High-throughput gene expression analysis

A number of powerful gene expression analysis techniques are now in existence to help accelerate drug discovery and development.

Dr Gregory Lennon, Gene Logic Inc

As the use of gene expression technologies for pharmaceutical R&D has accelerated over the past few years, several major new strategies for exploiting high throughput methods have become available for researchers. New DNA microarrays, refined differential display methods and large, searchable databases can greatly facilitate the identification and prioritisation of drug targets and lead compounds.

Imagine that in each of the cells in the human body there are 100,000 construction workers - but in any given cell, many of these aren't involved in the work at hand. Furthermore, of those that are working routinely in, say, brain cells, their performance can vary between cells from two individuals - for example, the brain cells of a normal individual compared with those of an individual with Alzheimer's disease. Such is the world of gene expression - or the study of which genes are working in different cells and how hard each is working in different conditions.

Methods for measuring gene expression can be roughly divided into two basic types - open and closed.

Methods for measuring gene expression can be roughly divided into two basic types: open methods, which examine samples to characterise previously unknown genes and their expression levels; and closed methods, which measure expression levels of known genes in a sample and compare them against levels previously found in diseased or normal tissue. These two types of method can be used to correlate mRNA expression patterns in many tissue types with specific diseases, to identify side effects of current and experimental treatments, and to determine the effects of compounds upon non-target tissues.

Differential display

Among open methods, differential display represented a major advance for comparing gene expression differences between cells or tissues when it was first reported in 1992 (1). It is perhaps the most widely used method involving gel electrophoresis for comparing gene expression, along with mRNA fingerprinting by arbitrarily primed PCR. Both methods amplify partial sequences of cDNA 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 measure both up- and down-regulation across tens of samples simultaneously. It works by non-selectively priming first strand cDNA with a one- or two-base anchor at the 3'-end, followed by amplification through PCR. Products showing significant differential expression are sequenced after size fractionation of the PCR sample with denaturing gel electrophoresis, generally after overnight autoradiographic exposure.

Differential display is one of several open methods that can be used for gene expression analysis; others include serial analysis of gene expression (SAGE), expressed sequence tags (EST) and filter arrays. We have chosen to use a modification of differential display - restriction enzyme analysis of differentially expressed sequences (READS) - because we believe it offers the best combination of sensitivity, accuracy and comprehensiveness for open analyses.
Refining differential display

READSTM is an important modification of the classical differential display technique, since it uses selective PCR amplification and display of the 3’-end restriction fragments of double-stranded cDNAs. The methodology, discovered by Prashar and Weissman in 1996, has been well-validated as a fingerprinting technique (2). READSTM works by ligating 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. Commonly, around 100 combinations of restriction enzymes and oligo (dT) primers (with specific he pale sequences) are used to selectively measure expression of approximately 100-200 transcripts per combination. READSTM provides: consistent, highly accurate, and reproducible patterns (including accurate measurement of up- and down-regulation); very comprehensive coverage of the mRNAs in a sample; adequate information for quantifying levels of gene expression; and the ability to distinguish highly homologous as well as alternatively polyadenylated transcripts.

Various differential display methods have proven helpful in detecting genes correlated with 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. Since only small amounts of total RNA are needed 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 homogeneous cell types, such as those isolated using laser-capture, micro-dissection or flow-sorting.

DNA microarrays

Unlike READSTM, DNA microarrays are used in closed analyses of differential expression - that is, to measure gene expression across a set of known samples. DNA microarrays measure expression by using templates containing many (usually hundreds or thousands of) probes that are exposed simultaneously to a target sample. Microarray techniques are so powerful that their uses are often limited largely by the challenge of managing and analysing the data they generate.

Because of their speed and breadth, microarrays will impact drug discovery in several ways - accelerating our understanding of the molecular basis of disease, improving our knowledge of model systems, exploring pathogens and pathogenic reactions in terms of gene expression, pinpointing new drug targets and examining efficacy and toxicity responses to new drugs. A few examples of each type of contribution are explored below.

Microarrays have already determined how some important genes are abnormally regulated in disease. For example, a microarray of about 100 genes that have a role in inflammation was used to examine rheumatoid tissue; it revealed up-regulation in the genes encoding interleukin-6 and a number of metalloproteinases (3). In another case, a novel gene involved in promoting tumours was discovered by using a 1,000-element microarray of unknown cDNAs to examine how treatment with phorbol testers affect expression levels (4).

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 our understanding of responses to drug treatments. For example, a recent study measured the effects of kinase inhibitors on the entire yeast genome by measuring changes in mRNA levels before and after treatment in this model system (5). In another example, microarray studies of yeast cells showed that the immunosuppressive drug, FK506, had the same effects upon gene expression level patterns as ablation of the gene suppressed by FK506. This study showed that, without the presence of the gene, FK506 affected expression levels in ways distinct from its effects upon the gene - suggesting that the drug may have more than one target (6). Finally, microarrays are already proving of assistance in the determination of drug toxicity (7).

Types of microarray technique

cDNA microarrays are made by arraying many gene-specific amplicons derived from the cDNA clones onto a single matrix. Two basic types of hybridisation are used, and each type as its advantages and disadvantages. With two-colour hybridisation methodology, cDNA representations of total RNA pools are created from test and reference cells, fluorescently tagged with two different colours, and then mixed together before being hybridised to the matrix. For each transcript, the resulting fluorescence signals reflect the difference in abundance between the two samples. Two-colour hybridisations lead to very rapid comparisons between the two samples, but they do not measure the absolute levels of gene expression for either sample.

By contrast, one-colour hybridisation technology is slower, since hybridisations of the two samples must be performed separately to obtain meaningful comparisons; however, each one-colour hybridisation measures actual, absolute levels of gene expression, rather than comparative levels. Once these actual levels are recorded into databases, they can be used to make meaningful comparisons with actual levels from other samples, without the need to perform comparative experiments. Although performing 1,000 two-colour hybridisations results in 1,000 pair-wise comparisons, conducting 1,000 one-colour
hybridisations yields close to half a million pair-wise comparisons, since the absolute values of one-colour hybridisations can be compared against any other.

Using either the one-colour or two-colour methods, cDNA microarray experiments need to be performed repeatedly to ensure accuracy of the data. However, computational averaging of the signals from multiple independent samples is more straightforward following a series of one-colour hybridisations. The choice between using one-colour versus two-colour methods depends on the number of transcripts under examination, the need for speedy results and cost differences, among other factors. One-colour methodology is often more useful for surveying a large number of genes, while two-colour methodology may be preferable for more restricted sampling.

**Approaches to gene expression analysis**

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

The earliest probe array on the market was the Affymetrix GeneChip; it is also the most popular. The GeneChip array owes its construction to two fundamental technological developments: the ability to make hundreds of thousands of oligonucleotides in very specific locations with high spatial resolution, and the capacity to measure molecular binding events precisely with the use of laser, confocal fluorescence scanning. The GeneChip array is based on light-directed synthesis, using both photolithography and solid-phase DNA synthesis. The recent use of semiconductor-like photosensit techniques has enabled the synthesis of arrays with features as small as two micrometers (8,9). Array sets are currently available covering 42,000 human, 19,000 mouse and 20,000 rat genes, and periodic updates will increase these numbers.

Gene Logic is pioneering a new approach to enhance the efficiency of DNA chips. Almost all DNA array technologies use mechanical or lithographic means to bind elements on a two-dimensional substrate. One way to improve upon this standard process is to maximise binding sites, using a three-dimensional substrate that increases the surface area available for binding probes. Gene Logic’s Flow-thru Chip uses this method. Microchannels connect the upper and lower faces of the chip so that target substances can move through it; the probes are then mobilised on the walls of the microchannels. Experiments have pointed to a number of distinct advantages for this type of chip over the two-dimensional type. These include: 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; and the potential for using smaller sample and reagent volumes because the reaction volume is reduced.

**Using gene expression databases**

The use of microarrays to prioritise drug targets and lead compounds is becoming a standard approach for pharmaceutical researchers. One strategy is to compare samples against large databases that contain gene expression data. Without presenting an exhaustive review of this subject, some illustrations can be provided from our experience in developing the GeneExpress™ databases. These types of database help identify disease, toxicity and drug response pathways by comparing known differential expression data, relevant data about biological pathways and tissues, and patient data.

Such databases are now increasingly effective since they include information about all the tissues of all the major organs of the body, and most of the significant disease states. Our database includes qualitative information on gene expression for the relevant genes in the sample tissues, based on results from GeneChip array expression studies. One GeneExpress database, called BioExpress™, is used primarily for target selection and prioritisation. 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 aberrantly regulated in a sample or set of samples, to discover what biological processes and biochemical pathways their products participate in, or to determine what their associated disease states are. One can also examine all of the genes in the database to discover which are differentially regulated in a given pathway. This type of search provides clues for determining drug targets likely to 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 prioritise or eliminate genes as potential drug targets.
A second major use for gene expression information is to assist with lead selection. In our case, we have developed a predictive toxicology database known as ToxExpress™. This database incorporates gene expression data from human and rat cells, and tissues that have been exposed to a variety of toxic compounds. Databases of this type seek to compare the gene expression profiles of test compounds with known toxins, helping to predict which compounds are least likely to induce the classes of toxicity exhibited by the reference compounds.

Conclusion

In conclusion, a number of powerful gene expression analysis techniques are now in existence to help shorten the drug discovery process. Open methods such as READS, and closed methods such as the Affymetrix GeneChip array and the Flow-thru Chip, will improve target discovery, lead prioritisation and compound testing. By working with a company that has developed gene expression databases for a wide range of normal and diseased tissues, drug researchers can avoid the time and tremendous expense involved in building in-house databases. Pharmaceutical companies should obtain enormous benefit in a number of ways - lower R&D costs, fewer failed drug candidates and a greater number of new products each year.

References


