing known differential expression data with relevant data about biological pathways and tissues, and with patient data. Such databases will become increasingly effective as more information concerning tissues of all the major organs of the body and most of the significant disease states becomes available.

One GeneExpress database (BioExpress™) is used primarily for target selection and prioritization. It contains information from normal tissues and cells, diseased tissues and tissues from patients who have been exposed to drug treatments, in addition to a wide variety of mouse and rat tissues. Searches can be performed to find genes that are downregulated in a sample or set of samples, discover what biological processes and biochemical pathways their products participate in, or determine their associated disease states. All of the genes in the database can also be examined to discover which of these genes are differentially regulated in a given pathway. This type of search provides clues for determining drug targets that counter differential regulation in diseased tissues. Such databases can also be queried to determine whether a group of genes associated with a particular disease is linked to other diseases, which might help find new uses for existing drugs. Additionally, the effects of deliberate over- or under-expression of particular genes, such as by antisense inhibition, gene knockout or over-expression and similar techniques, can be evaluated by microarray analysis to help prioritize or eliminate genes as potential drug targets.

A second major use for gene expression information is to assist with lead selection. Such databases can be used to

compare the gene expression profiles of test compounds against known toxins, which assists in the prediction of which compounds are least likely to induce the classes of toxicity exhibited by the reference compounds. The effects of test compounds can also be compared with the effects of marketed pharmaceutical products, to help the selection of compounds with similar (or different) modes of action.

The future of gene expression analysis

Companies engaged in studying the molecular basis for drug development are rapidly moving the emphasis from gene discovery to gene expression. The imminent conclusion of the first phase of the Human Genome Project will simply accelerate this transition.

In the future, microarrays should become cheaper, more sensitive and more comprehensive until they contain probes for the entire genome. In addition, major improvements are anticipated for measuring protein abundance with increased sensitivity and on a large scale. Microarrays are likely to drive the large-scale collection of transcription and translation information, providing complementary information for drug development.

The success of high-throughput gene expression systems will create another set of challenges, including the development of tools to manage the vast quantities of data. The pharmaceutical industry will need to examine models from other fields as diverse as the financial industry, NASA and the aerospace industry, meteorology, and the online industries that create search engines. With proper database management tools in place, high-throughput gene expression will help develop drugs with a speed and efficiency that was unthinkable a few years ago.

REFERENCES

1 Alwine, J.C. *et al.* (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5350–5354

2 Berk, A.J. and Sharp, P.A. (1977) Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell* 12, 721–732

3 Lee, J.J. and Costlow, N.A. (1987) A molecular titration assay to measure transcript prevalence levels. *Methods Enzymol.* 152, 633–648

4 Maniatis, T. *et al.* (1978) The isolation of structural genes from libraries of eucaryotic DNA. *Cell* 15, 687–701

5 Hedrick, S.M. *et al.* (1984) Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature* 308, 149–153

6 Swaroop, A. *et al.* (1991) A simple and efficient cDNA library subtraction procedure: Isolation of human retina-specific cDNA clones. *Nucleic Acids Res.* 25, 1954

7 Lisitsyn, N. *et al.* (1993) Cloning the differences between two complex genomes. *Science* 259, 946–951

8 Diatchenko, L. *et al.* (1996) Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. U. S. A.* 93, 6025–6030

9 Greenberg, M.E. and Ziff, E.B. (1984) Stimulation of 3T3 cells induces transcription of the *c-fos* proto-oncogene. *Nature* 311, 433–438

10 Marzluff, W.F. (1978) Transcription of RNA in isolated nuclei. *Methods Cell Biol.* 19, 317–331

11 Manley, J.L. and Gefter, M.L. (1981) Transcription of mammalian genes *in vitro*. *Gene Amplif. Anal.* 2, 369–382

12 Liang, P. and Pardee, A.B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257, 967–997

13 Ivanova, N. and Belyavsky, A. (1995) Identification of differentially expressed genes by restriction endonuclease-based gene expression fingerprinting. *Nucleic Acids Res.* 23, 2954–2958

14 McClelland, M. *et al.* (1995) RNA fingerprinting and differential display using arbitrarily primed PCR. *Trends Genet.* 11, 242–246

15 Liang, P. *et al.* (1994) Differential display using one-base anchored oligo-dT primers. *Nucleic Acids Res.* 22, 5763–5764

16 Ito, T. *et al.* (1994) Fluorescent differential display: Arbitrarily primed RT-PCR fingerprinting on an automated DNA sequencer. *FEBS Lett.* 351, 231–236

17 Prashar, Y. and Weissman, S. (1996) Analysis of differential gene expression by display of 3′ end restriction fragments of cDNAs. *Proc. Natl. Acad. Sci. U. S. A.* 93, 659–663

18 Zhang, L. *et al.* (1997) Gene expression profiles in normal and cancer cells. *Science* 276, 1268–1272

19 Polyak, K. *et al.* (1997) A model for p53-induced apoptosis. *Nature* 389, 300–305

20 Hillier, L. *et al.* (1996) Generation and analysis of 280,000 human expressed sequence tags. *Genome Res.* 6, 807–828

21 Lennon, G.G. *et al.* (1996) The I.M.A.G.E. consortium: An integrated molecular analysis of genomes and their expression. *Genomics* 33, 151–152

22 Gress, T. *et al.* (1992) Hybridization fingerprinting of high-density cDNA-library arrays with cDNA pools derived from whole tissues. *Mamm. Genome* 3, 609–619

23 Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98, 503–517

24 Schena, M. *et al.* (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470

25 Heller, R.A. *et al.* (1997) Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc. Natl. Acad. Sci. U. S. A.* 94, 2150–2155

26 Schena, M. *et al.* (1996) Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. U. S. A.* 93, 10614–10619

27 Gray, N.S. *et al.* (1998) Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science* 281, 533–538

28 Marton, M.J. *et al.* (1998) Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nat. Med.* 4, 1293–1301

29 Braxton, S. and Bedilion, T. (1998) The integration of microarray information in the drug development process. *Curr. Opin. Biotechnol.* 9, 643–649

30 Bowtell, D.L. (1999) Options available – from start to finish – for obtaining expression data by microarray. *Nat. Genet.* (Suppl.) 21, 25–32

31 Beecher, J.E. *et al.* (1997) Chemically amplified photolithography for the fabrication of high density oligonucleotide arrays. *Polym. Mater. Sci. Eng.* 76, 597–598

32 McGall, G. *et al.* (1996) Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13555–13560