Closed-loop healthcare data processing system: The use of proteomics and information technology to improve healthcare

Neil G. Siegel, Ph.D. (correspondence author)

Sector Vice-President & Chief Technology Officer

Northrop Grumman Corporation, 1800 Glenn Curtiss Street, DH6-2271, Carson, California 90746, USA

E-mail: neil.siegel@ngc.com

Larry Gold, Ph.D.

Chairman, Somalogic, Incorporated

2945 Wilderness Place, Boulder, Colorado 80301, USA

E-mail: lgold@somalogic.com

Abstract: Clinical care decision-making can be improved though the use of a "closed-loop" information processing cycle that integrates sensed and recorded data on individuals (initially centered around a combination of highly-multiplexed, longitudinal protein measurements, and electronic health-care records), and applies, analyzes, and visualizes those data using "causality cases" (authoritative findings about how sensed data relates to diagnosis), so as to enable personalized diagnostic and treatment guidance for consideration by the patient, the care team, and the healthcare administrator. We integrate scans of the proteome with a "learning" information processing system. The resulting system, soon capable of measuring thousands of proteins simultaneously from very small blood sample sizes, can be the basis for potentially significant improvement *at the system level* in the healthcare system, helping improve average outcomes while decreasing total costs. Advances in both data processing technology and biotechnology have reached a level to make this approach feasible, and if applied broadly and effectively, can provide a scalable and exportable mechanism for achieving the "triple aim" [1] of simultaneously improving care outcomes, lowering costs, and enhancing the patient experience.

Keywords: healthcare informatics, proteomics, information processing, systems engineering

1 Background

Currently, the United States spends over 17% of GDP ($2.6 trillion per year) on healthcare, and this amount is expected to grow at a nearly 6% annual rate[2]. Healthcare costs for the Department of Defense alone have nearly tripled in the last decade, from $19 billion in 2001, to $53 billion in 2011. Many attempts have been made to slow down, and eventually reverse, this increase in healthcare spending; thus far, most attempts have failed (or, at least, have not yet demonstrated the intended impact[3]). We believe that new insight into the healthcare system is needed in order to solve this problem. In this paper, we describe a *data-driven*, *personalized approach* to healthcare, which proposes an intervention at the *system level* of the healthcare system; significant business process improvements are also needed in order to leverage these emerging scientific and technical breakthroughs described in this paper to enable this new approach. We will indicate how such

system-level intervention can create significant benefits in both healthcare outcomes and cost reductions (through broad-scale adoption), at a level of materiality not achieved by previous techniques.

In other domains (e.g., military and large civilian aircraft), both better outcomes and significant cost savings have been achieved through the use of multiple sensors tied to a processing system, which leads to a "condition-based maintenance" approach (where preventative and restorative maintenance are based on measurements of actual, present condition); such condition-based maintenance has proven superior in both cost and outcomes as compared to traditional "schedule-based maintenance" approaches. In healthcare today, protocols used to diagnose and treat patients are, in large part, driven by population averages, and schedule-based (e.g., periodic) assessments. For example, many prevention approaches are currently based on limited knowledge of a patient's actual biomedical condition and history. We propose a method that implements condition-mediated healthcare "maintenance" through the personalization of healthcare protocols and treatments. We use multiplexed proteomics as the primary sensor, providing unprecedented biological insight, and powerful analytics to identify medical issues before they become critical problems.

Variation in Practice: Healthcare also is subject to huge (and unjustified) variation in practice. It is not enough to have personalized data that indicate a patient's condition; the course-of-action must reflect appropriate practices consistently, and incorporate evidence-based standards. Data indicate, however, that such unjustified variation accounts for between 30 and 50 percent of the total U.S. healthcare spend[4], in addition to causing harm and even death (some sources say > 100,000 deaths / year in the U.S.). Wheeler[5] and others have shown, however, that improving quality must start with ***reducing*** unjustified ***variation***, rather than with trying to improve the mean of a process outcome; efforts to improve the mean are appropriate only ***after*** unjustified variation has been reduced. Such reduction in unjustified variation is a major element that is addressed by our closed-loop system. At the scale of the U.S. healthcare system, reducing variation first entails creating evidence for best practices based upon new findings at the individual person-level, and incorporating that evidence into recommendations for patients and healthcare providers[6].

Induced Learning Load: In addition, there are thousands of new refereed medical journal articles and hundreds of new drugs / treatment protocols every year; without automation to provide up-to-date recommendations, how is a practitioner to keep up? And, while doctors may be well-versed in some fields, their knowledge and currency may be variable across other sub-disciplines. An evidence-based approach for providing physicians with the "right information at the right time", tailored to the patient they are treating, is needed to facilitate a more standard treatment of healthcare, so as to achieve that reduction in unjustified variation discussed above. As more insightful (i.e., "highly predictive") clinical evidence based upon phenotypic molecular measurements is incorporated into the screening, diagnostic, and treatment process, more useful and applicable guidelines can be provided to the clinicians to give patients the best care possible.

Growing Imbalance Between Provider Supply and Healthcare Demand: Primary-care providers (about one-third of the physician population) are a key element of the healthcare system in addressing heart disease, cancer, and diabetes – yet their percentage of the physician population is likely to not keep pace with the needs of individuals who seek care. For example, the number of elderly people (who have higher burden of chronic disease) is anticipated to double by 2030. In addition, increasing numbers of individuals may likely receive increased access to care through expanded insurance coverage in the coming years. New support models are therefore necessary to increase the clinical efficiency of the healthcare system, and thereby, to meet the ever-increasing healthcare demand among the population.

Improper Stratification of Populations: According to Lawrence[7] and others, the U.S. healthcare system in essence consists of a "sick-care" system – where most of the money is spent – and a prevention or "wellness" system. At the same time, the population of the U.S. at any given point in

time can also arguably be divided into two populations, those that can be characterized as essentially well, and those that can be characterized as sick. At present, people make their determination of whether they are sick (and hence, whether and when to seek sick care) with only a minimum of information, much of which is subjective (e.g., how they "feel," "what hurts," their temperature, etc.). A key goal for an improved healthcare system needs to be the improvement of the mechanism by which every member of the population determines which group they are a part of at the present moment, and when / if they intend to seek help from the "sick-care" and/or wellness systems. Improper decisions about when and where to enter the health-care system have significant adverse effects on both outcomes and costs. ***This is, in our view, the "systems-engineering" leverage point for improving the healthcare system***. Note, however, that these decisions are generally made by the ***individual members of the population themselves***, not by healthcare professionals! A key to remedying this problem therefore involves not only assessing the health status of individuals more effectively, but ***providing some subset of that assessment directly to the potential healthcare consumer***, in addition to providing this assessment (in a more detailed form) to the healthcare professional. At present, some people enter the sick-care system late, when they have more advanced illness, while others enter into the system prematurely, receiving costly and unnecessary intervention. This matter is particularly important in chronic disease, where individuals might be "less sick", would use fewer healthcare resources, and would achieve better healthcare outcomes if they were diagnosed earlier (e.g., currently undiagnosed diabetics), and thereby received the most effective treatment for their condition(s) at an earlier stage.

This is a radically different insight and approach than any methodology currently used at scale in the current healthcare system, and also radically different than the approaches used by most previous attempts to improve the healthcare system via automation alone, which might therefore explain why (as cited above[8]), previous efforts to improve the healthcare system via information technology have essentially failed.

2 Working hypothesis

We can improve the healthcare system, on all of the "axes" identified above, by developing a "closed-loop," personalized, learning healthcare system, using assays based upon synthetic molecules called Slow Offrate Modified Aptamers ("SOMAmers") as the key sensor to measure the concentration of thousands of proteins in blood longitudinally. By combining those data with data from electronic health records and other relevant sources of information (e.g., lab tests, imaging results, and so forth), and processing the information using a set of initial and evolved algorithms ("causality cases") that relate specific signals with specific diagnoses, personalized healthcare protocols can be generated. This information will:

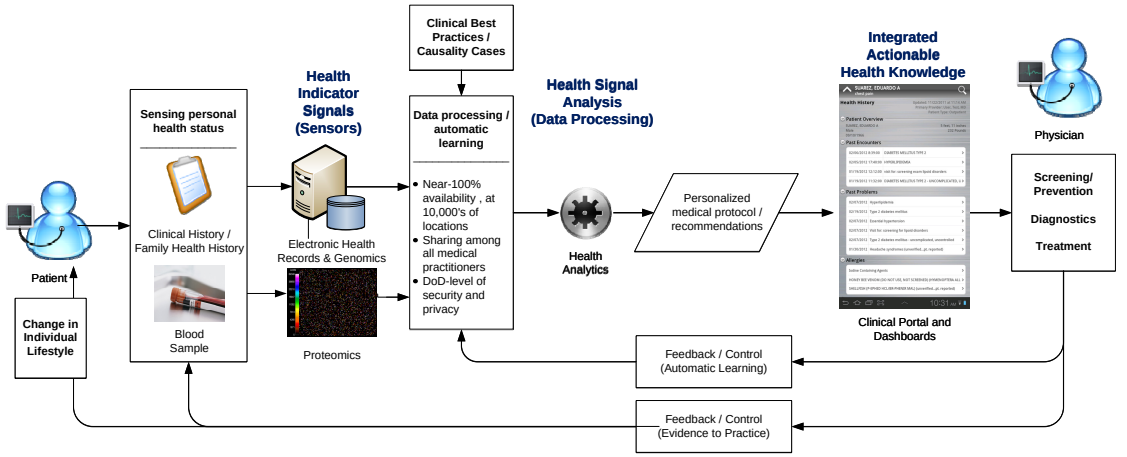
- Allow consumers to know accurately – based on objective data – *when* they need to enter the healthcare system (and *whether*, specifically, to enter the "sick-care" system or the wellness system) so as to seek detailed diagnosis and treatment.
- Permit the health-care practitioner at the patient's point-of-entry rapidly and more precisely to evaluate changes and off-nominal conditions in the patient across a wide range of factors, based on minimally-invasive technologies and data sources (combination of a one-drop blood test, electronic healthcare records, and other information), and with a high degree of certainty, route the patient to appropriate tests, screening, specialist practitioners, and procedures, thereby saving time, money, and frustration.
- Materially reduce unjustified variation in healthcare practice, by joining detailed diagnostic information with evidence of clinical effectiveness applied to specific patient

strata, thereby allowing providers and healthcare systems to improve and target delivery of care across both individuals and patient cohorts.

- Increase efficiency of healthcare system by more proactively identifying and monitoring sick people earlier in their disease course, so that they come into treatment more effectively and with reduced use of potentially suboptimal treatments.
- Enable practitioners and healthcare systems to become more efficient and effective in clinical practice through periodic incorporation of new "causality cases" (the latest information about measureable health indicators [e.g., proteomics] that indicate and predict health factors, diseases, and tendencies) into a computer database, which can then be linked automatically to personalized healthcare options.

Our vision of the future is the "closed-loop, condition-based" learning healthcare system that is summarized in Exhibit 1. This system represents a shift from the current medical model of being primarily reactive to disease, to one that takes a more proactive approach to treatment and prevention by changing the way massive amounts of clinical and other types of data are integrated, managed, analyzed, and used so as to improve decision-making. In addition to providing important and timely data to health-care providers, it explicitly also provides the data necessary to allow members of the general public better to understand when they need to enter the sick-care system. As a result, patients and healthcare practitioners are provided the information needed to improve outcomes, reduce costs, reduce wasted time, and reduce unjustified variation in diagnosis and treatment.

Exhibit 1: A closed loop system of "condition-based healthcare" will improve outcomes, lower costs, and achieve a better balance between prevention and acute treatment.



In what way is this different than today's practice? Today, the practitioner has to make an estimate of a candidate diagnosis based on very little information (e.g., blood pressure, temperature, a patient's description of "how they feel", and so forth), and then (since today's test are largely directed at identifying a single disease or condition) the practitioner orders tests to examine the patient for that one single condition. If this test comes back positive, the practitioner can either treat the patient, or refer the patient to another specialist practitioner, but their interpretation of the test results is still going to be informed largely by population averages; little knowledge of what might be "normal" for this particular patient is available. And in the case that the test comes back negative, the entire process essentially has to start all over again. Furthermore, this one-test-at-a-time methodology does little to correct for the sorts of statistical/diagnostic errors (e.g., false positive / false negative) that are in fact frequent in today's medical diagnostic tests.

By contrast, in our vision of the future, since we can measure thousands of biological markers in one inexpensive test from one drop of blood, the practitioner is not forced to make this initial estimate of a diagnosis based on little information; instead, the practitioner will have access to not only the patient's most-recent measurement of these thousands of biological markers, but in fact access to the entire longitudinal history of such measurements for this patient. A critical benefit is that the longitudinal measurements will markedly decrease the likelihood that the inevitable false-positive and false-negative from individual tests will confuse the diagnosis. Eventually, these could be combined with suitably-protected longitudinal measurements from people who are related to the patient (familial data), people who work at the same location as the patient (environmental data), and other correlative data. Furthermore, the set of causality-cases contained within the computer is constantly updated, assuring the practitioner and healthcare administrators that the computer's diagnosis is based on the newest accepted knowledge ("learning healthcare system"). The ability to "drill through" to the actual measurements, the causality cases, and even to the underlying reference literature build confidence in the recommendations of the computer system (which, of course, can always be overridden by the practitioner). The result is that the patient and the practitioner can select a highly personalized and likely-to-be-suitable diagnosis and treatment, without the "random walk" through the healthcare system that patients often face today.

3 The challenge

There are considerable technical and social factors that make this problem difficult: the need to sense and classify large-scale proteomic data consistently through the use of what in essence is a highly-distributed measurement instrument; those measurements and several other types of data must be combined so as to provide decision-making guidance in a fully-automatic, yet highly-credible manner (if not fully automatic, the processing would overwhelm personnel; if not credible, recommendations would not be used); often, there will be small signals embedded in lots of "noise," and actionable information is often not directly present, but must be synthesized by combining available data into higher-level abstractions; there must be highly secure data authorization and access controls due to the significant personally identifiable information that will be accessible across large patient populations; at the same time, the results must be made instantly available and accessible to a wide range of appropriately-authorized personnel at a large number of points-of-presence; the ingest rate at full scale will be immense; the system must have essentially 100% availability, and never lose data; there will be considerable adaptation over a very long system life-time; physicians and patients must respond to the information in beneficial ways; significant health system challenges to adoption must be overcome (provider acceptance, system adoption, reimbursement support, and so forth); data presentation to users will be a key factor in system acceptability.

While each of the above in isolation may seem tractable, the combination is formidable, and in some sense, unprecedented.

4 The approach, and what we can already accomplish

As formidable as the above seems, we in fact have successfully implemented many of the most difficult aspects already:

- From single-drop blood samples, we can accurately sample more than 1,000 proteins, over a 5-order-of-logarithm range of density – this is the SOMAscan assay[9]. Being able to do this from such a small sample size ought to enable the eventual collection to be done *without requiring a phlebotomist*, saving system time and money, as well as reducing risk and inconvenience to the patient.

- We are about to demonstrate (at a pilot level) that the huge volume of data generated by SOMAscan assays, other medical sensor systems, and electronic health records can be processed via clinically-relevant causality cases, so that they eventually could inform timely, accurate medical decision-making by both patients and healthcare providers.
- We have developed an initial set of causality cases (see Exhibit 2, below) that can allow for the automated generation of diagnoses from the measured data, while also allowing patient and health-care provider to "drill through" to the measurements, the triggering indications, and the underlying reference literature.
- We have demonstrated operation at suitable scale of computing systems that display the requisite characteristics of number of points of presence, security, availability, transaction rate, ingest rate, and other computer-science measures of system operations.
- We have demonstrated significant progress on the problems associated with display of the resulting information both to patients and healthcare providers, so that these data can facilitate effective, transparent, and timely decision-making.
- We have demonstrated impressive repeatability in the protein assay[10].
- We have demonstrated that the protein markers in the current version of the assay actually show rapid and consistent change in response to life-changes. This is an initial indication that a longitudinal time-history of these proteins may have clinical significance[11].

The current, initial version of the system makes the assumption that *proteins* are the key item to measure to enable the transition to the healthcare system model described above. It been long accepted throughout the medical community that *genomic information* instead would be the "sensor", but in fact one's genome changes relatively little as one ages and/or gets sick, and therefore, is actually a weak indicator of when a person might need to seek treatment (it ***is*** an indicator of a likelihood that one might develop a condition, but does not state whether one has in fact developed that condition).

The evidence from the tests to date[12] demonstrates that proteins are in fact good indicators of change and illness (whether these proteins are causal, or merely indicators, is actually not relevant to this approach).

There is of course nothing in the design of the closed-loop processing system that precludes the eventual incorporation of other "sensors"; the entire world of "omics" could be incorporated as they mature and achieve clinical significance, and when their operation is at a scale and price-point that make them effective contributors to the current protein-based analysis. We anticipate that other such "sensors" will be developed and added to the concept over time.

The mechanism that now makes this protein assay feasible at this time for this purpose is the use of SOMAmer reagents (reagents that can be identified to bind very specifically with designated target proteins). SOMAmers are synthesized, not raised in an animal, which allows for economical volume production with high consistency[13]. Combining SOMAmer-based protein concentration data with the data in electronic healthcare records and other data sources can provide both current and longitudinal information about each individual patient. Using a data base of current medical knowledge and best practices, in combination with SomaLogic's (and eventually, third parties') sets of "causality cases", which relate the sensed medical signals to current and – most importantly – predicted conditions and diagnoses (see Exhibit 2, SomaLogic Clinical Indications, below), can provide rapid, accurate, and personalized diagnoses and recommendations for thousands of conditions simultaneously, all from that single blood specimen (one drop!).

Exhibit 2: SomaLogic Clinical Indications

Condition	Utility	# Samples evaluated (by 4Q 2012)
Proof of Concept Achieved		
Lung Cancer	Screening, Monitoring	3700
Pancreatic Cancer	Screening, Monitoring	800
Renal Cell Carcinoma	Prognosis and Monitoring	550
Mesothelioma	Surveillance - asbestos	1500
Cardiovascular risk	Risk monitoring	2700
Diabetic retinopathy	Early detection	300
ALS (Lou Gehrig's)	Rule out	900
Pre-term birth	Prediction	1000
Pancreatitis	Diagnosis	500
Acute lung injury	Diagnosis, prognosis	400
COPD	Prognosis	304
NASH (liver fat/fibrosis)	Monitor in obesity	700
Studies Underway		
Diabetic heart disease	Diagnosis	400
Nutrition	Good vs bad	400
Prostate cancer	Gleason >6	400
Bladder cancer	Monitoring	300
Kidney function	Monitoring GFR	800
Diabetic kidney disease	Prognosis	800
Tuberculosis	Diagnosis (active)	300
C-difficile, Cholera toxins and other pathogens	Detection	Reagent development

The approach described above can improve outcomes by basing diagnosis and recommendations on far more data than are available or affordable today from any alternative diagnostic procedure, and by decreasing unjustified variation in protocol and treatment. We have demonstrated the capability to create a technical architecture that integrates, manages, and analyzes the huge volume of data generated by this project, and create – largely automatically – the actionable recommendations that will achieve the health outcomes and cost-reduction goals. The approach based on this capability can decrease costs by: (a) substituting an inexpensive, multi-diagnostic test for a series of expensive and single-purpose ones, (b) enabling earlier clinical detection and intervention, (c) mechanizing and accelerating the sharing of test results and new healthcare knowledge across specialists and institutions, (d) reducing variation in clinical decision practices, and (e) significantly reducing the broad range of individuals who are currently and unnecessarily screened, tested, and treated (e.g.,

providing a fact-based mechanism to allow patients to know when they need to enter the "sick-care" system). It combines the best of science, engineering, and medicine to create a paradigm shift in the way healthcare systems operate.

5 What remains to be accomplished?

This is a long and difficult journey, both on the technology and "social" sides of the problem. For example, important steps that remain to be solved include:

- Collect samples without having to freeze them. A preliminary approach to accomplishing this exists, but further work is probably required in order to do this at the scale required by a national-wide periodic collection program.
- Drive down the cost of each proteomic test. We are working towards a $150 / test target, and expect to get materially below that figure as volume scales up.
- Conduct pilot studies to explore and solve material issues associated with integration of the SOMAmer assay and processing capability into targeted healthcare systems, including pre-pilot identification and validation of appropriate metrics and collection of initial results.
- Evaluate clinical and systems data from pilots to determine effectiveness and assess potential generalizability for broad-scale healthcare system implementation

In addition to the technical problems cited above, incorporation into the healthcare sytem presents a number of complex social problems, for example:

- Where to insert the measurement actions? At doctor's offices? At retail sales establishments? At bespoke clinics?
- How do tests get associated with a particular patient without error? Ought we to collect fingerprints at the time of the assay?
- How do the tests get paid for? As a part of the membership fees for "closed" systems, such as Kaiser? Through conventional insurance plans?

These are merely indicative; the list is long. Bear in mind that if the goal is substantially to reduce healthcare costs (while improving outcomes), decreased costs means fewer healthcare jobs for someone – and so resistance will be mighty and prolonged.

6 Why is this the right time to solve this problem? Why has no one solved it before now?

We appear to be at a point in time where there is at least a "national conversation" on the subject of how to reduce healthcare costs, and at the same time there appears to be ever-increasing expectations on the part of the public for improved – and more consistent – healthcare outcomes. This may create a social context in which implementation and financing of an approach like the one indicated herein is feasible. At the same time, the convergence of the assay technology (SomaScan) and computing technology (enabling capacity, reliability, and privacy) create the potential for a solution at scale, and at a cost-per-test point, that could also be deemed feasible. Without this combination of (recent) developments, large-scale implementation was probably not technically or socially feasible. Consider the recent break-throughs:

- The SOMAscan assay has passed a point-of-inflection, showing feasibility of the core sensing concept of measuring 1,000's of proteins simultaneously, at scale, from a small blood sample (proteomics).

- In addition, these proteomic assays can now be tied to an initial set of medically-significant causality cases (e.g., metabolic syndrome, cancer, infectious diseases; see SomaLogic Clinical Indications, above). The causality cases link assay data to specific clinical dilemmas within a healthcare system, so that the impact on treatment, health economics, and operations can be measured.
- Previous approaches to personalized medicine centered primarily on genomics; however, one's genome does not significantly change to reflect age and health status. Genomic data will complement our proteomics-based approach by enhancing the assessment of an individual's background risk.
- Feasibility of solutions to many of the individual data processing challenges have been demonstrated, drawing primarily from our work in the intelligence community context, allowing this proposal to attempt a simultaneous solution at scale. There remain significant processing challenges, but Northrop Grumman has developed new techniques to address them. For example, we have developed a promising new concept for a revolutionary "privacy in the cloud" technique that links security services to business processes by relating security services directly to the data they implicitly protect. It employs techniques from reliable communications coding to distribute and retrieve data in the cloud in combination with granular, dynamic access control.
- Another example of a relevant processing challenge is the ability to manage effectively the escalating volume, velocity, variety, and complexity of "big data" is attracting significant attention and resources today. Northrop Grumman has pioneered an approach that supports the scale-up and scale-out of information management architectures based on critical operational performance metrics. This background is a key enabler to scalable information architecture for personalized medicine.
- The portion of the U.S. population served by "closed" healthcare systems – systems such as Kaiser, the Veteran's Administration, TRICARE for active-duty U.S. military personnel, etc., system that basically operate on a "fixed-price" basis per patient per year – has increased to the point where their business-case to adopt something like the capability described herein becomes attractive. It is likely that this approach only will be cost-effective if the scale of implementation is material.

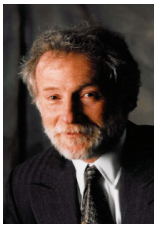
7 Summary

The radical and novel *data-driven, personalized approach* to healthcare described in this paper capitalizes on recent advances in both systems integration technology and biotechnology. Strategic partnerships bring the background and expertise that are required. The approach not only solves the technology problems in this domain, but also produces a systems architecture and "learning" information processing system that are scalable in a manner that makes them feasible for actual deployment to large-scale clinical use. This enables a true data-based approach to healthcare, which is likely necessary in order to achieve major improvement in outcomes and cost-reductions. By allowing ***both*** the provider and patient to assess health status, progression of disease, and responsiveness to therapy more efficiently, this approach can reduce variation in practice, improve provider efficiency, and provide better system support to meet increasing healthcare demand. Through broad-scale health system adoption, this new model of care will help foster and advance national health goals to deliver simultaneously enhanced patient experience, reduced costs, and improved population health – "the Triple Aim".

Approved for public release by Northrop Grumman. NGC release #14-2920.

8 Meet the authors

Neil Siegel, Ph.D., is sector vice-president & chief technology officer at Northrop Grumman. He has been responsible for the creation of many first-of-their-kind, large-scale, high-reliability data systems. He is a member of the National Academy of Engineering, a Fellow of the IEEE, and the recipient of the Simon Ramo Medal for systems engineering, among many other awards and honors.



Larry Gold, Ph.D. is Chairman, CEO of SomaLogic. He has founded several successful bio-tech companies, including Synergen, NeXstar, and SomaLogic. He discovered aptamers, which provide the basis for proteomic-based diagnosis. He is a member of the National Academy of Science, the American Academy of Arts and Sciences, among many other awards and honors.

References

1 Bisignano & Kenny, Pursuing the Triple Aim: Seven Innovators Show the Way to Better Care, Better Health, and Lower Costs, 2012.

2 Centers for Medicare & Medicaid Services, National Health Expenditure Data

3 "Why Health Information Technology Doesn't Work", Elmer V. Bernstam and Todd R. Johnson, The Bridge – the magazine of the U.S. National Academy of Engineering, fall 2009.

4 Lawrence, D. (former CEO of Kaiser Permanente Health System), personal communication, September, 2012.

5 Wheeler, Donald, Understanding Variation, SPC Press, 2000

6 Gawande, Big Med: Restaurant chains have managed to combine quality control, cost control, and innovation. Can healthcare?, The New Yorker, 13 August 2012

7 Lawrence, D., opere *citato*

8 e.g., Bernstam and Johnson, opere *citato*

9 Williams, Steve, Aphrodite and Medusa: Portraying the beauty of biology and the Gorgon of disease, GoldLab Symposium, May 2014

10 Williams, Steve, opere *citato*

11 Gold. Larry, Somascan longitudinal studies, August 2013.

12 Gold, Larry, opere *citato*

13 See "Aptamer-based Multiplexed Proteomic Technology for Biomarker Discovery," Public Library of Science [PLoS ONE], www.plosone.org, December 2010, for more information and additional references