Clinical Chemistry 65:2
342–344 (2019)

Citation Classic

A Blueprint for Systems Biology

Trey Ideker[1] and Leroy Hood[2,3*]

Featured Article: Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, et al. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 2001;292:929–34.[4]

By the mid-1990s, it was clear that we would soon be able to sequence the human genome and that this would enable many other types of molecular measurements in cells and tissues. Less apparent was how these different molecular technologies and their data should be integrated to map biological structure and understand function. That is, having systematically sequenced the DNA bases in a genome, could similar systematic concepts and tools be devised to understand the rest of the human biological informational system (RNA, proteins, metabolites, lipids) and their roles in biology?

Around that time, one of us (LH) had moved from Caltech to start a new department, called Molecular Biotechnology, in the medical school of the University of Washington. This department was based on the radical idea (at that time) that biology at its core was an information science, and that the path to understanding and integrating the genome and other types of biological information would be enabled only by joining the skills and expertise of biologists with those of leading investigators from engineering, chemistry, and mathematics, as well as the physical and computer sciences. This concept captured the imagination of the other of us (TI), who had studied computer science as an undergraduate at the Massachusetts Institute of Technology and, since that time, had been working in signal processing and circuit design in the defense industry. Trey was one of the first to sign up, having heard of Lee's new program that promised to train engineers in the life sciences.

In Seattle, Trey indeed began to accumulate biological knowledge and learn how to wield the increasingly powerful tools of experimental genetics and biochemistry. Importantly, however, he did not forget his engineering mind-set. Key questions emerged as to why, if cells were indeed biological computers in the midst of processing information, were the methods of biology and engineering so apparently different from one another? In engineering there is typically a circuit diagram to explain function. Likewise, in biology, we speculated that there must be biological circuits (later called biological networks) that provide the causal information flow to mediate development, physiology, and aging. Presumably, if these networks become perturbed, they could cause diseases. Although in biology these molecular circuits had not previously been delineated, given the right types and quantities of data and proper tools of data integration, biological circuits (or networks) might nonetheless be determined employing reverse engineering; that is, going from data to circuit delineation.

Guided by this philosophy, we first selected a model biological system to reverse engineer, and then we began to enumerate the types of systematic data we might collect. As a suitable target, we chose the galactose (*GAL*) gene regulatory circuit in the budding yeast *Saccharomyces cerevisiae (1)*, whose well-documented molecular interactions (e.g., protein/DNA and protein/protein connections) provided a critical positive control for any systematic approach we, or others, might develop.

The next question concerned the data. The yeast genome had just been sequenced *(2)*, allowing for the identification of all yeast genes and enabling the creation of DNA microarrays to measure the expression level of all gene transcripts. Knowledge of complete genomes was also leading to a revolution in protein-based mass spectrometry by enabling a direct search for and quantification of mass spectra matching every potential peptide encoded by the proteins of a genome *(3)*.

On the other hand, we were concerned that biological complexity was great enough that it would not be possible to infer a biological system simply from observational data such as transcript quantifications or protein abundances. For this reason, we designed a systematic panel of genetic knockouts (or biological perturbations) disrupting all genes individually known or suspected to be involved in *GAL* metabolism or pathway regulation, with global transcriptome and proteome profiles collected in response to each of these perturbations. Experimental perturbations always lie at the heart of systems biology's ability to decipher biological complexity. This causal data set was integrated with systematic maps of protein–DNA and protein–protein interactions that were also coming online at that time *(4, 5)*, to create a comprehensive and a dynamical model of the *GAL* regulatory pathway that explained much of its biology. Thus,

[1] Division of Genetics, Department of Medicine, UC San Diego, La Jolla, CA; [2] Institute for Systems Biology, Seattle, WA; [3] Providence Saint Joseph Health, Renton, WA.
* Address correspondence to this author at: Institute for Systems Biology, Seattle, WA 98109. Fax 206-732-1299; e-mail lhood@systemsbiology.org.
[4] This article has been cited more than 1350 times since publication.
Received August 23, 2018; accepted September 11, 2018.
Previously published online at DOI: 10.1373/clinchem.2018.291062

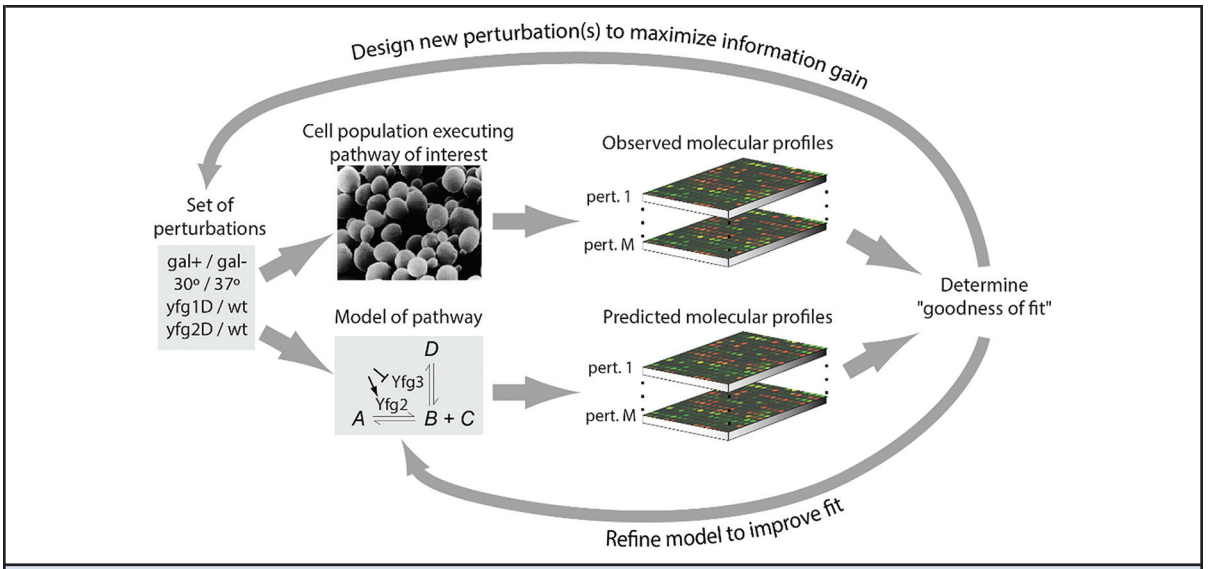

Fig. 1. Original conception of the systems biology approach, originally published in Ideker et al. *(6)*.

biological perturbations in conjunction with multiomic data were essential to understanding biological mechanisms as reflected by dynamical biological networks of *GAL* metabolism (Fig. 1).

The above systems approach, as originally used to study *GAL* metabolism, has since served as an archetype for many studies in the emerging discipline that would soon be called Systems Biology *(6)*. Networks and their dynamics are now a central concept in mainstream biological research along with network analysis software like Cytoscape *(7)*, which was originally developed to model the *GAL* gene regulatory network in our 2001 article in *Science*.

Systematic approaches to study gene regulatory circuits quickly spread, influencing the design of large consortium projects such as ENCODE (Encyclopedia of DNA Elements) *(8)* and GTEx (Genotype-Tissue Expression) *(9)*. We also moved toward elucidating the structure and function of disease-perturbed networks. For example, in the first reported study of disease-perturbed networks, prion-induced neurodegeneration in mice, disease-perturbed transcriptional networks from the brain were analyzed during disease progression. These networks explained virtually every aspect of the disease pathology *(10)*. Use of multiple-omics layers, initially integrated and modeled as a powerful explanatory network in the article discussed here, transitioned from the exception to the rule and were subsequently critical to the success of large integrated resources such as the Cancer Genome Atlas *(11)* and the Human Scientific Wellness Program *(12)*.

While our work was undergoing peer review, Lee left his post at the University of Washington to cofound an independent research institute, the first Institute for Systems Biology. Others followed his lead, starting a network of systems biology departments, institutes, and centers worldwide.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 4 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.*

L. Hood, administrative support.

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: T. Ideker, Data4Cure.
Consultant or Advisory Role: T. Ideker, Ideaya, Data4Cure.
Stock Ownership: T. Ideker, Ideaya, Data4Cure.
Honoraria: T. Ideker, Ideaya.
Research Funding: None declared.
Expert Testimony: None declared.
Patents: None declared.

References

1. Johnston M. A model fungal gene regulatory mechanism: the GAL genes of Saccharomyces cerevisiae. Microbiol Rev 1987;51:458–76.
2. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, et al. Life with 6000 Genes. Science 1996;274:546–67.
3. Dongré AR, Eng JK, Yates JR 3rd. Emerging tandem-mass-spectrometry techniques for the rapid identification of proteins. Trends Biotechnol 1997;15:418–25.
4. Ren B. Genome-wide location and function of DNA binding proteins. Science 2000; 290:2306–9.

5. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature 2000; 403:623–7.

6. Ideker T, Galitski T, Hood L. A new approach to decoding life: systems biology. Ann Rev Genomics Hum Genet 2001;2:343–72.

7. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.

8. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57–74.

9. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015;348:648–60.

10. Hwang D, Lee IY, Yoo H, Gehlenborg N, Cho JH, Petritis B, et al. A systems approach to prion disease. Mol Syst Biol 2009;5:252.

11. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 2008;455: 1061–8.

12. Price ND, Magis AT, Earls JC, Glusman G, Levy R, Lausted C, et al. A wellness study of 108 individuals using personal, dense, dynamic data clouds. Nat Biotechnol 2017; 35:747–56.