

Overview of biomarkers in disease, drug discovery and development

The pharmaceutical industry and the healthcare sector are both confronted with expensive technological innovation, escalating costs and pointed questions about productivity and efficiency. The parallels between the problems of producing new therapeutic agents and treating patients afflicted with poorly understood diseases are compelling. In part this is due to our limited ability to transform large datasets, (eg clinical data in diagnosis of disease) into meaningful information and knowledge. Knowledge of a disease process or impact of a drug in the treatment of disease is imperative for scientists, physicians and managers to make accurate and decisive decisions. One approach to enhance our understanding of such issues has emerged in the form of biomarker discovery, validation and utilisation. In this article an overview perspective of biomarkers is provided in the context of disease treatment and the drug discovery and development (DDD) process. A variety of issues are addressed including the sundry definitions and classifications of biomarkers. Furthermore, the practical matters to consider for biomarker discovery, the tools and technologies required, and what constitutes the optimal biomarker panel, are all discussed.

The Pharmaceutical Research and Manufacturers of America reported recently that their members had spent \$38.8 billion on R&D in 2004¹. This reflects a 12% increase from 2003 where the total budget was \$34.5 billion, and this is in concordance with the ~13% annual growth rates expended on biomedical research by both government and industry over the past decade². However, this trend of ever increasing R&D costs does not appear to have halted the continued decline in productivity, as seen in the decade-long decrease of new molecular entities (NMEs) and biologic licence applications (BLAs) submitted

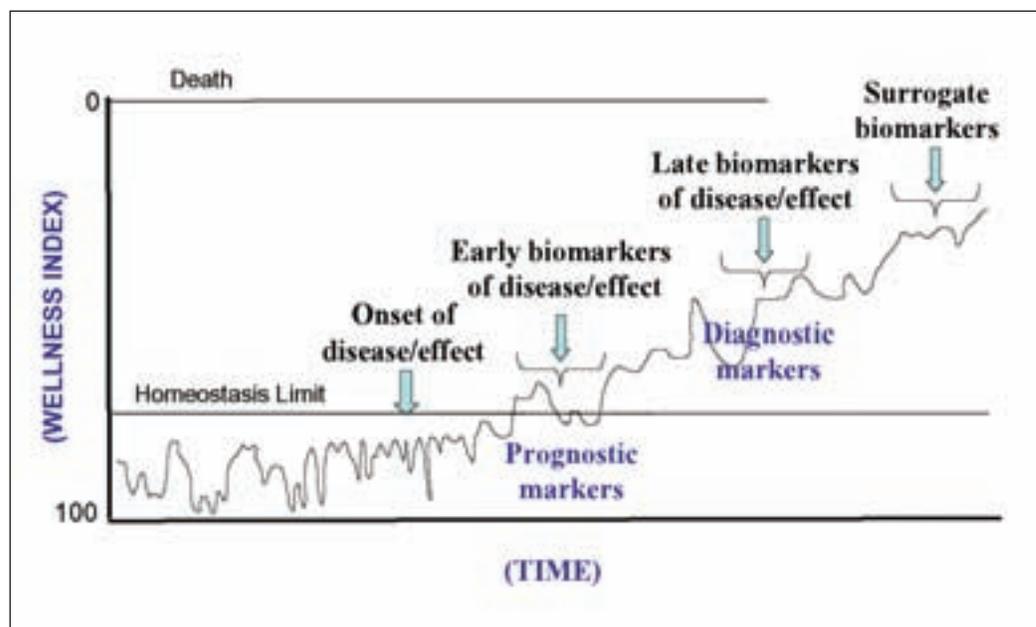
to regulatory agencies on an annual basis³. Furthermore, both DiMasi⁴ and Bains⁵ have noted the continued, rising cost of bringing a drug to market, and their estimates range from ~\$800 million to ~\$1.15 billion respectively.

Bains, in a topical and provocative article, highlighted a series of factors to consider for the drug discovery and development (DDD) process to become more efficient and cost-effective⁵. He argued that poor science, technology and medical understanding contribute significantly to the ballooning cost and time constraints of the process. However, he also made the salient point that poor

**By Professor
Stephen Naylor**

Figure 1

The role of biomarkers in determining the Wellness Index (WI) of an individual. The WI ranges from 0 (Death) through to 100 (Perfect Homeostatic Control). Individuals as a function of age, environment and health will strive to maintain a biological equilibrium of through homeostatic control. The latter will be viable over some limited WI range as shown in the figure. Various biomarkers can be used to determine the progression from homeostasis into disease onset, progression into disease and outcome



management decisions concerning borderline projects are a major contributing component. He audaciously announced that “implementing a ruthless success or die policy could half the cost and time to get a drug to market”. Scientists are also not spared in his analysis, and Bains suggested that another significant way to cut cost and time is for scientists to reduce “repeat” experimental steps at any stage in the DDD process⁵. Overall, he suggests that poor decision-making by both scientists and managers is at the heart of spiralling costs and decreasing productivity.

In a subsequent article, Naylor⁶ contends that both DDD managers and scientists must have high quality, accurate, reproducible and interpretable data in order to make unambiguous and decisive decisions. Unfortunately, like all of us in the age of the global communication village, managers and scientists are inundated each day with polybytes of data and information. They are ill-equipped to analyse such content, and efficiently utilise it in key decision making processes. Most of the data and information remains unfiltered, unprocessed and unused. Our ability to transform

Data → Information → Knowledge

is particularly limited, since we lack many of the appropriate tools. How does one go about interpreting and utilising such data and information in making informed decisions? A parallel argument can also be made about shortfalls in the healthcare sector and treatment of disease. The escalating cost

of the \$1.8 trillion healthcare industry in the US⁷, provides a good example. Innovative new technologies have not decreased costs, and disease treatments in crucial health areas including oncology, cardiovascular, CNS and immune-mediated diseases, have not improved dramatically over the past decade. The same issues of large datasets not being efficiently transformed into knowledge of the disease, hence inhibiting physicians from accurately diagnosing and efficiently treating disease has not been readily realised (Figure 1).

A critical question is how to adequately transform Data → Information → Knowledge and apply that to both healthcare and DDD decision-making processes. Many believe that part of the answer lies in the discovery, development and utilisation of biomarkers. The hope is that biomarker data will provide more predictive information and knowledge about the changes in the biological processes induced after perturbation with a therapeutic agent. This should allow better predictive capability and decision-making on the part of scientists and managers in the DDD process⁶. A similar argument could also be constructed around the diagnosis, treatment and management of disease⁸.

Definition and classification of biomarkers

This burgeoning field has attracted excitement, enthusiasm and confusion as it begins to impact on the DDD pipeline, as well many aspects of disease prediction, onset and progression⁹⁻¹³. In an attempt to bring some order to this diverse field,

the Biomarkers and Surrogate Endpoint Working Group (under the direction of the Office of the Director, National Institutes of Health), has agreed on both a definition as well as a classification system for biomarkers¹⁴. The definition of a biomarker is: “A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacological responses to a therapeutic intervention”. As this definition encompasses many elements of the pharmaceutical and biotechnology industries as well as much of the biomedical and conventional biological sciences, practitioners may still be excused from being lost in the morass and size of the biomarker space.

In a further attempt to bring clarity to the biomarker arena, a classification system has also been devised. Type 0 biomarkers purportedly measure the natural history of a disease and should correlate over time with known clinical indicators. Type I biomarkers indicate the intervention effect of for example a therapeutic drug, whereas Type II biomarkers are considered Surrogate Endpoint markers. This Working Group has also given much thought to defining Surrogate Endpoint as well as Clinical Endpoint. The latter is defined as: “A characteristic or variable that reflects how a patient feels, functions or survives. Clinical endpoints are distinct measurements or analyses of disease characteristics observed in a study or clinical trial that reflect the effect of therapeutic intervention”. Clinical endpoints are considered the most reliable indicators of disease or therapeutic response, however a biomarker can also rise to the status of Surrogate Endpoint. This is defined as: “A biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm) based on epidemiologic, therapeutic, pathophysiologic or other scientific evidence” and the reader if interested should peruse the original discussion by this group¹⁴, as well as the excellent review by Frank and Hargreaves⁹.

From a much more practical and focused perspective the US Food and Drug Administration has proposed that: “A surrogate endpoint or marker is a laboratory measurement or physical sign [sic] that is used in therapeutic trials as a substitute for a clinically meaningful endpoint that is a direct measure of how a patient feels, functions or survives and is expected to predict the effect of the therapy”¹⁵. In other words a biomarker is an indicator of change, and therefore fluctuates as a function of time and biological influence. Hence as pointed out by Zolg and Langen¹¹, under this strict definition single nucleotide polymorphisms

(SNPs) are not biomarkers. Finally, a group at Bayer Corporation has enunciated a pragmatic definition of the term biomarker from the pharmaceutical perspective¹⁶. A biomarker is “...a measurable property that reflects the mechanism of action of the molecule based on its pharmacology, pathophysiology of the disease, or an interaction between the two. A biomarker may or may not correlate perfectly with clinical efficacy/toxicity but could be used for internal decision-making within a pharmaceutical company”. For the interested reader there is a website, organised and regularly updated by Cambridge Healthtech that provides working definitions for all aspects of biomarker applications in the DDD process as well as disease biology and medicine¹⁷.

At present the biomarker arena can be divided into two broad subsets¹⁰. The pharmaceutical and biotechnology industries have adopted them as a wide ranging set of tools in monitoring and providing information feedback in DDD (Type I-like biomarkers)¹². They have impacted on determination of clinically relevant targets, high-throughput screening chemistries and preclinical ADME and toxicology, to clinical ‘decision making markers’ at Phase I-IV. They are pervasively used throughout the entire DDD pipeline. However, in most cases the biomarkers of indication are used primarily in pattern recognition mode, using a set of unidentified markers, which may be genomic, proteomic, metabolomic, or a combination dataset. In this instance it is not necessary *a priori* to determine biomarker constituent identities, since the pattern or signature alone denotes specific biological activity.

The second broad area where biomarkers are currently finding use is in the disease mechanism, monitoring and prediction arena¹³. One area of focus is the determination of specific biomarkers (gene, transcript, protein or metabolite) for either diagnosis of disease, or screening for a disease. Such compounds can be considered as Type II-like biomarkers. In these cases it is important that a single biomarker has been structurally identified and validated. However, a second approach is the discovery of biomarker panels to indicate specific disease states for predictive, early onset, progression, regression, treatment efficacy and diagnosis of disease (Type 0-like biomarkers). It is interesting to note that in this latter situation, the expectation of a ‘good’ biomarker can range from a molecular signature of structurally unidentified markers (similar in perspective to the pattern recognition mindset of the pharmaceutical industry), to a ‘panel of identified biomarkers’ specific for the disease process being evaluated.

Zolg and Langen¹¹ have recently described in some systematic detail, the biomarker discovery, validation and commercialisation process. They make the important point that the biomarker validation phase needs to be precise and accurate, and that it is time consuming and expensive. In part this is to ensure that statistical analysis is rigorous and that a receiver-operator characteristics (ROC) plot to determine specificity and sensitivity can be performed. Furthermore, other authors have also noted the underlying complexity and time-consuming nature of validating biomarkers for routine use. This has been discussed in some depth by Frank and Hargreaves⁹ as well as by De Meyer and Shapiro¹⁸. As noted by the former authors: “The standard concepts of test-retest reliability and validity apply with equal force to clinical biomarkers as they do in any [classical, standard] assay system”. They also note that rigorous standards and protocols are already in place for the latter and therefore provide a lattice framework for the former. They also ruefully and correctly note that: “The work required to establish the reliability and validity of a new biomarker should not be underestimated... and needs planning for each combination of clinical indication and mechanism of action”. For example Type 0 biomarkers can be validated longitudinally, in a well-defined patient population against a “gold standard clinical assessor”. Type I biomarkers should be validated in parallel with the drug candidate, and Type II biomarkers “must be relevant both to the mechanism of action of the drug and to the pathophysiology of the disease”.

Practical biomarker discovery and validation

As noted above, biomarker discovery and validation are active but emerging fields of endeavour. Definitions and classification of biomarkers are still being discussed and debated. In addition, the constituents of the optimal biomarker or biomarker panel are still controversial and not clearly defined. Furthermore, there are numerous practical issues and limitations that have to be considered. They include experimental design, biological sample quality and variability, technology platform capability, paucity of good ranking and predictive modelling algorithms, lack of context in disease and DDD process, limited use of knowledge assembly tools, lack of consideration of global initiatives in biomarkers of disease, company versus public databases, cost-benefit of technologies, and ultimately a poor understanding of the potential for reimbursement/or analysis of value of biomarkers.

1. **Experimental design.** It is imperative that one

considers what the intended outcome of the biomarker discovery experiment is designed to achieve. Experimental design is probably one of the most overlooked, least-understood components of biomarker discovery. You need to consider the appropriate number of samples to be analysed that will provide statistically significant data outcomes. Adequate controls are a very necessary element in the design of such studies. One needs to consider whether a global or targeted analysis is appropriate, as well as whether tissue or/and body fluid should be analysed.

2. **Sample quality.** The old adage of ‘garbage in, garbage out’ when applied to analysis of biological samples is particularly relevant in biomarker discovery. The quality of samples analysed will ultimately determine the quality of biomarkers produced. A number of factors must be taken into consideration. For example a clear lineage and adequate care for animals is necessary, whereas in the case of human samples, history, outcomes and storage conditions are all very important. In particular one must also consider whether to pool samples or analyse individual samples. Most practitioners today tend to agree that the intrinsic biological variability present in individual samples contains important information, and provided that you have the appropriate informatics and biostatistical tools, pooling is not appropriate. In the case of heterogeneous tissue (eg brain) one must determine if you will analyse the mix of cellular material, or specific cell populations which can be acquired using laser capture micro-dissection. At a more refined level, do you analyse the content of specific organelles by using biological sample preparation techniques? All such questions are determined by the focus of the study, and biological indication of the biomarkers being sought.

3. **Technology platforms.** There has been tremendous developments in -omic platform capability over the past decade. However there still remain a number of concerns. Expression profiling has matured into a stable, commercially available platform technology. But questions continue to arise about the precision and reproducibility of this approach. In the differential proteomic and metabolomic analyses of complex mixtures, a number of issues still need to be addressed. One of the major limitations of current technologies (predicated on chromatography and mass spectrometry) is the limited measurable dynamic range (typically 10⁴). Given that dynamic range can vary from 10⁶ to 10¹⁰ in biological tissue and fluids, this creates significant problems in terms of breadth of coverage and limited sensitivity.

Additional problems involve limited throughput capability and limited automation. Precision, and reproducibility as well as accurate quantitation are also issues that are still being addressed. In the case of imaging, the problem of limited throughput is still an issue. Finally, in systems biology approaches developing integrated platforms to carry out such analyses are still in their infancy²⁰.

4. Informatics and databases. The ability to integrate data from different platforms is not a straightforward procedure. To date only a limited number of Companies such as Ingenuity Systems (Mountain View, CA, USA), Gene Network Sciences (Ithaca, NY, USA), Entelos (Foster City, Ca, USA), BG Medicine (Waltham, MA, USA), and Icoria (Raleigh-Durham, NC, USA) as well as academic Institutions such as the Institute for Systems Biology (Seattle, WA, USA) and Max Planck Institute (Heidelberg, Germany) have such capability. The algorithms are proprietary and to date there is only a limited number of commercially available tools (see, for instance, Ingenuity, www.ingenuity.com). In the rapidly developing

world of biomarker ranking/prioritisation an alphabet soup of different approaches exist including SNR, SVM, t-test, POOF, Ecombo and stepwise LDA. Unfortunately, many of these are customised approaches, and hence there are no unifying standards in the biomarker field. Furthermore global databases are still only being considered and discussed, but most remain proprietary in the private sector. Finally, the data visualisation tools available today continue to develop at a pace, but are also still in their infancy.

5. Economics. Over the past two years there have been a number of conferences on biomarker discovery and validation (see, for example, Cambridge Healthtech Institute, www.healthtech.com; or IBC, www.IBCLifeSciences.com) Many focus on the important -omic platforms used to undertake biomarker discovery. However, a number of speakers have made the point that “while there is genomics, transcriptomics, pharmacogenomics, proteomics and metabolomics, the only really important -omics is ECON-omics”! This comment from a business perspective is most appropriate and timely. The

CALL FOR PAPERS

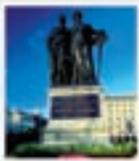
SBS 11th Annual Conference and Exhibition

Geneva






September 11-15, 2005
Geneva PalExpo
Geneva, Switzerland

Drug Discovery:
From Targets to Candidates

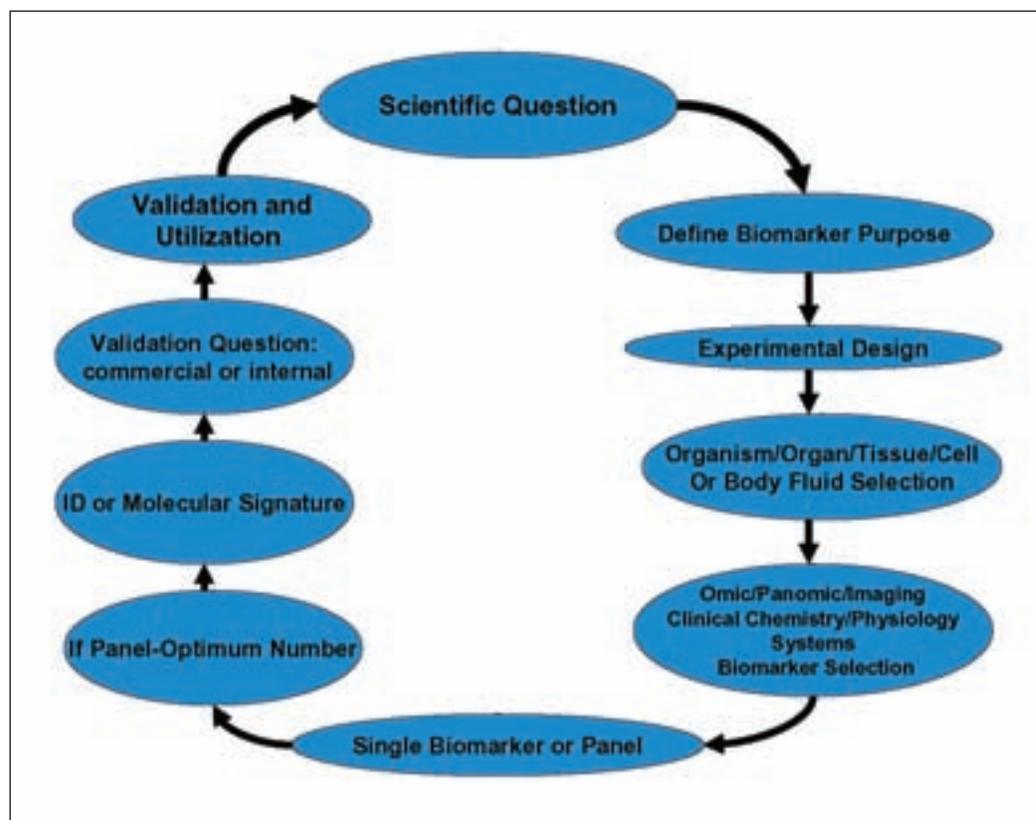


The Society for Biomolecular Screening,
36 Tamarack Avenue #348, Danbury, CT 06811, USA
Phone: +1 (203)743-1336 ♦ Fax: +1 (203)748-7557
email@sbsonline.org ♦ www.sbsonline.org

- ▲ Comparing HTS Technologies
- ▲ Advanced Technology for Drug Discovery
- ▲ Imaging & High Content Screening
- ▲ Information Technology
- ▲ New Approaches to ADME/Tox
- ▲ Target Discovery & Validation
- ▲ Target Biology & Screening: Ion Channels
- ▲ Target Biology & Screening: Kinases & Proteases
- ▲ Target Biology & Screening: Membrane Receptors & GPCR's
- ▲ Advancing Targets to HTS
- ▲ Medicinal Chemistry & Lead Optimization
- ▲ Biomarkers for Predicting Therapeutic Response
- ▲ Point/Counterpoint (TBA)

Figure 2

The Decision Wheel of Biomarker Discovery. The issue of what constitutes the optimal biomarker(s) is still an area of some strident debate and discussion, in part because biomarkers serve a multitude of purposes. However, one can envisage a relatively simple, compartmentalised series of modules that make up the 'Decision Wheel of Biomarker Discovery'. At each stage a careful and thoughtful decision has to be made to address the key issues raised within each module. The 'Decision Wheel' is designed to systematise the process of determining the optimal biomarker(s) for the scientific question under scrutiny



issue of reimbursement for biomarkers is a quagmire, mired in the political debate of escalating healthcare costs in both North America and Europe. The conventional diagnostics marketplace provides some background for consideration of the monetary value of biomarkers. However, who will actually bear the cost of discovery, and how that might be reimbursed is not so clear. In part it will be determined by the role of the biomarker and its use. For example, the use of biomarkers in the pharmaceutical industry is somewhat more straight forward, since the intrinsic value of biomarkers is to reduce the approximately \$800 million-\$1.15 billion needed to bring a drug to market. Unfortunately it is not so clear cut as to the value of biomarkers in monitoring disease processes. However, in that context of how one might go about actually valuing biomarkers has been elegantly discussed by Ferber¹⁹ in a recent paper. He discusses the use of the Pearson Index (a normalised measure of the financial value of a drug development project) in the context of using biomarkers as tools in acquiring additional information about the process. He concludes that: "Economy makes us try to obtain the most valuable, albeit still incomplete information with a limited investment. It is in this context... that biomarkers play their role".

-Omics and -ics of biomarker technologies

Disease biology and more specifically the DDD process have historically suffered, from a paucity of information. This has been predicated on the technical difficulties associated with obtaining meaningful measurements on biological systems (eg organisms, organs, tissue, cells or organelle) under investigation⁶. Ultimately, this has resulted in limited data output and hence information content. However, the advent of gene expression arrays pioneered by Brown²¹ in the early 1990s and commercialised by Affymetrix Inc (Santa Clara, CA, USA) forged the 'decade of measurements' which begat numerous high throughput analytical tools and technologies. The consequence of this 'Omics Revolution' has been the development of platforms that now routinely produce copious and substantial, genetic, genomic, transcriptomic, proteomic, functional proteomic and metabolomic datasets²². In a concomitant timeframe, there was also an explosive growth in the -ic technologies, such as informatics, bioinformatics and biostatistics. These tools enable the acquisition, manipulation and storage of large datasets, as well as mining them for new information and knowledge^{23,24}.

The development of such technologies has enabled

the emergence of biomarker discovery efforts. The platforms (hardware and software) now available include the existing -omic technologies, as well as the integrative analysis of systems (commonly referred to as systems biology, pathway or network biology, or panomics)^{6,20}. Such discovery platforms, which typically analyse molecular components, commonly utilise genetic linkage analysis, expression arrays, chromatography coupled with mass spectrometry, NMR and other sensitive detection devices such as electrochemical and laser-induced fluorescence detection. However, a wide variety of other approaches are also used that includes incorporation of conventional clinical chemistry measurements, all forms of imaging from immunohistochemical staining to NMRi, to whole cell analysis using flow-cytometry approaches. In order to mine and exploit the data acquired on such diverse platforms a panoply of data handling tools are required. They include data pre-processing software to subtract out baseline deviations, as well as align individual data files. A broad array of biostatistical tools is in use to identify specific cohorts of individual samples from a set of analyses and include principal component analysis (PCA) and discriminant analysis (PCDA). Prioritising individual biomarkers into panels based on fold-change and significance (Pearson Coefficient) requires a suite of conventional statistical approaches including ANOVA, t-test and Kolmogorov-Smirnov as well as more recent developments such as support vector machine analysis. Data visualisation, storage and retrieval packages are also critical to have in order to carry out such analyses. In addition, given the plethora of platforms used to acquire data, data integration and correlation (linear and non-linear) capability are essential features to have in the biomarker software toolbox. Finally, tools to extract knowledge from data are required. Such tools “renders knowledge derived from both structured and unstructured sources into a machine-readable format”²⁵. For an excellent overview of the technologies employed in biomarker discovery the interested reader should peruse the report written by Rubenstein entitled ‘Post-Genomic Biomarkers: Revolutionising Drug Development and Diagnostics’²⁶.

A number of companies offer -omic and -ic technologies and capabilities for biomarker discovery and they include Aclara (chemistry and microfluidics, www.aclara.com); Affymetrix (expression arrays, www.affymetrix.com); BG Medicine (integrated omic platform, www.bgmedicine.com); Biosite (phage display platform, www.biosite.com); Caprion (protein platform and software, www.caprion.com); Ciphergen (mass spectrometry-

based platform, www.ciphergen.com); diaDexus (oncology biomarker kits, www.diadexus.com); Genedata (software, www.genedata.com); Icoria (systems biology platform, www.icoria.com); Lipomics Technologies (lipid biomarkers, www.lipomics.com); Metabometrix (metabolite patterns, www.metabometrix.com), Molecular Staging (DNA amplification, www.molecularstaging.com); Rules-Based Medicine (multiplexed assays, www.rulesbasedmedicine.com) and Surromed (nanobeads and mass spectrometry, www.surromed.com). This is not a comprehensive list, but does provide some of the major private sector participants in biomarker discovery efforts.

The development of tools and technologies, as well as innovative new research in biomarker discovery is vibrant and active. This is in stark contrast to the actual validation and use of new molecular signatures, individual biomarkers or biomarker panels. In part this is simply due to temporal events and participant foci. Many of the necessary tools and technologies necessary for biomarker discovery have only recently become available, at least when used in a concerted manner. Furthermore the issue is compounded by the underlying complexity and time-consuming nature of validating biomarkers for routine use^{9,11,18}. It is paradoxical to note that the tools and technologies needed to undertake such tasks in the validation process are for the most part already available. They include expression arrays, protein arrays, high throughput immunoassays and conventional statistical and epidemiological analyses. Indeed, there is a reasonably well-defined paradigm in place to validate biomarkers, once they exit the discovery phase. The compounding issues are the fledgling state of biomarker discovery as well as the added complexity of analysing highly variable population samples.

What is the optimal biomarker?

As discussed above, the quality of biomarker discovery data has the potential to dramatically impact both the DDD process as well as disease biology. In the former case, the pecuniary effect could be significant if it leads to better information, knowledge and decision-making on the part of scientists and managers. In the latter case decisions affecting patient health and well being could also be improved dramatically if physicians had the appropriate tools and information on how to treat the complex and subtle infringement of disease on the patient (see Figure 1). In both cases, an efficient process flow of Data → Information → Knowledge should afford better decision making in both DDD as well as the treatment of

References

- 1 GenomeWeb Staff Reporter. Pharma spent \$38.8B on R&D in 2004, a 12% jump; breakdown on genome spending soon. GenomeWeb (www.genomeweb.com) 2/18/05.
- 2 Booth, B and Zimmel, R. Prospects for productivity. *Nat. Rev. Drug Dis.* 3: 451-456 (2004).
- 3 FDA Federal Drug Administration (FDA), Department of Human Health and Services. Challenge and opportunity on the critical path to new medical products. April 2004 (www.fda.gov/oc/initiatives/criticalpath/).
- 4 DiMasi, JA, Hansen, RW and Grabowski, HG. The price of innovation: New estimates of drug development costs. *J. Health Econ.* 22: 151-185 (2003).
- 5 Bains, W. Failure rates in drug discovery and development: will we ever get any better? *Drug Discovery World* 5: 9-18 (2004).
- 6 Naylor, S. Systems Biology, information, disease and drug discovery. *Drug Discovery World* 6: 23-33 (2005).
- 7 Mattera, MD. Memo from the editor: A way to curb healthcare costs? *Medical Economics* December 3rd, 2004, (www.memag.com/memag/article).
- 8 Morel, N et al. Introduction to Systems Biology-A new approach to understanding disease and treatment. *Mayo Clin. Proc.* 79: 651-658 (2004).
- 9 Frank, R and Hargreaves, R. Clinical biomarkers in drug discovery and development. *Nature Drug Discov.* 2: 566-580 (2003).
- 10 Naylor, S. Biomarkers: Current perspectives and future prospects. *Expert Rev. Mol. Diagn.* 3: 525-529 (2003).
- 11 Zolg, JW and Langen, H. How industry is approaching the search for new diagnostic markers and biomarkers. *Mol. Cell. Proteomics.* 3: 345-354 (2004).

Continued on page 29

Continued from page 27

disease. All this is predicated on the quality of the data produced in the biomarker discovery phase. Hence how does one go about determining what constitutes the optimal biomarker(s)?

The issue of what constitutes the optimal biomarker(s) is an area of considerable debate and discussion. There is no one widely accepted answer to this question, since biomarkers serve numerous purposes. However, one can envisage a relatively simple compartmentalised series of modules that make up the 'Decision Wheel of Biomarker Discovery'. This is shown in Figure 2, and consists of the following:

1. **Scientific question** – the biological context should be defined by the scientific question being posed. Also is it a hypothesis or discovery driven endeavour?

2. **Define biomarker purpose** – within the context of the scientific question under consideration, what is the required output from the biomarker dataset? For example, are the biomarkers being used in a simple go/no go decision making process, or are they being used to understand a mechanism of biological action?

3. **Experimental design** – predicated on modules 1 and 2, issues such as what is the statistically significant number of samples needed, appropriate controls (both positive and negative).

4. **Organism/tissue/cell or body fluid selection** – the scientific question and the information needed from the biomarker output, will determine the selection of biological system to be studied. For example, a simple prognostic test for pancreatic cancer might indicate a blood or urine analysis. However, disease mechanism might require either animal or human biopsy samples.

5. **-Omic/panomic/imaging/clinical chemistry/physiology/systems biomarker selection** – often-times, given sample size, cost and time factors, one must select which molecular class of markers or determine if imaging will provide more pertinent information about the process under investigation.

6. **Single biomarker or panel** – does a single biomarker or a combination of biomarkers provide the most accurate and useful information about the biological system?

7. **If panel: optimum number** – what is the optimal number of biomarkers in the panel? Is it less than 10 (economics) or more than 100 (information rich)?

8. **ID or molecular signature** – does a simple molecular signature suffice, where none of the biomarkers have been identified, or does a well characterised and identified panel afford more information-rich content?

9. **Validation question: commercial or internal** –

the rigour of validation is determined by whether this is an internal process (eg toxicity of drug in animal), or the biomarkers are used as part of a kit (eg disease diagnosis).

10. **Validation and utilisation** – the biomarkers are subject to validation predicated on the answer to module⁹.

The various component modules of the 'Biomarker Discovery Decision Wheel' (Figure 2) are still a source of vigorous debate. Many practitioners still argue that a molecular signature is perfectly acceptable, whereas others dismiss such an opinion as shortsighted. In part the latter group argues that biomarkers are markers (distinguishers) of biological processes, hence it is imperative that such markers be identified. Certainly, if biomarkers are to be used in a meaningful way to facilitate decision-making processes in DDD, as well as treatment of disease as discussed here, then it would appear that identification is of paramount importance to ensure the highest quality data and information. However, it is certain that such debate will continue over the next 1-2 years, as the field develops.

Conclusions

The discovery, validation, commercialisation and use of biomarkers continues unabated in 2004-5. Active programmes are in place across the DDD pipeline driven by pharmaceutical and biotechnology companies. There are an increasing number of well attended biomarker conferences that provide a forum for stimulating debate and discussion about the fundamentals as well as the practical aspects of biomarker discovery and validation. This augers well for the future of this fledgling but rapidly growing field of endeavour. In particular spirited debate over the past 2-3 years has clearly helped to refine working definitions of biomarkers as well as sub-classifications of various types of biomarkers¹⁷. However, at present the future of biomarkers appears to be tinged with a mixture of excitement and uncertainty. In part that uncertainty is predicated upon the fact that numerous disciplines and practitioners contribute to the biomarker effort. In order to provide direction, clarity of goals and continued fortification of the biomarker foundation, more organisation needs to be brought to bear. Such a diverse group of people and skill sets needs a variety of tools to hone and fashion this industry. Several initiatives need to be considered. For example, in order to build on the excellent progress of the Biomarkers and Surrogate Endpoint Working Group, a grass roots type organisation needs to be formed. This can take the

12 Colburn, W.A. Biomarkers in drug discovery and development. From target identification through drug marketing. *J. Clin. Pharmacol.* 43: 329-341 (2003).

13 Trull, A.K. et al (Eds). *Biomarkers of Disease. An evidence-based approach.* Cambridge University Press, Cambridge, UK. 2002.

14 Biomarker Definitions Working Group. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* 69: 89-95 (2002).

15 Temple, R. Are surrogate markers adequate to address cardiovascular disease drugs? *JAMA* 282: 790-795 (1998).

16 Lathia, C.D. Biomarkers and surrogate endpoints: How and when they might impact drug development. *Disease Markers* 18: 83-90 (2002).

17 Cambridge Healthtech; www.genomicglossaries.com/content/biomarkers.asp

18 De Meyer, G and Shapiro, F. Biomarker development: The road to clinical utility. *Current Drug Discovery May*: 23-27 (2003).

19 Ferber, G. Biomarkers and proof of concept. *Methods Find. Exp. Clin. Pharmacol.* 24 (Supplement C): 35-40 (2002).

20 Naylor, S and Cavanagh, J. Status of systems biology-does it have a future? *Drug Discovery Today-Biosilico* 2: 171-174 (2004).

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

22 Hood, L. A personal view of molecular technology and how it has changed biology. *J. Protome Res.* 1: 399-409 (2002).

23 Ilyin, S.E., Belkowski, S.M and Plata-Salaman, C.R. Biomarker discovery and validation: technologies and integrative approaches. *Trends Biotechnol.* 22: 411-416 (2004).

Continued on page 30

Continued from page 29

24 Ilyin, SE et al. Functional informatics: convergence and integration of automation and bioinformatics.

Pharmacogenomics 5: 721-730.

25 Neumann, E and Thomas, J. Knowledge assembly for the life sciences. Drug Discov. Today (Supplement) 7: S160-S162 (2002).

26 Rubenstein, K. Post-Genomic Biomarkers: Revolutionizing Drug Development and Diagnostics. (2003) DM&D Publications, Westborough, MA, USA. (<http://www.drugandmarket.com>).

form of a formal professional Society, or a more loose-knit structure, eg HUPPO-like (organisation focused on proteomics). This structure can provide a forum for broad based discussions on definitions, as well as defining key elements of new tools and technologies required to advance the field. An annual meeting or meetings need to be scheduled that augment the current meetings organised by professional conference companies such as CHI or IBC. Consideration of consortia formation is also something that needs to be discussed, particularly in regard to biomarker databases, nomenclature and data visualisation.

The development of new tools and technologies that impact on the biomarker space will for the most part be developed in the respective -omic arenas. For example the issue of reproducibility, quantitation, sensitivity enhancement and high throughput capability in both proteomic and metabolomic analyses will be addressed by those respective scientific communities. However, a key area that does need to be addressed by the biomarker practitioners is how to rapidly and reliably integrate data from different platforms. The value of a biomarker panel containing genes/transcripts, proteins and metabolites is at present unknown. We need to address and understand if and why a composite panel is more valuable (scientifically, information-content and economically) than individual panels of genes or proteins or metabolites alone. Furthermore there will be significant debate in the future as to the advantages of genes versus proteins versus metabolite panels. Which type of constituent component will provide the most information about the process being investigated? Finally and more near term, the issue of whether a molecular signature, an individual biomarker (diagnostic) or panel of identified biomarkers is the best approach as a final product clearly needs to be debated. The answer to this conundrum is at present not clear, since who you ask will dictate the expediency of the response, and clearly the regulatory agencies such as the FDA will certainly (and correctly) weigh in on this discussion.

As the field continues to develop one will see continued concerted efforts from individuals from very different disciplines. For example, as the biomarker discovery engine becomes more refined and capable, integration with the knowledge assembly team to put the biomarkers into biological context will be essential. This latter event is in effect a pre-validation step, since it places the biomarker components in the biology of the system under study. The pre-validation step should then provide a qualifier prior to sending on the biomarkers for the

time-consuming but well defined validation steps. Finally, the regulatory bodies in North America, Europe and Japan/SE Asia will play a significant role in how the biomarker space continues to develop. At present the value of biomarkers in the DDD pipeline, as well as indicators of disease is the subject of debate and scrutiny by such regulatory authorities. It is important for the biomarker community to continue to educate and debate the authorities, since the latter's decisions will impact significantly on the economics of value of biomarkers in the future. It is an exciting future, that needs help in defining where biomarkers go and the importance of them in the future. The potential role of biomarkers in helping the decision-making process in patient treatment, as well as DDD is a very real possibility. However, at present there are few concrete examples and there are still some questions as to their true value! **DDW**

Professor Stephen Naylor is currently Adjunct Professor of Genetics and Genomics at Boston University of Medicine (Boston, MA, USA), as well as a Visiting Faculty Member in the Division of Biological Engineering at MIT (Cambridge, MA, USA) and a Faculty Member of the Computational Systems Biology Initiative (CSBi) also at MIT. He is the former Chief Technology Officer, and Senior Vice-President for Research at Beyond Genomics where, in conjunction with his colleagues, he built the world's first integrated systems biology platform, consisting of both analytical, bioinformatic and knowledge assembly capability. Previously he was the founding Director of the Biomedical Mass Spectrometry and Functional Proteomics Centre at the Mayo Clinic. In addition he was Professor of Biochemistry and Molecular Biology and Professor of Molecular Pharmacology and Experimental Therapeutics. He was also Adjunct Professor of Clinical Pharmacology, as well as Biomedical Engineering (Molecular Biophysics) at the Mayo Foundation. Stephen received his PhD from Cambridge University (UK) in biological mass spectrometry, completed post doctoral work at MIT (USA) and served as Associate Director of Mass Spectrometry at the MRC Toxicology Institute in London. Professor Naylor also serves as a consultant to a number of analytical, pharmaceutical and biotechnology companies, has published more than 225 research papers, has filed a number of patents and made more than 600 presentations at seminars worldwide.