

**Spring 2020 – Systems Biology of Reproduction
Discussion Outline (Systems Biology)
Michael K. Skinner – Biol 475/575
Weeks 1 and 2 (January 23, 2020)**

Systems Biology

Primary Papers

1. Westerhoff & Palsson (2004) Nat Biotech 22:1249-1252
2. Joyner (2011) J Appl Physiol 111:335-342
3. Clarke, et al. (2019) Endocr Relat Cancer. 26(6):R345-368

Discussion

- Student 1 - Ref #1 above
- How does this support evolutionary systems biology?
 - What was the convergence discussed?
 - Give an example that supports this perspective.
- Student 2 - Ref #2 above
- What is the problem with reductionism?
 - What is the void?
 - What is the solution?
- Student 3 - Ref #3 above
- What is multiscale modeling?
 - What is the role of mathematical modeling?
 - What network analysis insights into endocrine cancers are described?

The evolution of molecular biology into systems biology

Hans V Westerhoff¹ & Bernhard O Palsson²

Systems analysis has historically been performed in many areas of biology, including ecology, developmental biology and immunology. More recently, the genomics revolution has catapulted molecular biology into the realm of systems biology. In unicellular organisms and well-defined cell lines of higher organisms, systems approaches are making definitive strides toward scientific understanding and biotechnological applications. We argue here that two distinct lines of inquiry in molecular biology have converged to form contemporary systems biology.

Whereas the foundations of systems biology-at-large are generally recognized as being as far apart as 19th century whole-organism embryology and network mathematics, there is a school of thought that systems biology of the living cell has its origin in the expansion of molecular biology to genome-wide analyses. From this perspective, the emergence of this 'new' field constitutes a 'paradigm shift' for molecular biology, which ironically has often focused on reductionist thinking. Systems thinking in molecular biology will likely be dominated by formal integrative analysis going forward rather than solely being driven by high-throughput technologies.

It is, however, incorrect to state that integrative thinking is new to molecular biology. The first molecular regulatory circuits were mapped out over 40 years ago. The feedback inhibition of amino acid biosynthetic pathways was discovered in 1957 (refs. 1,2), and the transcriptional regulation associated with the glucose-lactose diauxic shift led to the definition of the *lac* operon and the elucidation of its regulation³. With the study of these regulatory mechanisms, admittedly on a small scale, molecular biologists began to apply systems approaches to unravel the molecular components and logic that underlie cellular processes, often in parallel with the characterization of individual macromolecules. High-throughput technologies have made the scale of such inquiries much larger, enabling us to view the genome as the 'system' to study. Thus, the popular contemporary view of systems biology may be synonymous with 'genomic' biology.

This article discusses two historical roots of systems biology in molecular biology (Fig. 1). Although we briefly outline the more familiar first root—which stemmed from fundamental discoveries about the nature of genetic material, structural characterization of macromolecules and later developments in recombinant and

high-throughput technologies—more emphasis is placed on the second root, which sprung from nonequilibrium thermodynamics theory in the 1940s, the elucidation of biochemical pathways and feedback controls in unicellular organisms and the emerging recognition of networks in biology. We conclude by discussing how these two lines of work are now merging in contemporary systems biology.

Scaling-up molecular biology

In the decades following its foundational discoveries of the structure and information coding of DNA and protein, molecular biology blossomed as a field, with a series of breathtaking discoveries (Fig. 1). The description of restriction enzymes and cloning were major breakthroughs in the 1970s, ushering in the era of genetic engineering and biotechnology. In the 1980s, we began to see the scale-up of some of the fundamental experimental approaches of molecular biology. Automated DNA sequencers began to appear and reached genome-scale sequencing in the mid-1990s^{4,5}. Automation, miniaturization and multiplexing of various assays led to the generation of additional 'omics' data types^{6,7}.

The large volumes of data generated by these approaches led to rapid growth in the field of bioinformatics, again largely emanating from the reductionist perspective. Although this effort was mostly focused on statistical models and object classification approaches in the late 1990s, it was recognized that a more formal and mechanistic framework was needed to analyze multiple high-throughput data types systematically^{8,9}. This need led to efforts toward genome-scale model building to analyze the systems properties of cellular function.

Molecular self-organization

Even before the first key events in the history of molecular biology, several lines of reasoning revealed that integration of multiple molecular processes is fundamental to the living cell. Biochemical processes necessitate the production of entropy (chaos in the thermodynamic sense) as driving force. The paradox felt by many, but expressed by Schrödinger in his war-time lectures¹⁰, was how one could explain the progressive ordering that occurs in developmental biology (that is, the 'self-organization,' decrease in chaos) when entropy ('chaos') must be increased.

The answer was that one process could produce order (negative entropy or negentropy) provided it was coupled to a second process that produced more chaos (entropy): coupling, another word for integration of processes, is therefore essential for life. Onsager¹¹ provided the basis for this concept by stressing the significance of the coupling of dissimilar processes. He is also relevant because he discovered a law for such systems of coupled processes: close to equilibrium the dependence of the one process rate on the driving force of the other process should equal the dependence of the other process rate on the

¹Departments of Molecular Cell Physiology and Mathematical Biochemistry, BioCentrum Amsterdam, De Boelelaan 1085, NL-108, HV Amsterdam, the Netherlands. ²Department of Bioengineering, University of California-San Diego, 9500 Gilman Drive, La Jolla, California 92093-0412, USA. Correspondence should be addressed to H.V.W. (hw@bio.vu.nl) or B.O.P. (palsson@ucsd.edu).

Published online 6 October 2004; doi:10.1038/nbt1020

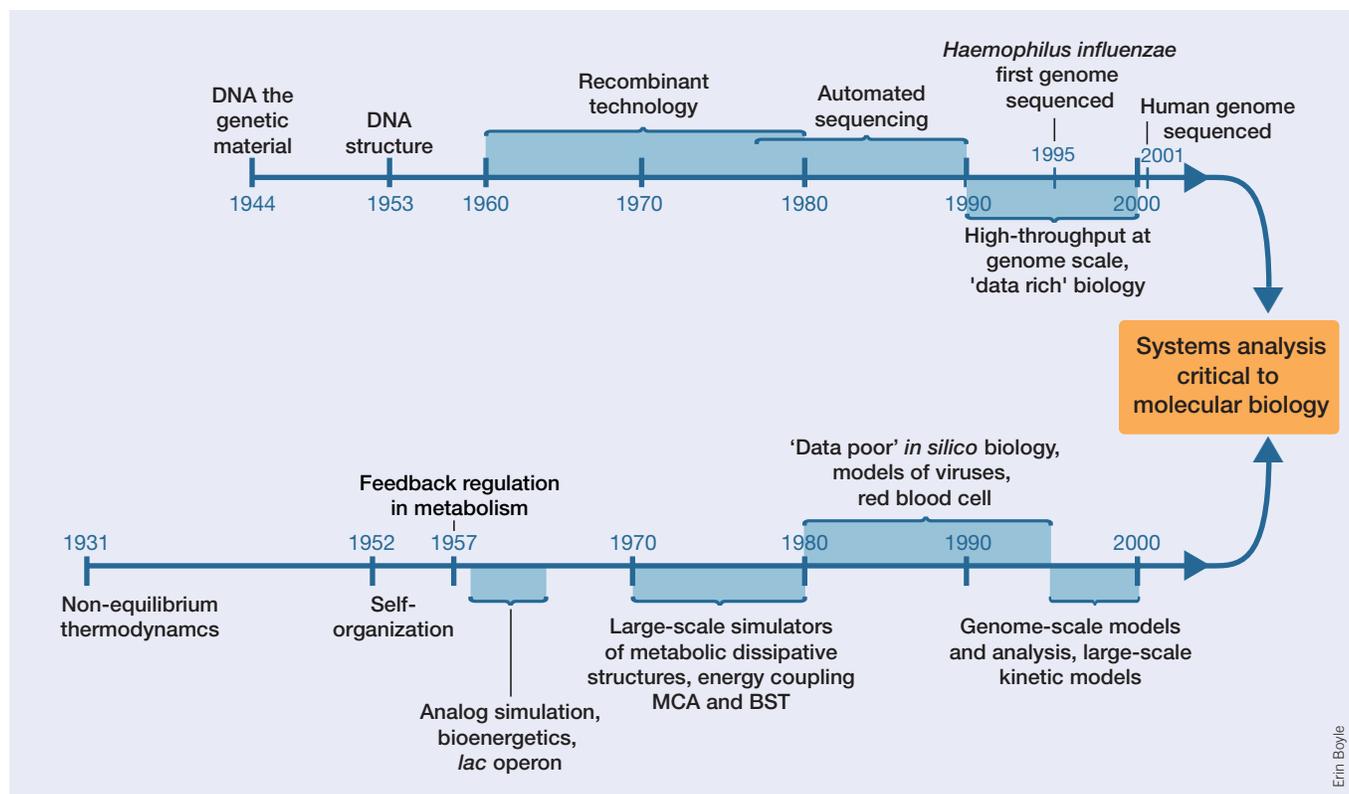


Figure 1 Two lines of inquiry led from the approximate onset of molecular biological thinking to present-day systems biology. The top timeline represents the root of systems biology in mainstream molecular biology, with its emphasis on individual macromolecules. Scaled-up versions of this effort then induced systems biology as a way to look at all those molecules simultaneously, and consider their interactions. The lower timeline represents the lesser-known effort that constantly focused on the formal analysis of new functional states that arise when multiple molecules interact simultaneously.

former driving force. Caplan, Essig and Rottenberg¹² later defined a coupling coefficient, which quantifies the extent to which two processes are coupled in a system and showed that this coefficient must range between 0 and 1.

These approaches were called nonequilibrium thermodynamics and constituted a prelude to systems biology at the cell and molecular levels in that they (i) dealt with integration quantitatively and (ii) aimed to discover general principles rather than just being descriptive. An improved procedure for describing ion movement and energy transduction in biological membranes, termed mosaic nonequilibrium thermodynamics, further progressed towards systems thinking in that it (iii) established a connection to molecular mechanisms and (iv) enabled the determination of the stoichiometry of membrane energy transduction from system data¹³. Peter Mitchell's¹⁴ chemiosmotic coupling principle was another early case of systems analysis in cell and molecular biology. It stated that ATP synthesis was coupled in quite an indirect way to respiration, involving an entire intracellular system, including a volume surrounded by an ion-impermeable membrane and proton movement across it. Indeed, for eukaryotes, this provided much of the rationale for the organization of the mitochondrion. In his calculations verifying that that the proposed chemiosmotic mechanisms transferred sufficient free energy to empower ATP synthesis, Mitchell demonstrated the sort of quantitative thinking that would eventually prove crucial to the study of biochemical systems¹⁴.

The problem of biological self-organization was to understand how structures, oscillations or waves arise in a steady and homogenous

environment, a phenomenon called symmetry breaking. Turing¹⁶ led the way, but the Prigogine school¹⁷ and others developed the topic from the perspective of nonequilibrium thermodynamics in molecular contexts such as biochemical reactions involved in sugar metabolism (glycolysis). They demonstrated how having a sufficient number of nonlinearly interacting chemical processes in a single system such as the Zhabotinski reaction, a developing tissue, or glycolysis, could lead to symmetry-breaking as a result of self-amplification of random fluctuations. Of course, more recent molecular developmental biology studies have shown that reality is even more complicated; pre-specification by external (maternally specified) gradients of morphogens may substitute for the random fluctuations, increasing the robustness of development¹⁸. Perhaps more importantly, Prigogine searched for and found a law (on minimum entropy production). Although it is strictly valid only in Onsager's near-equilibrium domain, it testified to the systems scientists' quest for the principles underlying systems, rather than just for their appearances.

Early on, oscillations in yeast glycolysis were the experimental systems of choice. Although intact cells were studied¹⁹, more often measurements were made using cell extracts²⁰. Reductionist biochemical thinking proclaimed that a single pacemaker enzyme should be responsible for the oscillations. Only relatively recently has systems-based analysis in one of our laboratories (H.V.W.) been used to reveal that the oscillations are simultaneously controlled by many steps in the intracellular network²¹ and how the oscillations in the individual cells synchronize actively²². Of course, with the more recent experimental capability to inspect single cells dynamically, more and more cells are

seen to exhibit asynchronous oscillations of all sorts and some of these cases are up for systems biology analysis. Slime mold aggregation was another early case where a network of reactions was shown to be essential for systems biology reaching one step beyond cell biology, again by combining mathematical modeling with experimental molecular information²³.

Building large-scale models

Following the events of the late 1950s and early 1960s, researchers undertook efforts that were not well publicized and formulated mathematical models to simulate the functions of newly discovered regulatory circuits in cells. Even before digital computers became available, simulations of integrated molecular functions were performed on analog computers²⁴. These efforts grew in scale to dynamic simulation of large metabolic networks in the 1970s^{25–27}. Following the pathway-centered kinetic models in the seventies²⁸, cell-scale flux models of the human red cell were published by the late 1980s (ref. 29), and by the early 1990s genome-scale models of viruses and large-scale models of mitosis were formulated³⁰. With the advent of genome-scale sequencing, the first genome-scale, constraint-based metabolic models for bacteria were constructed³¹. These models describe reconstructed networks and their possible functional states (phenotypes) and are now available at the genome-scale for a growing number of organisms. They treat the 'genome' as the 'system.'

Progress toward the development of detailed kinetic models at a large scale has proven to be slower. Some of these models approach computer replicas of pathways of metabolism, signal transduction and gene expression, and are active on the web, ready for experimentation and integration (compare <http://www.siliconcell.net/>). Obtaining *in vivo* numerical values for kinetic constants remains a key challenge.

Metabolic control analysis

We have agreed that contemporary systems biology has an historical root outside mainstream molecular biology, ranging from basic principles of self-organization in nonequilibrium thermodynamics, through large-scale flux and kinetic models to 'genetic circuit' thinking in molecular biology. 'Systems thinking' differs from 'component thinking' and requires the development of new conceptual frameworks.

Metabolic control analysis (MCA), developed in the early seventies^{28,32}, presented a key example of approaches to characterize properties of networks of interacting chemical reactions. At this time, thinking in biochemistry was dominated by the concept that there had to be a single 'rate-limiting' step at the beginning of all metabolic pathways. Criteria used to establish whether a given enzyme was rate-limiting referred to it as being far from equilibrium, strongly regulated by various metabolic factors or causing pathway flux to decrease when inhibited.

However, the application of these criteria to some metabolic pathways suggested that they contained more than a single rate-limiting step. Network thinking through MCA helped to resolve this paradox. First, mathematical models of metabolic pathways were developed both for inspiration and discovery, and subsequently used to check numerically the principles they conjectured^{28,32}. Second, quantitative definitions were developed to describe the extent to which a step limited the flux through a pathway. This 'flux-control coefficient' of a particular step corresponded to the sensitivity coefficient of the pathway flux with respect to the activity of the particular enzyme. Third, these investigators looked for proof of the concept that there should be a single rate-limiting enzyme in a pathway that should have a flux-control coefficient of unity, with all others having flux control coefficients of

zero. Instead, they found a theorem stating that all the flux-control coefficients must sum to unity^{28,32}. This result then suggested that there need not be a single rate-limiting step to a pathway and that instead many enzymes can contribute simultaneously to the control of the network. Thus, control was not a component property but a network property. The network nature of regulation was shown experimentally to be the case for mitochondrial ATP generation, where control was indeed distributed over more than three steps, and quite notably not particularly strong, neither for the first nor for the irreversible step of the pathway³³.

An important aspect of systems biology is to relate the system properties to the molecular properties of components that comprise a network. The kinetics-based sensitivity analysis by MCA, and its close relative, biochemical systems theory proposed by H.V.W and Chen³⁴, showed that by focusing on the properties of an individual component, one cannot properly decipher its role in the context of a whole network. The connectivity laws proven by MCA^{28,34} (see other references in ref. 35) pinpointed how that distribution of control relates to network structure and the kinetic properties of all network components simultaneously. Similarly, the topological analyses of network structure by our groups^{31,36} have revealed the existence of network-based definitions of pathways that can be used mathematically to represent all possible functional states of reconstructed networks³⁷. Thus, a growing number of methods now exist to analyze the properties mathematically of the large-scale networks that we are now able to reconstruct based on high-throughput data.

Convergence

Figure 1 presents our interpretation of the history of systems analysis in cell and molecular biology. Events in the upper timeline have been much more to the fore of scientific thinking than those in the lower timeline. In one sense, the dazzling stream of discoveries and exciting technologies (most recently with genome-wide data) provides the 'biology' root to contemporary systems biology. In contrast, scientific progress in the lower timeline has never gained much notoriety, although work in this area was much more prominent in European science throughout this period. This latter branch might be thought of as the 'systems' root of systems biology.

Systems modeling and simulation in molecular biology was once seen as purely theoretical and not particularly relevant to understanding 'real' biology. However, now that molecular biology has become such a data-rich field, the need for theory, model building and simulation has emerged. The systems-directed root always had the ambition of discovering fundamental principles and laws, such as those of nonequilibrium thermodynamics and MCA. This ambition should now extend to systems biology.

All too often, the field has been perceived as just pattern recognition and phenomenological modeling. Systems biology is a thorough science with its own quest for scientific principles at the interface of physics, chemistry and biology, with its remarkable mixture of functionality, hysteresis, optimization and physical chemical limitations. *In silico* analysis of complex cellular processes (whether for data description, genetic engineering or scientific discovery), with its focus on elucidating system mechanisms, has in fact become critical for progress in biology.

The historical dichotomy in approaches to molecular biology must now be reconciled with the need to corral resources and expertise in systems approaches. Although the reductionist molecular biological root has been the focus of a plethora of investigations, literature sources and curricula, the same is not true for the systems molecular biology root. There is now a need for development of theoretical and

analytical approaches, curricula and educational materials to advance understanding of the systems in cell and molecular biology. Unknown to many, the 'pre-online PDF' era contains answers to many of the current challenges and pitfalls facing the field. So although systems biology has an intellectually exciting future ahead of it, the leaders in the field should try to minimize rediscovery and focus on the newer challenges facing us, particularly those that come with the application of existing concepts to genome-scale problems and identification of the new issues that arise from the study of cellular functions on this scale.

Where has this history brought us? We now have the growing and general recognition that systems analysis is important to the future evolution of cell and molecular biology. Some reeducation of workers in the field may be in order (<http://www.systembiology.net/>). Over the near term, it is likely that successes with practical applications of systems biology will be confined to unicellular systems. We are now seeing successful applications of systems biology to microbes, including pathway engineering (e.g., see our recent publications^{37,38}), network-based drug design (e.g., H.V.W. and colleagues³⁹), and prediction of the outcome of complex biological processes, such as adaptive evolution (B.O.P. and colleagues⁴⁰). Although the mathematical modeling of whole-body human systems cannot yet be linked to genome-wide data and models, data analysis and modeling are likely to contribute to the success of realizing the goal of individualized medicine. Even if we have to rely on less precise models than the currently available genome-scale models of microorganisms, systems biology may soon lead to better diagnosis and dynamic therapies of human disease than the qualitative methodology presently in use.

ACKNOWLEDGMENTS

We thank Adam Arkin for comments and Timothy Allen for editing. B.O.P. serves on the scientific advisory board of Genomatica, Inc.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at <http://www.nature.com/naturebiotechnology/>

- Umbarger, H.E. & Brown, B. Threonine deamination in *Escherichia coli*. II. Evidence for two L-threonine deaminases. *J. Bacteriol.* **73**, 105–12 (1957).
- Yates, R.A. & Pardee, A.B. Control by uracil of formation of enzymes required for orotate synthesis. *J. Biol. Chem.* **227**, 677–692 (1957).
- Beckwith, J.R. Regulation of the *lac* operon. Recent studies on the regulation of lactose metabolism in *Escherichia coli* support the operon model. *Science* **156**, 597–604 (1967).
- Hunkapiller, T. *et al.* Large-scale and automated DNA sequence determination. *Science* **254**, 59–67 (1991).
- Rowen, L., Magharies, G. & Hood, L. Sequencing the human genome. *Science* **278**, 605–607 (1997).
- Scherf, M., Klingenhoff, A. & Werner, T. Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach. *J. Mol. Biol.* **297**, 599–606 (2000).
- Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
- Ge, H., Walhout, A.J. & Vidal, M. Integrating 'omic' information: a bridge between genomics and systems biology. *Trends Genet.* **19**, 551–560 (2003).
- Palsson, B.O. *In silico* biology through 'omics'. *Nat. Biotechnol.* **20**, 649–650 (2002).
- Schrödinger, E. *What is life? The physical aspects of the living cell*. Based on Lectures Delivered under the Auspices of the Dublin Institute for Advanced Studies at Trinity College, Dublin, in February 1943. (Cambridge University Press, Cambridge, UK, 1944). <http://home.att.net/~p.caimi/oremia.html>
- Onsager, L. Reciprocal relations in irreversible processes. *Phys. Rev.* **37**, 405–426 (1931).
- Rottenberg, H., Caplan, S.R. & Essig, A. Stoichiometry and coupling: theories of oxidative phosphorylation. *Nature* **216**, 610–611 (1967).
- Westerhoff, H.V. & Van Dam, K. Thermodynamics and Control of Biological Free-Energy Transduction (Elsevier, Amsterdam, 1987).
- Mitchell, P. Coupling of phosphorylation to electron and hydrogen transfer by a chemiosmotic type of mechanism. *Nature* **191**, 144–148 (1961).
- Mitchell, P. Chemiosmotic Coupling in Oxidative and Photosynthetic Phosphorylation. (Glynn Research Ltd., Bodmin, UK, 1966).
- Turing, A. The chemical basis of morphogenesis. *Phil. Trans. Roy. Soc. London, Ser. B* **237**, 37–72 (1952).
- Glandsdorff, P. & Prigogine, I. *Structure, Stabilité et Fluctuations* (Masson, Paris, 1971).
- Lawrence, P.A. *The Making of a Fly* (Blackwell, London, 1992).
- Chance, B., Estabrook, R.W. & Ghosh, A. Damped sinusoidal oscillations of cytoplasmic reduced pyridine nucleotide in yeast cells. *Proc. Natl. Acad. Sci. USA* **51**, 1244–1251 (1964).
- Hess, B. & Boiteux, A. Oscillatory phenomena in biochemistry. *Annu. Rev. Biochem.* **40**, 237–258 (1971).
- Teusink, B., Bakker, B.M. & Westerhoff, H.V. Control of frequency and amplitudes is shared by all enzymes in three models for yeast glycolytic oscillations. *Biochim. Biophys. Acta.* **1275**, 204–212 (1996).
- Wolf, J. *et al.* Transduction of intracellular and intercellular dynamics in yeast glycolytic oscillations. *Biophys. J.* **78**, 1145–1153 (2000).
- Tyson, J.J. & Murray, J.D. Cyclic AMP waves during aggregation of *Dictyostelium* amoebae. *Development* **106**, 421–426 (1989).
- Goodwin, B.C. *Oscillatory Organization in Cells, a Dynamic Theory of Cellular Control Processes* (Academic Press, New York, 1963).
- Garfinkel, D. *et al.* Computer applications to biochemical kinetics. *Annu. Rev. Biochem.* **39**, 473–498 (1970).
- Loomis, W. & Thomas, S. Kinetic analysis of biochemical differentiation in *Dictyostelium discoideum*. *J. Biol. Chem.* **251**, 6252–6258 (1976).
- Wright, B.E. The use of kinetic models to analyze differentiation. *Behavioral Sci.* **15**, 37–45 (1970).
- Heinrich, R., Rapoport, S.M. & Rapoport, T.A. *Progr. Biophys. Mol. Biol.* **32**, 1–83 (1977).
- Joshi, A. & Palsson, B.O. Metabolic dynamics in the human red cell. Part I—A comprehensive kinetic model. *J. Theor. Biol.* **141**, 515–528 (1989).
- Novak, B. & Tyson, J.J. Quantitative analysis of a molecular model of mitotic control in fission yeast. *J. Theor. Biol.* **173**, 283–305 (1995).
- Edwards, J.S. & Palsson, B.O. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J. Biol. Chem.* **274**, 17410–17416 (1999).
- Kacser, H. & Burns, J.A. *In Rate Control of Biological Processes* (ed., Davies, D.D.) 65–104 (Cambridge University Press, Cambridge, 1973).
- Groen, A.K., Wanders, R.J.A., Van Roermund, C., Westerhoff, H.V. & Tager, J.M. Quantification of the contribution of various steps to the control of mitochondrial respiration. *J. Biol. Chem.* **257**, 2754–2757 (1982).
- Savageau, M.A. *Biochemical Systems Analysis* (Addison-Wesley, Reading, MA, 1976).
- Westerhoff, H.V. & Chen, Y. How do enzyme activities control metabolite concentrations? An additional theorem in the theory of metabolic control. *Eur. J. Biochem.* **142**, 425–430 (1984).
- Westerhoff, H.V., Hofmeyr, J.H. & Kholodenko, B.N. Getting to the inside of cells using metabolic control analysis. *Biophys. Chem.* **50**, 273–283 (1994).
- Papin, J.A., Price, N.D., Wiback, S.J., Fell, D.A. & Palsson, B.O. Metabolic pathways in the post-genome era. *Trends Biochem. Sci.* **28**, 250–258 (2003).
- Kholodenko, B.N. & Westerhoff, H.V. (eds.) *Metabolic Engineering in the Post Genomics Era* (Horizon Bioscience, UK, 2004).
- Bakker, B.M. *et al.* Network-based selectivity of antiparasitic inhibitors. *Mol. Biol. Rep.* **29**, 1–5 (2002).
- Ibarra, R.U., Edwards, J.S. & Palsson, B.O. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* **420**, 186–189 (2002).

Edward F. Adolph Distinguished Lectureship

Giant sucking sound: can physiology fill the intellectual void left by the reductionists?

Michael J. Joyner

Department of Anesthesiology, Mayo Clinic, Rochester, Minnesota

Submitted 3 May 2011; accepted in final form 31 May 2011

Joyner MJ. Giant sucking sound: can physiology fill the intellectual void left by the reductionists?. *J Appl Physiol* 111: 335–342, 2011. First published June 2, 2011; doi:10.1152/jappphysiol.00565.2011.—Molecular reductionism has so far failed to deliver the broad-based therapeutic insights that were initially hoped for. This form of reductionism is now being replaced by so-called “systems biology.” This is a nebulously defined approach and/or discipline, with some versions of it relying excessively on hypothesis-neutral approaches and only minimally informed by key physiological concepts such as homeostasis and regulation. In this context, physiology is uniquely positioned to continue to provide impressive levels of both biological and therapeutic insight by using hypothesis-driven “classical” approaches and concepts to help frame what might be described as the “pieces of the puzzle” that emerge from molecular reductionism. The strength of physiology as a “bridge” between reductionism and epidemiology, along with its unparalleled ability to generate therapeutic insights and opportunities justifies increased attention and emphasis on our discipline into the future. Arguments relevant to this set of assertions are advanced and this paper, which was based on the 2011 Adolph Lecture, represents an effort to fill the intellectual void left by reductionism and improve scientific progress.

homeostasis; regulation; integrative

THIS PAPER REFLECTS IDEAS that were presented as part of the 2011 Adolph Lecture at the Experimental Biology meeting that was held in Washington, DC. The goal of the talk was to share a physiologist’s perspective on what reductionism in general and the “omic” revolution in particular has or has not done for biomedical research and associated therapeutic insights or advances. The main ideas highlighted in the lecture were the following.

1) Reductionism via various flavors of molecular biology and “omics” has so far failed to deliver its self-promoted revolution in clinical medicine.

2) Systems biology has a cell-centric focus that is marked by a limited understanding of and application to biology beyond the cell.

3) The failure of systems biology to recognize and use key concepts from physiology about homeostasis, regulation, redundancy, feedback control, and acclimation/adaptation are major limitations to this poorly defined approach.

4) While all the attention has been focused on reductionism and more recently systems biology, physiology continues to provide important biomedical insights that lead to therapeutic advances.

As the title demonstrates, my goal in the Adolph Lecture and in this paper was and is to be intentionally provocative and hopefully generate a dialogue with the reductionists. In this context, and because I am “taking sides”, I have adopted what might be called a conversational approach to this paper.

BIOLOGICAL ORTHOPEDIC SURGERY

A key idea or theme that seems to underpin the impetus for reductionism and various flavors of “omics” as applied to biomedical problems might be described as biological orthopedic surgery: “the gene is broken → fix the broken gene → cure the patient.” This thinking clearly seems to explain the enthusiasm about gene therapy that emerged after the discovery of the genetic defect responsible for the most common form of cystic fibrosis and more recently ideas about a limited number of common gene variants explaining the risk for common conditions like atherosclerosis and diabetes (10–12, 43, 51). The line of thinking described above flows from what Denis Noble has critically termed “Neo-Darwinian” thinking about the relationship between genes and phenotype (45, 46). It is exemplified by two quotes, the first from 1989 and second from Francis Collins (the current director of NIH), one of the people involved in the cystic fibrosis gene discovery.

The implications of this research are profound; there will be large spin offs in basic biology, especially cell physiology, but the largest impact will be biomedical (51).

Address for reprint requests and other correspondence: M. J. Joyner, 200 First St. SW, Dept. of Anesthesiology, Mayo Clinic, Rochester, MN 55905 (e-mail: joyner.michael@mayo.edu).

Here we are in 1997, eight years later, and the management of her disease has not changed. . . . But I will predict that in the course of the next 10 years management of CF will change. . . . The healthy form of the gene itself may even be used in so-called gene therapy (12).

What is interesting to note is that while gene therapy for cystic fibrosis has failed to materialize in the 20+ years since the gene defect was identified, there are traditional ion channel-based drugs that target the CFTR protein in clinical trials that show promise in cystic fibrosis (18, 66). At one level, the development of these drugs was likely facilitated by the genetic discoveries because they permitted the development of models that advanced the understanding of the biophysics and ultimately pharmacology of the defective channel. However, one is tempted to speculate, for cystic fibrosis and perhaps other diseases, that much faster therapeutic progress might have been made if traditional physiological and pharmacological approaches had been a bigger area of focus. Perhaps the optimism and drive for gene therapy was an example of what might be termed “silver bullet” thinking that I will discuss below.

REDUCTIONISM IS SEDUCTIVE

The type of reductionism that I have termed “biological orthopedic surgery” has a number of attractive features and is at some level very seductive. It is easy to understand, and when it delivers it is associated with a heroic narrative by a lone scientist or team of scientists making a fundamental discovery that solves a problem. This is the sort of silver bullet thinking mentioned above. However, it has been known for some time that both the easy to understand elements and heroic narratives associated with reductionism are mirages. In this context, when the factors that contribute to biomedical breakthroughs were subjected to analysis by Comroe and Drips (13) in the late 1960s and early 1970s via the “retrospectroscope,” biomedical breakthroughs are in fact more nuanced, incremental, and associated with a more serendipitous view of progress vs. the heroic narrative of reductionism.

HEMOGLOBIN IS A SHIFTY MOLECULE

Homeostasis—the ability to regulate key bodily functions within a narrow range in response to either internal (e.g., exercise) or external (e.g., harsh environmental conditions)—is one of the fundamental (perhaps the fundamental) concept in physiology (7). Homeostasis is also subserved by ideas about regulated systems, feedback control, redundant control mechanisms, and adaptation and acclimation over time. These physiological concepts and mechanisms contribute to what might be described as emergent properties, so that the behavior of the system is far more complex and (and likely more robust) than might be predicted on the basis of a single reductionist property (35).

A good, and early, example of this concept comes from the textbook description about the right shift in the oxygen-hemoglobin dissociation curve that occurs at high altitude or during other forms of hypoxia. The standard teaching is that under these conditions there is a rise in 2–3 DPG that allosterically modifies oxygen-hemoglobin dissociation curve and creates a right shift that facilitates the unloading of oxygen at the tissues. However, when measurements of the oxygen-hemoglobin dissociation curve are made in humans who have traveled to high

altitude (Fig. 1), under many circumstances there is in fact a net left shift in the oxygen hemoglobin dissociation curve. This left shift is facilitated by the rise in pH and fall in CO₂ caused by the hyperventilation driven by systemic hypoxia. Additionally, under some circumstances, it is driven further leftward by a fall in body temperature (68). Furthermore, it is of interest to note that all genetically adapted high altitude animals and the human fetus in the hypoxic intrauterine environment also have left shifted oxygen-hemoglobin dissociation curves, some with P50 values in the teens.

These observations make it seem likely that the main adaptive strategy is to shift the oxygen-hemoglobin dissociation curve to the left to facilitate the “loading” of oxygen at the lung in conditions (altitude) where oxygen availability is limited. This strategy also takes advantage of the fact that the mitochondria in the tissues can work efficiently at very low P_{O₂} values (and that under specific needs such as muscular exercise in hypoxia local increases in [H⁺] and temperature will reduce the leftward shift in muscle capillaries so that “unloading” of oxygen and tissue O₂ levels can be facilitated). It is also of note that the left shift in the oxygen-hemoglobin dissociation curve has been “known” since at least the 1920s. Along these lines, the sequencing of hemoglobin and the understanding of its biophysical properties was one of the earliest triumphs of what has come to be described as molecular biology (55). However, when the interpretation of such discoveries is too narrow, key physiological insights can be missed. The 2–3 DPG story is also an excellent and early example of how physiology trumps reductionist molecular biology as multiple systems and regulatory strategies interact to regulate homeostasis for the whole organism.

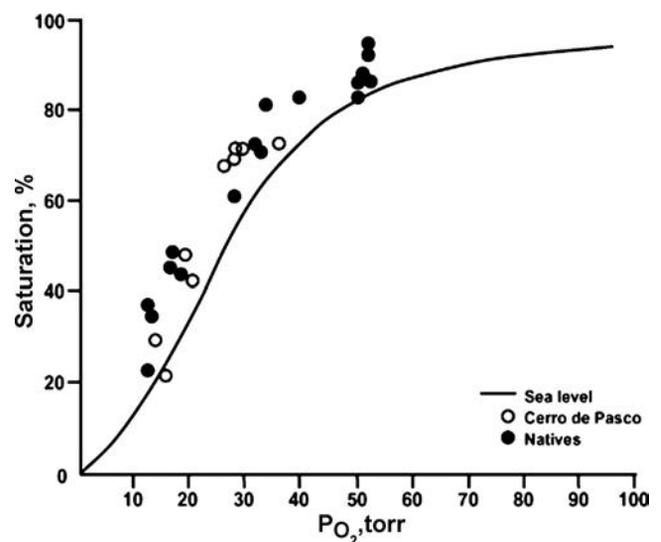


Fig. 1. Oxygen-hemoglobin dissociation curve demonstrating a left shift among sojourners (○) to high altitude and natives. The left shift in the oxygen-hemoglobin dissociation curve under these circumstances demonstrates that the combined effects of hypocapnia, increased pH, and cold override the simple effects of 2–3 DPG on the oxygen-hemoglobin dissociation curve. These data are an outstanding example of the limits of single mechanism reductionism. They are also consistent with the left shift seen in many genetically adapted animals that are native to high altitude. [Reprinted from Ref. 68, with permission from Elsevier.]

PREDICTIVE POWERS OF GENES?

In addition to gene therapy and other molecular treatments for rare diseases, reductionism also made promises about its ability to provide insight about who gets what complex disease like atherosclerosis, diabetes, hypertension, etc. As the quote below demonstrates, this idea became extremely popular after the sequencing of the human genome, and scientific funding agencies like the National Institutes of Health have invested huge sums of money in so-called “genome wide association studies” (GWAS) and other efforts to determine if a few genetic variants are harbingers of future disease in the population as a whole (10, 12, 43).

... because it been known all along that virtually every disease tends to track in families. What has changed is that... we are now beginning to see possible therapeutic approaches based on gene discoveries that will change the way medicine is practiced (12).

One attractive element of this paradigm was that if a few common variants explained much of the risk for disease like diabetes, then it should be possible to identify those at risk and target them for early intervention. So far, the data from many, if not most or even all of these studies, have been underwhelming (43). First, a large number of variants seem to cause a significant increase in risk, but this increase is small compared with behavioral and environmental factors. An increased risk of several percent seems also likely to fall below what might be described as a phenotypic signal-to-noise ratio. Second, when the gene variants (single nucleotide polymorphisms, SNPs) that have been identified via GWAS or other experimental approaches are tested in large populations, the distribution of risk SNPs is typically strikingly similar in populations with and without disease (50, 63; Fig. 2). Third, when so-called genetic risk scores for disease are compared with predictive algorithms based on traditional risk factors (family history, lifestyle, age, etc.), the genetic risk scores are far less predictive than traditional phenotype-based risk scores. Furthermore, addition of genetic risk elements to phenotypically based scores adds little or no additional predictive power (50, 63). Finally, the idea that identifying prospective genetic risks for complex diseases that include a number of lifestyle and environmental factors (and increasingly even prenatal factors) is fundamentally wishful thinking, because behavioral health issues and culture play such a dominant role in determining who gets what disease when, and it is unclear if people will change their behavior in a positive way if they know prospectively they are at increased

risk (24). Paradoxically, perhaps those at reduced genetic risk would pay less attention to behavioral risks.

SUCCESS IN PHARMACOGENOMICS AND ANTHROPOLOGY

So far, this paper has offered a sharp critique of the reductionists and taken the position that they over-sold what their technology had to offer on both the individual (gene therapy) basis and also in terms of population risk and intervention. However, there have been some notable successes stemming from application of this technology and two that seem especially worthy of comment. For example, there has been success in so-called pharmacogenomics. It has been well-known for some time that there are “responders” and “non-responders” to many forms of drug therapy. In many cases, this is related to how rapidly drugs are metabolized. In the case of tamoxifen, which had a dramatic effect on the recurrence of breast cancer, individuals with decreased drug metabolism appear to be at increased risk for recurrence. This is especially important for drugs like tamoxifen, which are ingested as pro-drugs with one or more metabolites that are active (56).

Another field where “omic” approaches have yielded dividends is anthropology. Two good examples include discoveries related to the independent development of lactase persistence into adulthood in areas of the world that were early adopters of herding (23, 34). In this context, one can imagine that the ability to digest lactose into adulthood provided the affected individuals a significant survival advantage and thus became the dominant genotype in only a few generations. Another good example that is perhaps counterintuitive relates to the individuals who migrated to the Tibetan plateau. These individuals do not develop chronic mountain sickness even with lifelong living at 3–4,000 m of elevation. These responses contrast to the high altitude natives in the Andes Mountains, who do develop chronic mountain sickness (58, 61, 70). Along these lines, those who migrated to the Tibetan plateau appear to have had selection pressure that favored a less functional variant of the hypoxia-inducible factor that, among other things, prevents them from developing excessive polycythemia, which plays a critical role in chronic mountain sickness.

INTERIM SUMMARY

So far, I have provided a general critique of what might broadly be termed “molecular reductionism”. I have presented evidence that its failure to live up to its self-generated hype is in reality a failure to recognize larger ideas about homeostasis

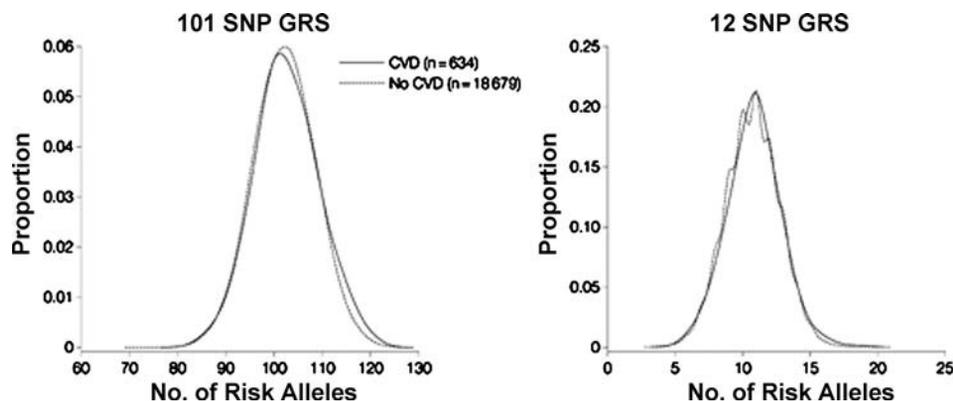


Fig. 2. Distribution of so-called high risk genes for cardiovascular disease in women with and without known coronary artery disease. The distribution of risk genes is similar, and construction of a genetic risk score for cardiovascular disease is thus problematic. This is just one example of the limited predictive power of “genomics” as it relates to the ability of relatively common gene variants to predict common diseases. [Borrowed with permission from Ref. 50. Copyright © 2010 American Medical Association. All rights reserved.]

and regulation that are central to physiology. This includes the specific example of the idea of gene therapy for relatively common genetic disorders like cystic fibrosis and also the limited predictive power of gene variants for common diseases. The question now is whether there is some way out of this problem and a better way to use potentially powerful technologies championed by the reductionists in a biomedical context.

IS SYSTEMS BIOLOGY THE ANSWER?

One idea to address the “failure” of molecular reductionism described above is to use a new approach called systems biology. The idea is that if powerful modeling tools and other data analysis techniques could be applied to the data generated via high throughput molecular reductionism, then somehow more meaningful insights would be generated and ultimately exploited for predictive or therapeutic purposes. The rationale for systems biology comes from a sampling of the comments on www.systemsbiology.org web site (34a).

Systems biology is the study of an organism, viewed as an integrated and interacting network of genes, proteins and biochemical reactions which give rise to life. Instead of analyzing individual components or aspects of the organism, such as sugar metabolism or a cell nucleus, systems biologists focus on all the components and the interactions among them, all as part of one system. These interactions are ultimately responsible for an organism's form and functions.

Traditional biology—the kind most of us studied in high school and college, and that many generations of scientists before us have pursued—has focused on identifying individual genes, proteins and cells, and studying their specific functions. But that kind of biology can yield relatively limited insights about the human body.

Biologists, geneticists, and doctors have had limited success in curing complex diseases such as . . . diabetes because traditional biology generally looks at only a few aspects of an organism at a time.

To a physiologist, there are obvious problems with systems biology. The problems start with the fact that physiology has been attempting for hundreds of years to understand the integrated function of organs and whole organisms that culminated in unifying big ideas about homeostasis and regulation discussed earlier. It is also clear that the type of biology that physiologists have been interested in starting with Harvey and the circulation has been about systems and has used modeling and computational techniques (1, 32, 57). Additionally, at this time the concept of systems biology and how it is defined remains very nebulous (52). Is systems biology a new discipline, an approach, a collection of tools, or merely a new name for integrative physiology generated by individuals who are generally unaware that our field exists (2, 28, 34a, 36, 40, 41, 45, 57)? Clearly physiology has provided and continues to provide insight about human disease, including insight that has led to vast therapeutic advances in recent years (37). Perhaps, the obvious question for the advocates of the cell-centric view of systems biology is did they skip physiology as part of their course work as students?

The concerns about systems biology outlined above at some level are about definitions and perhaps intellectual ownership. However, it also seems fair to ask what the long-term outlook for cell-centric systems biology is as an approach to making sense out of the vast amounts of data that can be generated

using modern “omic” technology. In this context, there are key intellectual issues related to how data elements are generated, their spatial and temporal relationships, and how many ways they might interact (Fig. 3) that question the very fundamental assumptions about systems biology and its reliance on “bottom up” or “hypothesis neutral” modeling (2, 6, 15, 27, 35, 36, 38, 48, 67). It seems to me that without a narrative approach that includes hypothesis testing and key concepts like homeostasis, systems biology runs the risk of becoming scientific “Abstract Expressionism”. Given the issues discussed earlier with gene therapy and GWAS approaches and the hype that surrounds systems biology, these concerns raise questions about what kind of science and scientific approaches deserve our future attention and funding (2, 24, 35).

REDUCTIONISM STALLS PHYSIOLOGY PROGRESSES

This is not the place for a comprehensive review of the contributions of physiology to biomedical research and therapeutic progress over the last 20–30 years. However, a few highlights that were initially seen as counterintuitive seem warranted. An obvious one is the discovery of EDRF and nitric oxide (25). This observation, which challenged the idea of the endothelium as merely a barrier, led to the discovery of gas-based signaling mechanisms and new therapeutic targets for conditions as diverse as erectile dysfunction and pulmonary hypertension. Would gas-based signaling mechanisms have been discovered by sequencing genes? Physiology has also helped redefine the optimal strategy used during mechanical ventilation in patients with adult respiratory distress syndrome (ARDS; 26). This has led to abandonment of strategies associated with high airway pressures and maintenance of arterial blood gases toward so-called permissive hypercapnia, alternate forms of mechanical ventilation and pressure support. Importantly, these new strategies that emphasize the avoidance of barotrauma have been associated with significant reductions in morbidity and mortality for ARDS. While part of the conventional wisdom now, this strategy was initially seen as counterintuitive.

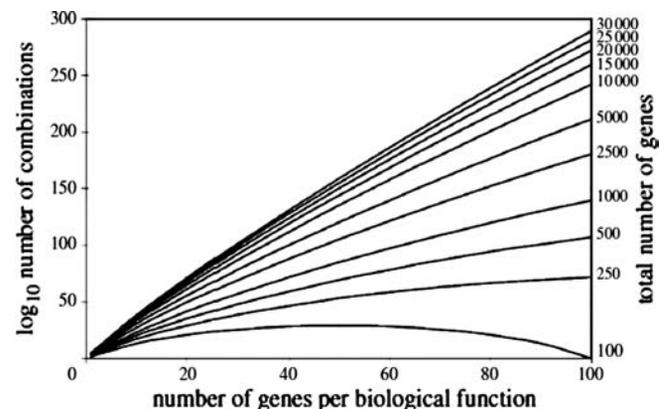


Fig. 3. Simulation of a number of possible combinations of genes gene interactions depending on the number of genes per biological function (x-axis) and the total number of genes in the organism. For biological functions with roughly 50 genes, $\sim 10^{150}$ possible combinations exist for most mammals. This figure shows the immense challenge associated with hypothesis-neutral systems biology and “bottom up” modeling. [Borrowed with permission from Ref. 46.]

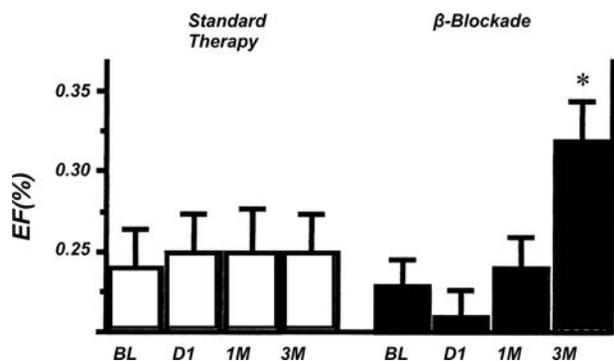


Fig. 4. Demonstration that beta-blockade can improve ventricular function (%EF) in humans with congestive heart failure over time. Standard therapy was associated with stable ventricular ejection fraction over 3 mo. By contrast, metoprolol (β -blockade) increased ventricular ejection fraction by $>50\%$ over 3 mo ($* < 0.05$ vs. baseline). This finding, while initially counterintuitive, was based on sound physiological reasoning and along with other therapies has improved outcomes for patients with congestive heart failure. [Adapted from Ref. 20, with permission from Wolters Kluwer Health.]

Another example of a counterintuitive physiologically based clinical strategy was the use of beta-blockers in congestive heart failure. For many years these drugs were contraindicated in congestive heart failure (CHF) because it was felt that high sympathetic drive to the heart was required to maintain an adequate cardiac output in CHF. In reality, high sympathetic activity to the heart over time contributed to the progression of the disease and promoted a downward spiral of cardiac remodeling and reduced function (20). Thus the use of beta-blockers along with vasodilator therapy has been revolutionary and can interrupt or slow the downward spiral noted above in patients with congestive heart failure (Fig. 4). Again, the conventional wisdom was turned on its head and provided new insights that ultimately led to improved therapy. In the case of ARDS and congestive heart failure there has also been a two-way street between observations from clinical research conducted “at the bedside” to more fundamental observations in the laboratory.

Three other examples of more straight forward physiologically based therapeutic successes in recent years include the long story of improved outcomes for premature infants cared for in the neonatal ICU including altered ventilatory strategies, avoidance of oxygen toxicity, and surfactant therapy (9, 60). These improved outcomes, in the littlest ICU survivors, continue to seem miraculous to individuals who care for these patients and practiced medicine or nursing prior to their use. A second example has been oral rehydration solutions that are life saving in infants and children with diarrheal disease, especially in developing countries where it is a primary and frequent cause of death (8). Finally, in the developed world, where obesity and physical inactivity are leading to a pandemic of type 2 diabetes, physical activity (especially walking training in middle-aged people) has been proven to be highly effective in preventing, limiting, and in some cases reversing type 2 diabetes (16, 29). Each of these therapeutic successes is based on a foundation of physiologically based experimental evidence and insights.

REDUNDANCY, FEEDBACK, AND ACCLIMATION/ADAPTATION

Why has physiology continued to contribute in the era of reductionism? Physiologists are well versed in the overall

concept of homeostasis, regulation, feedback, redundancy, and acclimation/adaptation. A classic example of redundancy comes from coronary circulation where coronary vasodilation is tightly linked to myocardial oxygen demand. In this context, a number of vasodilator systems likely contribute to this response. However, pharmacological blockade of one system, or in fact multiple systems, fails to alter this fundamental relationship between coronary vasodilation and myocardial oxygen demand in most species (19, 64; Fig. 5) This suggests that multiple redundant pathways contribute to this critical physiological response so that when one is blocked or absent, oxygen supply to the heart is not threatened when demand rises.

The fundamental relationship between coronary vasodilation and myocardial oxygen demand is also an observation that has had vast therapeutic implications and explains in large part why age specific death rates for cardiovascular disease have fallen dramatically over the last 30–40 years. There are drugs that reduce myocardial oxygen demand, mechanical therapy like stents, bypass surgery that improves myocardial oxygen delivery, and other drugs and lifestyle interventions that can affect both elements of the equation over time (30, 44). This physiological narrative and the progress that has flowed from it is in stark contrast to the relative lack of progress against cancer where there does not seem to be a unifying physiologically based story or model that can be exploited to address the general problem of cancer.

One of the classic feedback control mechanisms in physiology is the arterial baroreflex. While barodenervated animals have relatively normal blood pressure over a given 24 h period, their blood pressure becomes much more variable (14). The relative stability of blood pressure in the long run shows the power of redundant control via renal regulation of arterial pressure. However, for short-term adaptations, essential for things like exercise or changes in posture, feedback control

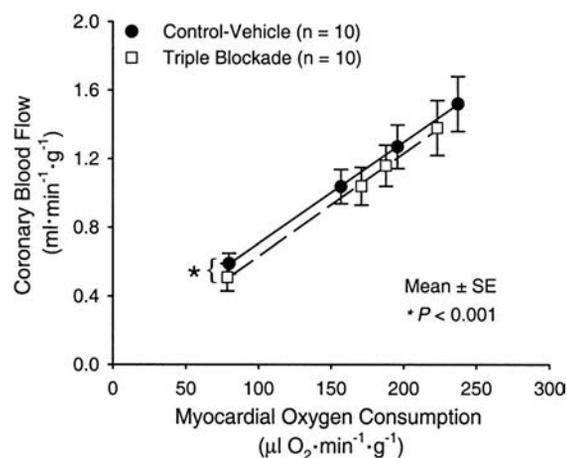


Fig. 5. Myocardial oxygen demand on the x-axis and coronary blood flow on the y-axis. Note that coronary blood flow rises in proportion to myocardial oxygen demand and that this rise is unaffected by triple inhibition of kATP^+ channels, nitric oxide synthase, and adenosine receptors. This is a classic example of the concept of physiological redundancy. This well-known phenomenon may also explain why the absence of many so-called critical genes or proteins has limited impact on overall organ or organism function. This is because so-called redundant systems are able to alter their function and “upregulate” when one or more systems is blocked. [Borrowed with permission from Ref. 64.]

from arterial baroreflexes is essential for normal physiological responses.

An outstanding example of how humans acclimatize and adapt to physiological stress comes from studies that demonstrate that the ability of individuals to exercise in the heat can be remarkably improved by a few weeks of training in the heat (54). This improved exercise tolerance in the heat is associated with expanded plasma volume, increased sweating, and altered thermoregulatory skin blood flow. Another outstanding example is what might be called the adaptability of insulin sensitivity and glucose uptake in skeletal muscle. These variables are extremely sensitive to exercise and changes in daily activity and seem especially relevant in the era of the physical inactivity/obesity pandemic (29, 49, 53, 65).

Ideas about redundancy, feedback control, and acclimation/adaptation are also why physiologists are not that surprised by the ability of various gene knockout animals to survive and thrive (33). At some level this approach is conceptually similar to the classic denervation or high dose pharmacological blockade studies used by physiologists for generations and primarily show the power of the regulatory mechanisms highlighted above to preserve both long term phenotype and homeostasis despite the loss of one or more critical pathways or mechanisms (17). In this context, it is not surprising the yeast can survive without 80% of their genes and the function of these genes only becomes apparent when the organism is stressed (33). Is it too cynical to point out that knockout animals are essentially a “can’t lose” experimental approach? If the knockout is lethal or leads to significant phenotypic dysfunction it is essential. If it survives then genetic or other compensatory mechanisms were upregulated to overcome the absence of the essential gene.

Physiology or physiologically based tests can also provide insight into the risk of future disease and/or predictive outcomes. For example, the blood pressure responses to common sympathoexcitatory stress can be used to define those at risk for future hypertension in a way that is potentially much more predictive than any current genetic test. Additionally, tests of autonomic function are strong predictors of outcomes in large populations of humans, and cardiorespiratory fitness is an especially good predictor of all-cause mortality.

TOOLS VS. BIG IDEAS

At some level molecular reductionism and systems biology are at existential cross roads. Are they in fact real disciplines informed by big ideas like homeostasis and regulation, or are they essentially tools and approaches that will facilitate the work of disciplines informed by bigger ideas and more importantly bigger questions and more comprehensive strategies? Based on the concepts and examples highlighted in this paper I would argue that until the vast amounts of data generated by modern “omic” techniques are put in a physiological context it will be an exercise in what Sydney Brenner has deemed “low input, high throughput, no output biology” (6). Along these lines, I want to end on an optimistic note with examples of how physiology is making a difference by applying reductionist tools as part of a more comprehensive approach to important questions. Because the Adolph lecture is sponsored by the Exercise and Environmental Physiology section of the American Physiological Society, relevant examples from related

areas will be used. In each case there seems to be an overall hypothesis and a strategy that exploits what might be called responders and non-responders to an intervention.

Britton and Koch and colleagues (39, 69) have used selective breeding strategies to develop rats with vastly different inherent aerobic endurance capacities (Fig. 6). These animals have been used in a variety of studies to better understand the gene environment interactions. In many instances the animals selected for low intrinsic aerobic capacity seem to be at increased risk for complex diseases like diabetes, obesity, and heart disease. Additionally, studies using these animals have begun to identify genetic and transcriptional factors and networks that explain in part this increased risk (39).

Another example of how physiologists are using tools from the “new biology” is the HERITAGE study, which broadly seeks to understand the genetic basis for the differing physiological responses to exercise training in a large number of humans exposed to a standard protocol (3–5). This is an excellent example of how what might be called “high resolution” physiologically based phenotyping in conjunction with genetics. This hypothesis-driven approach also includes uses

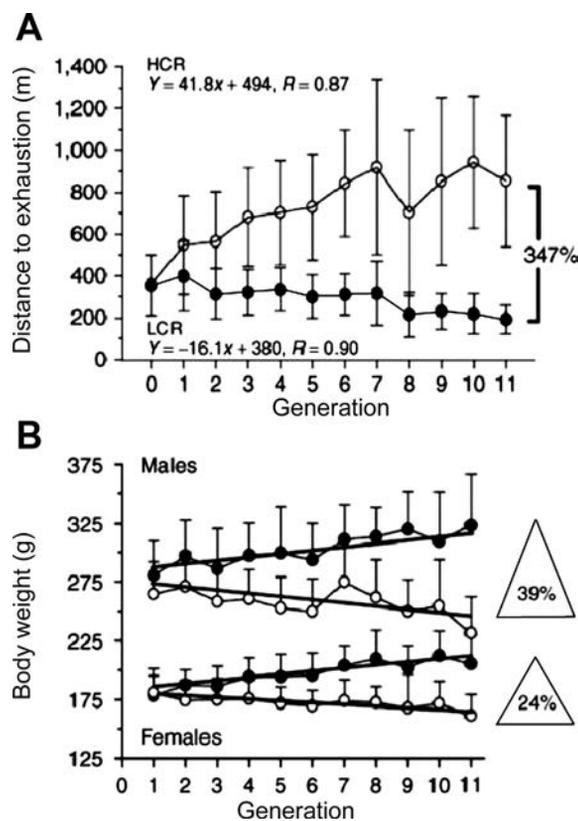


Fig. 6. Selective breeding of rats with divergent aerobic capacities. These data show that animals selected for their running capacity diverge dramatically after a few generations and is sustained for many generations. Importantly, at the same time body weight also began to diverge as did a number of risk factors for cardiometabolic disease. Phenotypic studies conducted on these animals in conjunction with more targeted forms of “omic” approaches and other types of molecular reductionism are providing new insights about gene environment interactions. These findings may also have applicability to physically active and inactive humans. The approach of Britton and Koch is a classic example of using reductionist tools in a physiological context to gain new insights with direct applicability to human health and disease. [Reprinted from Ref. 39 with permission from Macmillan Publishers Ltd. *Obesity Suppl.* copyright 2008.]

various “omic” and systems biology approaches and was initiated by physiologists before the terms genomics or systems biology existed. Additionally, like the examples from pharmacogenomics and anthropology discussed earlier, it takes advantage of the fact that there are responders and non-responders in response to a given intervention or environmental stressor.

Finally, my collaborator John Eisenach and I along with our colleagues have performed carefully controlled studies on how common genetic variants in the β_2 -adrenergic receptor influence a number of physiological responses and how any genotype-based differences might be influenced by dietary sodium (21, 22, 31, 59). These studies were initiated because epidemiological evidence suggested that genetic variation in the β_2 -adrenergic receptor influenced blood pressure in large populations. In our studies only homozygotes for the genetic variant of interest were recruited in an effort to see the maximum potential physiological effect of the variants. Using this approach, it appears that there are genotype-specific patterns associated with increased cardiac output responses to exercise that may interact with NO-mediated β_2 -adrenergic receptor peripheral vasodilation. These responses clearly link and mechanistically define how a common gene variant in a key regulatory system can influence a physiological response in humans. They may also provide physiological explanations relevant to the original epidemiological observations on blood pressure and other outcomes, including those in patients with the acute coronary syndrome (42).

SUMMARY

In this paper and in the Adolph Lecture I have highlighted some of the claims associated with molecular reductionism and more recently systems biology. In both cases I have argued that the apparent inability and/or unwillingness of the advocates of these approaches to use key concepts from physiology and ultimately use their tools in a physiological context has limited the contribution of the approaches they advocate. By contrast physiology has continued to use new tools in the service of its big ideas and also continued to provide biomedical insight and therapeutic advances. As the final examples show, it is possible to incorporate reductionist tools in a physiological context to gain broader biomedical insights. Hopefully these insights will fuel the next wave of physiologically inspired therapeutic advances.

ACKNOWLEDGMENTS

The author thanks Drs. Nisha Charkoudian, Doug Seals, and Jerry Dempsey for critical reviews of the manuscript.

GRANTS

My laboratory has been funded continuously by National Institutes of Health since the early 1990s (HL-46493, HL-83947), and our work has also been facilitated by the former Mayo GCRC grant and more recently the CTSA grant (RR-024150). I have also been supported by the American Heart Association, the Mayo Foundation, and the Frank and Shari Caywood Professorship. I would also like to thank the fellows and collaborators who have worked with me, my outstanding technical support staff, and the many subjects who have volunteered for our studies.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author.

REFERENCES

1. **Auffray C, Noble D.** Origins of systems biology in William Harvey's masterpiece on the movement of the heart and the blood in animals. *Int J Mol Sci* 10: 1658–1669, 2009.
2. **Beard DA, Kushmerick MJ.** Strong inference for systems biology. *PLoS Comput Biol* 5: 1–10, 2009.
3. **Bouchard C, An P, Rice T, Skinner JS, Wilmore JH, Gagnon J, Perusse L, Leon AS, Rao D.** Familial aggregation of $\dot{V}O_{2\max}$ response to exercise training: results from the HERITAGE Family Study. *J Appl Physiol* 87: 1003–1008, 1999.
4. **Bouchard C, Leon AS, Rao DC, Skinner JS, Wilmore JH, Gagnon J.** Aims, design, and measurement protocol. *Med Sci Sports Exerc* 27: 721–729, 1995.
5. **Bouchard C, Sarzynski MA, Rice TK, Kraus WE, Church TS, Sung YJ, Rao DC, Rankinen T.** Genomic predictors of maximal oxygen uptake response to standardized exercise training programs. *J Appl Physiol* 110: 1160–1170, 2011.
6. **Brenner S.** Sequences and consequences. *Phil Trans R Soc B* 365: 207–212, 2010.
7. **Cannon WB.** Organization for physiological homeostasis. *Physiol Rev* 9: 399–431, 1929.
8. **Cheng AC, McDonald JR, Thielman NM.** Infectious diarrhea in developed and developing countries. *J Clin Gastroenterol* 39: 757–773, 2005.
9. **Clements JA, Avery ME.** Lung surfactant and neonatal respiratory distress syndrome. *Am J Respir Crit Care Med* 157: S59–S66, 1998.
10. **Collins FS.** Contemplating the end of the beginning. *Genome Res* 11: 641–643, 2001.
11. **Collins FS.** Cystic fibrosis: molecular biology and therapeutic implications. *Science* 256: 774–779, 1992.
12. **Collins FS.** The human genome project and the future of medicine. *Ann NY Acad Sci USA* 882: 42–55, 1999.
13. **Comroe JH Jr, Dripps RD.** Ben Franklin and open heart surgery. *Circ Res* 35: 661–669, 1974.
14. **Cowley AW Jr, Liard JF, Guyton AC.** Role of the baroreceptor reflex in daily control of arterial blood pressure and other variables in dogs. *Circ Res* 32: 564–576, 1973.
15. **Csete ME, Doyle JC.** Reverse engineering of biological complexity. *Science* 295: 1664–1669, 2002.
16. **Diabetes Prevention Program Research Group.** Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 346: 393–403, 2002.
17. **Donald DE, Milburn SE, Shepherd JT.** Effect of cardiac denervation on the maximal capacity for exercise in the racing greyhound. *J Appl Physiol* 19: 849–852, 1964.
18. **Donaldson SH, Boucher RC.** Sodium channels and cystic fibrosis. *Chest* 132: 1631–1636, 2007.
19. **Duncker DJ, Bache RJ.** Regulation of coronary blood flow during exercise. *Physiol Rev* 88: 1009–1086, 2008.
20. **Eichhorn EJ, Bristow MR.** Medical therapy can improve the biological properties of the chronically failing heart. *Circulation* 94: 2285–2296, 1996.
21. **Eisenach JH, Barnes SA, Pike TL, Sokolnicki LA, Masuki S, Dietz NM, Rehfeldt KH, Turner ST, Joyner MJ.** Arg 16/Gly beta-2 adrenergic receptor polymorphism alters the cardiac output response to isometric exercise. *J Appl Physiol* 99: 1776–1781, 2005.
22. **Eisenach JH, McGuire AM, Schwinger RM, Turner ST, Joyner MJ.** The Arg16/Gly β_2 adrenergic receptor polymorphism is associated with altered cardiovascular responses to isometric exercise. *Physiol Genomics* 16: 323–328, 2004.
23. **Enattah NS, Jensen TGK, Nielsen M, Lewinski R, Kuokkanen M, Rasinpera H, El-Shanti H, Seo JK, Alifrangis M, Khalil IF, Natah A, Ali A, Natah S, Comas D, Mehdi SQ, Groop L, Vestergaard EM, Imtiaz F, Rashed MS, Meyer B, Troelsen J, Peltonen L.** Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am J Hum Genet* 82: 57–72, 2008.
24. **Evans JP, Meslin EM, Marteau TM, Caulfield T.** Genomics. Deflating the genomic bubble. *Science* 331: 861–862, 2011.
25. **Furchgott RF.** The discovery of endothelium-derived relaxing factor and its importance in the identification of nitric oxide. *JAMA* 276: 1186–1188, 1996.

26. Gillette MA, Hess DR. Ventilator-induced lung injury and the evolution of lung-protective strategies in acute respiratory distress syndrome. *Respir Care* 46: 130–148, 2001.
27. Golub TR. Counterpoint: Data first. *Nature* 464: 679, 2010.
28. Greenhaff PL, Hargreaves M. “Systems biology” in human exercise physiology: is it something different from integrative physiology? *J Physiol* 589: 1031–1036, 2011.
29. Henriksen EJ. Exercise effects of muscle insulin signaling and action. Invited review: effects of acute exercise and exercise training on insulin resistance. *J Appl Physiol* 93: 788–796, 2002.
30. Heron M, Hoyert DL, Murphy SL, Xu J, Kochanek KD, Tejada-Vera B. Deaths: Final data for 2006. *National Vital Statistics Reports*; vol 57, no 14. Hyattsville, MD: National Center for Health Statistics, 2009, pp 1–117.
31. Hesse C, Schroeder DR, Nicholson WT, Hart EC, Curry TB, Penheiter AR, Turner ST, Joyner MJ, Eisenach JH. Beta-2 adrenoceptor gene variation and systemic vasodilatation during ganglionic blockade. *J Physiol* 588: 2669–2678, 2010.
32. Hester RL, Ilescu R, Summers R, Coleman TG. Systems biology and integrative physiological modelling. *J Physiol* 589: 1053–1060, 2011.
33. Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St. Onge RP, Tyers M, Koller D, Altman RB, Davis RW, Nislow C, Giaever G. The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* 320: 362–365, 2008.
34. Ingram CJE, Mulcare CA, Itan Y, Thomas MG, Swallow DM. Lactose digestion and the evolutionary genetics of lactase persistence. *Hum Genet* 124: 579–591, 2009.
- 34a. Institute for Systems Biology. (Online). <http://www.systemsbio.org> [June 22, 2011].
35. Joyner MJ, Pedersen BK. Ten questions about systems biology. *J Physiol* 589: 1017–1030, 2011.
36. Joyner MJ. Physiology: alone at the bottom, alone at the top. *J Physiol* 589: 1005, 2011.
37. Joyner MJ. Why physiology matters in medicine. *Physiology* 26: 72–75, 2011.
38. Kell DB, Oliver SG. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *Bioessays* 26: 99–105, 2004.
39. Koch LG, Britton SL. Development of animal models to test the fundamental basis of gene-environmental interactions. *Obesity* 16, Suppl 3: S28–S32, 2008.
40. Kohl P, Crampin EJ, Quinn TA, Noble D. Systems biology: an approach. *Clin Pharmacol Ther* 88: 25–33, 2010.
41. Kuster DWD, Merkus D, Van Der Velden J, Verhoeven AJM, Duncker DJ. Integrative Physiology 2.0: integration of systems biology into physiology and its application to cardiovascular homeostasis. *J Physiol* 589: 1037–1045, 2011.
42. Lanfear DE, Jones PJ, Marsh S, Cresci S, McLeod HL, Spertus JA. Beta2-adrenergic receptor genotype and survival among patients receiving beta-blocker therapy after an acute coronary syndrome. *JAMA* 294: 1526–1533, 2005.
43. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TFC, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature* 461: 747–753, 2009.
44. Nelson RR, Gobel FL, Jorgensen CR, Wang K, Wang Y, Taylor HL. Hemodynamic predictors of myocardial oxygen consumption during static and dynamic exercise. *Circulation* 50: 1179–1189, 1974.
45. Noble D. New-Darwinism, the Modern Synthesis and selfish genes: are they of use in physiology? *J Physiol* 589: 1007–1015, 2011.
46. Noble D. Differential and integral views of genetic in computational systems biology. *Interface Focus* 1: 7–15, 2011.
47. Noble D. Genes and causation. *Phil Trans R Soc A* 366: 3001–3015, 2008.
48. Nurse P, Hayles J. The cell in an era of systems biology. *Cell* 144: 850–854, 2011.
49. Olsen RH, Krogh-Madsen R, Thomsen C, Booth FW, Pedersen BK. Metabolic responses to reduced daily steps in healthy nonexercising men. *JAMA* 299: 1261–1263, 2008.
50. Paynter NP, Chasman DI, Pare G, Buring JE, Cook NR, Miletich JP, Ridker PM. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* 303: 631–637, 2010.
51. Pearson H. One gene, twenty years. *Nature* 460: 165–169, 2009.
52. Powell A, O’Malley MA, Muller-Wille S, Calvert J, Dupre J. Disciplinary baptisms: a comparison of the naming stores of genetics, molecular biology, genomics, and systems biology. *Hist Phil Life Sci* 29: 5–32, 2007.
53. Rogers MA, Yamamoto C, King DS, Hagberg JM, Ehsani AA, Holloszy JO. Improvement in glucose tolerance after 1 wk of exercise in patients with mild NIDDM. *Diabetes Care* 11: 613–618, 1988.
54. Robinson S, Turrell ES, Belding HS, Horvath SM. Rapid acclimatization to work in hot climates. *Am J Physiol* 140: 168–176, 1943.
55. Rossmann MG. Chapter 3: Recollection of the events leading to the discovery of the structure of haemoglobin. *J Mol Biol* 392: 23–32, 2009.
56. Schroth W, Goetz MP, Hamann U, Fasching PA, Schmidt M, Winter S, Fritz P, Simon W, Suman VJ, Ames MM, Safgren SL, Kuffel MJ, Ulmer HU, Bolander J, Strick R, Beckmann MW, Koelbl H, Weinsilbum RM, Ingle JN, Eichelbaum M, Schwab M, Brauch H. Association between CYP2D6 polymorphisms and outcomes among women with early stage breast cancer treated with tamoxifen. *JAMA* 302: 1429–1436, 2009.
57. Secomb TW, Pries AR. The microcirculation: physiology at the mesoscale. *J Physiol* 589: 1047–1042, 2011.
58. Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, Bai Z, Lorenzo FR, Xing J, Jorde LB, Prchal JT, Ge R. Genetic evidence for high-altitude adaptation in Tibet. *Science* 329: 72–75, 2010.
59. Snyder EM, Beck KC, Dietz NM, Eisenach JH, Joyner MJ, Turner ST, Johnson BD. Arg16Gly polymorphism of the β_2 -adrenergic receptor is associated with differences in cardiovascular function at rest and during exercise in humans. *J Physiol* 571: 121–130, 2006.
60. Sol RF. Current trials in the treatment of respiratory failure in preterm infants. *Neonatology* 95: 368–372, 2009.
61. Storz JF. Genes for high altitudes. *Science* 329: 40–41, 2010.
63. Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, Brunner EJ, Kumari M, Kivimaki M, Humphries SE. Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. *Br Med J* 340: b4838, 2010.
64. Tune JD, Richmond KN, Gorman MW, Feigl EO. K(ATP)(+) channels, nitric oxide, and adenosine are not required for local metabolic coronary vasodilation. *Am J Physiol Heart Circ Physiol* 280: H868–H875, 2001.
65. van Dieren S, Beulens JW, van der Schouw YT, Grobbee DE, Neal B. The global burden of diabetes and its complications: an emerging pandemic. *Eur J Cardiovasc Prev Rehabil* 17, Suppl 1: S3–S8, 2010.
66. Verkman AS, Galiotta LJV. Chloride channels as drug targets. *Nat Rev* 8: 153–171, 2009.
67. Weinberg RA. Point: Hypotheses first. *Nature* 464: 678, 2010.
68. Winslow RM. The role of hemoglobin oxygen affinity in oxygen transport at high altitude. *Respir Physiol Neurobiol* 158: 121–127, 2007.
69. Wisloff U, Najjar SM, Ellingsen O, Haram PM, Swoap S, Al-Share Q, Fernstrom M, Rezaei K, Lee SJ, Koch LG, Britton SL. Cardiovascular risk factors emerge after artificial selection for low aerobic capacity. *Science* 307: 418–420, 2005.
70. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliusen TS, Zheng H, Liu T, He W, Li K, Luo R, Nie X, Wu H, Zhao M, Cao H, Zou J, Shan Y, Li S, Yang Q, Asan Ni P, Tian G, Xu J, Liu X, Jiang T, Wu R, Zhou G, Tang M, Qin J, Wang T, Feng S, Li G, Huasang Luosang J, Wang W, Chen F, Wang Y, Zheng X, Li Z, Bianba Z, Yang G, Wang X, Tang S, Gao G, Chen Y, Luo Z, Gusang L, Cao Z, Zhang Q, Ouyang W, Ren X, Liang H, Zheng H, Huang Y, Li J, Bolud L, Kristiansen K, Li Y, Zhang Y, Zhang X, Li R, Li S, Yang H, Nielsen R, Wang J, Wang J. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329: 75–78, 2010.

REVIEW

Systems biology: perspectives on multiscale modeling in research on endocrine-related cancers

Robert Clarke¹, John J Tyson², Ming Tan³, William T Baumann⁴, Lu Jin¹, Jianhua Xuan⁵ and Yue Wang⁵

¹Department of Oncology, Georgetown University Medical Center, Washington, District of Columbia, USA

²Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA

³Department of Biostatistics, Bioinformatics & Biomathematics, Georgetown University Medical Center, Washington, District of Columbia, USA

⁴Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA

⁵Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Arlington, Virginia, USA

Correspondence should be addressed to R Clarke: clarker@georgetown.edu

Abstract

Drawing on concepts from experimental biology, computer science, informatics, mathematics and statistics, systems biologists integrate data across diverse platforms and scales of time and space to create computational and mathematical models of the integrative, holistic functions of living systems. Endocrine-related cancers are well suited to study from a systems perspective because of the signaling complexities arising from the roles of growth factors, hormones and their receptors as critical regulators of cancer cell biology and from the interactions among cancer cells, normal cells and signaling molecules in the tumor microenvironment. Moreover, growth factors, hormones and their receptors are often effective targets for therapeutic intervention, such as estrogen biosynthesis, estrogen receptors or HER2 in breast cancer and androgen receptors in prostate cancer. Given the complexity underlying the molecular control networks in these cancers, a simple, intuitive understanding of how endocrine-related cancers respond to therapeutic protocols has proved incomplete and unsatisfactory. Systems biology offers an alternative paradigm for understanding these cancers and their treatment. To correctly interpret the results of systems-based studies requires some knowledge of how *in silico* models are built, and how they are used to describe a system and to predict the effects of perturbations on system function. In this review, we provide a general perspective on the field of cancer systems biology, and we explore some of the advantages, limitations and pitfalls associated with using predictive multiscale modeling to study endocrine-related cancers.

Key Words

- ▶ systems biology
- ▶ mathematical biology
- ▶ computational biology
- ▶ predictive modeling

Endocrine-Related Cancer
(2019) **26**, R345–R368

Introduction

Over the past few decades, many advances in endocrine-related cancers have come from the experimental fields of cellular and molecular biology and from their translation into clinical applications. Generally speaking, cellular and molecular studies have taken a mostly reductionist

approach, focusing on mechanistic studies of specific genes and proteins, linear signaling pathways, and particular anticancer drugs and other interventions. A systems-based approach builds on this important work by providing a more holistic account of the complex networks

of interacting genes, proteins and metabolites that determine how a cancer cell survives and thrives within the tumor microenvironment and how the host responds to the tumor. From this viewpoint, molecular networks and the subcellular processes they regulate are seen to interact with activities occurring within the tumor cell, its microenvironment and the cancer-bearing organism. A holistic view, where interactions can have both local and distant effects, is nothing new for endocrinologists and experts in some other fields. However, in what is now often referred to as the 'post-genomic era', the tools and technologies available to effectively study any cancer as a systems-disease have changed dramatically. In concert with these advances has come greater insight into the remarkable complexity of signaling, its integration and the coordination evident in controlling and executing cellular functions.

In this review article, we hope to introduce a broad readership to the potentials and limitations of a systems approach to improve our understanding and treatment of endocrine-related cancers. The scope of endocrine-related cancer systems biology is large and complex, and we acknowledge that some issues in this field are addressed here at a relatively simplistic level. Nonetheless, we believe

that a systems approach, including computational and mathematical modeling of new data streams, is essential to transform data into actionable knowledge that leads to fundamental improvements in human health. An overview of the organization of this review is provided in Fig. 1. We begin with a section on why models are needed, how modelers generally approach building their models, and some considerations regarding the specific goals of modeling. Next, we describe how models may be based on a modular structure, and how modularity can lead to emergent behaviors, as consequences of the dynamical properties of signaling networks. We discuss deterministic, stochastic and Bayesian models, and how their parameters are estimated from data and provided with error bounds. We then discuss model performance, potential sources of error, the importance of independently validating model predictions and modeling drug interactions. Subsequent sections discuss examples of a knowledge-guided computational tool for building networks, a mathematical model of the estrogen receptor landscape and some insights into interpreting models.

For our purposes in this review, a system is a collection of interacting components that produces a defined biological output in response to specific inputs.

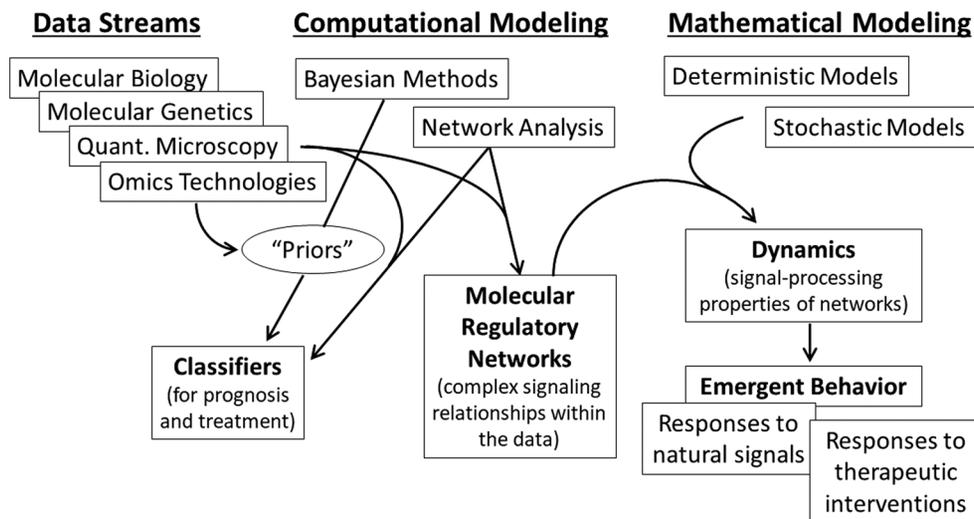


Figure 1

Representation of data streams and how these relate to computational and mathematical modeling in the context of systems biology. The four primary sections of this review contain specific insights into different aspects of modeling that reflect how modeling uses data streams to build multiscale models. We first describe why models are needed in 'Why build models'. The second section 'Multiscale modeling' introduces several critical aspects of modeling, from some basic goals of modeling, then describing how models can use a modular structure that can explain the emergent properties of biological systems. Deterministic, stochastic, and Bayesian models are then presented, as is the critical feature for cancer therapies of strategies to model drug interactions. These subsections are followed by a discussion of types of error in models, assessing model performance, and validating model predictions. The final two subsections within the section on multiscale modeling provide specific examples of tools or approaches to modeling: a knowledge-guided computational tool for building networks, and a mathematical model of the estrogen receptor landscape. The penultimate section 'Interpreting models' provides some insights into the challenges and pitfalls of interpreting model solutions. The final section 'Future directions' offers some brief insights into where the authors see the field going in the next few years.

To be useful, such input–output models must adequately capture the complexity of the system. Complexity does not necessarily mean ‘big’ (many nodes and edges). Relatively small networks can exhibit non-intuitive signal-processing capabilities due to inherent feedforward and feedback loops and non-linear kinetic rate laws, for which small changes in input produce disproportionately large changes in output.

Most biological systems are open, complex, dynamic and adaptive. While these fundamental properties may be missed in work that adopts a solely reductionist perspective, there would be little for systems biologists to model without the data and insights obtained from reductionist studies. Systems biologists acknowledge both the complexity of biological systems and the fact that much of what must be modeled and interpreted is still poorly understood. Computational and mathematical models are often used to analyze and integrate data from multiple technological platforms into new representations of system function. These new representations can expand our understanding of complex regulatory systems (Lavrik & Zhivotovsky 2014, Wang & Deisboeck 2014, Altrock *et al.* 2015, Peng *et al.* 2016, Janes *et al.* 2017, Ji *et al.* 2017). Ultimately, systems-based insights into the biology of endocrine-related cancers may lead to better treatments and outcomes for patients (Werner *et al.* 2014, Jinawath *et al.* 2016, Ji *et al.* 2017).

While the idea of generating mathematical models of signal flow in a biological system is not new (Le 2007, Ji *et al.* 2017), the sources and magnitude of data for multiscale modeling, and many of the computational/mathematical tools available, have changed dramatically in recent years. Many of the newer technologies fall into the rapidly developing fields of omics (genomics, transcriptomics, proteomics, metabolomics), an increasing number of sub-omic technologies and quantitative microscopy including gene expression in single cells (Sandberg 2014, Buettner *et al.* 2015, Kanter & Kalisky 2015). Central to our ability to analyze and integrate these new data streams and to build new mathematical models and computational representations of the data, are the analytical approaches and software tools that continue to be developed by computer scientists, mathematicians and statisticians. Rather than being identified with any of these particular specializations, systems biology sits uniquely at their nexus.

We will focus our discussion on the use of computational and mathematical approaches to model system function in the context of endocrine-related cancer biology. For the purposes of this review,

we consider a ‘mathematical model’ as using differential equations and stochastic algorithms to create dynamic, semi-mechanistic models of control networks of limited scope (dozens of genes and their products). Of course, such dynamical models must ultimately be simulated on a digital computer, but we consider a ‘computational model’ as something different: as using machine-learning tools to explore high-dimensional data (hundreds or thousands of genes and/or proteins).

Mathematical models may be deterministic or stochastic in nature, depending on the role of random events in the system being modeled. In either case, all models ultimately entail a statistical evaluation of how well the model’s output fits the available experimental data. Both stochastic and deterministic models can be useful when used appropriately (Twycross *et al.* 2010). At present, deterministic models are usually the initial approach taken to provide a description of molecular events in cellular control systems. However, considering the paucity of informative data within the flood of omics results, the unavoidable noise in biological measurements, and our ignorance of latent variables in regulatory networks, stochastic (Wilkinson 2009) or hybrid models (Twycross *et al.* 2010) are being applied more widely. Some of the general limitations in modeling have been discussed elsewhere (Di *et al.* 2006, Wilkinson 2009, Twycross *et al.* 2010) and will not be reiterated here.

From a clinical perspective, useful *in silico* models will have to be multiscale. For example, drug action at the molecular scale must be linked to clinical outcomes at the tissue or organism scale. Multiscale models use many different data types from multiple sources, spanning scales from DNA to RNA to protein, from metabolites to cells to tissues, from tissues to organisms and even to interacting populations. Modeling based only on genome and/or transcriptome data can be limited because approximately 50% of changes found in the transcriptome may not be present in the proteome (Vogel & Marcotte 2012); an even smaller percentage of changes in the genome may filter through to the proteome. Hence, spanning scales (provided necessary data are available) may improve the models and provide new insights into cancer physiology (Deisboeck *et al.* 2011).

In this review, we explore some of the basic concepts and challenges in applying computational and mathematical modeling to endocrine-related cancer research. Rather than providing detailed descriptions of tools-of-the-trade, we discuss a variety of computational and mathematical approaches that are often applied, the advantages and limitations of each, and the specific challenges for using

them correctly and usefully. Since we will not discuss specific experimental designs here, readers interested in exploring the many tools, workflows and frameworks and emerging standards for systems-based research may find the following sources useful (Brazma *et al.* 2006, Swertz & Jansen 2007, Gehlenborg *et al.* 2010, Ghosh *et al.* 2011, Wu & Stein 2012, Hofree *et al.* 2013, Sedgewick *et al.* 2013, Wen *et al.* 2013, Cheng *et al.* 2014a,b, Hoadley *et al.* 2014, Creixell *et al.* 2015, Leiserson *et al.* 2015, Dimitrova *et al.* 2017, Nam 2017, Keenan *et al.* 2018, Miryala *et al.* 2018). Similarly, there are many sources of cancer omics data in the public domain that are too numerous to capture here. However, we provide examples of some widely used large omics datasets that include data from breast and other endocrine-related cancers in Table 1.

Given clear evidence of a significant lack of reproducibility in biomedical research (Begley 2013, Mobley *et al.* 2013, Hatzis *et al.* 2014) and the potential for systems approaches to both reduce and exacerbate this problem, an appreciation of some of the key challenges – for which there may or may not be adequate current solutions – is timely. While we cannot address all the major issues in such an interdisciplinary subject, we hope that our perspective will be pertinent to using systems biology to attain a better understanding of endocrine-related cancers.

Why build quantitative models of biological systems?

‘The statistician knows, for example, that in nature there never was a normal distribution, there never was a straight line, yet with normal and linear assumptions, known to be false, he can often derive results which match, to a useful approximation, those found in the real world.’

George E P Box (1919–2013)

To extract new insights and build integrated, predictive models, particularly from experiments that generate ‘big data’, requires some form of *in silico* analysis to deal with the complexity of the data. For biological systems, complexity can arise from dimensionality (many genes and their interactions) and from general properties of the system that reflect its topology (feedforward and feedback loops), adaptability (redundancy, degeneracy), multimodality (concurrent performance of multiple integrated and coordinated tasks) and dynamism (changes in time and space) (Clarke *et al.* 2008, Tyson *et al.* 2011). Complexity can also arise at the cellular level. A notable feature of several

endocrine-related cancers is their cellular heterogeneity, which creates a dynamic microenvironment of many cell types in addition to the cancer cell component and can also affect a tumor’s response to treatment (Junttila & de Sauvage 2013, Meacham & Morrison 2013, Martelotto *et al.* 2014). Often, models are built with transcriptome data that reflect averaged expression values, since tissue microdissection prior to collecting omic data remains relatively uncommon. When applying computational and mathematical modeling to study cell type/tissue type in data from complex tissue samples, data deconvolution using either supervised or unsupervised approaches is a prerequisite. Supervised data deconvolution can be performed by integrating tissue-specific gene or protein expression profiles (Newman *et al.* 2015) from the Gene-Tissue Expression program (GTEx Consortium 2015) and The Human Protein Atlas (Ponten *et al.* 2011). Alternatively, in the more challenging case of intra-tumor heterogeneity where subclone-specific markers are often unknown, an unsupervised data deconvolution approach such as Convex Analysis of Mixtures can be exploited to uncover the hidden subclone specificity (Wang *et al.* 2015, 2016, Herrington *et al.* 2018). While tools for supervised (Zuckerman *et al.* 2013, Hart *et al.* 2015) and unsupervised deconvolution of averaged data from heterogeneous tissues (Chen *et al.* 2011, Wang *et al.* 2016) can be used as a data processing step prior to modeling – this preprocessing step remains uncommon.

The properties of high-dimensional data, particularly data from omics technologies, present unique challenges (Clarke *et al.* 2008) that are often inadequately addressed or fully appreciated. Nonetheless, the purpose of *in silico* analysis is to apply tools to extract meaningful results from high-dimensional data for the purposes of generating and testing biological hypotheses (Tyson *et al.* 2011). For instance, we may wish to understand and predict under what conditions a cancer cell will begin to proliferate *in situ* or migrate to a new location. Extracting such knowledge from large datasets by intuitive reasoning alone can be difficult or impossible and is often associated with a high risk of operator bias and/or error. Thus, new tools and approaches continue to emerge to deal with the challenges of working in high-dimensional data spaces and to enable integrating the spatial, temporal and cell context-specific nature of regulatory networks (Hoadley *et al.* 2014, Leiserson *et al.* 2015, Masoudi-Nejad *et al.* 2015, Tape 2016, Barberis & Verbruggen 2017, Dimitrova *et al.* 2017). New concepts, such as ‘master regulator proteins’ that may determine the transcriptional state of a cancer cell, also continue to arise (Califano & Alvarez 2017).

Table 1 Examples of the most commonly used endocrine-related breast cancer public omic datasets.

Database	URL	Data spaces
CPTAC	https://proteomics.cancer.gov/data-portal	Proteome
EGA	https://ega-archive.org/datasets	Genome, Transcriptome
EMBL-EBI	https://www.ebi.ac.uk/services/all	Genome, Transcriptome, Proteome, Metabolome
GNPS/Massive	https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp	Metabolome
ICGC	https://dcc.icgc.org/	Genome, Transcriptome
MassIVE	https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp	Proteome
Metabolomics Workbench	https://www.metabolomicsworkbench.org/	Metabolome
NCBI-GEO	https://www.ncbi.nlm.nih.gov/gds	Genome, Transcriptome
ONCOMINE	https://www.oncomine.org/resource/login.html	Genome, Transcriptome
ProteomeXchange (PX) Consortium	http://www.proteomexchange.org/	Proteome
ProteomicsDB	https://www.proteomicsdb.org/	Proteome
TCGA	https://portal.gdc.cancer.gov/	Genome, Transcriptome

Primary data and metadata quality vary across and within these sites. For example, clinical metadata for human subjects are often limited. The platform used for data collection in each omics space also can vary across and within these sites. While most provide access to the raw (unprocessed) data, ONCOMINE primarily exposes only processed data; the method of data processing can vary across individual studies.

Computational modeling can provide unbiased results from large data sets, allowing us to visualize complex signaling relationships within the data (Gehlenborg *et al.* 2010). Some of the more useful approaches in this area come from applications of graph theory. Graphs are mathematical structures that represent pairwise relationships between nodes. Each gene/protein is a node (or vertex) and each connection with another gene/protein is an edge. Graphical representations of molecular signaling are readily available on the web. For example, signal transduction pathways may be found at the community-based Kyoto Encyclopedia of Genes and Genomes (KEGG; <http://www.genome.jp/kegg>) or the commercially supported Biocarta Pathways Project (<http://www.biocarta.com/genes/index.asp>). These graphical representations are mostly assembled intuitively from the literature to provide a static reflection of the topological features of mostly canonical signaling networks.

Static maps are widely used to represent complex signaling networks and to guide largely intuitive interpretations of signaling, but they are of limited use for predicting signal flows through edges of the network in a living cell responding to signals received from its environment. Limited dynamic information may be evident in the directionality of signal flow (such as, protein A upregulates the production of protein B), but the consequences of many such interactions in a complex, interconnected network are challenging to predict by intuitive reasoning alone. Appropriate computational models can help to uncover complex associations hidden in the data and often may provide a statistical assessment of the strength of any predicted association. For example,

gene set enrichment analysis can rapidly probe a large database of genes and their hierarchically annotated functions to suggest signaling pathways closely affiliated with a list of differentially expressed genes (Subramanian *et al.* 2005); for example, see <http://software.broadinstitute.org/gsea/index.jsp>. A pathways database and a search tool is also provided by the Gene Ontology Consortium (see <http://geneontology.org/page/go-enrichment-analysis>).

Given adequate data, both computational and mathematical models can make quantitative predictions of the biological state under investigation. One of the primary uses of quantitative models is to perform *in silico* experiments where the values of specific nodes or edges are changed and the model is used to predict how the change affects other nodes in the network. It is possible to run hundreds or thousands of such simulations to explore both model performance and how specified changes in node/edge values affect the distribution of predicted outcomes. For example, *in silico* modeling can be used to compare multiple drug combinations including the effects of scheduling and dosing that would be very difficult in animal models or even in some cell culture models (Tang & Aittokallio 2014, Ryall & Tan 2015, Ledzewicz & Schaettler 2016). Appropriate quantitative models, when effectively applied to sufficient, high-quality data, can enable investigators to explore questions in ways that would otherwise not be possible. Visualization of the outputs from computational analysis of high-dimensional data can be an indispensable aid in interpreting the biological significance of the data (Gehlenborg *et al.* 2010, Cirillo *et al.* 2017, Pavlopoulos *et al.* 2017, Robinson *et al.* 2017). Thus, multiscale modeling enables investigators to

explore complex datasets and signaling in new ways that are both tractable and productive.

Multiscale modeling

'Numquam ponenda est pluralitas sine necessitate'
(Plurality should not be proposed unnecessarily)

William of Occam (c. 1287–1347)

'Since all models are wrong the scientist cannot obtain a 'correct' one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena.'

George E P Box (1919–2013)

All models are abstract representations of the system they are built to portray. The types of models we consider here are not intended to explain all of cancer biology. Rather, we use models to learn something new about how a specific function may operate, be controlled and interact with other cellular functions to affect a specific biological outcome. For example, we may wish to understand how estrogens affect the decision of some breast cancer cells to enter and complete a turn of the cell cycle. Understanding this function could then lead to addressing larger goals, such as developing new therapeutic interventions to block cell cycling or predicting which patients would receive the greatest benefit from blocking this action of estrogens. Thus, the primary goals of modeling are to give insights into how a control system works at the molecular level and to make robust, reliable predictions about how the system responds to a variety of natural situations and medical interventions.

For molecular signaling studies, the latter goal can be achieved by changing the values of parameters in the model and experimentally validating the predicted outcomes. Given a perturbation or rewiring of a control network, the output of a model is a prediction of the changed state of the cell (for example, alive or dead; proliferating or growth arrested). When simulations of a model under a variety of realistic conditions inadequately reflect what is already known to occur in cells, the model must be modified or extended. For example, a model may predict that reducing the expression of one gene should increase the expression of another, but the observed result of this experiment (perhaps using an RNAi approach) is the opposite. By considering how to resolve this discrepancy between model and experiment, new insights may be gained into how the control system works,

and new predictions will be generated that can be used to test the modified assumptions.

A suitable framework to guide the modeling effort is a key starting point. The framework describes, at a high level, what is generally known about the system in the context of integrated modules that perform specific cellular functions. Thus, a modular function, such as cell death, may be explained by a model of the signaling that controls and executes one or more forms of cell death, such as apoptosis. Where there is sufficient knowledge of an individual module, a reasonably detailed influence diagram of known or predicted signaling relationships can be created to guide construction of the mathematical equations. This knowledge can be gained from specific experimental data available in the laboratory, from the literature, or perhaps based on a static canonical model as might be obtained from KEGG or Biocarta. Where a canonical model does not exist (or there is good reason to believe that canonical signaling is inadequate), computational modeling can be used to formulate new hypotheses about the topology of a control module from high-dimensional data (Clarke *et al.* 2011). Where there is sufficient knowledge of the components and interactions of a control system, the interaction diagram can be translated into a set of mathematical equations that quantitatively represent dynamical fluxes through the network (readers interested in exploring specific *in silico* models can find examples in several databases including JWS Online, available at <http://jjj.biochem.sun.ac.za/index.html>, and the Biomodels Database <http://www.ebi.ac.uk/biomodels-main/>). An example of such a framework can be seen in our roadmap for systems modeling of endocrine responsiveness in breast cancer (Tyson *et al.* 2011).

At some level, useful models need to address the open, complex, dynamic and adaptive nature of biological systems. While we do not intend to provide a detailed description of the concepts and methods of model building, we can mention some general, widely applicable principles. First of all, we must keep our end-goal in mind (what aspect of cancer cell physiology are we trying to understand) as well as our starting point (what is our working hypothesis about the underlying control system). Then, ideally, we would like to get from the working hypothesis to accurate predictions of cell behavior with a model that is as simple as possible, but not so simple as to leave out crucial features of the molecular biology or cell physiology. Of course, these are vague and often antithetical requirements (what is simple? what is crucial?), but it is the job of the modeler

to make informed decisions about how much detail can and should be included in the mathematical model. Often these uncertainties can be addressed by an iterative approach, involving knowledge-guided trial-and-error or the use of multiple feature selection tools (as an example see the Feature Selection functions by MathWorks, <http://www.mathworks.com/help/stats/feature-selection.html>).

To highlight issues that may be useful for the non-expert wishing to evaluate published models and/or to collaborate with modelers, we next address the utility and methodology of mathematical and computational modeling. In our studies, we use computational tools to extract small, robust and information-rich topological features from high-dimensional data sets. These features can then be tested and validated experimentally, and at this stage, a simple mathematical model may be useful in capturing this knowledge, working out its implications, and making predictions to guide further laboratory experiments (Clarke *et al.* 2011). This iterative approach requires a modeling framework (a network diagram), some relevant experimental data, and a basic understanding of how components of the network may interact to produce observed physiological responses of cells. The network diagram guides the construction of the mathematical model, which can be used to compute the expected behavior of the simulated cells. To carry out simulations, we must first estimate the values of the parameters (such as rate constants and binding constants) in the mathematical model. Parameter estimation is a difficult problem, but it can (and must) be carried out in light of existing experimental data (Tyson *et al.* 2011). There would be no rationale to include a parameter without some data or direct evidence of its involvement in reactions, and these data can provide bounds on parameter values in the mathematical model. Once an initial model adequately accounts for the existing data, it can be used to predict specific outcomes of new experiments that can be run to confirm, extend or adjust the model. Thus, iterative modeling with the addition of new data allows both testing and refining of the model, which leads to new biological insights (Clarke *et al.* 2011).

Examples of modeling goals

Cancer systems biology studies tend to focus either on classification, where the goal is to predict a phenotype or outcome based on data, or on mechanistic modeling, where the goal is to learn something new about how the system (a tumor, a cancer cell or a signaling network within the cell) functions (Clarke *et al.* 2011).

An example of the classification task would be the use of gene expression data from a patient's tumor to predict the patient's prognosis and/or to determine the best choice of treatment. Among the simplest examples is the heuristic guide for the treatment of breast cancer patients based on a three-gene classification scheme: estrogen receptor alpha (ER), progesterone receptor (PGR) and HER2. Knowledge of the expression of these three genes defines three molecular subgroups: ER and/or PGR-positive (can be treated with an endocrine therapy), HER2-positive (can be treated with an anti-HER2 therapy, approximately half of these also express ER and/or PR and may also receive an endocrine therapy) and absence of expression of all three – often referred to as triple-negative breast cancer (TNBC) – which is usually treated with cytotoxic chemotherapy. A similar goal is exemplified by using a panel of clinical/pathological measures to predict prognosis in breast cancer; an example being the semi-quantitative assessment that produces the Nottingham Prognostic Index (Galea *et al.* 1992). Classifiers based on omics data are also available and in common clinical use, including the 70-gene signature that comprises the MammaPrint prognostic predictor (Bedard *et al.* 2009) and the prognostic PAM50 gene signature (Parker *et al.* 2009). Signatures that have not yet become adopted widely in the clinic continue to emerge (Wu & Stein 2012, Cheng *et al.* 2013). The output from these types of models is a prediction of the future behavior of the cancer – a clinical outcome such as an estimate of patient survival (prognosis) – often within a defined time period.

Omics-based classifiers (most frequently transcriptomic) are usually built using a supervised approach, where a training set of data from samples with known outcomes is used and the predictive model is subsequently validated in independent datasets. Classification models often rely primarily on the statistical properties of each measurement/input variable and do not require that these properties derive specifically from any biological function of the system (Clarke *et al.* 2008). The literature contains many different attempts to build classification schemes in breast cancer but often with varying results and robustness, even for some of the most widely used tools (Mackay *et al.* 2011, Venet *et al.* 2011). While some schemes produce comparable outcomes on a common dataset, the features selected for classification by each scheme often have little overlap (Imamov *et al.* 2005). Given the complexities in molecular signaling and the selection of genes based on their statistical properties to support classifier performance, it is not clear whether this observation reflects different genes representing

similar underlying processes (Imamov *et al.* 2005) or a lack of robustness in feature selection unrelated to biology.

Network-based classification can also be performed on individual patient data (Creixell *et al.* 2012). The key is to develop a quantitative metric based on the topology of a learned network that can be applied to new observations to determine if the new observation is likely to share the same topology. For example, once phenotype-specific networks are learned, a model-based likelihood measure can be calculated to determine which topological hypothesis is more likely generating the new observation, where the learned variance of network topology is used to support such likelihood-based hypothesis testing.

The second goal of a systems analysis of data is to generate new insights into mechanistic aspects of the cancer phenotype. For example, the model may be used to understand why patients respond differently to a specific therapy or how molecular signaling regulates or executes a specific phenotype. Hence, the analysis may be structured to test if a series of proposed features might be true (hypothesis testing) or to discover new features that might explain mechanism (hypothesis generation). While these models also frequently use the statistical properties of the measurements to find signaling features of interest, there is an explicit assumption that the measurements, and any changes in their values across phenotypes, are derived from relevant biological properties of the system.

Among the more common approaches for mechanistic studies is the use of transcriptome data to build gene regulatory networks, as exemplified by a network of transcription factors (TFs) and the target genes that they are known, or predicted, to regulate. Insights from models built primarily from transcriptome data can be limited by the often low frequency with which transcriptome changes translate into similar expression changes in the proteome (Vogel & Marcotte 2012). The target genes for TFs are identified either *in silico* (predicted using DNA sequence data; see MotifDb at (<http://www.bioconductor.org/packages/release/bioc/html/MotifDb.html>) as an example of a tool for performing this function) or experimentally (chromosome immunoprecipitation-based methods; ChIP). These studies often produce small and mostly unidirectional maps (TF→target) and they can be noisy. For example, *in silico* predictions of targets based only on promoter sequences do not account for DNA structure/accessibility and are often incomplete. Experimentally measured promoter occupancy (such as by ChIP) does not always reflect functional regulation of the adjacent gene. Correlations of measured (ChIP/ChIPseq) or predicted promoter sequence binding with differential

mRNA regulation in microarray data are often used to validate these signatures. Studies with RNAi or cDNA overexpression, mostly done using cell lines growing *in vitro*, may also be used to further establish the influence of gene expression on target gene regulation.

These approaches may not account fully for the complexity of a given target gene's transcriptional regulation, such as whether factors other than the protein complex that is detected as being bound to a specific promoter element are driving the measured differential expression of the target gene. For example, TF1→Target Gene could still be driven through a latent variable(s), since the same experimental outcomes could be seen if TF1 was knocked down and the true relationship was TF1→TF2→Target Gene or even TF1→TF2←Target Gene. Hence, both false-positive and false-negative regulatory events may be obtained in addition to true events. For *in silico* modeling, including data on TF2 may or may not affect model function. Where it does not, the measurements of TF2 are superfluous and, in the interests of parsimony, can be eliminated from the model. Alternatively, there may be technical reasons that make the measurements of TF2 more reproducible than those of TF1. In this case, when TF1 and TF2 capture the same information, the model may perform better with TF2 measurements than using those for TF1.

Modules and emergent behavior

System models can be constructed as a network of integrated and interacting modules that perform the system's component operations in a coordinated manner (Tyson *et al.* 2011). The topology of signaling for a module can be extracted *de novo* from the data, with functions being implied from any known activities of their member nodes (Wu & Stein 2012). However, for modeling known functions where there is significant data and prior knowledge, modules can be viewed more discretely as integrated network components that regulate and/or execute a specific function (Tyson *et al.* 2011). For example, apoptosis could be considered as a module that performs a cell death function; apoptosis can then be modeled as a discrete process, perhaps as a closed, input–output device. Cells have other modules that perform similar functions, including autophagy (which can produce prodeath or prosurvival outcomes). These modules represent biological redundancy because if an irreversible cell fate decision is made in favor of death, one of several differently constituted modules can execute that decision. Some genes may play key, but not

necessarily similar, functions in more than one of these modules. For example, BCL2 can regulate the activation of the autophagy module through its ability to sequester BECN1, while also affecting execution of the apoptosis module through its effects on mitochondrial membrane permeability (Clarke *et al.* 2012). Cell fate may depend on the amount of BCL2 present and its subcellular location. For example, BCL2 bound to BECN1 may be unable to protect the mitochondria, with BCL2:BECN1 complexes effectively preventing the initiation of pro-survival autophagy (BECN1) and concurrently not preventing apoptosis (BCL2). Since other pro-survival BCL2 family members can also bind to BECN1, the balance of pro-survival-to-pro-death BCL2 family members (there is potentially significant signaling degeneracy within apoptosis), the concentration of free BECN1 remaining available to activate autophagy, and their respective subcellular localization(s) may all contribute to the final cell fate decision. The potential for cell context-specific wiring (and rewiring in response to stress) is evident.

A clear understanding of these interactions in ER+ breast cancer cells requires both significant insight and quantitative data from wet laboratory studies. Predicting cell fate outcomes robustly in the presence of various endocrine stressors (estrogen withdrawal, exposure to SERMs/SERDs) is unlikely to be successful without adequate *in silico* modeling. An effective dynamic model of these relationships could also be used to predict optimal drug dosing and scheduling to drive maximal cell death and potentially limit the emergence of drug resistance (Tang & Aittokallio 2014, Ryall & Tan 2015).

Integration of modular functions allows a cancer cell to coordinate and execute the activities it needs to proliferate, survive, move and invade locally, respond to stress and manage its metabolism to support these actions. Modules can be combined differentially in time and space, creating some of the phenotypic diversity that is characteristic of breast cancer cells. When modules interact in complex feedback and feedforward loops, they can exhibit redundancy (different modules performing similar functions), degeneracy (different signaling routes allowing a module to perform the same function in different ways) and novelty (the ability to perform new functions or old functions in new ways). This plasticity of the response characteristics of modular networks is the origin of their 'emergent' properties (Bhalla & Iyengar 1999). For example, an apoptosis module may be blocked in a cell but the cell death decision may now be executed by an autophagy module. The ability to recombine signaling features in complex regulatory networks

in response to specific stresses is likely the emergent property that drives both the phenotypic plasticity often attributed to cancer cells and the development of resistance to anticancer drugs. From an intuitive point-of-view, emergent properties are challenging because they are difficult to deduce from a knowledge of the individual components of the system, and the relationships between the emergent property and its component parts may be non-linear and dynamic (changing over time). To deal reliably with these complexities requires comprehensive and accurate mathematical models to guide our thinking and predictions.

Emergence may underlie many novel behaviors of cancer cells that cannot easily be foreseen from knowledge of the system's individual components. In evolutionary biology, emergence can reflect the development of larger or more complex functions or behaviors derived from the interactions among, but not shared with, individual smaller or less complex features (Okasha 2012, Gho & Lee 2017). New behaviors in tumors likely arise through changes that affect interactions within and among modules. For example, changes in signaling from within the tumor microenvironment (adaptive) or the acquisition of a genetic/epigenetic change (such as activating or inactivating mutations) could alter the level of expression, function or subcellular location of a molecule or the activity of a pathway in a network. Consequently, this pathway may now connect different modules that perform a new cellular function or continue to perform an existing function in a different manner. Where these new emergent properties confer a biological advantage, they are expected to experience positive selection (in a Darwinian sense) (Enriquez-Navas *et al.* 2015). Acquired drug resistance may be an example of a new emergent property that is not evident in the initial cell population. Such resistance could be mutational (ER mutations that confer resistance to aromatase inhibitors in breast cancer) or adaptive (activation and integration of the unfolded protein response module with a pro-survival autophagy module that act together to confer resistance to antiestrogens) (Clarke *et al.* 2011, 2012).

The emergent properties of cells in a system like an ER+ breast tumor likely explain, in part, the phenotypic heterogeneity of some breast tumors and also the diversity of responses that confer drug resistance (Clarke *et al.* 2012). The property of emergence with respect to acquired multiple drug resistance (a function that is likely subject to positive selection), and the potential that some complex functions may never stabilize (the rate of appearance of

new metastatic foci may continue to increase throughout the disease process), may underlie the high prevalence of distant recurrences that are poorly responsive to available systemic therapies, and so are generally fatal.

Dynamics

One of the major strengths of quantitative mathematical modeling is the ability to capture the dynamic nature of a system (Aldridge *et al.* 2006, Anderson & Quaranta 2008, Toettcher *et al.* 2009, Spencer & Sorger 2011, Molinelli *et al.* 2013). In particular, models of endocrine-related cancers have provided new insights into the temporal development of invasive, metastatic cells (Quaranta *et al.* 2008, Gallaher *et al.* 2014), drug-treatment responses and drug-resistant states (Chen *et al.* 2013, 2014, Parmar *et al.* 2013, McKenna *et al.* 2017) and the origins of network plasticity (Tavassoly *et al.* 2015, Picco *et al.* 2017). Examples of some of the methods used in mathematical modeling are provided in Table 2 (Tyson *et al.* 2019).

Despite their evident utility, dynamic models in molecular cell biology must be interpreted cautiously. Model predictions can be very accurate when restricted to conditions close to the experimental conditions on which the model was built, but less reliable when extrapolated far beyond the range for which they have been verified.

Table 2 Methods of mathematical modeling.

Method	Dynamic variables	Time	Example
Boolean networks	$X(t) = 0$ or 1 $Y(t) = 0$ or 1	$t = \text{integer}$ (0, 1, 2, ...)	X inhibits synthesis of Y and Y inhibits synthesis of X $X(t+1) = -Y(t)$ $Y(t+1) = -X(t)$
Ordinary differential equations	$X(t) = \text{positive real number}$ $Y(t) = \text{positive real number}$	$t = \text{real number}$ ($t \geq 0$)	X inhibits synthesis of Y and Y inhibits synthesis of X $\frac{dX}{dt} = \frac{k_{sx}}{1+Y^p} - k_{dx}X$ $\frac{dY}{dt} = \frac{k_{sy}}{1+X^q} - k_{dy}Y$
Stochastic models	$M(t) = \text{positive integer}$	$t = \text{real number}$ ($t \geq 0$)	Propensity of mRNA synthesis = k_{sm} Propensity of mRNA degradation = $k_{dm}M$ Probability density function for number of mRNA molecules in the cell is $P(M) = e^{-\lambda} \frac{\lambda^M}{M!}, \text{ where } \lambda = \frac{k_{sm}}{k_{dm}}$
Hybrid deterministic-stochastic models	$M(t) = \text{positive integer}$ $P(t) = \text{positive real number}$	$t = \text{real number}$ ($t \geq 0$)	Genetic regulatory network: Simulate mRNA fluctuations, $M(t)$, with a stochastic model and protein dynamics, $P(t)$, with ordinary differential equations

Additional information can be found in Tyson *et al.* (2019). Reprinted from *Journal of Theoretical Biology*, Vol 462, Tyson JJ, Laomettachit T & Kraikivski P, Modeling the dynamic behavior of biochemical regulatory networks, Pages 514–527, Copyright (2019), with permission from Elsevier.

Nonetheless, like weather prediction, mathematical models of cellular regulatory systems can be very useful for short-term forecasting of local activity without being reliable predictors of long-term ‘weather’ patterns on a ‘global’ scale.

Parameters

To simulate a mathematical model, we must first estimate the values of all kinetic parameters from experimental observations. Examples of parameters include reaction rate constants (such as protein synthesis and degradation, or phosphorylation and dephosphorylation) and binding or dissociation constants (for example, Michaelis constants for enzyme-catalyzed reactions). Estimation of these parameter values is often the most difficult aspect of building a useful mathematical model (Liepe *et al.* 2014, Kimura *et al.* 2015). The goal of parameter estimation is often not to find the ‘optimal’ set of parameter values for fitting a selection of experimental results but rather to find a representative collection of parameter sets that all provide an ‘acceptable’ fit to the data (Tavassoly *et al.* 2015).

When faced with the dimensionality of data from an omics platform, a mathematical model with thousands of variables would be difficult to formulate and almost impossible to parametrize. Currently, high-dimensional data are more effectively explored using computational

modeling where the assumptions of the model are higher level and less demanding of detailed kinetic information. For example, machine-learning techniques can learn the features of molecular networks and their relationships from the data. Bayesian approaches are common in this regard and are discussed below.

Deterministic and stochastic models

Deterministic models, defined usually by differential equations, produce specific outcomes for a given set of parameter values and initial conditions, without any evidence of randomness. In contrast, stochastic models evolve in time with significant random fluctuations (Singhania *et al.* 2011, Barik *et al.* 2016). For example, a gene regulatory network, where TFs regulate specific targets, could be modeled deterministically or stochastically. In a deterministic model, the rate of gene transcription would have a definite value determined by the activity of the transcription factor. In a stochastic model, the activity of the TF would determine only the propensity (probability per unit time) of transcribing the gene into an mRNA molecule. In this case, a stochastic model represents more accurately the noisy process of gene transcription in individual cells, but a deterministic model may capture adequately the average rate of expression of the gene over a population of cells responding to an external stimulus that is activating the TF. If we have data on the noise associated with gene transcription in individual cells, then a stochastic model may be warranted and needed. Stochastic models have been useful for exploring the dynamic responses of endocrine-related cancers (Jain *et al.* 2011, Chen *et al.* 2014, Morken *et al.* 2014). A deterministic model is simpler and more appropriate if we have only gross transcriptome data on populations of cells under constant conditions.

Bayesian models

A general objective of computational tools is to find patterns (correlation structures) within data. For example, with transcriptomic data an algorithm may look for patterns of changes in gene expression that are correlated with each other and with the phenotype(s) or function(s) of interest (Dutta *et al.* 2016, Anafi *et al.* 2017, Califano & Alvarez 2017). Some measure of the statistical strength of these correlations, using either a Bayesian (conditional probabilistic) or frequentist (parametric or non-parametric probabilistic) approach, is usually applied to help identify the associations most likely to be correct.

Whichever approach is selected, statistical models (Bayesian or frequentist) have assumptions that can be violated and parameters (even non-parametric probabilistic tools have parameters; these are not fixed in advance but obtained from the data) that can be affected by the data structure and that can influence performance. While it is not always evident which statistical model is most appropriate for the data being analyzed, understanding what the model outputs represent is important for correctly inferring biological meanings or appreciating the uses and limitations of the output.

An increasingly common approach for computational modeling is to build models that incorporate prior knowledge of the system (Tian *et al.* 2014b, 2015). Prior knowledge can be as simple as looking at the expression levels of genes already known to contribute to the phenotype, at known interactions among molecules such as protein–protein or protein–DNA interactions (PPIs or PDIs) or at relationships reported in canonical signaling pathway representations. Incorporation of prior knowledge, depending on the quality of the knowledge, can greatly improve the performance of algorithms to build Bayesian networks. Indeed, a major challenge in constructing Bayesian networks is the selection of appropriate prior probability distributions (priors) for the variables in the model. How these parameters are estimated for a Bayesian approach affect its outcomes (Lampinen & Vehtari 2001). Poorly estimated priors (relative to ground truth – which is often unknown) may provide fits to the data that are statistically acceptable and intuitively logical, but solutions that are, nonetheless, noisy and lead to incorrect biological interpretations. Influence of the prior can be reduced using Bayesian hierarchical models and robust priors (Berger 2010).

In Bayesian networks, the edges are directed but the sign is not specified. Consequently, whether the edge is positive (such as driving) or negative (such as inhibiting) must be inferred from sources external to the model and/or established experimentally. A further limitation is that edges cannot be interpreted as necessarily reflecting direct interactions. While some interactions may well be direct, latent variables can also create direct edges in the model solution where none exist in the biological system. For example, the predicted edge of $A \rightarrow B$ in the model may really be $A \rightarrow C \rightarrow B$ (see also the discussion of modules and emergent behavior, above). Inferring feedback loops can also be difficult, such as $A \rightarrow C \rightarrow B \rightarrow A$.

For gene network modeling, the quality of the knowledge and its incorporation into the selection of priors will improve the predictions. Two implications

follow from this observation. Firstly, a team with better biological understanding of a system may build a Bayesian-based algorithm that outperforms others on the analysis of this specific system (because the model's priors are more correctly defined by the team's existing knowledge) but produces less robust/accurate predictions than other algorithms when it is applied to related systems. Secondly, detailed prior knowledge of a system limits what new knowledge can be discovered. The more that is understood about the system ahead of time, the better the model will perform. However, the model will be making predictions in a shrinking space where there is less new knowledge to be discovered. In reasonably well understood systems, these latter models may have most utility in building our confidence that what we believe to be true may indeed be true. In systems that are inadequately known, the new knowledge space can be large and the predictions noisy; the extent to which something is now believed to be true may require careful evaluation. Overall, the primary advantages of modeling include the ability to integrate significant amounts of knowledge, to help researchers to understand confounding events seen in the data and to answer questions of combinatorial complexity for which experimentation within the wet laboratory is prohibitive.

Error, performance and validation

Some workflows may include the output of one algorithm as a means to guide parameter estimation for another. For example, in building a gene regulatory network from expression data, an investigator could take the output predictions from a tool that predicts a TF and its targets as a means to define the priors for a Bayesian network modeling analysis of how these molecules are related in the data from a gene expression study. Intuitively, even if the TF output is statistically noisy, it might be expected to outperform a model with uninformed priors where equal probabilities are assigned to each outcome. Nonetheless, some of the predictions will be wrong and represent errors in the prior that may be worse than uninformative; these types of errors will be propagated from the output of one tool to the output of the next. Since the variables and their relationships (as captured in their priors) were thought to be intuitively correct, if these incorrect variables persist as key features of the Bayesian model solution, they could create the trap of self-fulfilling prophecy (Clarke *et al.* 2008). Predictions from one tool will also be associated with a level of error (variability), and this type of error will also propagate when the outputs are used as input variables for another tool in a workflow.

Here, error propagation represents the effects of the variability in the input variables on their respective model functions and on model output (Mangado *et al.* 2016). Estimating (and reporting) uncertainty propagation and its implications is an important consideration in assessing model calibration and interpretation (Vanlier *et al.* 2012). Methods to estimate uncertainty propagation continue to be developed and applied (Ades & Lu 2003, Welton & Ades 2005, Dubois 2010, Moseley 2013, Mangado *et al.* 2016). In his discussion of error propagation in metabolomics studies, Moseley notes that both derived and propagated uncertainty should be reported along with the results (Moseley 2013).

Measurement errors, as they apply to the relationship between a measured variable and its covariate, are additive (Eckert *et al.* 1997). Integrative analyses across workflows in multiscale modeling, as may occur when combining data from DNA sequence, RNA sequence/abundance and/or PPI studies, include many relationships between the measured variables (such as mRNA and protein expression levels) and covariates (such as a clinical outcome or changes in phenotype). Such analyses may be prone to error propagation and to error additivity or even amplification. For example, agglomerative techniques (such as some hierarchical clustering), growing decision trees (such as some random forest methods) or the network propagation algorithms that have begun to attract increased attention (Cowen *et al.* 2017) may be sensitive to error propagation. Once an error (node-edge connection) is made during the graph build, it may remain and affect the accuracy of subsequent local connections and of the overall model solution. A build error that remains can lead to a model solution that reaches convergence and appears 'globally correct' but contains features that are 'locally wrong'. The challenge here is that it is the local connections that are used to guide individual wet laboratory experiments.

Studies that apply bioinformatic/biostatistic tools to solve problems in large data spaces are likely to be at greatest risk of experiencing the various types of errors described above. The 'hairball' models often produced are rarely robustly tested for local error, especially when the global model fit provides an apparently miniscule *P* value. For example, independent datasets showing the same topologies are often not shown, frequently because the data are not available to do independent validation. The internal topology of individual cliques is rarely tested, even using a simple *n*-fold cross-validation. Global solutions are also rarely tested by an analogous *n*-fold cross-validation, such as removing entire cliques at random. Since the overall topology of the solution

is likely to be influenced by the relationships among discrete discovered features, without testing the effects of removing features on the remaining structures, there are few ways to determine topological robustness. While these 'hairballs' will likely have met the statistical requirements for global algorithmic convergence, how many of the local structures are correct, either internally within each feature or externally within the global solution, is often left to human intuition and the risks therein (Clarke *et al.* 2008).

Appropriate assessments of model robustness and validation are critical to the successful use of a systems biology approach (Steyerberg *et al.* 2001). There are many tools to assess model performance and validation and a detailed technical discussion is beyond our scope. Here, we use performance to denote assessments of the robustness or reproducibility of model predictions. For performance, biostatistical assessments of model fit are usually incorporated into the workflow. Examples of approaches to assess performance include use of a receiver operating characteristic analysis and estimates of the positive predictive value and negative predictive value. An internal n -fold cross-validation is commonly used, particularly when data are limited (Waljee *et al.* 2014). A random portion of the data is withheld at each interaction as a 'validation set', and the remaining data are used as a 'training set' for running the model. Multiple iterations are run and the performance for each iteration is compared to assess the overall model performance. A model can be tuned by adjusting its parameters until the predictions from the training and internal validation sets become sufficiently comparable. Since this approach can lead to model overfitting, the most informative assessment of model performance is obtained from the use of independent datasets not used in model building and any internal performance analyses. A robust model is expected to produce broadly similar predictions in all comparable data sets. For classification studies using human tumors, the use of independent datasets may also be the only tractable option for validating model predictions.

For models that are used to predict system function in a biological context, mechanistic or functional validation of a prediction is almost always required. Here, validation refers to experimental validation in the form of appropriate wet laboratory studies. These validation studies are often done in cell lines and/or animal models and can include applying perturbations to the experimental system and then measuring whether the changes predicted by the model occur. A common

approach is to knockdown a target gene in cells where it is overexpressed, overexpress the gene in cells where its expression is low and then determine if the biological function(s) is altered as the model predicts. Knockdown is commonly achieved by an RNAi method such as siRNA or shRNA transfection. A gene may also be eliminated using CRISPR (Yin *et al.* 2019). How often a cell totally loses a gene or its expression likely requires careful consideration. Total loss of a protein's expression, as would usually occur with CRISPR, could alter a signaling feature in a manner that does not occur when expression is lowered but not eliminated in the phenotype(s) of interest. While CRISPR is often preferred over RNAi, for genes where downregulation rather than total loss is the primary biological observation, RNAi may offer a more physiologically relevant validation approach. A similar caveat applies to the use of cDNA transfection to produce overexpression of a gene. The level of overexpression may be outside the range seen in the phenotype(s) under study, and so also produce changes in network features that are not physiologically relevant. These types of events could lead to misinterpretations of the validation experiments. For example, the phenotype predicted by the *in silico* model is not observed or further studies to determine the effects of the manipulation of a gene on signaling identifies new relationships that are signaling artifacts from a physiological relevance perspective.

As an example of a biological validation strategy, consider a prediction by an *in silico* model that an antiestrogen should induce autophagy through altering expression of BECN1 in ER+ breast cancer cells. One approach to mechanistic validation of this prediction could be to apply the drug and its vehicle control to ER+ and ER- cells (negative control), measure changes in BECN1 and autophagy and then use a molecular approach to study if BECN1 knockdown or overexpression altered the regulatory effects of the antiestrogen on autophagy. An underappreciated challenge with these types of studies is that the experimental validation may be frustrated by a high proportion of intuitively rational, statistically significant, but biologically incorrect *in silico* model predictions (the wet lab validation experiments show the predictions to be invalid).

Modeling drug interactions

Another area of significant potential for a systems approach is the search for drug combinations for treating a specific cancer in the context of a multicomponent signaling network within the cancer cells (Tang & Aittokallio 2014,

Ryall & Tan 2015). Effective combination therapy, which is a hallmark of current cancer treatment, requires an adequate understanding of signal complexity. Developing and evaluating drug combinations is difficult because the complexity of the problem increases combinatorially with the number of constituent drugs proposed to address an integrated driver pathway of the cancer. When the possibility of sequencing drugs at different times relative to one another is added to the mix, complexity again increases dramatically. Progress has been made using a systems biology approach. For example, the joint effects of multidrug combinations can be evaluated based on the mechanisms of action of the drugs (Fitzgerald *et al.* 2006). If the constituent drugs in a combination therapy exert their effects through known mechanisms that feed into common pathways, the joint effect of the combination may be assessed by the 'Loewe additivity'. If the drugs act non-exclusively on multiple targets, the effect may be assessed by the 'Bliss additivity' (Baeder *et al.* 2016). Knowledge of the biological system can be used for experimental design and data analysis. Thus, drugs with different mechanisms of action, as revealed by systems biology modeling, may exhibit different shapes of their dose–response relationships. Such information can be augmented by experimental data on a single drug to optimally design the experiments on the joint effect of the drug combinations.

Because the complexity of the problem increases rapidly with the number of constituent drugs, even the development of systems-based methods for the design and analysis of three-drug combinations has been only recent (Fang *et al.* 2017). The case of three-drug combinations is fundamentally more difficult than two-drug combinations. Finding doses of the combination, number of combinations and replicates needed to detect departures from additivity depend on the dose–response shapes of each of the constituent drugs. Thus, different classes of drugs with different dose–response shapes must be treated as separate cases. We designed and analyzed a combination study of three anticancer drugs (PD184, HA14-1 and CEP3891) that inhibit the H929 myeloma cell line. The three-drug combinations study used the original 4D dose–response surface formed by the dose ranges of the three drugs (Fang *et al.* 2017).

Methods for screening large numbers of drug combinations are being developed to reduce the problem to one that is more experimentally manageable by using the experimental data from dose–response studies of single drugs and from a few combinations along with a systems analysis of pathway/network information to obtain an estimate of

the signaling network model parameters and the functional structure of the dose–response relationship (Fang *et al.* 2016). This model comprises a Hill equation for signals arriving at each receptor, a generic enzymatic rate equation to describe the transmission of signals among connecting genes, and a logistic equation to represent the cumulative effect of genes implicated in the onset of the cell death machinery. These statistical models generate a global drug sensitivity index based on the joint dose–response characteristics. Only the few terms with large global-sensitivity indices, much like principal components, are kept and subject to further experimental validation. Recently, the experimental design required for such subsequent experimentation has also been worked out (Fang *et al.* 2016, Huang *et al.* 2018).

An example of computational modeling: KDDN

Cancers are often characterized by dysregulation of molecular signaling (Barabasi *et al.* 2011, Tyson *et al.* 2011, Creixell *et al.* 2012). Significant rewiring of molecular networks can drive key phenotypic transitions that can occur in both a tumor and its microenvironment (Califano 2011, Roy *et al.* 2011, Ideker & Krogan 2012). The impact of a treatment can spread through the network and alter the activity of functionally relevant gene products (Roy *et al.* 2011, Creixell *et al.* 2012). Most molecular components exert their functions through interactions with other molecular components (Li *et al.* 2008, Gong & Miller 2013). How cancer cells differ from each other in their responses to environments or treatments is intrinsically context specific (Mitra *et al.* 2013) and identifying such differences may represent a 'wicked' problem for the research community (Rittle & Webber 1973, Courtney 2001, Clarke *et al.* 2011). Changes in molecular interdependencies across cancer phenotypes may reveal novel hub genes and pathways, which may be suitable targets for drug development. Instead of asking 'which genes are differentially expressed?' the question here is 'which genes are differentially connected?' (Hudson *et al.* 2009). Studies on network-attacking events will shed new light on whether network rewiring is a general principle of cancer cell responses, as most molecular therapies target proteins and their networks but not genes (Califano 2011). Novel hypotheses inferred from the rewired TFs and their distal enhancers or partners can be proposed and examined (Creixell *et al.* 2012, Mitra *et al.* 2013).

While multiscale omics data and the prior knowledge that provide insight into complex interactions are increasingly available, models and analysis methods to

functionally integrate this information are still sorely needed. In particular, systematic efforts to characterize selectively activated regulatory components and mechanisms must effectively distinguish significant network rewiring from random background fluctuations. Most published biological network inferences were obtained from molecular datasets acquired under a single condition, for which the statistically significant network rewiring across different conditions is unknown or unreported (Mitra *et al.* 2013). The inability to identify significant rewiring in biological networks represents a major limitation on the use of these results for molecular signaling studies. The Knowledge-fused Differential Dependency Network (KDDN) method has been developed to infer significant rewiring of complex biological dependency networks, via sparse modeling and data-knowledge integration (Zhang *et al.* 2009, 2011, Tian *et al.* 2013, 2014a,b, 2015). Specifically, KDDN formulates the inference of differential dependency networks (Zhang *et al.* 2009, 2011, Tian *et al.* 2014a) that incorporate both conditional data and prior knowledge as a convex optimization problem (Zhang & Wang 2010, Tian *et al.* 2011) and uses an efficient learning algorithm to jointly infer the conserved biological network and significant rewiring across different conditions (Tian *et al.* 2014b, 2015). KDDN uses a minimax strategy to maximize the benefit of prior knowledge while confining its negative impact under the worst-case scenario. Furthermore, KDDN matches the values of model parameters to the expected false-positive rates on network edges at a specified significance level and assesses edge-specific P values on each of the differential connections.

Tests on synthetic data have shown that KDDN produces biologically plausible results (Zhang *et al.* 2009, 2016, Herrington *et al.* 2018) and can reveal statistically significant rewiring in biological networks. The utility of KDDN is evident following its application to a variety of real gene and protein expression datasets including yeast cell lines (Tian *et al.* 2014b), breast cancer (Tian *et al.* 2014b), ovarian cancer (Zhang *et al.* 2016) and medulloblastoma (Tian *et al.* 2014a). The method efficiently leverages data-driven evidence and existing biological knowledge while remaining robust to false-positive edges in the prior knowledge. The network rewiring events identified by KDDN reflect previous studies in the literature and provide new mechanistic insight into the biological system(s) that extends beyond this earlier work.

To study how gene networks may rewire during the transition from normal to neoplastic breast cells, we have focused on understanding how ER+ breast cancer

cells adapt to the stresses of endocrine-based therapies. Our central hypothesis invokes a gene network that coordinately regulates those functions of a cell module that determine and execute the cell's fate decision. Using the KDDN tool, we identified three small topological features and then overlaid these onto the canonical apoptosis pathway from KEGG (Fig. 2). The largest of the three features reflected much of our prior knowledge, despite not explicitly incorporating this knowledge into the models (Zhang *et al.* 2009). Following the predictions of this topology, we uncovered some fundamentally new insights into molecular signaling; for example, the direct regulation of BCL2 by XBP1 and the requirement of NF κ B for XBP1 signaling to regulate the pro-survival cell fate outcome in the context of antiestrogen treatment and resistance (Clarke *et al.* 2011, Tyson *et al.* 2011, Hu *et al.* 2015). In applying KDDN to data from a rodent model, we found that exposure to estrogens *in utero* induces a rewired network in the mammary glands of the offspring that predicts for resistance to endocrine therapies in tumors that arise in these glands during adulthood. Subsequent studies showing that tumors in these mammary glands are less responsive to tamoxifen (TAM) provided the first direct demonstration of why many ER+ breast cancers may be pre-programmed to fail to respond to TAM treatment or respond and later recur (Hilakivi-Clarke *et al.* 2017).

We further pursued the functional evidence of the hidden dependencies/crosstalk inferred by KDDN. For example, KDDN analysis of global protein expression data from 122 TCGA ovarian cancer samples (selected based on homologous recombination deficiency, HRD, a phenotype with distinct prognosis and response to therapies) resulted in a number of phenotype-dependent modules of co-expressed proteins. Several of the member proteins in the modules were known to be involved in histone modification. With the additional evidence of HRD status-dependent acetylation or deacetylation of histone proteins in the same samples, we were able, using patient population data, to support what has been shown in cells (Gong & Miller 2013, Tang *et al.* 2013) that histone protein acetylation affects the choice of DNA double-strand break repair pathways (between homologous recombination and non-homologous end-joining) (Zhang *et al.* 2016).

An example of mathematical modeling: ER landscape

Dynamic mathematical models track a system as it evolves in time. A key use of such models is to optimize

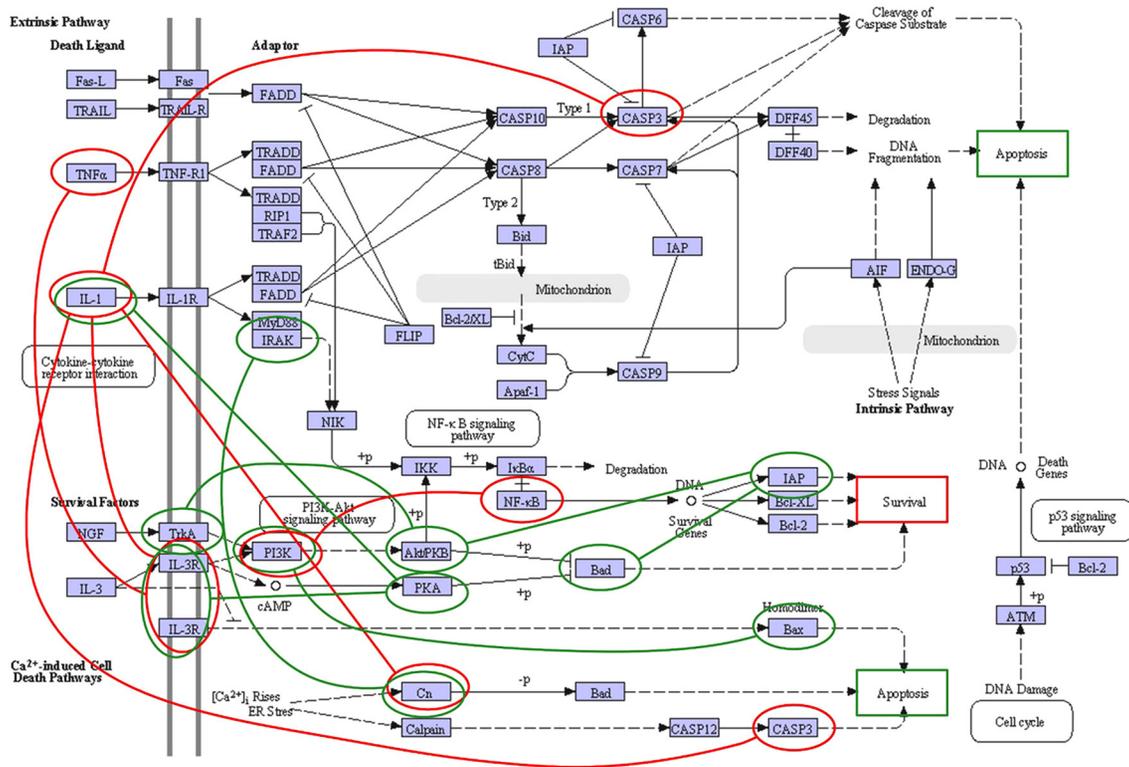


Figure 2

Differential dependency network focused on the KEGG apoptosis pathway (Kanehisa & Goto 2000). Recurrent breast cancers (uniquely featured by red edges) showed the imbalance between apoptosis and survival with only one route into the cell through IL1B-induced inhibition of proapoptotic CASP3. Non-recurrent breast cancer (uniquely featured by green edges) had a cascade of signaling pathways inside the cell that provides the balance between apoptosis and survival. Copyright Kanehisa Laboratories. Reproduced with permission from KEGG.

therapeutic protocols. For example, instead of applying a given drug or combination of drugs continuously for a specified overall duration, the drug(s) can be applied for fixed durations with rest intervals in between. Alternatively, several drugs can be applied in a repeating sequence for fixed durations. Optimizing the durations and dosing of drugs is a combinatorial problem that is difficult to solve experimentally, but relatively simple to solve via computer simulation, assuming an accurate dynamical model is available. Impressive results have been obtained in prostate cancer and glioblastoma using two-compartment models that simulate the temporal development of the sensitive and drug-resistant populations of cancer cells (Jain *et al.* 2011, Leder *et al.* 2014, Morken *et al.* 2014).

In the case of ER+ breast cancer and antiestrogens, the resistance character of the cells changes with time in response to the drugs. Hence, it is necessary first to model the dynamics of development of drug resistance in individual cells, then to model the dynamics of a population of treated cells by linking the cellular scale to the population scale, and finally to consider strategies for optimizing drug therapy. A proof of concept of this

idea considered estrogen deprivation therapy (Chen *et al.* 2014). ER+ cells were presumed to exist in three different states: an estrogen-sensitive state (growth driven by the estrogen receptor bound to estrogen), an estrogen-hypersensitive state (growth driven by membrane-associated estrogen receptor (ERM) bound to estrogen) and an estrogen-independent state (growth driven by growth factor receptors (GFRs)). Transitions between the states were governed by the estrogen level (high, low, trace) in which the cells were grown. If cells were growing in a high (physiological) concentration of estrogen, most cells would transition to the estrogen-sensitive state. If the estrogen concentration dropped to a low level, sensitive cells would begin to die, but some would transition to a hypersensitive state and continue growing.

To model the transitions among these states, we developed a stochastic differential equation model of an individual cell. States were characterized in the model by ERM activity (high or low) and GFR activity (high or low). The model qualitatively matched observations in the literature concerning sensitivity transitions in breast cancer cells as the estrogen level was varied.

The fact that resistance to estrogen deprivation was reversible if resistant cells were transferred back to estrogen-rich medium for a sufficiently long time was also captured. Using techniques from statistical physics, it is possible to visualize this model as a landscape upon which the system makes spontaneous transitions among three low-lying basins (Fig. 3A), which represent the three states of estrogen sensitivity. Random fluctuations in the cells can occasionally cause transitions from one basin to another, representing the natural heterogeneity seen in a cell population. However, the system typically resides in the lowest basin, as determined by the estrogen level.

It is not efficient to simulate large numbers of these ‘model cells’ for long periods of time in order to compute how a population would evolve in response to changes in estrogen dose. To circumvent this problem, a cell-level model was used to compute the transition probabilities among states as a function of estrogen concentration. These probabilities were then used to create a population model that efficiently tracked the number of cells in each state. A treatment regimen consisting of cycles of estrogen deprivation followed by a drug holiday was considered, and the deprivation and break durations were optimized to drive the cancer cell population as low as possible. Results are shown in Fig. 3B and C for the situation where the cancer population is initially 1000 cells. For the parameters in the model, the cancer cannot be eradicated. However, over a suitable range of therapeutic parameters,

the disease can be kept in check (similar to increasing duration of the recurrence-free survival period).

This example provides a possible roadmap for how modeling a molecular understanding of the response of a cancer cell to a drug can be transitioned to a tissue-level model and used for therapy optimization. While the situation in patients is certainly more complicated than the model systems described here, the success of simple compartment models to guide therapy in simulated tumors provides hope that more complicated, molecularly-based, multiscale models will ultimately be useful in guiding therapy in the clinic.

Interpreting models: caveat lector

‘A little learning is a dang’rous thing; drink deep, or taste not the Pierian spring: there shallow draughts intoxicate the brain, and drinking largely sobers us again.’

Alexander Pope (1688–1744)

The qualitative and quantitative models that we have described above produce results that can be difficult to interpret correctly and usefully. Correct interpretation is important, of course, because no one wants to spend time and precious experimental resources failing to validate an incorrect understanding of the results of a computational and/or mathematical analysis of a cellular

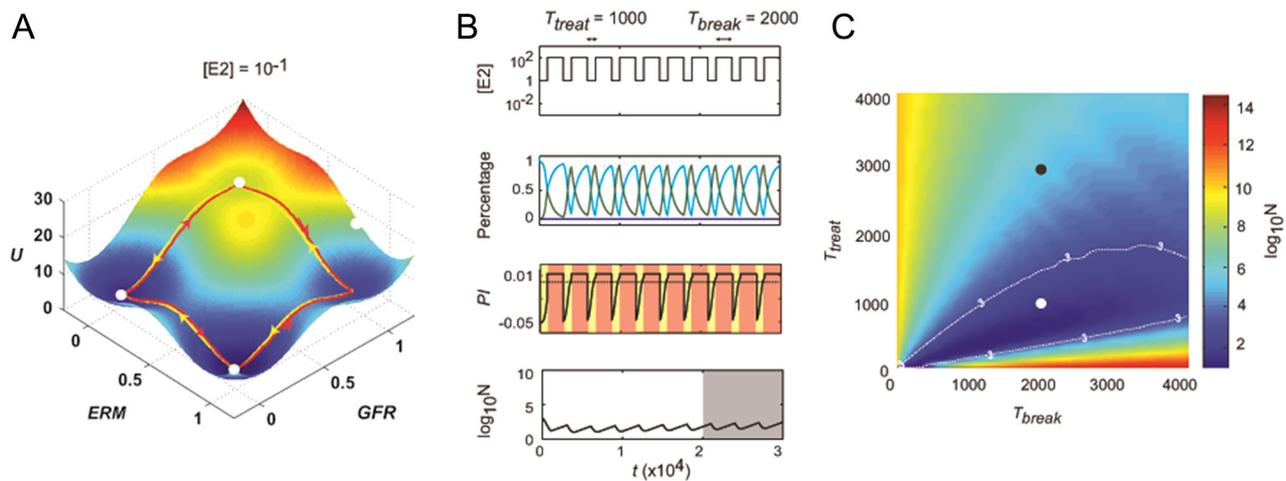


Figure 3

(A) The estrogen-response landscape for a particular level of estrogen stimulation. There are four basins of attraction for the cell state corresponding to sensitive (ERM–/GFR–), hypersensitive (ERM+/GFR–) and independent (GFR+). (B) A sample intermittent treatment regimen (top panel) produces varying proportion of cells in different states (second panel; cyan = sensitive, green = hypersensitive, blue = independent), a varying proliferation index of the overall cell population (third panel; yellow indicates death and red indicates growth). The overall population level, starting from 1000 cells, is shown in the bottom panel. (C) Plot of the average value of cell number $\langle \log_{10} N \rangle$ over the interval $t \in (2 \times 10^4, 3 \times 10^4)$ as a function of T_{treat} and T_{break} . The white dot indicates the case in (B). Any combination of T_{treat} and T_{break} that puts the system within the $\log_{10} N = 3$ contour will suppress cancer growth. This figure is adapted, with permission, from Fig. 3 and 6 of Chen *et al.* (2014).

control network. For example, gene set enrichment is a powerful tool to explore high-dimensional data sets (Subramanian *et al.* 2005), but how its results are interpreted and used requires a thorough understanding of what the results do and do not imply. In this case, gene set enrichment analysis provides a static representation of canonical signaling pathways, which are highly idealized views of the most frequently observed events in a signal transduction pathway. However, once an event is identified and reported, it is more likely to be studied further and eventually to be considered as being canonical. Moreover, these canonical signaling maps (examples include KEGG and Biocarta) are often assembled from a variety of sources (cell types, tissues, and species). Consequently, these graphical representations may be relevant only in part to the signaling processes under consideration in the specific cell context that a researcher is studying experimentally and trying to model computationally.

Researchers are often limited to applying reductionist wet laboratory technologies to validate the predictions of models that attempt to explain some or all of the complexity in the biological system under investigation. Often, the cost in time and resources needed to validate experimentally the predictions of multiscale models can be prohibitive, making the ability to select among multiple solutions a necessity. Most algorithms, given input variables in the correct format, will produce outputs/predictions, but these often include false positives and false negatives that are not easily identified. While model outputs are usually associated with probability estimations, the results of any significance tests generally provide an evaluation only of how well the model fits the available data. This statistical evaluation is not necessarily an estimate of how well the predictions reflect biological truth. Moreover, when sorting through multiple apparently statistically significant predictions, an investigator can be left relying on subjective intuition, perhaps guided by an incomplete, inadequate, or incorrect understanding of the system. Since model predictions should generally be consistent with the experimental data and/or the (sometimes) limited knowledge of the system currently available, the trap of self-fulfilling prophesy becomes almost unavoidable (Clarke *et al.* 2008).

A gene set enrichment algorithm may produce several predicted pathways and functions associated with a single set of differentially expressed genes. While some genes can certainly participate in more than one pathway or regulate more than one cellular function, the investigator must determine which output(s) (which pathway, module, or function) is most likely to represent the truth. A statistical

assessment, usually a P value, accompanies each model prediction, and understanding what these assessments represent is important in evaluating the results. It is not unusual for a model to provide many predictions for which the P values are small (highly statistically significant), but how easily or appropriately these statistical estimates can be used to guide biological interpretation is not always clear. Primarily, P values reflect how well each model output fits the data input, subject to the parameters and assumptions in the statistical model used. Thus, the use of different statistical tools with the same input gene list and the same database may give different outputs, or the same outputs with different P values, because the parameters and assumptions in each statistical model are different. Also, if some pathways in the database are larger, better annotated, or more fully (and correctly) understood than others, the P values associated with these pathways could be smaller (implying a more statistically significant fit) than less well represented pathways that may be a better reflection of the underlying biological truth.

When the decision as to which is likely to be the correct solution is left entirely to intuition, it is not surprising that the solution that best supports the current hypothesis, or that is most easily explained by the operator's existing knowledge, is often selected over other statistically significant outputs that are not easily understood or may even refute the hypothesis. In such cases, the investigator is likely to fall into the trap of self-fulfilling prophesy (Clarke *et al.* 2008). To be able to interpret model outputs appropriately, it is often critical to understand both what the data represent and some of the basic principles of how the model works. This prior knowledge is particularly important when the correct interpretation is counterintuitive or inconsistent with the hopes or expectations of the study designers.

Cell context, by which we mean the unique patterns of genes, proteins, and metabolites that are expressed in a cell and that interact to influence the physiology of that cell (Clarke & Br nner 1996), is one of the central determinants of how signaling and function are related in biological systems. Context is clearly related to the cell/tissue type, local microenvironment, status of the host, and other external and internal influences. Some aspects of cellular context may be highly conserved, as can be seen from the DNA sequence of some genes through to the basic signaling topology of some highly conserved functions. Nonetheless, there can also be substantial diversity, even within closely related species, tissues, or cells. For cancer research, the differences between the normal and neoplastic state in the same tissue or cell type

is where we look most often for molecular targets that can be diagnostic, prognostic, and/or therapeutic. Here, even small changes in cell context can have substantial implications for the ability to address a specific hypothesis. Despite the often fundamental importance of cellular context, it is frequently ignored.

Some modeling approaches are of limited utility either scientifically or clinically and need to be re-addressed. The complex hairball models often generated by some computational tools may (or may not) contain valid insights into regulatory biology. However, their complexity can be so high and the noise sufficiently extensive and the errors undefined that these models cannot be tested meaningfully or interpreted reliably. Employment of a razor to shave away those components that do not add to the utility, robustness, or accuracy of a model may be desirable, but only if its application is tractable and it is evident what 'whiskers' can be removed without a significant loss of predictive power. Proactively incorporating feature elimination tools during modeling (for example, applying a support vector machine with recursive feature elimination for classification; [Guyon *et al.* 2003](#)) may help to address this concern by attempting to arrive at the smallest model that meets predetermined requirements of convergence and statistical significance. Nonetheless, the need for human intuition to interpret outcomes remains central to many study designs, and, consequently, the risk of falling into the trap of self-fulfilling prophesy must be carefully avoided ([Clarke *et al.* 2008](#)).

Future directions

The properties of high-dimensional data spaces, and the challenges and opportunities these provide ([Clarke *et al.* 2008](#)), remain central to the performance of many computational modeling approaches and bioinformatic tools and workflows. Tools designed to manage these properties explicitly, such as support vector machines, and workflows to address high dimensionality, such as including dimensionality reduction as a preprocessing step in data analysis, are likely to remain in use. New and more powerful tools and workflows are likely to continue to emerge, increasing the power and accuracy of predictive models, the quality and accuracy of data interpretation, and the utility of the new knowledge gained. Deep learning, a subset of approaches within the broader field of machine learning that generally applies neural network-based modeling, has gained recent attention as a

potentially powerful approach to extracting new features from high-dimensional data spaces ([Hosny *et al.* 2018](#), [Zou *et al.* 2019](#)). It is likely that deep-learning approaches will be more commonly applied in the near future to guide knowledge discovery within the framework of cancer systems biology.

Another area that has attracted renewed interest is the heterogeneity arising from the presence of multiple cell types and the consequent complexity of interactions within tumor microenvironments. For molecular signaling, a key issue in this context is whether the events identified as being associated with a biological outcome or phenotype are intrinsic or extrinsic to the cancer cells and/or other cells within the microenvironment. While most therapeutic interventions attempt to induce cell death programs that are executed within the cancer cell (intrinsic), many of the signals that initiate this intrinsic activity are generated by activities originating in stromal or immune cells (extrinsic). Single-cell RNAseq can address some of these issues, but this is not always feasible and the technology has its own limitations ([Cheng *et al.* 2014a](#), [Saliba *et al.* 2014](#)). Moreover, many public omics datasets are populated with data representing averaged signals from multiple cell types, as might be expected from a study that used tumor biopsies as the primary material. Some form of data deconvolution is then required. Tools to achieve deconvolution continue to emerge but for many of these datasets the tools must be effective when applied in an unsupervised manner because data that could supervise the analysis is often absent. Tools that can accurately and robustly perform unsupervised data deconvolution are likely to become more widely used in the near future.

The application of systems biology approaches to critical questions in endocrine-related and other cancers may provide new insights into cancer biology and lead to new treatments. Many signaling networks and the biological processes that they regulate often prove to be too complex for biostatistics, bioinformatics or mathematical biology alone to unravel. However, the integrated use of these approaches can support the building of predictive multiscale models from a systems perspective. The virtuous cycle of *in silico* model prediction, validation in appropriate wet laboratory experiments, with validated results feeding back to improve model predictions, can then drive new discovery of complex systems in a manner that often outstrips intuitive reasoning. In those cancers where hormone and growth factor receptors and their signaling play a major role, systems approaches may offer the best means to address the complexity and dynamic nature of signaling and how it responds to therapeutic

interventions that affect the cancer cells and their interactions within their microenvironments.

Declaration of interest

Robert Clarke is an Associate Editor of *Endocrine-Related Cancer*. Robert Clarke was not involved in the review or editorial process for this paper, on which he is listed as an author. The other authors have nothing to disclose.

Funding

This work was supported in part by Public Health Service Awards U01-CA184902, U54-CA149147, and DoD-BCRP-CA171885 to R Clarke, the Georgetown-Lombardi Comprehensive Cancer Center grant (P30-CA51008-19), R01-CA164717 to M Tan, and R01-CA201092 to WT Baumann.

References

- Ades AE & Lu G 2003 Correlations between parameters in risk models: estimation and propagation of uncertainty by Markov chain Monte Carlo. *Risk Analysis* **23** 1165–1172. (<https://doi.org/10.1111/j.0272-4332.2003.00386.x>)
- Aldridge BB, Burke JM, Lauffenburger DA & Sorger PK 2006 Physicochemical modelling of cell signalling pathways. *Nature Cell Biology* **8** 1195–1203. (<https://doi.org/10.1038/ncb1497>)
- Altrock PM, Liu LL & Michor F 2015 The mathematics of cancer: integrating quantitative models. *Nature Reviews: Cancer* **15** 730–745. (<https://doi.org/10.1038/nrc4029>)
- Anafi RC, Francey LJ, Hogenesch JB & Kim J 2017 CYCLOPS reveals human transcriptional rhythms in health and disease. *PNAS* **114** 5312–5317. (<https://doi.org/10.1073/pnas.1619320114>)
- Anderson AR & Quaranta V 2008 Integrative mathematical oncology. *Nature Reviews: Cancer* **8** 227–234. (<https://doi.org/10.1038/nrc2329>)
- Baeder DY, Yu G, Hoze N, Rolff J & Regoes RR 2016 Antimicrobial combinations: Bliss independence and Loewe additivity derived from mechanistic multi-hit models. *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences* **371** 20150294. (<https://doi.org/10.1098/rstb.2015.0294>)
- Barabasi AL, Gulbahce N & Loscalzo J 2011 Network medicine: a network-based approach to human disease. *Nature Reviews: Genetics* **12** 56–68. (<https://doi.org/10.1038/nrg2918>)
- Barberis M & Verbruggen P 2017 Quantitative systems biology to decipher design principles of a dynamic cell cycle network: the ‘Maximum Allowable mammalian Trade-Off-Weight’ (MAmTOW). *NPJ Systems Biology and Applications* **3** 26. (<https://doi.org/10.1038/s41540-017-0028-x>)
- Barik D, Ball DA, Peccoud J & Tyson JJ 2016 A stochastic model of the yeast cell cycle reveals roles for feedback regulation in limiting cellular variability. *PLoS Computational Biology* **12** e1005230. (<https://doi.org/10.1371/journal.pcbi.1005230>)
- Bedard PL, Mook S, Piccart-Gebhart MJ, Rutgers ET, Van't Veer LJ & Cardoso F 2009 MammaPrint 70-gene profile quantifies the likelihood of recurrence for early breast cancer. *Expert Opinion on Medical Diagnostics* **3** 193–205. (<https://doi.org/10.1517/17530050902751618>)
- Begley CG 2013 Six red flags for suspect work. *Nature* **497** 433–434. (<https://doi.org/10.1038/497433a>)
- Berger JO 2010 *Statistical Decision Theory and Bayesian Analysis*. New York, NY, USA: Springer-Verlag.
- Bhalla US & Iyengar R 1999 Emergent properties of networks of biological signaling pathways. *Science* **283** 381–387. (<https://doi.org/10.1126/science.283.5400.381>)
- Brazma A, Krestyaninova M & Sarkans U 2006 Standards for systems biology. *Nature Reviews: Genetics* **7** 593–605. (<https://doi.org/10.1038/nrg1922>)
- Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, Teichmann SA, Marioni JC & Stegle O 2015 Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nature Biotechnology* **33** 155–160. (<https://doi.org/10.1038/nbt.3102>)
- Califano A 2011 Rewiring makes the difference. *Molecular Systems Biology* **7** 463. (<https://doi.org/10.1038/msb.2010.117>)
- Califano A & Alvarez MJ 2017 The recurrent architecture of tumour initiation, progression and drug sensitivity. *Nature Reviews: Cancer* **17** 116–130. (<https://doi.org/10.1038/nrc.2016.124>)
- Chen L, Chan TH, Choyke PL, Hillman EM, Chi CY, Bhujwala ZM, Wang G, Wang SS, Szabo Z & Wang Y 2011 CAM-CM: a signal deconvolution tool for in vivo dynamic contrast-enhanced imaging of complex tissues. *Bioinformatics* **27** 2607–2609. (<https://doi.org/10.1093/bioinformatics/btr436>)
- Chen C, Baumann WT, Clarke R & Tyson JJ 2013 Modeling the estrogen receptor to growth factor receptor signaling switch in human breast cancer cells. *FEBS Letters* **587** 3327–3334. (<https://doi.org/10.1016/j.febslet.2013.08.022>)
- Chen C, Baumann WT, Xing J, Xu L, Clarke R & Tyson JJ 2014 Mathematical models of the transitions between endocrine therapy responsive and resistant states in breast cancer. *Journal of the Royal Society, Interface* **11** 20140206. (<https://doi.org/10.1098/rsif.2014.0206>)
- Cheng WY, Ou Yang TH & Anastassiou D 2013 Development of a prognostic model for breast cancer survival in an open challenge environment. *Science Translational Medicine* **5** 181ra50. (<https://doi.org/10.1126/scitranslmed.3005974>)
- Cheng F, Jia P, Wang Q, Lin CC, Li WH & Zhao Z 2014a Studying tumorigenesis through network evolution and somatic mutational perturbations in the cancer interactome. *Molecular Biology and Evolution* **31** 2156–2169. (<https://doi.org/10.1093/molbev/msu167>)
- Cheng F, Jia P, Wang Q & Zhao Z 2014b Quantitative network mapping of the human kinome interactome reveals new clues for rational kinase inhibitor discovery and individualized cancer therapy. *Oncotarget* **5** 3697–3710. (<https://doi.org/10.18632/oncotarget.1984>)
- Cirillo E, Parnell LD & Evelo CT 2017 A review of pathway-based analysis tools that visualize genetic variants. *Frontiers in Genetics* **8** 174. (<https://doi.org/10.3389/fgene.2017.00174>)
- Clarke R & Br unner N 1996 Acquired estrogen independence and antiestrogen resistance in breast cancer: estrogen receptor-driven phenotypes? *Trends in Endocrinology and Metabolism* **7** 291–301. ([https://doi.org/10.1016/S1043-2760\(96\)00127-0](https://doi.org/10.1016/S1043-2760(96)00127-0))
- Clarke R, Resson HW, Wang A, Xuan J, Liu MC, Gehan EA & Wang Y 2008 The properties of very high dimensional data spaces: implications for exploring gene and protein expression data. *Nature Reviews: Cancer* **8** 37–49. (<https://doi.org/10.1038/nrc2294>)
- Clarke R, Shajahan AN, Wang Y, Tyson JJ, Riggins RB, Weiner LM, Bauman WT, Xuan J, Zhang B, Facey C, *et al.* 2011 Endoplasmic reticulum stress, the unfolded protein response, and gene network modeling in antiestrogen resistant breast cancer. *Hormone Molecular Biology and Clinical Investigation* **5** 35–44. (<https://doi.org/10.1515/hmbci.2010.073>)
- Clarke R, Cook KL, Hu R, Facey CO, Tavassoly I, Schwartz JL, Baumann WT, Tyson JJ, Xuan J, Wang Y, *et al.* 2012 Endoplasmic reticulum stress, the unfolded protein response, autophagy, and the integrated regulation of breast cancer cell fate. *Cancer Research* **72** 1321–1331. (<https://doi.org/10.1158/0008-5472.CAN-11-3213>)
- Courtney JF 2001 Decision making and knowledge management in inquiring organizations: towards a new decision-making paradigm for DSS. *Decision Support Systems* **31** 17–38. ([https://doi.org/10.1016/S0167-9236\(00\)00117-2](https://doi.org/10.1016/S0167-9236(00)00117-2))

- Cowen L, Ideker T, Raphael BJ & Sharan R 2017 Network propagation: a universal amplifier of genetic associations. *Nature Reviews: Genetics* **18** 551–562. (<https://doi.org/10.1038/nrg.2017.38>)
- Creixell P, Schoof EM, Erler JT & Linding R 2012 Navigating cancer network attractors for tumor-specific therapy. *Nature Biotechnology* **30** 842–848. (<https://doi.org/10.1038/nbt.2345>)
- Creixell P, Reimand J, Haider S, Wu G, Shibata T, Vazquez M, Mustonen V, Gonzalez-Perez A, Pearson J, Sander C, *et al.* 2015 Pathway and network analysis of cancer genomes. *Nature Methods* **12** 615–621. (<https://doi.org/10.1038/nmeth.3440>)
- Deisboeck TS, Wang Z, Macklin P & Cristini V 2011 Multiscale cancer modeling. *Annual Review of Biomedical Engineering* **13** 127–155. (<https://doi.org/10.1146/annurev-bioeng-071910-124729>)
- Di Ventura B, Lemerle C, Michalodimitrakis K & Serrano L 2006 From in vivo to in silico biology and back. *Nature* **443** 527–533. (<https://doi.org/10.1038/nature05127>)
- Dimitrova N, Nagaraj AB, Razi A, Singh S, Kamalakaran S, Banerjee N, Joseph P, Mankovich A, Mittal P, DiFeo A, *et al.* 2017 InFlo: a novel systems biology framework identifies cAMP-CREB1 axis as a key modulator of platinum resistance in ovarian cancer. *Oncogene* **36** 2472–2482. (<https://doi.org/10.1038/onc.2016.398>)
- Dubois D 2010 Representation, propagation, and decision issues in risk analysis under incomplete probabilistic information. *Risk Analysis* **30** 361–368. (<https://doi.org/10.1111/j.1539-6924.2010.01359.x>)
- Dutta A, Le Magnen C, Mitrofanova A, Ouyang X, Califano A & Abate-Shen C 2016 Identification of an NKX3.1-G9a-UTY transcriptional regulatory network that controls prostate differentiation. *Science* **352** 1576–1580. (<https://doi.org/10.1126/science.aad9512>)
- Eckert RS, Carroll RJ & Wang N 1997 Transformations to additivity in measurement error models. *Biometrics* **53** 262–272. (<https://doi.org/10.2307/2533112>)
- Enriquez-Navas PM, Wojtkowiak JW & Gatenby RA 2015 Application of evolutionary principles to cancer therapy. *Cancer Research* **75** 4675–4680. (<https://doi.org/10.1158/0008-5472.CAN-15-1337>)
- Fang HB, Huang H, Clarke R & Tan M 2016 Predicting multi-drug inhibition interactions based on signaling networks and single drug dose-response information. *Journal of Computational Systems Biology* **2** 101.
- Fang HB, Chen X, Pei XY, Grant S & Tan M 2017 Experimental design and statistical analysis for three-drug combination studies. *Statistical Methods in Medical Research* **26** 1261–1280. (<https://doi.org/10.1177/0962280215574320>)
- Fitzgerald JB, Schoeberl B, Nielsen UB & Sorger PK 2006 Systems biology and combination therapy in the quest for clinical efficacy. *Nature Chemical Biology* **2** 458–466. (<https://doi.org/10.1038/nchembio817>)
- Galea MH, Blamey RW, Elston CE & Ellis IO 1992 The Nottingham prognostic index in primary breast cancer. *Breast Cancer Research and Treatment* **22** 207–219. (<https://doi.org/10.1007/BF01840834>)
- Gallaher J, Babu A, Plevritis S & Anderson ARA 2014 Bridging population and tissue scale tumor dynamics: a new paradigm for understanding differences in tumor growth and metastatic disease. *Cancer Research* **74** 426–435. (<https://doi.org/10.1158/0008-5472.CAN-13-0759>)
- Gehlenborg N, O'Donoghue SI, Baliga NS, Goesmann A, Hibbs MA, Kitano H, Kohlbacher O, Neuweger H, Schneider R, Tenenbaum D, *et al.* 2010 Visualization of omics data for systems biology. *Nature Methods* **7** S56–S68. (<https://doi.org/10.1038/nmeth.1436>)
- Gho YS & Lee C 2017 Emergent properties of extracellular vesicles: a holistic approach to decode the complexity of intercellular communication networks. *Molecular BioSystems* **13** 1291–1296. (<https://doi.org/10.1039/c7mb00146k>)
- Ghosh S, Matsuoka Y, Asai Y, Hsin KY & Kitano H 2011 Software for systems biology: from tools to integrated platforms. *Nature Reviews: Genetics* **12** 821–832. (<https://doi.org/10.1038/nrg3096>)
- Gong F & Miller KM 2013 Mammalian DNA repair: HATs and HDACs make their mark through histone acetylation. *Mutation Research* **750** 23–30. (<https://doi.org/10.1016/j.mrfmmm.2013.07.002>)
- GTex Consortium 2015 Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348** 648–660. (<https://doi.org/10.1126/science.1262110>)
- Guyon J, Weston J, Barnhill MD & Vapnik V 2003 Gene selection for cancer classification using support vector machines. *Machine Learning* **46** 389–422.
- Hart Y, Sheftel H, Hausser J, Szekely P, Ben-Moshe NB, Korem Y, Tendler A, Mayo AE & Alon U 2015 Inferring biological tasks using Pareto analysis of high-dimensional data. *Nature Methods* **12** 233–235. (<https://doi.org/10.1038/nmeth.3254>)
- Hatzis C, Bedard PL, Birkbak NJ, Beck AH, Aerts HJ, Stem DF, Shi L, Clarke R, Quackenbush J & Haibe-Kains B 2014 Enhancing reproducibility in cancer drug screening: how do we move forward? *Cancer Research* **74** 4016–4023. (<https://doi.org/10.1158/0008-5472.CAN-14-0725>)
- Herrington DM, Mao C, Parker SJ, Fu Z, Yu G, Chen L, Venkatraman V, Fu Y, Wang Y, Howard TD, *et al.* 2018 Proteomic architecture of human coronary and aortic atherosclerosis. *Circulation* **137** 2741–2756. (<https://doi.org/10.1161/CIRCULATIONAHA.118.034365>)
- Hilakivi-Clarke LA, Warri A, Bouker KB, Zhang X, Cook KL, Jin L, Zwart A, Nguyen N, Hu R, Cruz MI, *et al.* 2017 Effects of *in utero* exposure to ethinyl estradiol on tamoxifen resistance and breast cancer recurrence in a preclinical model. *Journal of the National Cancer Institute* **109** djw188. (<https://doi.org/10.1093/jnci/djw188>)
- Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, Leiserson MDM, Niu B, McLellan MD, Uzunangelov V, *et al.* 2014 Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158** 929–944. (<https://doi.org/10.1016/j.cell.2014.06.049>)
- Hofree M, Shen JP, Carter H, Gross A & Ideker T 2013 Network-based stratification of tumor mutations. *Nature Methods* **10** 1108–1115. (<https://doi.org/10.1038/nmeth.2651>)
- Hosny A, Parmar C, Coroller TP, Grossmann P, Zeleznik R, Kumar A, Bussink J, Gillies RJ, Mak RH & Aerts HJWL 2018 Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. *PLoS Medicine* **15** e1002711. (<https://doi.org/10.1371/journal.pmed.1002711>)
- Hu R, Warri A, Jin L, Zwart A, Riggins R & Clarke R 2015 NFκB signaling is required for XBP1 (U and S) mediated effects on antiestrogen responsiveness and cell fate decisions in breast cancer. *Molecular and Cellular Biology* **35** 390. (<https://doi.org/10.1128/MCB.00847-14>)
- Huang HZ, Fang HB & Tan MT 2018 Experimental designs for multidrug combination studies using signaling networks. *Biometrics* **74** 538–547. (<https://doi.org/10.1111/biom.12777>)
- Hudson NJ, Reverter A & Dalrymple BP 2009 A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. *PLoS Computational Biology* **5** e1000382. (<https://doi.org/10.1371/journal.pcbi.1000382>)
- Ideker T & Krogan NJ 2012 Differential network biology. *Molecular Systems Biology* **8** 565. (<https://doi.org/10.1038/msb.2011.99>)
- Imamov O, Shim GJ, Warner M & Gustafsson JA 2005 Estrogen receptor beta in health and disease. *Biology of Reproduction* **73** 866–871. (<https://doi.org/10.1095/biolreprod.105.043497>)
- Jain HV, Clinton SK, Bhinder A & Friedman A 2011 Mathematical modeling of prostate cancer progression in response to androgen ablation therapy. *PNAS* **108** 19701–19706. (<https://doi.org/10.1073/pnas.1115750108>)
- Janes KA, Chandran PL, Ford RM, Lazzara MJ, Papin JA, Peirce SM, Saucerman JJ & Lauffenburger DA 2017 An engineering design approach to systems biology. *Integrative Biology* **9** 574–583. (<https://doi.org/10.1039/C7IB00014F>)
- Ji Z, Yan K, Li W, Hu H & Zhu X 2017 Mathematical and computational modeling in complex biological systems. *BioMed Research International* **2017** 5958321. (<https://doi.org/10.1155/2017/5958321>)

- Jinawath N, Bunbanjerdusuk S, Chayanupatkul M, Ngamphaiboon N, Asavapanumas N, Svasti J & Charoensawan V 2016 Bridging the gap between clinicians and systems biologists: from network biology to translational biomedical research. *Journal of Translational Medicine* **14** 324. (<https://doi.org/10.1186/s12967-016-1078-3>)
- Junttila MR & de Sauvage FJ 2013 Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature* **501** 346–354. (<https://doi.org/10.1038/nature12626>)
- Kanehisa M & Goto S 2000 KEGG: Kyoto Encyclopedia of genes and genomes. *Nucleic Acids Research* **28** 27–30. (<https://doi.org/10.1093/nar/28.1.27>)
- Kanter I & Kalisky T 2015 Single cell transcriptomics: methods and applications. *Frontiers in Oncology* **5** 53. (<https://doi.org/10.3389/fonc.2015.00053>)
- Keenan AB, Jenkins SL, Jagodnik KM, Koplev S, He E, Torre D, Wang Z, Dohlman AB, Silverstein MC, Lachmann A, *et al.* 2018 The library of integrated network-based cellular signatures NIH program: system-level cataloging of human cells response to perturbations. *Cell Systems* **6** 13–24. (<https://doi.org/10.1016/j.cels.2017.11.001>)
- Kimura A, Celani A, Nagao H, Stasevich T & Nakamura K 2015 Estimating cellular parameters through optimization procedures: elementary principles and applications. *Frontiers in Physiology* **6** 60. (<https://doi.org/10.3389/fphys.2015.00060>)
- Lampinen J & Vehtari A 2001 Bayesian approach for neural networks – review and case studies. *Neural Networks* **14** 257–274. ([https://doi.org/10.1016/S0893-6080\(00\)00098-8](https://doi.org/10.1016/S0893-6080(00)00098-8))
- Lavrik IN & Zhivotovskiy B 2014 Systems biology: a way to make complex problems more understandable. *Cell Death and Disease* **5** e1256. (<https://doi.org/10.1038/cddis.2014.195>)
- Le NN 2007 The long journey to a systems biology of neuronal function. *BMC Systems Biology* **1** 28. (<https://doi.org/10.1186/1752-0509-1-28>)
- Leder K, Pitter K, Laplant Q, Hambardzumyan D, Ross BD, Chan TA, Holland EC & Michor F 2014 Mathematical modeling of PDGF-driven glioblastoma reveals optimized radiation dosing schedules. *Cell* **156** 603–616. (<https://doi.org/10.1016/j.cell.2013.12.029>)
- Ledzewicz U & Schaeffler H 2016 Optimizing chemotherapeutic anti-cancer treatment and the tumor microenvironment: an analysis of mathematical models. *Advances in Experimental Medicine and Biology* **936** 209–223. (https://doi.org/10.1007/978-3-319-42023-3_11)
- Leiserson MD, Vandin F, Wu HT, Dobson JR, Eldridge JV, Thomas JL, Papoutsaki A, Kim Y, Niu B, McLellan M, *et al.* 2015 Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nature Genetics* **47** 106–114. (<https://doi.org/10.1038/ng.3168>)
- Li H, Xuan J, Wang Y & Zhan M 2008 Inferring regulatory networks. *Frontiers in Bioscience* **13** 263–275. (<https://doi.org/10.2741/2677>)
- Liepe J, Kirk P, Filippi S, Toni T, Barnes CP & Stumpf MP 2014 A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nature Protocols* **9** 439–456. (<https://doi.org/10.1038/nprot.2014.025>)
- Mackay A, Weigelt B, Grigoriadis A, Kreike B, Natrajan R, A'Hern R, Tan DS, Dowsett M, Ashworth A & Reis-Filho JS 2011 Microarray-based class discovery for molecular classification of breast cancer: analysis of interobserver agreement. *Journal of the National Cancer Institute* **103** 662–673. (<https://doi.org/10.1093/jnci/djr071>)
- Mangado N, Piella G, Noailly J, Pons-Prats J & Ballester MÁ 2016 Analysis of uncertainty and variability in finite element computational models for biomedical engineering: characterization and propagation. *Frontiers in Bioengineering and Biotechnology* **4** 85. (<https://doi.org/10.3389/fbioe.2016.00085>)
- Martelotto LG, Ng CK, Piscuoglio S, Weigelt B & Reis-Filho JS 2014 Breast cancer intra-tumor heterogeneity. *Breast Cancer Research* **16** 210. (<https://doi.org/10.1186/bcr3658>)
- Masoudi-Nejad A, Bidkhorji G, Hosseini Ashtiani S, Najafi A, Bozorgmehr JH & Wang E 2015 Cancer systems biology and modeling: microscopic scale and multiscale approaches. *Seminars in Cancer Biology* **30** 60–69. (<https://doi.org/10.1016/j.semcancer.2014.03.003>)
- McKenna MT, Weis JA, Barnes SL, Tyson DR, Miga MI, Quaranta V & Yankeelov TE 2017 A predictive mathematical modeling approach for the study of doxorubicin treatment in triple negative breast cancer. *Scientific Reports* **7** 5725. (<https://doi.org/10.1038/s41598-017-05902-z>)
- Meacham CE & Morrison SJ 2013 Tumour heterogeneity and cancer cell plasticity. *Nature* **501** 328–337. (<https://doi.org/10.1038/nature12624>)
- Miryala SK, Anbarasu A & Ramaiah S 2018 Discerning molecular interactions: a comprehensive review on biomolecular interaction databases and network analysis tools. *Gene* **642** 84–94. (<https://doi.org/10.1016/j.gene.2017.11.028>)
- Mitra K, Carvunis AR, Ramesh SK & Ideker T 2013 Integrative approaches for finding modular structure in biological networks. *Nature Reviews: Genetics* **14** 719–732. (<https://doi.org/10.1038/nrg3552>)
- Mobley A, Linder SK, Braeuer R, Ellis LM & Zwelling L 2013 A survey on data reproducibility in cancer research provides insights into our limited ability to translate findings from the laboratory to the clinic. *PLoS One* **8** e63221. (<https://doi.org/10.1371/journal.pone.0063221>)
- Molinelli EJ, Korkut A, Wang W, Miller ML, Gauthier NP, Jing X, Kaushik P, He Q, Mills G, Solit DB, *et al.* 2013 Perturbation biology: inferring signaling networks in cellular systems. *PLoS Computational Biology* **9** e1003290. (<https://doi.org/10.1371/journal.pcbi.1003290>)
- Morken JD, Packer A, Everett RA, Nagy JD & Kuang Y 2014 Mechanisms of resistance to intermittent androgen deprivation in patients with prostate cancer identified by a novel computational method. *Cancer Research* **74** 3673–3683. (<https://doi.org/10.1158/0008-5472.CAN-13-3162>)
- Moseley HN 2013 Error analysis and propagation in metabolomics data analysis. *Computational and Structural Biotechnology Journal* **4** e201301006. (<https://doi.org/10.5936/csbj.201301006>)
- Nam S 2017 Databases and tools for constructing signal transduction networks in cancer. *BMB Reports* **50** 12–19. (<https://doi.org/10.5483/BMBRep.2017.50.1.135>)
- Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M & Alizadeh AA 2015 Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods* **12** 453–457. (<https://doi.org/10.1038/nmeth.3337>)
- Okasha S 2012 Emergence, hierarchy and top-down causation in evolutionary biology. *Interface Focus* **2** 49–54. (<https://doi.org/10.1098/rsfs.2011.0046>)
- Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, *et al.* 2009 Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology* **27** 1160–1167. (<https://doi.org/10.1200/JCO.2008.18.1370>)
- Parmar JH, Cook KL, Shajahan-Haq AN, Clarke PA, Tavassoly I, Clarke R, Tyson JJ & Baumann WT 2013 Modelling the effect of GRP78 on anti-oestrogen sensitivity and resistance in breast cancer. *Interface Focus* **3** 20130012. (<https://doi.org/10.1098/rsfs.2013.0012>)
- Pavlopoulos GA, Paez-Espino D, Kyripides NC & Iliopoulos I 2017 Empirical comparison of visualization tools for larger-scale network analysis. *Advances in Bioinformatics* **2017** 1278932. (<https://doi.org/10.1155/2017/1278932>)
- Peng H, Tan H, Zhao W, Jin G, Sharma S, Xing F, Watabe K & Zhou X 2016 Computational systems biology in cancer brain metastasis. *Frontiers in Bioscience* **8** 169–186. (<https://doi.org/10.2741/s456>)
- Picco N, Gatenby RA & Anderson ARA 2017 Stem cell plasticity and niche dynamics in cancer progression. *IEEE Transactions on Bio-Medical Engineering* **64** 528–537. (<https://doi.org/10.1109/TBME.2016.2607183>)
- Ponten F, Schwenk JM, Asplund A & Edqvist PH 2011 The Human Protein Atlas as a proteomic resource for biomarker discovery. *Journal*

- of *Internal Medicine* **270** 428–446. (<https://doi.org/10.1111/j.1365-2796.2011.02427.x>)
- Quaranta V, Rejniak KA, Gerlee P & Anderson AR 2008 Invasion emerges from cancer cell adaptation to competitive microenvironments: quantitative predictions from multiscale mathematical models. *Seminars in Cancer Biology* **18** 338–348. (<https://doi.org/10.1016/j.semcancer.2008.03.018>)
- Rittle HWJ & Webber MM 1973 Dilemmas in a general theory of planning. *Policy Sciences* **4** 155–169. (<https://doi.org/10.1007/BF01405730>)
- Robinson JT, Thorvaldsdottir H, Wenger AM, Zehir A & Mesirov JP 2017 Variant review with the integrative genomics viewer. *Cancer Research* **77** e31–e34. (<https://doi.org/10.1158/0008-5472.CAN-17-0337>)
- Roy S, Werner-Washburne M & Lane T 2011 A multiple network learning approach to capture system-wide condition-specific responses. *Bioinformatics* **27** 1832–1838. (<https://doi.org/10.1093/bioinformatics/btr270>)
- Ryall KA & Tan AC 2015 Systems biology approaches for advancing the discovery of effective drug combinations. *Journal of Cheminformatics* **7** 7. (<https://doi.org/10.1186/s13321-015-0055-9>)
- Saliba AE, Westermann AJ, Gorski SA & Vogel J 2014 Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Research* **42** 8845–8860. (<https://doi.org/10.1093/nar/gku555>)
- Sandberg R 2014 Entering the era of single-cell transcriptomics in biology and medicine. *Nature Methods* **11** 22–24. (<https://doi.org/10.1038/nmeth.2764>)
- Sedgewick AJ, Benz SC, Rabizadeh S, Soon-Shiong P & Vaske CJ 2013 Learning subgroup-specific regulatory interactions and regulator independence with PARADIGM. *Bioinformatics* **29** i62–i70. (<https://doi.org/10.1093/bioinformatics/btt229>)
- Singhania R, Sramkoski RM, Jacobberger JW & Tyson JJ 2011 A hybrid model of mammalian cell cycle regulation. *PLoS Computational Biology* **7** e1001077. (<https://doi.org/10.1371/journal.pcbi.1001077>)
- Spencer SL & Sorger PK 2011 Measuring and modeling apoptosis in single cells. *Cell* **144** 926–939. (<https://doi.org/10.1016/j.cell.2011.03.002>)
- Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y & Habbema JD 2001 Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology* **54** 774–781. ([https://doi.org/10.1016/S0895-4356\(01\)00341-9](https://doi.org/10.1016/S0895-4356(01)00341-9))
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, *et al.* 2005 Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* **102** 15545–15550. (<https://doi.org/10.1073/pnas.0506580102>)
- Swertz MA & Jansen RC 2007 Beyond standardization: dynamic software infrastructures for systems biology. *Nature Reviews: Genetics* **8** 235–243. (<https://doi.org/10.1038/nrg2048>)
- Tang J & Aittokallio T 2014 Network pharmacology strategies toward multi-target anticancer therapies: from computational models to experimental design principles. *Current Pharmaceutical Design* **20** 23–36. (<https://doi.org/10.2174/13816128113199990470>)
- Tang J, Cho NW, Cui G, Manion EM, Shanbhag NM, Botuyan MV, Mer G & Greenberg RA 2013 Acetylation limits 53BP1 association with damaged chromatin to promote homologous recombination. *Nature Structural and Molecular Biology* **20** 317–325. (<https://doi.org/10.1038/nsmb.2499>)
- Tape CJ 2016 Systems biology analysis of heterocellular signaling. *Trends in Biotechnology* **34** 627–637. (<https://doi.org/10.1016/j.tibtech.2016.02.016>)
- Tavassoly I, Parmar J, Shajahan-Haq AN, Clarke R, Baumann WT & Tyson JJ 2015 Dynamic modeling of the interaction between autophagy and apoptosis in mammalian cells. *CPT: Pharmacometrics and Systems Pharmacology* **4** 263–272. (<https://doi.org/10.1002/psp4.29>)
- Tian Y, Zhang B, Shih IM & Wang Y 2011 Knowledge-guided differential dependency network learning for detecting structural changes in biological networks. In *BCB '11: Proceedings of the 2nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, pp 254–263. New York, NY, USA: ACM. (<https://doi.org/10.1145/2147805.2147833>)
- Tian Y, Chen L, Zhang B, Zhang Z, Yu G, Clarke R, Xuan J, Shih IM & Wang Y 2013 Genomic and network analysis to study the origin of ovarian cancer. *Systems Biomedicine* **1** 55–64. (<https://doi.org/10.4161/sysb.25313>)
- Tian Y, Wang SS, Zhang Z, Rodriguez OC, Petricoin E, III, Shih IeM, Chan D, Avantaggiati M, Yu G, Ye S, *et al.* 2014a Integration of network biology and imaging to study cancer phenotypes and responses. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **11** 1009–1019. (<https://doi.org/10.1109/TCBB.2014.2338304>)
- Tian Y, Zhang B, Hoffman EP, Clarke R, Zhang Z, Shih IeM, Xuan J, Herrington DM & Wang Y 2014b Knowledge-fused differential dependency network models for detecting significant rewiring in biological networks. *BMC Systems Biology* **8** 87. (<https://doi.org/10.1186/s12918-014-0087-1>)
- Tian Y, Zhang B, Hoffman EP, Clarke R, Zhang Z, Shih IeM, Xuan J, Herrington DM & Wang Y 2015 KDDN: an open-source cytoscape app for constructing differential dependency networks with significant rewiring. *Bioinformatics* **31** 287–289. (<https://doi.org/10.1093/bioinformatics/btu632>)
- Toettcher JE, Loewer A, Ostheimer GJ, Yaffe MB, Tidor B & Lahav G 2009 Distinct mechanisms act in concert to mediate cell cycle arrest. *PNAS* **106** 785–790. (<https://doi.org/10.1073/pnas.0806196106>)
- Twycross J, Band LR, Bennett MJ, King JR & Krasnogor N 2010 Stochastic and deterministic multiscale models for systems biology: an auxin-transport case study. *BMC Systems Biology* **4** 34. (<https://doi.org/10.1186/1752-0509-4-34>)
- Tyson JJ, Baumann WT, Chen C, Verdugo A, Tavassoly I, Wang Y, Weiner LM & Clarke R 2011 Dynamic modeling of oestrogen signalling and cell fate in breast cancer cells. *Nature Reviews Cancer* **11** 523–532. (<https://doi.org/10.1038/nrc3081>)
- Tyson JJ, Laomettachit T & Kraikivski P 2019 Modeling the dynamic behavior of biochemical regulatory networks. *Journal of Theoretical Biology* **462** 514–527. (<https://doi.org/10.1016/j.jtbi.2018.11.034>)
- Vanlier J, Tiemann CA, Hilbers PA & van Riel NA 2012 An integrated strategy for prediction uncertainty analysis. *Bioinformatics* **28** 1130–1135. (<https://doi.org/10.1093/bioinformatics/bts088>)
- Venet D, Dumont JE & Detours V 2011 Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS Computational Biology* **7** e1002240. (<https://doi.org/10.1371/journal.pcbi.1002240>)
- Vogel C & Marcotte EM 2012 Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews: Genetics* **13** 227–232. (<https://doi.org/10.1038/nrg3185>)
- Waljee AK, Higgins PD & Singal AG 2014 A primer on predictive models. *Clinical and Translational Gastroenterology* **5** e44. (<https://doi.org/10.1038/ctg.2013.19>)
- Wang Z & Deisboeck TS 2014 Mathematical modeling in cancer drug discovery. *Drug Discovery Today* **19** 145–150. (<https://doi.org/10.1016/j.drudis.2013.06.015>)
- Wang N, Gong T, Clarke R, Chen L, Shih IeM, Zhang Z, Levine DA, Xuan J & Wang Y 2015 UNDO: a Bioconductor R package for unsupervised deconvolution of mixed gene expressions in tumor samples. *Bioinformatics* **31** 137–139. (<https://doi.org/10.1093/bioinformatics/btu607>)
- Wang N, Hoffman EP, Chen L, Chen L, Zhang Z, Liu C, Yu G, Herrington DM, Clarke R & Wang Y 2016 Mathematical modelling of transcriptional heterogeneity identifies novel markers and subpopulations in complex tissues. *Scientific Reports* **6** 18909. (<https://doi.org/10.1038/srep18909>)

- Welton NJ & Ades AE 2005 Estimation of Markov chain transition probabilities and rates from fully and partially observed data: uncertainty propagation, evidence synthesis, and model calibration. *Medical Decision Making* **25** 633–645. (<https://doi.org/10.1177/0272989X05282637>)
- Wen Z, Liu ZP, Liu Z, Zhang Y & Chen L 2013 An integrated approach to identify causal network modules of complex diseases with application to colorectal cancer. *Journal of the American Medical Informatics Association* **20** 659–667. (<https://doi.org/10.1136/amiajnl-2012-001168>)
- Werner HM, Mills GB & Ram PT 2014 Cancer systems biology: a peek into the future of patient care? *Nature Reviews: Clinical Oncology* **11** 167–176. (<https://doi.org/10.1038/nrclinonc.2014.6>)
- Wilkinson DJ 2009 Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews: Genetics* **10** 122–133. (<https://doi.org/10.1038/nrg2509>)
- Wu G & Stein L 2012 A network module-based method for identifying cancer prognostic signatures. *Genome Biology* **13** R112. (<https://doi.org/10.1186/gb-2012-13-12-r112>)
- Yin H, Xue W & Anderson DG 2019 CRISPR-Cas: a tool for cancer research and therapeutics. *Nature Reviews Clinical Oncology* **16** 281–295. (<https://doi.org/10.1038/s41571-019-0166-8>)
- Zhang B & Wang Y 2010 Learning structural changes of gaussian graphical models in controlled experiments. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence, UAI 2010*, pp 701–708. Waterloo, ON, Canada: AUAI Press. (available at: https://event.cwi.nl/uai2010/papers/UAI2010_0221.pdf)
- Zhang B, Li H, Riggins RB, Zhan M, Xuan J, Zhang Z, Hoffman EP, Clarke R & Wang Y 2009 Differential dependency network analysis to identify condition-specific topological changes in biological networks. *Bioinformatics* **25** 526–532. (<https://doi.org/10.1093/bioinformatics/btn660>)
- Zhang B, Tian Y, Jin L, Li H, Shih IeM, Madhavan S, Clarke R, Hoffman EP, Xuan J, Hilakivi-Clarke L, *et al.* 2011 DDN: a caBIG(R) analytical tool for differential network analysis. *Bioinformatics* **27** 1036–1038. (<https://doi.org/10.1093/bioinformatics/btr052>)
- Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE, Zhou JY, Petyuk VA, Chen L, Ray D, *et al.* 2016 Integrated proteogenomic characterization of human high grade serous ovarian cancer. *Cell* **166** 755–765. (<https://doi.org/10.1016/j.cell.2016.05.069>)
- Zou J, Huss M, Abid A, Mohammadi P, Torkamani A & Telenti A 2019 A primer on deep learning in genomics. *Nature Genetics* **51** 12–18. (<https://doi.org/10.1038/s41588-018-0295-5>)
- Zuckerman NS, Noam Y, Goldsmith AJ & Lee PP 2013 A self-directed method for cell-type identification and separation of gene expression microarrays. *PLoS Computational Biology* **9** e1003189. (<https://doi.org/10.1371/journal.pcbi.1003189>)

Received in final form 25 March 2019

Accepted 8 April 2019

Accepted Preprint published online 8 April 2019