

# The next decade of gene expression profiling

The art and science of prediction is a risky and humiliating exercise. As bad as we are at it, predicting things remains one of the fundamental human instincts, undoubtedly linked to a Darwinian selective advantage for our species. So no matter that the umbrella was left behind because rain was not predicted or the bridge falls down because we did not know enough about harmonics, our predictions for the next decade of gene expression profiling will undoubtedly be accurate.

**T**he last decade has seen a remarkable transformation in biology based principally on three technological advances: automated sequencing, gene expression microarrays and information technology. The root of each of these technologies has, in reality, many decades of development. This is because technology evolution typically shows exponential development curves, which is often exciting when the inflection point for accelerated rate of change occurs, but the long and slow preceding development is frequently forgotten (Figure 1). More importantly, although the rate of change in the middle part of the curve is frequently associated with quick and remarkable advances along with very high expectations, it is the latter part of the curve – when the rate of change flattens out – that the biggest value is created. This phase of the curve represents the commodity phase, when a technology becomes ubiquitous. The term commodity has an unfortunate negative connotation in marketing and sales circles, but this is a valuable phase in technology. Cell phones and computers are currently in commodity phases and this is a testament to the robustness of the technology. In the commodity phase, the technology is then no longer the next big new thing that will solve all problems, but it does then, in fact, yield real advances in understanding simply because it has been widely adopted over time and has been combined with other advances and accumulated effort.

Gene expression technologies are beginning to move to the upper parts of this technology development curve, though there are many improvements still to be made.

While the advances in sequencing technology and information technology have been, in and of themselves, very exciting and valuable, these technologies have also greatly contributed to the advances in gene expression technology. The development of microarrays was a tremendous advance that would not have succeeded without detailed knowledge of the sequence of the human genome and the ability to handle and integrate large amounts of data. While the effort to determine the sequence of the human genome was a tremendously expensive, highly visible effort that easily captured the attention of the scientific world as well as the general public, the actual sequence is static and its function, for the most part, is not understood (although much can be inferred from work in comparative genomics and evolutionary genomics). Sequence data gives us a crude blueprint of biology, but certainly does not reveal how or why it works. In contrast, the big advantage of gene expression study is that it truly helps to reveal how the genome works and unlocks basic biological function.

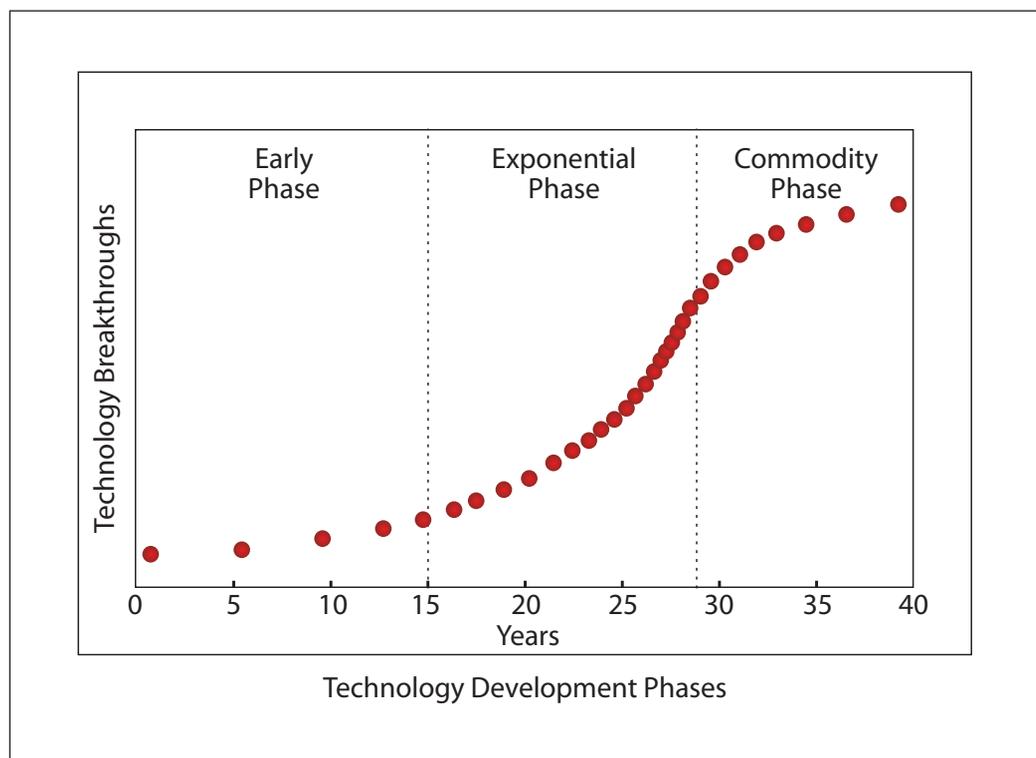
## Where gene expression is today

Ten years ago, gene expression was understood one gene at a time. Scientists could with confidence

By Dr Y. Douglas  
Dolginow

---

Figure 1



point to a handful of genes involved in diseases such as Alzheimer's or diffuse large B cell lymphoma. Today we can identify thousands of genes that are associated with a specific disease or biological process. Five years ago, biology researchers in academia and pharmaceutical settings had varying degrees of acceptance of the importance of gene expression microarrays in their work. Today, the major pharmaceutical companies and academic centres have established internal microarray programmes and routinely utilise gene expression data generated by microarrays. Most major institutions conducting biological research have established gene expression core facilities and many are now in the process of upgrading to high-throughput systems to apply these technologies to clinical trials, basic research and biomarker discovery programmes. Clearly, gene expression study is now integrated into life science research and drug development in a variety of productive ways.

### What is gene expression good for?

Gene expression has a wide variety of current applications and future possibilities. First and foremost, it is a sensitive indicator of biological activity, processes and change. Almost any biological activity is reflected in a changing gene expression pattern that can be measured with microarrays. Researchers use gene expression analysis to identi-

fy key genes correlated with biological processes of interest or groups of genes (profiles). For example, there has been a wealth of targets identified as part of the target validation process wherein gene expression is used to identify candidate genes and help prioritise those genes. Gene profiles are also successfully used to predict the toxicity of novel compounds in humans as part of the routine pre-clinical drug candidate assessment. However, it can be difficult to distinguish key regulatory genes directly involved in a biological process from those that simply correlate with the biological process. However, gene expression is increasingly used to identify mechanisms of toxicity and mechanisms of action of novel compounds.

Another area of application of gene expression technology is to facilitate the value of other technologies. For example, when combined with iRNA or knockout models, gene expression is an excellent tool for identifying mechanisms of biological activity and creating putative pathway associations. Gene expression data are also used to help focus more in-depth evaluations of biology via proteomic technology and, to some extent, single nucleotide polymorphisms (SNP) discovery and disease association studies<sup>19</sup>. A growing number of efforts are under way to link gene expression profiles with chemical profiles via chemigenomic efforts<sup>20</sup>. As the cost of gene expression data generation declines

and the efficiency of analysing the massive amounts of data increases, it becomes increasingly possible to evaluate the extremely large number of chemicals and biological systems involved.

Gene expression has taken us beyond mechanisms of action and pathway association. Microarray-based gene expression data is also proving to be a very useful predictor of biological outcomes, such as disease-related prognosis or the response of organisms to chemical compounds or other external perturbations. Most outcome predictions require large data sets to establish reliable predictive pattern algorithms and assure high specificity in the prediction.

As exciting and dramatic as the last 10 years have been, we are still only at the beginning of a 20 year development cycle, not dissimilar from information technology as it stood in 1980 versus where it is now in 2005. What do we expect to see over the next 10 years?

### The next 10 years: the new frontiers

#### Reclassified biology

One of the near certainties over the next 10 years is the persistent application of gene expression data to reclassify disease and biological processes. There has been an explosion of articles published over the last two years using gene expression data to look at disease classification and prognosis<sup>5,6,13-18,21,22,26,29,30</sup>. Today there are more than 200 articles published demonstrating the effectiveness of this approach to biological stratification of patients in a clinical setting.

We can draw a number of conclusions from these publications. First, gene expression is a very sensitive and accurate biosensor. In most cases it is better for this purpose than the traditional pathology-based models. For example, when there has been a discrepancy in comparing gene expression data versus traditional data, the expression profile is more likely to be correct. It is the rule rather than the exception that informative gene sets will be identified that segregate the biology of interest when gene expression data are analysed. In fact, it is difficult to find an example of a studied biological process where there is not an informative expression profile.

Historically, new technologies that are validated and adopted as part of the standard tools of research biology have led to significant changes in disease classification. Examples of this are numerous with each new technology advance, as evidenced by, for instance, advances in tissue chemistry (Hematoxylin and Eosin stain), immunophenotyping of leukemia, cytogenetic analysis in lym-

phoma, thyroid disease, adrenal disease, ovarian function with the advent of radioimmunoassay analysis to name a few. In each case, technology has changed the way we define disease, predict outcomes, choose biomarkers to assess and determine treatment. It is almost a certainty that molecular gene expression profiles will redefine a wide spectrum of diseases and biological processes.

Ultimately, this will lead to new diagnostic biomarkers and clinical diagnostic assays routinely employed in the hospital laboratory. However, this transformation will likely occur over a 10-20-year period rather than a 10-year period. Many scientists claim that diagnostic molecular markers will be a reality within 10 years. However, experience suggests that a combination of factors will hinder rapid availability and widespread use of molecular diagnostics being widely in the marketplace. Some of the roadblocks include:

- Requirement for extensive, detailed clinical diagnostic trials including those showing that information can be associated with a positive intervention strategy
- Needed technology improvements that further reduce cost and establish broad, accepted controls and standards, which is typically a difficult and time consuming process
- Improved sample collection and handling
- The slow pace of establishing insurance reimbursement
- Physician education about the use of new assays
- Traditionally limited economics within the diagnostic market

Many ideas in biotechnology are over-exposed and perhaps over-promised due to factors such as those listed above. Alternatively, two key factors – efficiency and quality – help a technology recognise its potential.

#### Improved efficiency and quality

Moore's law, that the number of transistors per integrated circuit will double every few years, is cited so frequently that its truth may seem easily dismissed. In fact it is quite relevant to gene expression study, a field still in relative infancy. The last 10 years have seen a persistent increase in the density of relevant information that can be generated along with a steady decrease in the cost of actually generating that data. This trend will continue. The factors driving this include: 1) the development of higher density arrays; 2) automation of the entire analysis process; and 3) competitive pressures.

High density arrays will become increasingly dense due to improvement of manufacturing

Figure 2



technologies. The capabilities of probes used to scan the arrays will also increase for the same reason. The curve of increasing density over the past 10 years for Affymetrix is a good example of the progress in this field and there is every reason to believe that it will continue for the next 10 years (Figures 2 and 3).

Surprisingly, automation of gene expression data generation is lagging. The process of generating gene expression data is, at best, semi-automated at this time. Sample preparation, sample loading and scanning continue to be mostly manual processes. The next decade will bring fully automated, high-throughput, stand-alone systems with ever-increasing quality control and better calibration methods. It is very likely that a common metric will be established that allows conversion of a measured signal to a standardised measure of quantitative gene expression, such as copy per cell. This type of progress has been seen over and over in technology development cycles over the last 50 years (eg chemistry analysis, hormone assays, ELISA assays and RIA assays, to name a few), culminating in accepted standard reagents as defined by govern-

ment agencies such as the National Institute for Standards and Technology (NIST) in the US, federal regulatory agencies, or professional and manufacturing associations.

Finally, there are multiple technology platforms now available commercially. While the market is still relatively small and dominated by a single company, commercial competition will, by its nature, push technical improvements and focus more and more attention on efficient data generation, reduced costs, improved standardisation and improved sensitivity. These competitive driving pressures are extremely important and will greatly benefit the scientists utilising these technologies and aid in the general expansion of the market size.

The general notion propagated over the last 10 years is that high density arrays will only be used for limited discovery efforts and the results migrated on to higher throughput, lower cost systems. This has been driven more from a desire to reduce costs than anything else. In reality, higher density platforms result in generating more data that have higher predictive value, better sensitivity and better quality. If the cost of data generation is low, the

benefits of more data rather than less will be recognised and high density arrays will continue to be utilised for the next decade in increasing numbers, rather than decreasing.

Inherent in Moore's Law is that not only will more information be generated but better information will be generated. Gene expression information content will continue to improve qualitatively because of better genome annotation, development of alternative splice arrays and creation of whole genome arrays. Despite the announcement of the completion of the human genome project in 2001, annotation of the human genome still has far to go. There is still no consensus of how many genes are in the human genome and a detailed, reliable map of the human genome will continue to evolve over the next decade. This is partly dependent upon continued advances in sequencing technology so that many more complete genomes can be generated. Beyond the simple need to identify what sequences correlate to what genes, almost every gene that has been studied has hundreds of SNP variations and multiple alternative splice variants. All of these variants must be identified and incorporated into the next generations of microarray design to improve the quality and content of the data generated. Equally important is the Encyclopedia of DNA Elements (ENCODE) joint effort between Affymetrix and the National Human Genome Research Institute (NHGRI) to validate a full genome microarray. This effort has already identified a wealth of relevant gene expression data and is likely to lead to a new generation of whole genome arrays that will further elucidate key, fundamental and still unknown functions and mechanisms of the genome.

#### Shift to a commodity phase of the technology cycle

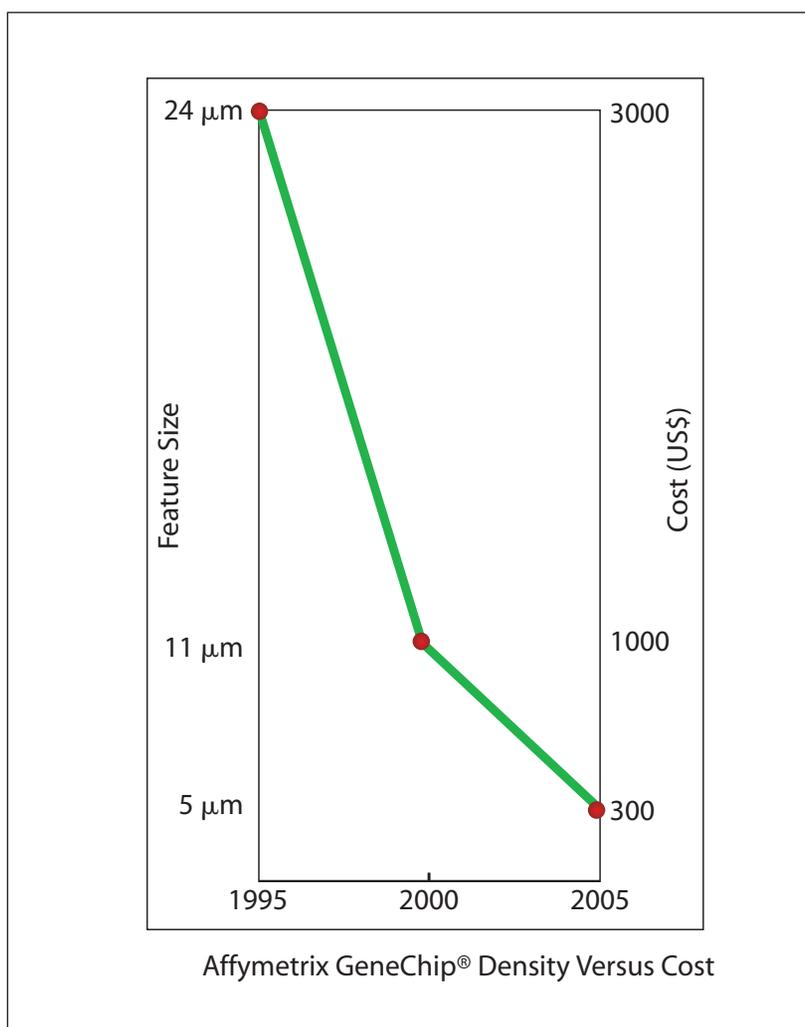
The combination of improved efficiency, lower cost and repeated changes in the technology format will likely result in a shift to large production centres to take advantage of scale. Again, this has been seen with almost every technology innovation as it evolves. When something is new, it is generally complex and costly. There are a few sophisticated early adaptors willing to take on the cost and complexity. As the technology takes hold, it is viewed as the salvation to a wide number of problems and a must-have technology in order to be perceived as being state of the art. Many groups will invest in these technologies simply to have them and be at the forefront. The new technology capability is prominently displayed and highlighted in communication. It appears that one has the next great thing finally in hand.

However, as others gain access to the technology, the shortfalls as well as the benefits become clear, the technology improvement waves become tiresome, and other technologies come along that become the exciting new thing, even as the current one becomes widely established. The emphasis shifts to high quality and low cost and generating the data internally becomes less important. When high-throughput, dedicated facilities that take advantage of economies of scale become available, the work becomes more centralised or at least regionalised. In addition, continued technology changes in instruments, algorithms, data formats and array versions will push groups to conclude that they would rather let somebody else sort out those problems, let them generate the data and avoid the hassle.

#### Other technologies will improve

Over the next decade, SNP technology and proteomic technology will improve to production

Figure 3



### References

- 1 Turecki, G. Molecular Characterization of Suicide by Microarray Analysis. *American Journal of Medical Genetics*, 2005.
- 2 Loscher, W et al. The antiepileptic drug levetiracetam selectively modifies kindling-induced alterations in gene expression in the temporal lobe of rats. *European Journal of Neuroscience*, 2004.
- 3 Gao, W. Expression Profiling of a human cell line model of prostatic cancer reveals a direct involvement of interferon signaling in prostate tumor progression. *PNAS* March 2002.
- 4 Shen, Grace. Discovery of Novel Tumor Markers of Pancreatic Cancer using Global Gene Expression Technology. *American Journal of Pathology* April 2002.
- 5 Getzenberg, H. Symptomatic and asymptomatic benign prostatic hyperplasia: Molecular differentiation by using microarrays. *PNAS*, May 2002.
- 6 Vockley, J. Identification of Differentially Expressed Genes in Hepatocellular Carcinoma and Metastatic Liver Tumors by Oligonucleotide Expression Profiling. *Cancer*, July 2001.
- 7 Scherf, U. Large-scale gene expression analysis in molecular target discovery. *Leukemia*, 2002.
- 8 Prakash, K, Munger, W. Future Molecular Approaches to the Diagnosis and Treatment of Glomerular Disease. *Seminars in Nephrology*, January 2000.
- 9 Recupero, A. Streamlining Drug Discovery: The Use of Gene Expression Analysis Methodologies. *Pharmaceutical Discovery and Development* 2000.
- 10 Topaloglou, T, Kosky, A, Markowitz, V. *American Association for Artificial Intelligence* 1999.

Continued on page 61

## Examples of gene expression in therapeutic development

Rapidly accelerating the identification of candidate therapeutic targets

- Genentech identified candidate therapeutic targets for invasive prostate cancer (*PNAS*, March 2002)<sup>3</sup>

Improving the ability to prioritise potential therapeutic targets

- Stanford validates new targets in autoimmune encephalomyelitis (*Nature Medicine*, May 2002)<sup>27</sup>

Understanding mechanism of drug action

- UCB Pharma determined mechanism of action of effective anti-epileptic drug, levetiracetam (*European Journal of Neuroscience*, 2004)<sup>2</sup>

Predicting the human toxicity of novel compounds

- Identification of gene profiles that predict human specific toxicity using rat gene expression data (*Toxicology Science*, 2002)<sup>31</sup>

Understanding the mechanism of toxicity of compounds

- Mechanism of canine-specific hepatic sclerosis determined for novel compound (Mattes, *Gene Logic*, submitted for publication)

Identifying biomarkers used to assess therapeutic response

- Blood biomarkers identified to follow response to anti-depressants in treating patients with major depression syndrome (*American Journal of Medical Genetics*, 2005<sup>1</sup> and unpublished communication)

Improving clinical trial outcomes by selecting appropriate patients for investigational new drugs

- Clinical stratification of tomosifen-treated, node negative breast cancer (*NEJM*, December 2004)<sup>13</sup>

Improving disease classification for developing targeted therapeutics

- Diffuse large B-cell lymphoma redefined based on expressions profile (*Nature Medicine*, January 2002)<sup>25</sup>

Predicting biological response to novel chemical structures

- NIH correlates molecular chemical substructure with biological response predicted from gene expression data (*The Pharmacogenomics Journal*, 2002)<sup>28</sup>

scale similar to that of gene expression technology today, and the combination of these technologies will become increasingly important. The importance of including SNP variants in the sequence content of the gene expression microarrays has already been described. In addition, it can be argued that since gene expression profiles are one of the more sensitive indicators of phenotype, they will therefore greatly improve the quality of SNP association studies as well as define phenotypes of various SNP genotypes. However, significant

advances will have to be made, and probably will be made in SNP technology for this to occur for the following described below.

There has been a growing interest in the identification of SNPs that could be used to improve drug development. The basic concept involves identification of genetic variations of a gene and correlation of these variations to some clinical effect. This approach has been tried for over 10 years with limited success. One recent example of this approach is the differential response of non-small cell lung

Continued from page 60

cancers to erlotinib and gefitinib. Patients who have a response to these drugs have been shown to have mutations in epidermal growth factor receptor (EGFR), which is the target of these drugs.

SNP analysis can most successfully be utilised when the functional activity of a gene product that is directly involved in drug response can be measured. Examples of this include evaluating SNP variants associated with metabolising and transport genes involved in toxicity, and SNPs specifically located in the binding domain of a specific drug, such as the example cited above.

When the specific biology is not known, then one must discover the genes and their associated SNP variants related to drug response. There are two general approaches to this type of discovery. The first – called the Candidate Gene Approach – involves screening selected genes for SNPs thought to play a role in the biology of the disease. The second – Genome Wide Screening – involves screening the whole genome to discover a SNP pattern associated with a particular response profile.

Both of these approaches have generally struggled to produce meaningful results. In the candidate gene approach, the SNPs driving drug response may not be in the genes selected. The Genome Wide Screen approach has been a challenge because the technology to efficiently look at all the millions of SNPs in the genome is still under development. Recently, technology has improved so that 100,000 to 500,000 of the millions of discovered SNPs can be evaluated. The problem is that statistically one needs to evaluate 500 to 1,000 patients per response group. This becomes very difficult both in terms of cost and adequate trial design to enlist enough patients. It is hoped that the requirement for large numbers of patients required for SNP association studies could be reduced in the future by developing new technology approaches such as combining SNP data with expression studies.

More importantly, the analytical tools to associate a SNP to a therapeutic effect are limited. Almost all known SNPs have other associated SNPs that significantly modulate their phenotype (for example in cystic fibrosis). Some of these SNPs are located within the same genes, some in genes associated with the relevant pathway, and some are entirely outside of genes coding for proteins. While the analytical tools to identify independent single SNPs are well established, the technology to identify multiple interacting SNPs has not been developed.

Along the same lines, even once an SNP is found that is associated with a specific phenotype, it can not be certain that other SNPs exist that are also

responsible for the same phenotype or may neutralise the phenotype. This is the case for the EGFR receptor mutation described above. Recent publications have identified two associated mutations that neutralise the beneficial effect of the originally described SNP mutation. These discoveries generally required extensive additional clinical trial work, outside the scope and possibility of a drug development programme.

As mentioned above, in situations where there is a biological assay available to assess the function of an SNP, then SNP analysis can be much easier. An example of this would be analysis of genetic variations of the metabolising and transport genes involved in drug metabolism. In this case, SNP variants of metabolising and transport genes can be cloned and used to develop microsomal-based assays to directly measure the enzymatic efficiency of each of the variants. In cases like this, where enzyme efficiency or binding efficiency can be directly assessed, the development of SNP markers is greatly facilitated.

Clearly, over the next decade technology will improve such that millions of SNPs will be validated and studied without high cost. It is likely that SNP analysis will be improved so that more SNPs and fewer samples will be required to identify relevant SNP correlations across the entire genome.

Inherent in this is the continued development of true integrated genomic databases that combine large-scale gene expression, proteomics and SNP data. Such databases will improve the efficiency of evaluating and interpreting data generated in biological experiments.

### The big challenges

So what are the big hurdles facing gene expression and, to some degree, genomics in general?

First and foremost is the complexity of biology. The unified field theory in physics is an attractive concept but has been the subject of intense work for nearly 100 years without success. Biology may be even more complicated. Francis Crick gave some perspective to the problem 25 years ago. He pointed out that his goal of understanding the function of the visual system was to understand how we actually see. The problem was, even after careful and detailed mapping of the visual cortex and all of its physical and electrochemical connections had been completed, a comprehensive understanding of how we actually see something had not been achieved. That is still basically true today.

In the gene expression arena, there are many uncomfortable examples to consider. For example, for nearly 15 years the complete gene sequence and

**11** Laroco, L. Extending traditional query-based integration approaches for functional characterization of post-genomic data. Oxford University Press, Bioinformatics, 2001.

**12** Lennon, G. Cystatin B-deficient mice have increased expression of apoptosis and glial activation genes. *Human Molecular Genetics*, 2001.

**13** Wolmark, N.A. Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node Negative Breast Cancer. *NEJM*, Jan 2005.

**14** Haferlach, T. Gene Expression Profiling in Acute Myeloid Leukemia. *NEJM* April 2004.

**15** Delwel, R. Prognostically Useful Gene-Expression Profiles in Acute Myeloid Leukemia. *NEJM* April 2004.

**16** Pollack, J. Use of Gene-Expression Profiling to Identify Prognostic Subclasses in Adult Acute Myeloid Leukemia. *NEJM* April 2004.

**17** Said, J. Gene Expression Profile of Serial Samples of Transformed B-cell Lymphomas. *Laboratory Investigation* 2003.

**18** Evans, W. Treatment-specific changes in gene expression discriminate in-vivo drug response in human leukemia cells. *Nature Genetics* May 2003.

**19** Friend, S.A. A new paradigm for drug discovery: integrating clinical, genetic, genomic and molecular phenotype data to identify drug targets. *Biochemical Society* 2003.

**20** Covell, D. Mining the National Cancer Institutes Tumor-Screening Database: Identification of Compounds with Similar Cellular Activities. *J. Med. Chem* 2002.

**21** Shaughnessy, J. Global gene expression profiling of multiple myeloma, monoclonal gammopathy of undetermined significance, and normal bone marrow plasma cells. *Blood*, March 2002.

Continued on page 62

Continued from page 61

- 22** Hanash, S. Accurate Molecular Classification of Human Cancers Based on Gene Expression Using a Simple Classifier with a Pathological Tree-Based Framework. *American Journal of Pathology* November 2003.
- 23** Hacohen, N. Plasticity of Dendritic Cell Responses to Pathogens and Their Components. *Science* 2001.
- 24** Jansen, R. The current excitement in bioinformatics-analysis of whole-genome expression data: how does it relate to protein structure and function. *Current Opinion in Structural Biology* 2000.
- 25** Golub, T. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nature Medicine* 2002.
- 26** Petersen I. Diversity of gene expression in adenocarcinoma of the lung. *PNAS* 2001.
- 27** Steinman, L. Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis. *Nature Medicine* May 2002.
- 28** Weinstein, J. Pharmacogenomic analysis: correlating molecular substructure classes with microarray gene expression data. *The Pharmacogenomics Journal*, April 2002.
- 29** Hanash, S. Gene Expression profiles predict survival of patients with lung adenocarcinoma. *Nature Medicine*, August 2002.
- 30** Meyerson, M. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *PNAS*, November 2001.

most of the function has been known for cystic fibrosis, the human immunodeficiency virus and haemoglobin defects. Yet, in none of these disorders is there an effective cure. Again, it is the rule rather than the exception that just because we can map out an organism we do not fully understand how an organism works. This is the primary thrust of systems biology today. With the correct approach, systems biology is no more likely to be successful in the short run than physicists have been in solving the unified field theory.

Second, gene expression technology has to improve by many orders of magnitude in detecting very low levels of gene expression in individual cells. The key regulatory events occur in a low copy number, often in a limited cell population. While laser capture microdissection has improved our ability to look at the cellular level rather than the tissue level, the technology is expensive and the data not necessarily comparable. Significant advances are needed to improve efficacy and comparability. More importantly, the relevant cell-signalling events can currently only be inferred. They must be directly and quantitatively measured across thousands of genes. In short, some improvement of low-end sensitivity must occur over the next 10 years.

Third, improvement must be made in technologies that allow new biological pathways to be constructed from expression data. Currently, genes can only be grouped into associated clusters. These clusters probably represent many different pathways that are co-regulated and interact. The type of detailed biochemical pathways typical in biochemistry are lacking in gene expression study, and to develop these today is very laborious. We require relevant biological systems that can be manipulated, for example with iRNA, so that dynamic data can be assembled into detailed gene interaction pathways with associated positive and negative feedback loops and cross pathway interactions. This will also require advances in mathematical modelling for bioinformatic analysis of the data. One of the biggest hurdles is finding relevant and accessible biological systems in which these studies can be performed. And, it assumes the cost of data generation will be very low because the amount of data that will need to be analysed will probably need to be very high.

Fourth is improvements in the efficiency of gene sequencing. Sequencing is important because of the previously described need to improve sequence annotation and identify relevant SNP and alternative splice variations. Very high throughput, very low cost sequencing will significantly advance the

pace of biological research and especially improve the gene expression studies.

Finally, improved proteomics is important to evaluate gene expression data. Gene expression can go up because a gene is active or inactive. Only direct measurement of protein function is helpful in resolving such issues. Current proteomic methods cannot effectively address this type of problem on a comprehensive, organism wide scale. In addition, post-translational modification of proteins can only be inferred from gene expression data. Again, direct and comprehensive methods of systematic, efficient proteome analysis are needed. And, once generated, the protein data need to be correlated with gene expression data. We are still 10 to 15 years away from commercial-grade proteomic systems for this purpose.

### Conclusion

It is wrong to think of the world today as the post-genomic world. It is in fact still the early genomic world with most of the productivity, benefit and advance still to come. The number of biological processes that are still to be studied, assessed and understood utilising today's sequencing and gene expression tools is almost unlimited. As cost goes down and efficiency and quality go up, large-scale programmes should and will be initiated. The results of these programmes will certainly be therapeutics designed around novel biological targets, improved productivity in therapeutic development, a complete rethinking about how we define biology and disease processes and a significantly changed and enhanced set of diagnostic tools and tests. And, we will be able to use the umbrella we brought along for the predicted rain to keep off the brilliant sunshine. **DDW**

---

*Y. Douglas Dolginow, MD is currently Executive Vice President for Gene Logic Inc and has served the company in executive roles since 1998, including as Senior Vice-President, Pharmacogenomics. Prior to September 1998, Dr Dolginow served as President, Chief Operating Officer and as a director of Oncormed, Inc, a gene therapy biotechnology company. Prior thereto, he served as medical director for several clinical laboratories and since March 1997 he has been an active member of the Clinical Faculty at the University of California, San Francisco. Dr Dolginow received an MD degree from the University of Kansas.*