TRANSCRIPTOMICS
realising the promise with a new era of drug discovery and diagnostics

The advent of modern genomics has provided new ways to link targets to diseases and validate these targets. The result is the potential for a new era of drug discovery – transcriptomics or the discovery of drugs that control disease through the regulation of gene expression, mimicking the mechanism by which our bodies exert regulatory control over a range of normal and disease functions.

The transcriptome is the molecular compartment of transiently expressed genes, RNA, its regulation and control of the synthesis of proteins that make up the 'other half' of molecular functionality of the cells in our body. While the genome is the same in every cell and influences the propensity for disease, it is the transcriptome of expressed RNA and proteome of proteins that mediates most diseases and disease responses. **Figure 1** depicts three major compartments of functional molecules, the genome of DNA, the transcriptome of expressed RNA and the proteome of proteins. To date, the industry has focused on only the proteome or half of the disease mediating molecules. This is largely because biochemistry has provided the industry with the quantitative methods and tools for protein ignoring RNA which is ultimately necessary to discover active compounds and optimise the activity of those compounds.

The hallmarks of such assays are reproducibility and repeatability. Reproducibility is the ability to reach the same result each time an assay is run. It shows how variable measurements are and must be less than the typical step in improvement of activity, improvements that may only be 20-50% better or different from other analogs. Repeatability is the ability to compare a drug and results obtained on one day, in one lab, to those obtained another day, on another compound and if necessary, in another lab. The medicinal chemist must be able to

compare the data obtained on compounds tested at the beginning of a programme to those tested at the end, possibly two to five years later.

Despite the absence of drug discovery programmes evaluating the transcriptome, the industry has marketed successful drugs, discovered by serendipity or through receptor-based programmes that act on the transcriptome. These include steroids, retinoids (acutane) and hormones (tamoxifen, estrogen). These classes of compounds provide a glimpse into the power of such transcriptome-based drugs. True gene-to-screen transcriptomics drug discovery will provide a myriad of other regulatory points for intervention and discrimination of desirable effects providing therapeutic efficacy from undesirable side-effects.

A major fundamental difference between transcriptomics and proteomics is the accessibility to a broad scope of activity through measurement of a single class of molecules. A single assay can be used to measure different expressed RNA molecules. Traditional methods require different assays not just to measure enzymes and receptors, but to measure different enzymes and different receptors due to the highly variable nature of proteins.

Therefore, transcriptomics bridges the gap between genomics and proteomics. Testing and evaluating the transcriptome can provide the industry with new drug discovery programmes at a lower cost and in general provide higher quality
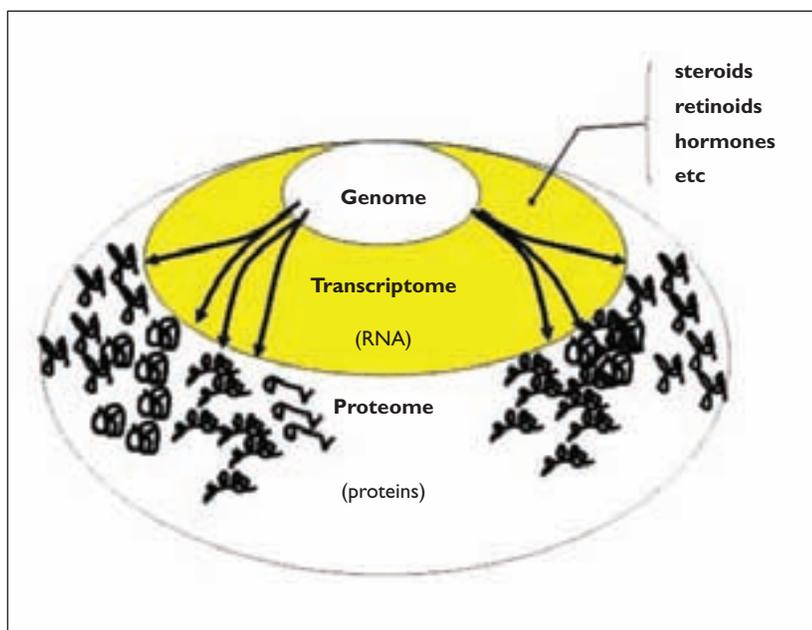
By Dr Bruce Seligmann

Transcriptomics

clinical candidates of all types (transcriptome and protein based drugs) for testing in humans.

Limitations in studying the genome

Genomics is the study and sequencing of an organism's genetic material – its DNA genome. The primary objective of genomics is to understand the organisation of genetic information and what it reveals about the biology and health of the organism. More specifically, it offers insight into the potential for response or susceptibility to disease. The genome of DNA does not provide any information regarding current function, state or stage of disease and disease response and is composed of a relatively small number of genes. While some diseases are caused directly by abnormal genes, most diseases are related to what genes are expressed and how they are expressed through the complex regulation occurring in the transcriptome.

Post-genomics research is focused on the selective and highly regulated expression of genes as RNA: the transcriptome. There is significant attention on identifying disease-specific signatures or patterns of gene expression that characterise the disease state and stage, referred to as functional genomics.

One of the major issues with genomics research today is that the functional genomics methods used to measure gene expression do not produce the quantity or quality of data required by pharmaceutical researchers to validate drug discovery targets and discover drugs acting upon gene expression. As a direct result, the industry is forced to make the expensive, time consuming and labour-

intensive leap from gene to protein based on relatively poor validation. To make this leap, one or a number of genes are selected from the target validation process to proceed to the next level. This is typically a 12 to 36-month process (**Figure 3**).

Before the transcriptome

RNA measurement technologies

RNA measurement techniques such as PCR (polymerase chain reaction), gel-based ribonuclease protection assays, reporter gene assays and branched DNA (bDNA) assays have been used with varied success (**Figure 2**). These methods have been useful for target identification and validation but generally they have not been successful for drug discovery, nor have they impacted drug time-to-market. In the past year, new transcriptomics technologies have emerged to not only identify targets but to speed up the drug discovery timeline.

One technology being used currently is PCR (polymerase chain reaction). PCR assays have been a fundamental enabling tool for genomics. When used to measure gene expression, or RNA, a host of problems are encountered that have limited its utility. It is necessary to first extract RNA from a sample, purifying it from DNA and then reverse transcribe the RNA target into a cDNA (complementary DNA) template before performing PCR amplification and measurement. This is a complex process that typically requires large samples that can be difficult or costly to produce or treat making this type of assay unsuitable for the first stage of drug discovery. Whenever extraction or purification is required, the process typically decreases a sample size even when extreme care is taken. No matter how reliable PCR may be, there is an inherent variability and artifact introduced by extraction.

Another approach used in target validation is the gel-based ribonuclease protection assay. Long considered the 'gold standard' assay for measuring RNA, this assay uses synthetic DNA oligonucleotides to hybridise with and protect a portion of each target RNA molecule from nuclease digestion. The 'protected' RNA sequences are then detected by performing analysis on gels. This technique is inexpensive, capable of measuring the expression levels of multiple genes from the same sample and avoids the artifact of extraction, reverse transcription and amplification. However, this method is labour-intensive, making it difficult to test many samples. Its utility has generally been limited to validation of targets rather than the high throughput screening of hundreds of thousands of compounds or rational drug design and optimisation.

Reporter gene assays were the first solutions introduced for transcriptomics-based drug discovery but have generally failed due to several critical flaws. Reporter gene assays were based on the notion that a single promoter controlled both the expression and expression lifetime of each gene. We now know this is not the case, but at the time it became possible to engineer a gene containing this promoter regulating the expression of a surrogate 'reporter' gene, one that produced a measurable protein or effect, such as an enzyme that produced a coloured, fluorescent or chemiluminescent product and could be easily measured in place of the actual target gene or gene product. Those compounds that suppressed or induced the reporter could be identified using this whole cell assay.

Unfortunately, the reporter gene assay approach is riddled with problems. These included oversimplification of the complex mechanism of the transcriptome, instability of cloned cell lines and high false positive hit rates. The labour-intensive and time-consuming process of cloning each gene, identifying the promoter, creating an engineered gene construct and then transfecting and validating each cell line made these assays difficult to establish.

bDNA (branched DNA), adapted from a diagnostics platform, can be used to examine the lysates of native cells without RNA extraction or amplification. bDNA uses hybridisation capture of the native RNA on to the surface of a microplate, one gene or well. Capture is achieved by using capture probes (10 to 50), each of which hybridises to a different portion of the target RNA. Detection is similarly achieved using a large number (10 to 50) of branched DNA detection molecules to amplify the assay signal, each capable of binding a second generic probe that binds multiples of a third generic probe. In addition, each is able to bind multiple copies of a luminescent enzyme detection probe. This leads to a highly 'branched' and amplified detection complex. Probe design is a complex and problematic approach to establish the assay for one gene target. These can support the high throughput screening of a single target. However, besides being very difficult to establish, this technology has not been widely adopted due to two other major drawbacks. For one, bDNA lacks multiplexed measurement which means the precise mechanism of action needs to be known and limited to a single target, making this unsuitable for the most common multigenic disease example. In fact, multiplexing would further aggravate the task of establishing a new HTS, since it would multiply the difficulty of establishing an assay. Second, as a single target whole cells assay, bDNA suffers from a high 'false positive' hit rate.

Target identification and validation
Target identification using functional genomics, the identification of which genes are expressed or suppressed in diseased cells and tissues, has proven to be a powerful breakthrough in
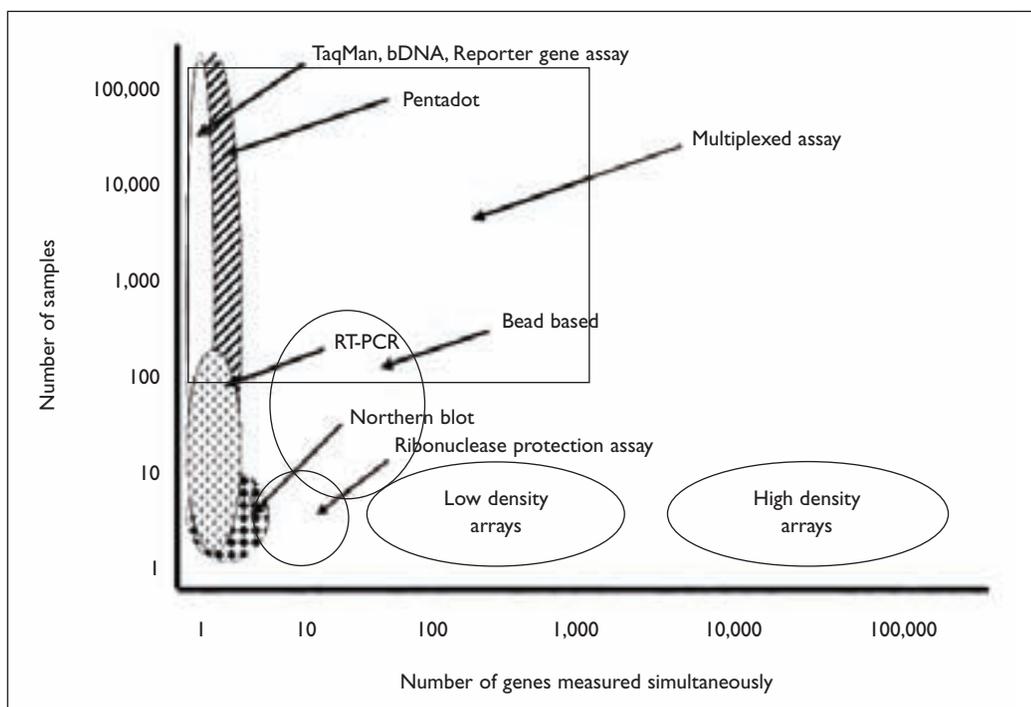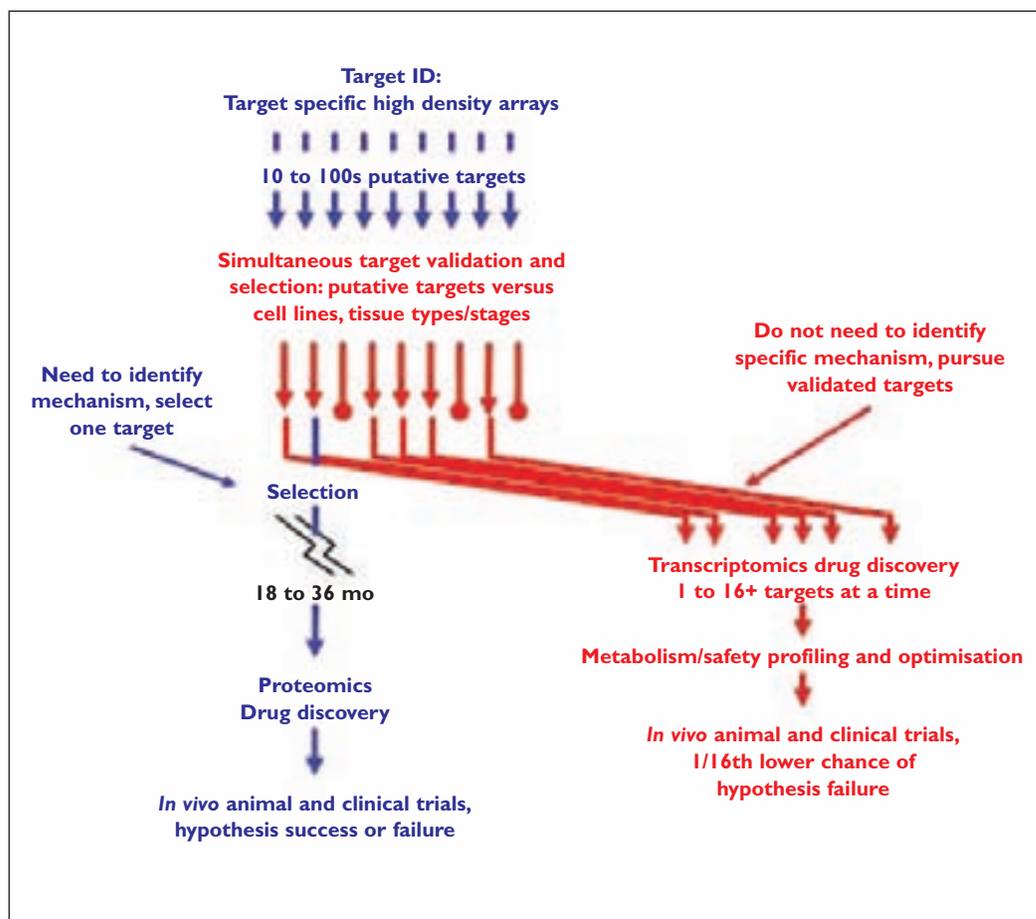
Transcriptomics

Figure 3
The new transcriptomics era drug discovery paradigm

genomics. This is typically achieved using high density arrays that measure the expression level of 10,000 to 30,000 genes at a time. However, samples are typically subject to extraction or amplification and labelling steps. These arrays typically require the use of multiple elements, gene and complex bioinformatics to identify changes in expression level and to cluster the data for useful interpretation. Also, the cost and difficulty of high density array experiments limits their use to the testing of a handful of samples for each disease indication. Consequently high density array data needs to be confirmed and validated by independent experiments.

Target validation involves a variety of methods designed not only to verify the high density array results, but to further characterise disease mechanisms. Expression changes show causality of the disease process and good targets for drug discovery. Typically methods that measure one or a few genes at a time but which are more reproducible than high density arrays are used at this stage, but only a handful of samples are tested.

Success of transcriptomics drug discovery

Transcriptomics technology must enable the multiplexed measurement of the expression of one to several hundred genes at a time. Only a multiplexed assay can fully enable the gene signature of multigenic diseases or functional responses to be used to discover broadly acting, potentially more effective drugs. In addition, only a multiplexed assay can reduce the risk of hypothesis failures by permitting the simultaneous pursuit of many targets without having to absolutely define the mechanism of action and handle the balance between the number of target genes and the number of samples for applications ranging from target validation, high throughput screening and rational drug design, lead optimisation, metabolism and safety profiling, to metabolism and safety optimisation. The first of a new generation of assays to offer these capabilities, enabling drug discovery at the level of the transcriptome, has recently become available to the industry (**Figure 2**).

Transcriptomics technology for drug discovery must fulfil the following criteria:

● Permit samples of 10,000 to 20,000 cells to be reliably tested to quantify low and moderately high expressed genes.

● Automation friendly.

● Permit high sample throughput with the reproducibility to discriminate differences in expression levels.

● Provide repeatable results from compound to compound, operator to operator, day to day, lab to lab.

Market potential for transcriptomics technology

There are three major trends propelling efforts to increase the productivity of drug discovery programmes. One is the sequencing of the human genome and the concurrent advances in genomics, which have resulted in an explosion of potential new targets for drug discovery. A second trend is the growing number of chemical compounds that have been archived for testing in screening efforts. The third trend is computational chemistry and biology methods that encompass a range of computationally intensive tools and databases.

Despite these trends and new capabilities, the decreased output of marketed drugs during the past two years suggests that genomics might be doing more to exacerbate the pharmaceutical industry's productivity issues than solve them. The influx of new drug targets does not address the age-old bottlenecks created by safety, clinical trials and the need to identify protein targets amenable to the use of traditional high throughput screening assays.

The influx of new drug targets does not address the risk of hypothesis failure but actually contributes to drug candidate failure. Traditional methods do not provide the tools needed to pursue a multiplexed drug discovery paradigm different from the single target, 'magic bullet' approach the industry has pursued for the past 20 or more years. The industry now spends its own resources on both identifying and validating new targets, where, in the pre-genomics era, institutions and the academic community carried out these tasks.

Transcriptomics technology offers a solution to break the bottlenecks imposed by these factors and provide a new solution of multi-target drug discovery that will reduce the risk of hypothesis failure, lead to the rational discovery of drugs that modulate the multigenic basis of many diseases and provide a new gene-to-screen drug discovery platform that does not require the expensive, time-consuming steps to identify protein targets (**Figure 3**). Early adopters of transcriptomics-based drug discovery

will have the opportunity to leverage their rational drug design and medicinal chemistry skills to gather new drug leads, such as the discovery of structurally diverse mimics of steroids, retinoids and marketed drugs discovered by serendipity, but now recognised to act through regulation of transcription, the transcriptome. Transcriptomics technology for drug discovery addresses all the hopes for the industry to reverse the current trends in rising costs and to fully exploit genomics.

Transcriptomics – The future

Researchers have needed a new assay method that can provide accurate, highly reproducible and multiplexed RNA quantification, as well as enable the expression of a select set of relevant genes in thousands of samples. Such assay technologies will enable the industry to fully exploit the complexity and regulatory control that lies in the transcriptome and gene expression, ushering in a new era of drug discovery. As discussed, many approaches have been deployed to validate and quantitatively measure RNA. However, to date, none of these methods has worked well enough to enable companies to launch transcriptomics drug discovery programmes.

Transcriptomics assay technology provides researchers with a rapidly implemented, sensitive, quantitative, reproducible, repeatable and robust assay method for measuring the expression of up to several hundred genes across many samples in an easily automated microplate format. The technology is highly versatile and can measure RNA, DNA and protein from the same sample in an industry standard 96-well microplate. The transcriptomics assay is built on the use of a novel, reagent-programmable array that provides the user with a custom tailored assay design, implementation control, solution-phase hybridisation and modified ribonuclease protection assay that only requires the sample to be lysed. The tedious research gel is replaced by an industry strength high throughput array, and the protection probes, not the RNA, are measured. To date, out of 1,500 genes there have been no design failures, unlike PCR where the failure rate can be 20-30%. Avoiding the extraction, reverse transcription and amplification steps of PCR not only simplifies the assay but makes it highly reproducible and repeatable.

Using transcriptomics, low expressed genes can be measured from samples as small as 1,000 cells without additional signal amplification. Samples as diverse as cultured cells, tissues and whole organisms (including plants and fruit flies) have been used for testing. This advancement in technology is also well-suited to find treatments for multi-factorial diseases and therefore multiple targets on which a drug must act. It is uniquely able to analyse the interplay of multiple active genes, which allows researchers to identify signatures or fingerprints that are important in diseases for which the cause is rooted in the relationship between multiple genes. Multiplexing enables all the targets in a profile to be pursued simultaneously as a set which is important because many diseases have more than one cause. This group includes most cancers, inflammatory diseases and heart disease, where traditional testing against individual targets is insufficient.

After the failures of other technologies, the industry can now execute gene-to-drug discovery it had originally planned with the use of genomics. By enabling true gene-to-drug performance, the new wave of transcriptomics technologies will enable drug discovery and clinical diagnostic applications to keep pace with the rapid advances in investigating the human genome, as well as reduce the costs of new drugs and lower risk of hypothesis failure. Not all compounds acting on gene expression that are discovered and optimised using this recently enabled transcriptomics technology for drug discovery will make it to market, but the earlier testing of such compounds will be invaluable. Transcriptomics promises a new era from discovery to the clinic. **DDW**

Bruce Seligmann is President, CEO and Chairman of High Throughput Genomics in Tucson, Arizona, USA. Bruce Seligmann is responsible for High Throughput Genomics' overall strategy and business direction. Prior to High Throughput Genomics, he founded combinatorial chemistry pioneer SIDDCO and served as its President, CEO and Chairman through its sale to Discovery Partners International and the resulting divestiture of HTG. Prior to SIDDCO, Seligmann was Center Director of Selectide. During his tenure with that company, he tripled its size and staged it for purchase by Marion Merrill Dow (now Aventis). Before Selectide, he was a Senior Research Fellow at Ciba-Geigy (now Novartis) for seven years. Seligmann also spent seven years as a Senior Staff Scientist at the National Institute of Health, National Institute of Allergy and Infectious Diseases and Laboratory of Clinical Investigation, where he became internationally known for his research. Seligmann earned his doctorate from the University of Maryland and holds a BS in chemistry from Davidson College.