

Decoding the Folding Patterns of Serum Proteins: An Alternative Strategy for Cancer Biomarker Validation?

Dai J. Li · Daniel W. Chan

Published online: 20 July 2010
© Springer Science+Business Media, LLC 2010

Not everything that counts can be counted, and not everything that can be counted counts—Albert Einstein

Medicine is a science of uncertainty and an art of probability—Sir William Osler

The past decade has witnessed revolution in biomedical sciences, among them are the completion of human genome projects and coming of age of the systems biology [1, 2]. Their impacts on medicine are continuously unraveling [1]. With the developments of high-throughput measurement platforms, such as mass spectrometry (MS), and the next generation DNA sequencing technology, we are now capable of systematically categorizing almost all of the biomolecules of humans in physiology and in illness. Many candidate biomarkers potentially useful in early diagnosis, prediction of natural history, and response to therapy have been found even though their clinical usefulness remains to be determined. Subsequently, it becomes clear that the translation of the early success in biomarker discovery into clinical practice is a much daunting task than most of us have expected with validation a major hurdle.

Measurements are fundamental to scientific discovery in which observation and data collection are the very first steps. To ensure the reliable scientific inference, analytical, statistical, and clinical validations are essential so that clinical usefulness of the candidate biomarkers can be fully understood. However, validation has been extremely

challenging in cancers and other complex human disease alike. This is evident by the fact that few biomarkers have been convincingly validated and approved for clinical use despite tremendous efforts invested in recent years [3].

We are at the crossroad where validation is the barrier in bringing these biomarkers into clinical use. The dilemma we are facing is that there are too many candidate biomarkers. We have to be highly selective with respect to which biomarkers are potentially useful clinically. We believe that we need in-depth understanding of the information contents of the candidate biomarkers in order for us to commit the limited resources to bring the effective cancer biomarkers to patient care. We need to rethink the rationales for biomarkers discovery and our working assumptions on which successful validations will be hinged.

The question is why biomarker validation is so difficult to tackle under the current strategy for biomarker evaluation?

Serum biomarker discovery has been largely based on a working assumption that the concentrations of biomolecules would be up- or down-regulated in illness compared with the homeostasis of human physiology. For example, we are searching for changes in serum concentration of cancer biomarkers by comparing matching samples from healthy and disease population. The study design usually consisted of cases and controls or with a set of patient samples at multiple time points. Vast amount of data in clinical diagnostic devices have been collected in such fashion. In addition, by default we also assume that there is inherent information in the quantitative clinical observations, even for data collected from the studies with poorly defined study population. However, these assumptions may be flawed. The heterogeneity of the study populations aside, it

D. J. Li (✉) · D. W. Chan
Center for Biomarker Discovery,
Johns Hopkins Medical Institutions,
Baltimore, MD 21231, USA
e-mail: Dli9@jhmi.edu

is arguable that the signaling derived from the up and down movements of biomarkers may be merely indicative of early stages of the disease spectrum. The up or down movements in concentration of the target biomolecules is likely due to feedback regulation of the system's attempt to compensate for the gain or loss of functions incurred by the underlying pathophysiological changes. Such changes could be the result of nucleotide structural variations from the accumulation of somatic mutations, or other transient signaling events caused by posttranslational modifications such as phosphorylation, glycosylation, and others. Under these circumstances, the quantitative signal is probably not pathological. In another word, they are mostly physiological noises, which are not robust, even irrelevant for disease marker discovery.

Complex human disease networks are composed of many interconnecting pathways [4]. We believe that human cancer is likely caused by constitutively activated signaling pathways. However, these signaling pathways may not be present in the same configuration as that in the homeostasis. It is likely that the accumulated somatic mutations could cause protein conformational or other subtle structural changes leading to gradually decompensation of the physiological functions of tissues or organs. The result of these changes in protein folding patterns could facilitate the aberrant protein–protein interactions in the vastly complex disease networks. The conformation change could expose novel antigenic determinants to the pattern recognition receptors of the innate immunity triggering systemic immune response and inflammatory reactions [5]. These types of biomarkers are not differentiable by their mass values as an upward or downward movement in concentration.

Furthermore, we should not assume that all serum biomarkers carry equal amount of information as implied by their intrinsic values. Information is the useful data [6]. We should recognize that the data or observations do not change with time but our interpretation could. For example, with different binding partners, a candidate biomarker may have different function. More importantly, data needed to be interpreted in the context of the dynamic process of disease development which requires integrating all the temporal and spatial information from the biomarkers. The usefulness of biomarkers is defined by the clinical context. Therefore, understanding of the target population, prevalence, and disease parameters of the target population is vital for the validation. Biomarkers cannot be effectively validated if the underlying clinical entity is not clearly defined.

In biology, information flows from DNA to RNA and to proteins. Proteins are the effectors of biological functions. The protein folding patterns cannot be inferred by the underlying DNA sequence. At the protein level, the information is encoded in the corresponding variations in amino acid sequence and modulated by an array of the

posttranslational modifications [7–9]. If a protein in its linear form has no biological function, the information content must be represented in the overall patterns of protein folding. Disease phenotypes in humans could reflect the loss of homeostasis at least in part by the changes in protein conformation as demonstrated by many neurodegenerative diseases. The relevance of the protein conformation to cancer development is unknown largely because the information represented by the variations in conformation is not readily measurable and understood. If the data are noninterpretable in clear biological languages, there is no information associated with the data [10].

Currently, most of the clinical measurements are not designed to capture the information presented in protein conformation changes. The common practice in biomarker validations of adopting an artificial cutoff threshold in the quantitative assays to demarcate the “disease” from the “non-disease” is problematic for effective biomarker validation. Cancer involves probably stochastically connected aberrant signaling pathways [11]. The protein conformation and complex structures are complicated. The information pertaining to disease status cannot be deduced reliably from the data that have been reduced to two distinctive subsets denoted as “Yes” and “No” (1 and 0) because disease spectrum is a continuous process with progressive, accumulative, and qualitative changes. Unless the population distribution of the markers in disease and nondisease is clearly bimodal with minimal overlap, the imposed cutoff is intuitively counterproductive.

Cancer is a complex disease. Multiple cancer markers are likely needed to delineate its pathophysiology and its molecular phenotypes. This is especially true when the complexity in human disease generates emerging phenotypes at the systems level [12]. In this sense, there are probably no perfect biomarkers if used alone, but with multiplexing, they can contribute to incremental gain of information. In this case, data processing with the cutoff threshold approach could be too simplistic to be meaningful biologically and clinically.

Diagnostics is playing increasingly important roles in triage of different molecular subtypes of a cancer under a common name to appropriate therapeutic regimens. In clinical settings, effective and safe clinical measurements are intended to answer a seemingly simple question; i.e., should the patient be treated? Validated and interpretable signature biomarkers for the molecular phenotypes are in great demands. Toward this goal, we should be mindful that whether our assumptions about the illness are defensible, and the rationales of biomarkers discovery are valid to ensure the candidate biomarkers can be effectively decoded for useful information.

It is necessary to explore the alternative approach in biomarker discovery for unmet medical needs for actionable

information. The variations in folding patterns of the functional proteins such as enzymes and cell signaling molecules like the interleukins and oncoproteins could offer a new paradigm in identifying the candidate biomarkers for the complex diseases that can be efficiently validated. Technical advancement in measurements and thoughtful interpretation of the information should allow the vision of tailored treatment of personalized medicine within our reach in the foreseeable future.

References

1. The New York Times Editorial: The Genome, 10 Years Later. June 21, 2010
2. Weston AD, Hood L. Systems biology, proteomics, and the future of health care: toward predictive, preventive, and personalized medicine. *J Proteome Res.* 2004;3:179–96.
3. Chan DW. Will cancer proteomics suffer from premature death? *Clin Proteomics.* 2010;6:1–3.
4. Goh KL, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. *PNAS.* 2007;104(21):8685–90.
5. Chen R, Alvero AB, Silasi DA, Steffensen KD, Mor G. Cancer take their Toll—the functions and regulation of Toll-like receptors in cancer cells. *Oncogene.* 2008;27:225–33.
6. Antezana E, Kuiper M, Mironov V. Biological knowledge management: the emerging role of the Semantic Web technologies. *Brief Bioinform.* 2009;10(4):392–407.
7. Dobson CM. Protein folding and misfolding. *Nature.* 2003;426:884–90.
8. Yang S, Banavali NK, Roux B. Mapping the conformational transition in Src activation by cumulating the information from multiple molecular dynamics trajectories. *PNAS.* 2009;106(10):3776–81.
9. Shank EA, Cecconi C, Dill JW, Marqusee S, Bustamante C. The folding cooperativity of a protein is controlled by its chain topology. *Nature.* 2010;465:637–41.
10. Altschuler S, Wu L. Cellular heterogeneity: do differences make a difference? *Cell.* 2010;141:559–63.
11. McClellan J, King MC. Genetic heterogeneity in human disease. *Cell.* 2010;141:210–17.
12. Plsek PE, Greenhalgh T. The challenge of complexity in health care. *BMJ.* 2001;323:625–8.