

GENE EXPRESSION analysis and drug R&D

It is clear that gene expression profiling systems have a role to play in many facets of drug discovery and development but several questions remain. We take a look at the use of DNA microarray for gene expression or expression profiling and how DNA chips are used for drug discovery and development as a whole.

The pharmaceutical industry demands drug research and development that is faster, better and cheaper. The post genomic era of research and development has the potential to deliver on some of these desired attributes. Up until now, the number of targets used for drug discovery has been limited to several hundred (300-500). The completion of the sequencing of the human genome has uncovered 32,000 to 35,000 predicted genes, offering the prospect for many more drugable targets. Many investigators are now confident that the number of targets will rise into the thousands over the next several years. Recent kinase targeted drugs like Gleevac® are achieving very promising results in a number of types of cancer, so kinases have become an even more popular target for drug discovery. The sequence analysis of the 32-35,000 predicted genes indicates there may be as many as 1,000 kinase or kinase-like genes. Many of these may be legitimate targets for drug discovery and development. Questions that remain are: 1) How many of these predicted kinases are real genes? 2) How many are in key regulatory pathways and thus good targets? 3) How many are drugable? Gene expression analysis may uncover or facilitate answers to these questions as well as assist research and development and health care in several other ways.

DNA microarrays

In today's research laboratory, two common ways that gene expression is assessed is by using quantitative polymerase chain reaction techniques (QTPCR) for monitoring expression of smaller numbers of genes, and by employing DNA microarrays for the highly parallel monitoring of thousands of genes (or the whole transcriptome). DNA microarrays take advantage of the major feature of the DNA double helix, the sequence complementarity of the two-paired strands, by using DNA capture probes which are the complement of the expressed target sequence (mRNA, cRNA or cDNA made from the mRNA). Two of the most common uses of the DNA microarrays are genetic analysis and the analysis of gene expression. Genetic analysis includes procedures for genotyping, SNP (single nucleotide polymorphism) detection, strain identification and various other procedures. Analysis of gene expression can include measuring expression levels of a small set of genes to whole genome expression monitoring. This report will focus on using DNA microarrays for the analysis of gene expression or expression profiling, and how DNA chips are used for drug discovery and development.

Origin of DNA chips

Early studies of DNA melting and reformation were carried out in aqueous solutions and yielded

By Dr Ralph Sinibaldi

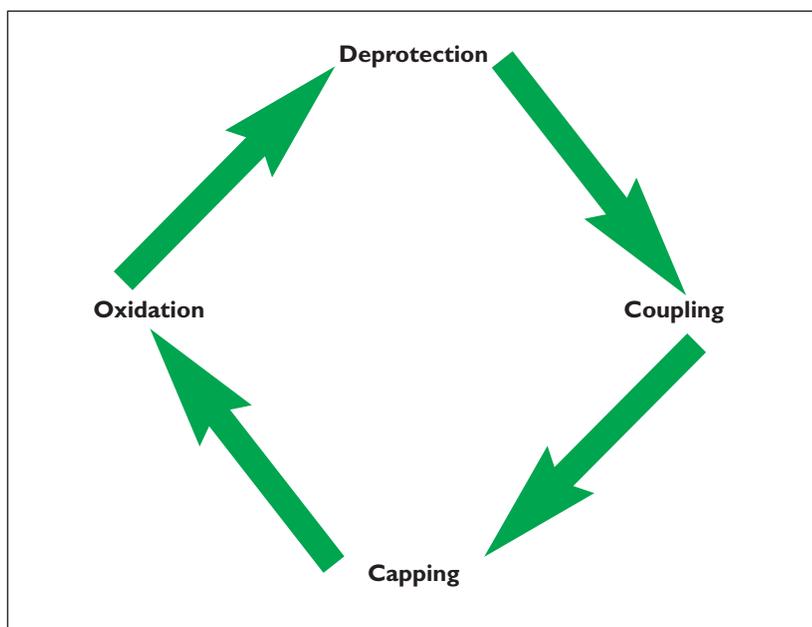


Figure 1
Four steps of DNA synthesis chemistry

important information about the dependence of melting temperature (T_m) on the G+C composition and salt concentration, as well as information on the dependence of the rate of reassociation on the sequence complexity of the nucleic acid. The introduction of solid supports for DNA hybridisation/reassociations greatly broadened the range of applications of nucleic acid hybridisations, and provided the basis for solid-based methods being used today. Gillespie and Spiegelman¹ observed that single stranded DNA binds strongly to nitrocellulose membranes in a

manner that minimises the two strands reassociating with each other, but allows the hybridisation to complementary RNA. This method was used to measure the number of copies of repeated genes such as rRNA genes² and to measure whether specific genes were under-replicated during the replication process used in forming polytene chromosomes³. Dot blotting and dot hybridisation⁴ evolved out of the filter hybridisation technique and provided the basic concept for DNA arrays. The difference between dot blots and today's DNA microarrays lies in a smaller spot size and the use of a non-porous rigid solid support such as glass, which has its advantages over a porous membrane. The membranes require much large volumes of hybridisation solutions to hybridise to the immobilised DNA in the porous substrate. The non-porous support requires very small volumes of hybridisation solution and more rapid rates of hybridisation are evident when compared to filter hybridisations. Non-porous support also facilitates the washing step of the hybridisation and provides a matrix of very low inherent fluorescence so fluorescently labelled probes can be effectively utilised in the hybridisation. Furthermore, the glass substrate provides an amenable substrate to which DNA or oligonucleotides can be stably attached by several types of chemistry.

DNA array types

In situ synthesised

The two most common types of DNA microarrays are those in which the DNA (in the form of a single stranded oligonucleotide of 25nt) is actually synthesised *in situ*⁵ and those where the DNA (usually in the form of a cDNA or full length ORF) is post-synthetically attached to a glass support⁶. The *in situ* synthesis of capture probes is performed with a photosensitive chemistry using a series of photo masks⁵ or the photo labile protecting group can be removed by selective exposure to UV light using an array of tiny mirrors to direct the light. The *in situ* synthesis can be also performed with standard DNA synthesis chemistry using ink jet or piezo-electric delivery of the reagents in a four step process (Figure 1). The first type is very useful for genetic analysis, which requires relatively short oligonucleotides but can also be used for expression analysis by using many short oligonucleotides (12-20) to cover a gene. The length of the synthesised oligonucleotide is limited by the relatively low coupling efficiency (92-96%) of *in situ* DNA synthesis (see Figure 2).

Figure 2: Coupling efficiency and % full length oligonucleotide probes

COUPLING EFFICIENCY %						
LENGTH	90	92	94	96	98	99.4
10	35	43	54	66	82	94
20	12	19	29	44	67	89
25	6.4	11.4	20	34	59	85.5
30	4.2	8.2	16	29	55	83
40	1.4	3.6	8.4	20	45	79
50	0.52	1.5	4.5	13	36	74
60	0.18	0.67	2.4	8.6	30	70
70	0.063	0.29	1.3	5.7	24	66

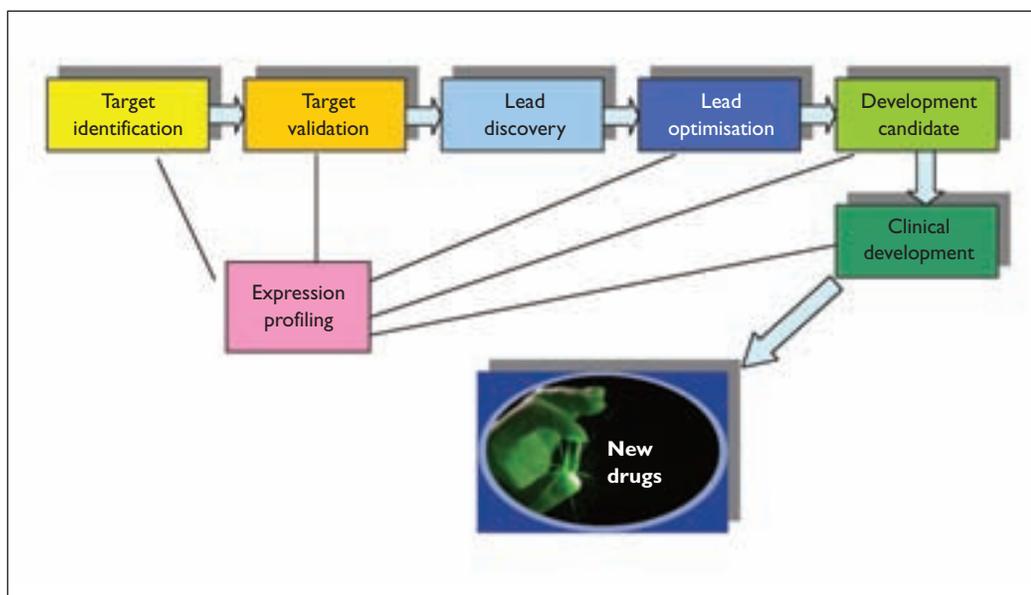


Figure 3
The drug discovery and development pathway-expression profiling can be at several steps

ORF or cDNA-based microarrays

Historically, the deposition of pre-made nucleic acid probes has involved the synthesis of full-length ORFs or cDNAs and printing/depositing them using pin spotters on 1in x 3in microscope slides. This homebrew type of microarray is used for the analysis of gene expression, but has certain limitations. The ORFs are extremely variable in their length and localised T_m , which can lead to several hybridisation issues. Experiments are usually designed using the dual fluorescent label approach, where the cDNA made from the control RNA (or treatment 1) is labelled with one fluorophore, and the cDNA made from the experimental RNA (or treatment 2) is labelled with a distinct contrasting fluorophore. Both labelled cDNA targets (or RNAs) are hybridised to the same array and results are tallied by comparing the ratios of the two fluorescent emissions of the fluorophore containing cDNA targets hybridised to each of the printed DNA capture probes. A fluorescent scanner scans the fluorescent pattern of each fluorophore; and the two patterns can be overlaid to assess which genes have been up-regulated or down-regulated by the experimental treatment.

A more important limitation of the ORF-based microarrays is the issue of cross-hybridisation of related or overlapping sequences or genes. Proteins from genes may have common features (ie ATP binding sites) and can have some degree of sequence identity or homology with other genes. Most organisms contain a number of genes in gene families, and in many cases these genes have a great degree of sequence identity to each other and

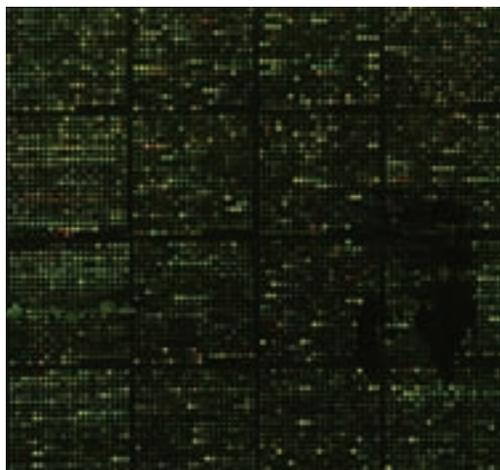
can only be distinguished from each other by designing and using shorter gene-specific hybridisation probes. Thirdly, many organisms have overlapping genes where one gene is on one strand of the DNA duplex and another gene is found on the complementary strand of the DNA duplex. For example in *S. cerevisiae*, 728 of the ~6,000 genes have 100% sequence identity over a distance of 101nts. These genes' expression levels cannot be accurately measured using full-length ORF-based DNA microarrays. Lastly, many organisms employ alternative RNA splicing of genes in response to differentiation and other signals, and these alternative forms of gene expression cannot be distinguished on ORF-based arrays.

Long oligonucleotide-based microarrays

Other investigators have utilised longer oligonucleotides (70mers) to give increased sensitivity over the *in situ* synthesised 25mers, and post-synthetically attaching them a glass matrix using a stable linkage⁷. These oligonucleotides are synthesised in a closed system, typically using a CPG column, where coupling efficiencies range from 98-99.4%. The offline synthesised oligonucleotides can be assessed by mass spectrometry, and then be further purified, resulting in capture probes that are >98% full length. In contrast, *in situ* synthesis coupling efficiencies can range from 92-96% and the truncated products remain attached to the chip surface. The design of these longer oligonucleotide capture probes uses a complex computer program where the melting/hybridisation temperatures are normalised, secondary structure is avoided and cross

Figure 4

A overlay of a two-colour experiment performed with more than 6,000 genes in response to a change in growth conditions of the cells



References

1 Gillespie, D and Spiegelman, SA (1965). A quantitative assay for DNA-RNA hybrids with DNA immobilized on a membrane. *J. Mol. Biol.* 12; 829-842.

2 Ritossa, F, Malva, C, Boncinelli, E, Graziani, F and Polito, L (1971). The first steps of magnification of DNA complementary to ribosomal RNA in *Drosophila melanogaster*. *Proc. Nat. Acad. Sci. USA* 68; 1580-1584.

3 Sinibaldi, RM and Cummings, MR (1981). Localization and characterization of rDNA in *Drosophila tumiditarsus*. *Chromosoma* 81;655-671.

4 Kafatos, FC, Jones, CW and Estradiatis, A (1979). Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. *Nuc. Acids Res.* 24; 1541-1552.

5 Lockhart, DJ, Dong, H, Byrne, MC, Follettie, MT, Gallo, MV, Chee, MS, Mittmann, M, Wang, C, Kobayashi, M, Horton, H and Brown, EL (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* 14;1675-1680.

6 Schena M, Shalon, D, Davis, RW and Brown, PO (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270;467-470.

hybridisation to related sequences is minimised by an extensive BLAST search of all the genomic sequences. The single-stranded 70mer oligonucleotides require no heat denaturation and are of sufficient length that shredding of the target RNA (to better match the capture probe length) is not required. Therefore, standard cDNA synthesis methods can be employed to make the labelled hybridisation target. The 70mers can be designed to examine alternative or differential RNA splicing. This longer expression analysis technology has been tested in various biological systems including diauxic shift experiments in yeast, heat shock and oxidative stress⁸. QIAGEN/Operon has made available genomic sets of 70mers for human, mouse, yeast, Arabidopsis, *Drosophila*, *Candida*, *Mycobacterium* and other organisms using this probe design, and this technology has yielded significant results in many laboratories studying expression profiling and systems biology. These genomic sets allow the researcher maximum flexibility in the layout of capture probes on the chip and to include additional capture probes of their own design and purpose. The probes are deposited using either a pin spotter or a modified inkjet or piezo dispenser. The printing process requires an investment in equipment, clean room space and personnel to perform the tasks.

How are DNA microarrays used in drug discovery?

The drug discovery and development pathway is outlined in Figure 3.

Target identification and validation

Expression profiling is used at several points in drug discovery. First, it can help to identify and validate new drug targets. As previously men-

tioned, there are potentially more than 1,000 kinase targets and expression profiling can determine whether these genes are expressed by detecting specific hybridisation of the target sequence to the capture probes on the chip (assuming a specific capture probe can be designed). Secondly, we can begin to locate where these potential kinase genes may be found in various biochemical pathways by determining how these genes are regulated under various conditions, and by determining which genes are being co-regulated with our target gene. A two-colour microarray image is shown in Figure 4 where two sets of conditions were imposed on the biological system and mRNA was isolated from each of the two treatments to the cells. One population of mRNA was reverse transcribed and subsequently labelled with the Cy3 fluorophore and mRNA from the other treatment was isolated, reverse transcribed and labelled with the Cy5 fluorophore. Both populations of target cDNAs were hybridised to a DNA microarray spotted with long oligonucleotides (70mers) specifically designed for each potential gene in this biological system. Hybridisation was for 16 hours and the DNA microarray was washed and dried before being scanned with a dual laser scanner. The results from this scan are illustrated in Figure 4. This figure is an overlay of both the Cy3 and Cy5 images with the following colour-coding. Red indicates a gene that is being up-regulated when comparing treatment A to treatment B, whereas green indicates genes that are down-regulated and yellow indicates genes whose expression remains unchanged. The fluorescent emission wavelengths of these two fluorophores do not overlap, so each channel can be accurately measured with both targets hybridised to the chip. Thus, the two colour experimental results are stated in ratios of one channel to the other channel. DNA chip platforms that utilise only one colour unveil hybridisation signals of varying intensities that require two separate chips for comparison between treatments or for comparing control levels to experimental levels.

After several experiments have been completed using varied treatments to the biological system (cells), a clustering algorithm can be used to illustrate which genes are being co-regulated during the various treatments. These genes are displayed in a clustering diagram shown in Figure 5. Depending on the treatment, a specific molecular signature of gene expression will become apparent. DNA microarrays can only reproducibly distinguish gene expression levels that vary by greater

Continued on page 43

than 50%. More subtle increases or decreases in expression levels such as 10% or 20% are not resolved. The genes that are up-regulated are illustrated in shades of red on the left and genes that are down regulated are clustered on the right and are various shades of green. Genes who are co-regulated may lie in the same biochemical pathway or serve a similar biological function. This can be analysed by using a number of different pathway software that is commercially available. If we know which biochemical pathways certain unknown genes are mapped to, we may get an idea to what these genes may be doing biochemically. It may be necessary to perform gene knockout experiments to fully validate and identify the function of an unknown gene. This can be done in gene knockout vectors with transgenic animals or with siRNA or miRNA in cellular systems.

Lead optimisation

The second point in the drug discovery pathway where expression profiling is done is during lead optimisation. After a compound has illustrated the

desired drug activity during the high throughput-screening step, its activity needs to be further evaluated and optimised. Typically, the specificity and efficacy of the lead compound needs to be determined. Efficacy determination is performed with standard dose response experiments, whereas specificity can be examined by several methods including expression profiling. Exposure to certain drugs will result in a pattern of gene expression that is indicative of that class of drug. If a more complex pattern or signature is observed where multiple biochemical pathways and genes are affected, this may be a red flag. This can indicate that the drug will have significant side-effects or may be toxic. The lead optimisation scientist may have additional compounds made by combinatorial chemistry that are similar to the lead compound and investigate the structure activity relationship (SAR) of all the related compounds. By examining the expression profile of each of the derivatives, the researcher will be able to more accurately choose a compound that has the desired activity and affects the least number of other genes. This process should lead to having

MipTec
The Leading European Event on Enabling Technologies for Drug Discovery

3–6 May 2004
Convention Center Basel, Switzerland
www.miptec.com

The International Conference and Exhibition on Drug Discovery

- Target Identification and Validation
- Biomolecular Screening Strategies and the Hit-to-Lead Process
- Novel Assay Technologies and Orphan Target Approaches
- Structure-Based Drug Discovery and Chemical Genomics
- Chips, Arrays and Microfluidics
- Leveraging the Power of Informatics in Drug Discovery
- Maximizing Value from compound Collections
- Biomarker Discovery, Proteomics, and Pharmacodynamics
- ADME-Tox and Pre-Clinical Profiling

Pre-Conference
Innovations in Anti-Cancer Drug Discovery
4 May 2004

Sponsored by

bucher biotec

Molecular Devices

QIAGEN

mich messe schweiz

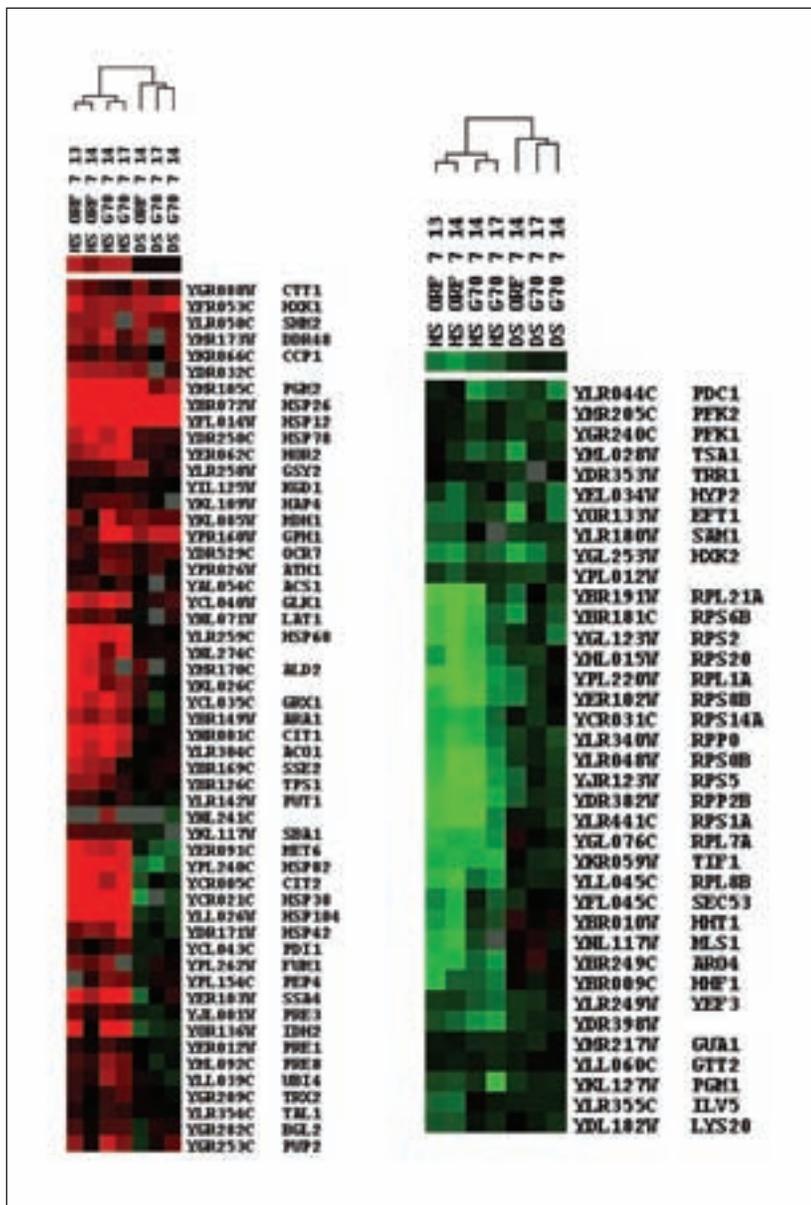


Figure 5
 A clustering diagram of co-regulated genes. On the left is a cluster of co-regulated genes that are being up-regulated for the most part. The genes are plotted on the Y-axis and the experimental treatments are plotted on the X-axis across the top. Many of these genes are involved in stress response. On the right is a clustered set of genes that are being down-regulated for the most part during the same experimental conditions shown on the left. Many of these genes are ribosomal protein genes that are co-regulated

fewer lead compounds being rejected on the basis of toxicity or secondary effects.

Development

Toxicity is a major area where expression profiling cannot only save time and money, but can lead to better drugs. From a cost standpoint the earlier a drug candidate can be eliminated from further consideration the better. As mentioned in the last paragraph, lead optimisation already deals with specificity and the lack of sufficient specificity can be indicative of toxicity. One would like to be able to confidently remove lead compounds from the development pathway that may be toxic or have too many secondary effects. Many companies have begun to expression profile lead compounds, and develop sets of criteria that eliminate offending lead compounds from further development. This is an ongoing process that requires a significant amount of expression profiling data along with correlative toxicological data to develop expression criteria that will predict the toxicology. One company, Iconix Pharmaceuticals, developed a subscription-based expression profile database for all the FDA registered drugs, as its business. They have illustrated that the expression profile analysis of a drug family leads to a better understanding of side effects and to the drug mechanism. They are confident that the drug signatures they observe result in classification models that can be used to predict drug effects, a very powerful tool. This work shows that expression profiling can be used for determining the mode of action of drugs as well as serving as a predictor of toxicological problems.

Clinical development

Expression profiling may also facilitate the phase IV clinical trials where new indications, including new conditions or diseases for drugs may be sought.

Diagnostics

The use of DNA chip technology has extended beyond research and development efforts into other areas of medicine and healthcare. DNA microarrays and expression profiling are being increasingly used for the diagnosis of various diseases, the classification of cancers and for prognosis for several forms of cancer. Cancer investigation with microarrays has dominated the research efforts with more than 70% of all microarray papers investigating diseases published since 1995 dealing with cancer in some form. One landmark paper appeared in 2000 where the investigators were able to classify previously unclassifiable forms of B-cell lymphoma⁹. Since then many

Continued from page 40

papers have appeared including two where the investigators prove that expression profiling can be used as a predictor of survival in breast cancer patients^{10,11}. It is only a matter of time before expression profiling will be used in personal medicine applications to guide the most appropriate drug therapy for the patient.

Summary

There are a number of commercially available expression profiling systems to which researchers have access, as well available reagents (genomic sets of probes) that the investigator can use to fabricate their own DNA chips. Equipment and software to process and analyse these chips is also available from many vendors. Some researchers prefer the consistency of commercial DNA microarray systems, whereas others prefer the flexibility of the chips that can be fabricated by the investigator. The number of published expression profiling research papers has grown logarithmically since 1995, yielding significant new data about gene expression. Comparison of expression data obtained from the different DNA chip platforms remains difficult, as each platform has its own characteristics, strengths and weaknesses.

Several questions remain. Does expression profiling technology result in cost or time saving for the pharmaceutical industry? The simple answer is no, as there has been little if any reduction in drug development times since the 1980s¹² and costs have risen significantly. The number of new drug approvals has actually declined since 1996, so what has happened? Even though the number of potential drug targets has risen significantly, identifying and validating these new targets is consuming significant time and resources. It is apparent that expression profiling has been a key technology in this effort, and that many new targets have been identified and validated. However, this is a long process and these targets are being gradually incorporated into high throughput screening schemes. Reliable and robust assay development used for the screening remains a significant bottleneck. Expression profiling also has a role to play in predictive toxicology, but FDA acceptance is needed in order to have its greatest impact. FDA approval will be needed before this technology can be adapted for diagnostic and prognostic use.

It is clear that expression profiling is playing an even more important role in lead optimisation. By examining the effect a lead compound has on the expression of any gene in the genome gives the investigator a more complete picture of potential effects a drug may have than he has ever had

before. This is helping researchers to more thoroughly investigate structural variations of a lead compound. Investigators can select structural variations that have good efficacy and are directed to the specific target without altering the expression of genes found in other biochemical pathways. This is clearly leading to better and more specific drugs. Better and more specific drugs lead to better health and ultimately to lower costs for healthcare. **DDW**

Biotechnology Consultant and Educator Ralph Sinibaldi, PhD has been involved with molecular biology research and development for more than 30 years. He consults in the areas of genomics, gene expression, DNA microarrays and drug discovery at the University of California, Berkeley Extension, as well as developing and offering courses in biotechnology at Ohlone College in Fremont, CA. Ralph has more than 22 years of experience in biotechnology management. He has been involved in several start-ups including a high throughput screening company and a gene expression technology company. He was Vice-President of Scientific Affairs at Operon Technologies, the leading DNA synthesis company. He also served as the Associate Research Director/Senior Staff Scientist for the Sandoz (Novartis) biotechnology department in Palo Alto, CA for 14 years. Ralph has turned over major pieces of technology to Sandoz and Operon for product development and has published numerous scientific articles on original research. He was directly involved in the development of more than 30 products at Sandoz and Operon. Ralph has received organisational and management training from UC Berkeley. He has BS, MS and PhD degrees in biology from the University of Illinois. He received postdoctoral training at the University of Illinois Medical Center, where he was the recipient of a Damon Runyon-Walter Winchell Cancer Fund postdoctoral fellowship and at the University of Chicago where he was an NIH trainee in Developmental Biology.

- 7** Sinibaldi, RM, O'Connell, C, Seidel, C and Rodriguez, H (2001). Gene Expression Analysis on Medium-Density Oligonucleotide Arrays. In "Methods in Molecular Biology" Vol. 170, Ed. J.B. Rampil, Humana Press.
- 8** Seidel, C and Sinibaldi, R (2002). Genechips for Age/Stress-Related Gene Expression and Analysis. In "Oxidative Stress and Aging: Advances in Basic Sciences, Diagnostics and Intervention" Ed. R.G. Cutler and H. Rodriguez. World Scientific Publishing Co.
- 9** Alizadeh AA, Eisen, MB, Davis, RE, Ma, C, Lossos, IS, Rosenwald, A, Boldrick, JC, Sabet, H, Tran, T, Yu, X, Powell, JI, Yang, L, Marti, GE, Moore, T, Hudson, J Jr, Lu, L, Lewis, DB, Tibshirani, R, Sherlock, G, Chan, WC, Greiner, TC, Weisenburger, DD, Armitage, JO, Warnke, R, Levy, R, Wilson, W, Grever, MR, Byrd, JC, Botstein, D, Brown, PO, Staudt, LM. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000 Feb 3;403(6769):503-11.
- 10** van't Veer, LJ, Dai, H, van de Vijver, MJ, He, YD, Hart, AA, Mao, M, Peterse, HL, van der Kooy, K, Marton, MJ, Witteveen, AT, Schreiber, GJ, Kerkhoven, RM, Roberts, C, Linsley, PS, Bernards, R, Friend, SH. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002 Jan 31;415.
- 11** van de Vijver, MJ, He, YD, van't Veer, LJ, Dai, H, Hart, AA, Voskuil, DW, Schreiber, GJ, Peterse, JL, Roberts, C, Marton, MJ, Parrish, M, Atsma, D, Witteveen, A, Glas, A, Delahaye, L, van der Velde, T, Bartelink, H, Rodenhuis, S, Rutgers, ET, Friend, SH, Bernards, R. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002 Dec 19;347(25):1999-2009.
- 12** Burrill, GS. Fewer drugs approved, more money spent, Where's the Beef? *Drug Discovery World Winter 2003/4*: 9-11.