Editorial

Systems biology and biomarker discovery

Karin D. Rodland*
Guest-Editor

Medical practitioners have always relied on surrogate markers of inaccessible biological processes to make their diagnosis, whether it was the pallor of shock, the flush of inflammation, or the jaundice of liver failure. Obviously, the current implementation of biomarkers for disease is far more sophisticated, relying on highly reproducible, quantitative measurements of molecules that are often mechanistically associated with the disease in question, as in glycated hemoglobin for the diagnosis of diabetes [1] or the presence of cardiac troponins in the blood for confirmation of myocardial infarcts [2]. In cancer, where the initial symptoms are often subtle and the consequences of delayed diagnosis often drastic for disease management, the impetus to discover readily accessible, reliable, and accurate biomarkers for early detection is compelling. Yet despite years of intense activity, the stable of clinically validated, cost-effective biomarkers for early detection of cancer is pathetically small and still dominated by a handful of markers (CA-125, CEA, PSA) first discovered decades ago.

It is time, one could argue, for a fresh approach to the discovery and validation of disease biomarkers, one that takes full advantage of the revolution in genomic technologies and in the development of computational tools for the analysis of large complex datasets. This issue of *Disease Markers* is dedicated to one such new approach, loosely termed the 'Systems Biology of Biomarkers'. What sets the Systems Biology approach apart from other, more traditional approaches, is both the types of data used, and the tools used for data analysis – and both reflect the revolution in high throughput analytical methods and high throughput computing that has characterized the start of the twenty first century.

The first article in this series, 'Systems Biology and the Discovery of Diagnostic Biomarkers' by Kai Wang, Inyoul Lee, Leroy Hood and David Galas, provides an eloquent description of the concept of 'systems biomedicine', and how this new approach can be used to support a predictive and personalized approach to medical practice that may revolutionize health care. The power of this approach is demonstrated in an analysis of prion diseases using mouse models and dynamic measurements of gene expression changes over the course of the disease. Significant changes in gene expression, mapping to biologically relevant pathways, are detected long before the onset of clinical symptoms, providing support for the concept that preclinical diagnosis through biomarkers is possible.

In their discussion of 'Systems Biology Approaches to Disease Marker Discovery', Sharon, Chen and Snyder provide an overview of the most prevalent methodologies currently in use for biomarker discovery, including protein microarrays, high throughput sequencing technologies for RNA and DNA, and mass-spectrometry-based proteomics. Using high density protein microarrays, the Snyder group has had significant success identifying biomarkers for SARS-coronavirus infection and ovarian cancer by devising protein microarrays that focus on the host immune response to infection or oncoproteins.

'Reverse Phase Protein Microarrays: Applications in biomarker discovery/validation, disease understanding, and high throughput clinical screening' by Wil-

*Corresponding author: Karin D. Rodland, PNNL Laboratory Fellow and Chief Scientist – NIH Programs Biological Sciences Division, Pacific Northwest National Laboratory, 902 Battelle Boulevard, P.O. Box 999, Richland, WA 99352, USA. Tel.: +1 509 371 6935; Fax: +1 509 371 6933; E-mail: karin.rodland@pnl.gov.

son, Liotta and Petricoin describes a novel technology that literally stands protein arrays on their head, by printing high density microarrays of the target (tumor biopsies, cell lysates, etc) and probing these arrays in a multiplex fashion for phosphoproteins indicative of activated signal transduction pathways. As described by the authors, the resulting profiles provide specific information about the disruption of critical signaling networks in disease, thus facilitating the identification and characterization of promising targets for molecular therapeutics.

Novel strategies for the application of mass-spectrometry based proteomics are delineated by Pitteri and Hanash in 'A systems approach to the proteomic identification of novel cancer biomarkers'. The focus of this group is on the application of sophisticated biochemical and physical subfractionation methods to samples derived from mouse models and *in vitro*, stable isotope-labeled cell cultures to identify low abundance proteins that may serve as serum biomarkers of cancer. Knowledge of cancer biology is exploited in the emphasis on secreted and cell surface proteins, and in the application of network analysis tools such as Metacor, Ingenuity and others. This approach has led to identification of 21 up-regulated proteins with known roles in cell adhesion and/or motility.

In 'Alternative Splice Variants, a New Class of Protein Cancer Biomarkers Candidates: Findings in pancreatic cancer and breast cancer with systems biology implications', Omenn and Menon discuss how opening biomarker discovery efforts to a new class of biomolecules, the protein products of alternatively spliced transcripts, can lead to both the identification of novel tumor-specific proteins, and to new insights into tumor-associated processes. One specific example from their study, the observation of a novel splice variant of pyruvate kinase in pancreatic cancer along with multiple new splice variants of glucose 3 phosphate dehydrogenase (GP3D), may provide insight into the mechanisms underlying the well known increase in glycolysis observed in tumors.

The final article in the series, 'Separating the Drivers from the Driven: Integrative network and pathway approaches aid identification of disease biomarkers from high-throughput data' by McDermott, Costa, Janszen, Singhal and Tilton, provides an in depth analysis of the new computational tools available for integration and analysis of the high content data resulting from omics studies described in the preceding articles. The abstract concept of 'networks' and the inter-relationships between network components is discussed, as a way of deriving mechanistic insight from the very large lists of genes and proteins generated in biomarker discovery experiments. It is the application of these abstract mathematical analyses to go beyond mere statistics that essentially defines the difference between a 'systems biology' approach to biomarkers, and more traditional brute force approaches that may tend to focus on the obvious and the known.

But what is a 'Systems Biology' approach? Wang et al. in this issue provide a very comprehensive description of the systems biology approach, as comprising five features: 1) measuring and quantifying biological information on a global scale; 2) integrating distinct modalities of biological information such as DNA, RNA and proteins; 3) capturing the dynamics of biological systems and networks, 4) using these sources of information to model the system, and then 5) iteratively testing and refining the model. Although this description of Systems Biology does prominently feature the generation of large quantitative datasets using a variety of 'omics technologies (features 1 through 3), numbers alone do not constitute a systems biology study. Systems Biology goes beyond the documentation of global changes in gene expression or protein abundance to model the flow of information in the system, and this process requires the seasoned application of 'expert knowledge' in the relevant biology. The parent study of prion disease in mouse models referenced in Hwang et al. [3] provides an illustrative example of the process. In that study, Hwang et al. used co-expression data from their comprehensive measurements of differential gene expression to build hypothetical protein networks that incorporated public interaction databases (BIND and HPRD), annotations in the GO ontology, and known pathological features of prion disease [3]. Similar strategies based on the use of published ontologies and protein interaction databases have been used in the other studies reported in this issue, with demonstrated success in providing reasonable candidate biomarkers for cancer [4–7] and other diseases.

Identifying functional modules within complex datasets is a very important first step in building useful biological models, but not sufficient for harnessing the full power of systems biology in biomarker discovery. In fact, Pepe et al. have shown that adding co-expressed genes to a multimarker panel does not add substantially to the information content of the panel [8,9]. Several investigators are beginning to emphasize network topology, specifically the flow of information from module to module within biological systems, with the object

of identifying 'bottlenecks' between hubs [11]. Bottleneck genes identified by topological analysis have proven to be highly enriched in essential processes associated with growth and virulence [10,11] and perform well as therapeutic targets [12].

The process of identifying biomarkers for cancer and other diseases can be likened to the process of threat detection in the national defense arena. There are many parallels – the need to identify susceptibilities, the need to respond adaptively as the threat changes over time, the problem of distinguishing signal from noise in highly complex datasets [13]. The defense community has responded by adopting 'composite signatures' of threat, which employ multidimensional datasets in which each dimension is a different measurement or technique [14, 15]. Perhaps it is time for biomedical scientists to adopt the same strategy – and Systems Biology provides the appropriate tools for building, testing, and validating a composite signature of disease.

References

[1] C.D. Saudek and J.C. Brick, The clinical use of hemoglobin a1c, *J Diabetes Sci Technol* **3** (2009), 629–634.

[2] J.E. Adams, 3rd., G.A. Sicard, B.T. Allen, K.H. Bridwell, L.G. Lenke, V.G. Davila-Roman, G.S. Bodor, J.H. Ladenson and A.S. Jaffe, Diagnosis of perioperative myocardial infarction with measurement of cardiac troponin i, *N Engl J Med* **330** (1994), 670–674.

[3] D. Hwang, I.Y. Lee, H. Yoo, N. Gehlenborg, J.H. Cho, B. Petritis, D. Baxter, R. Pitstick, R. Young, D. Spicer, N.D. Price, J.G. Hohmann, S.J. Dearmond, G.A. Carlson and L.E. Hood, A systems approach to prion disease, *Mol Syst Biol* **5** (2009), 252.

[4] R. Menon, Q. Zhang, Y. Zhang, D. Fermin, N. Bardeesy, R.A. DePinho, C. Lu, S.M. Hanash, G.S. Omenn and D.J. States, Identification of novel alternative splice isoforms of circulating proteins in a mouse model of human pancreatic cancer, *Cancer Res* **69** (2009), 300–309.

[5] S.J. Pitteri, L. JeBailey, V.M. Faca, J.D. Thorpe, M.A. Silva, R.C. Ireton, M.B. Horton, H. Wang, L.C. Pruitt, Q. Zhang, K.H. Cheng, N. Urban, S.M. Hanash and D.M. Dinulescu, Integrated proteomic analysis of human cancer cells and plasma from tumor bearing mice for ovarian cancer biomarker discovery, *PLoS One* **4** (2009), e7916.

[6] M.E. Hudson, I. Pozdnyakova, K. Haines, G. Mor and M. Snyder, Identification of differentially expressed proteins in ovarian cancer using high-density protein microarrays, *Proc Natl Acad Sci USA* **104** (2007), 17494–17499.

[7] C. Gulmann, K.M. Sheehan, R.M. Conroy, J.D. Wulfkuhle, V. Espina, M.J. Mullarkey, E.W. Kay, L.A. Liotta and E.F. Petricoin, 3rd, Quantitative cell signalling analysis reveals downregulation of mapk pathway activation in colorectal cancer, *J Pathol* **218** (2009), 514–519.

[8] M.S. Pepe, Z. Feng, Y. Huang, G. Longton, R. Prentice, I.M. Thompson and Y. Zheng, Integrating the predictiveness of a marker with its performance as a classifier, *Am J Epidemiol* **167** (2008), 362–368.

[9] H. Janes and M.S. Pepe, Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: An old concept in a new setting, *Am J Epidemiol* **168** (2008), 89–97.

[10] H. Yu, P.M. Kim, E. Sprecher, V. Trifonov and M. Gerstein, The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics, *PLoS Comput Biol* **3** (2007), e59.

[11] J.E. McDermott, R.C. Taylor, H. Yoon and F. Heffron, Bottlenecks and hubs in inferred networks are important for virulence in salmonella typhimurium, *J Comput Biol* **16** (2009), 169–180.

[12] L. Yao and A. Rzhetsky, Quantitative systems-level determinants of human genes targeted by successful drugs, *Genome Res* **18** (2008), 206–213.

[13] T.M. Mitchell, Mining our reality, *Science* **326** (2009), 1644–1645.

[14] C.M. Boner, Automated detection of terrorist activities through link discovery and massive datasets.

[15] C.M. Boner, G. Godfrey, T. Mifflin and J. Skokan, A random graph model for terrorist transactions, *IEEE Aerospace Conference Proceedings* (2004).