SYSTEMS BIOLOGY: USING SYSTEMS APPROACHES TO MANAGE BIOLOGICAL COMPLEXITY

Joel F. Moxley and Gregory Stephanopoulos*
Massachusetts Institute of Technology
Cambridge, MA 02139

Isidore Rigoutsos
IBM TJ Watson Research Center
Yorktown Heights, New York 10598

Abstract

Excitement surrounds systems biology because we hope the emerging field will interpret and contextualize large, diverse sets of biological measurements and, in doing so, elucidate the mechanisms behind complex phenomena not apparent without an integrated perspective. However, we can easily misinterpret the true nature of systems biology and the likely current impacts on advancing the state of knowledge in biotechnology and biomedical research. In this paper, we offer one view of this exciting area and accompanying realistic near-term expectations.

Keywords

Systems biology.

Introduction

As recently as a decade ago, the core paradigm of molecular biology followed an established path. Namely, a well-defined hypothesis spawned a well-defined experiment measuring a few known genes, proteins, and perhaps metabolites of the target system. However, genome sequencing and derived technologies such as gene expression arrays expanded our cellular view with a bevy of new data. Suddenly with a key to the digital code at biology's core, we became able to identify and measure important classes of intracellular molecules like gene transcripts and proteins.

These developments revealed the obvious: hundreds or thousands of molecules that originally were not the focus of the traditional experiment were also found to vary significantly during the experiment. Lee Hood popularized the term “systems biology” to describe an updated, global approach to cellular understanding that considers these now accessible genome-scale measurements and proceeded to form an institute dedicated to the pursuit.

Here we define systems biology as the field aiming to establish connections between and among important classes of molecules so as to aid the mechanistic explanation of cellular processes. More specifically, systems biology identifies concrete molecular relationships for targeted analysis through the interpretation of cellular phenotype in terms of integrated biomolecular networks.

At the logical starting point of the systems biology hierarchy, one finds the output from a host of other computational biology methods and processes that are possible in our post-genomic era; for example, bioinformatic sequence analysis and comparative genomics continue to advance our understanding of gene sequence data. At higher levels of the hierarchy, data from gene expression experiments and analysis of experimental protein-protein interaction screens, and

* To whom all correspondence should be addressed.

protein-DNA binding information allow us to shed light upon biomolecular networks. These networks denote possible mechanistic avenues that may be compared across organisms to highlight common and differing pathways. Similarly, large-scale kinetic models may reproduce macroscopic cellular behavior such as the cell cycle. In each case, the insight gained from this work diffuses into the public literature, and upon scrutiny, may be used to refine databases of biomolecular networks and more effectively measure states.

System biology as practiced at the biomolecular network level differs from other computational biology counterparts through the simultaneous, systems-level integration of both networks and states. Thus we must possess both network and state data to identify specific molecular relationships, and much of this information is provided by genomic, proteomic and transcriptomic levels of the hierarchy.

Traditionally, we have understood cells as elegant coordinated systems, but the sequenced genomes made possible the systems-level probing of both concerted molecular relationships and the resulting cellular states. Far from replacing specific low-level understanding explored piecewise in the past, these systematic probes rather hasten the discovery of unknown molecular relationships and place the existing knowledge in a greater context.

Thus one can see that systems biology is a framework in which practitioners can leverage all existing targeted analysis techniques and accumulated data. But if one puts metabolism aside, gene expression states and biomolecular interaction screens produce large, noisy data sets even for relatively simple organisms and at the same time present countless avenues for analysis. How do we approach this data, and what can we hope to accomplish?

As a framework, systems biology aims at providing greater understanding for specific cellular mechanisms. Traditionally, we have judged our understanding by the ability to forward-engineer specific desired responses. For example, in humans, we hope to prevent and combat disease. Analogously, in microbes, we hope to maximize product generation. But when the only thing that is available is large, noisy data sets, how can we select a specific analysis so as to gain insight for specific mechanisms?

To answer this last question, we must begin by anchoring systems biology analyses to specific questions and build upon the existing, strong core infrastructure of targeted analyses. In the context of a well-studied system, it is much easier to identify potentially fruitful data characteristics that, upon detailed study, can lead to a gain of our knowledge. Batteries of induced, well-defined perturbations (such as gene deletions or environmental changes) to a base system, represent an effective way to attribute specific responses to specific mechanisms.

In summary, systems biology offers a powerful paradigm for probing cellular processes, and eventually gaining mechanistic insight. As engineers with a rich heritage of systems analysis, we stand poised to take full advantage of this opportunity.

Prior Systems Approaches to Biological Research

Although the term “systems biology” only recently entered the popular lexicon, engineers have applied integrated, systemic approaches to understand cellular processes for many years. From our laboratory's perspective, we employ the following approaches that use systems-level analysis to forward-engineer desired responses. Although the approaches encompass comparatively smaller systems, we may incorporate the lessons from each as we move forward at a larger scale.

Metabolic Control Analysis

Metabolic pathways and, in general networks of reactions, are characterized by substantial stoichiometric and (mostly) kinetic complexity in their own right. A commonly applied assumption to these systems is that of a single rate limiting step leading to great simplification of the reaction network and often yielding analytical expressions for the conversion rates. This assumption, however, is not justified for most biological systems for which kinetic control is not concentrated in a single step but rather distributed among several enzymatic steps. As a result, instead of a single rate limiting step, a more appropriate measure of the kinetic effect of each enzyme in a pathway is the flux control coefficient, simply defined as the change impacted on the pathway flux in response to an infinitesimal change in the concentration of the enzyme. Flux control coefficients are properties of the entire system, and theorems of Metabolic Control Analysis (MCA) show how these coefficients can be determined from experiments or models of the local network elements, i.e kinetics of the pathway enzymes. The concepts of MCA and distribution of kinetic control in a reaction pathway have had profound impact on the identification of target enzymes for genetic modification for the purpose of amplifying the product flux through a pathway (Kacser and Burns, 1973; Stephanopoulos et al., 1998). Most importantly, they rationalized the need for multiple enzyme amplification to achieve the goal of flux amplification, as verified experimentally recently (Koffas et al., 2003).

Signaling Pathways

Signal transduction is the process by which cells communicate with each other and their environment. They involve a multitude of proteins that may be present in active and inactive states. In their active (phosphorylated) state they act as catalysts for the phosphorylation (i.e., activation) of subsequent steps in the signaling cascade. The end result is the activation of a transcription factor that, in turn, activates gene transcription events. Despite the fact that several of the known proteins participate in more than one signaling cascades, such systems were

being studied until recently in isolation from one another. A corollary of this approach was, of course, that a single gene can be activated by a single ligand, or that a single ligand leads to the activation of a single gene. Neither conclusion is true, and signaling pathways branch and interact with one another creating a rather intricate and complex signaling network. The usual approaches for studying such systems based on mass action kinetics of the individual activation reactions are rather inadequate for the treatment of the overall network, both from a parameter identification and data validation standpoint. A different approach, making use of activation ratios for the analysis of integrated signaling networks was proposed recently (Femenia, 2004). Clearly this is an area of immense importance in biomedical mechanism identification and drug discovery and more tools are required both in the computational and experimental areas. These are the goals of the recently formed Alliance for Cellular Signaling, a NIH funded project involving several laboratories and research centers (www.signaling-gateway.org).

Flux Map Reconstruction

Metabolic pathway fluxes, defined as the actual rates of metabolite interconversion in metabolic network, are most informative measures of the actual physiological state of cells and organisms. As they depend on enzymatic activities as well as metabolite concentrations, such fluxes provide an accurate representation of carbon and energy flows through the various pathway branches and they are important in identifying critical reaction steps that impact flux control for the entire pathway. Flux determination is thus an essential component of strain evaluation and metabolic engineering (Stephanopoulos, 1999).

Intracellular flux determination requires the enumeration and satisfaction of all intracellular metabolite balances along with the use of sufficient measurements typically derived from the introduction of isotopic tracers and metabolite and mass isotopomer measurement by Gas Chromatography-Mass Spectrometry. It is essentially a problem of constrained parameter estimation in very over-determined systems. Over-determination provides the requisite redundancy for reliable flux estimation. These methods are basically methods of network reconstruction and the obtained fluxes represent properties of the entire system. As such, they accurately reflect changes introduced through genetic or environmental modifications and therefore can be used for the assessment of their impact on cell physiology and product formation and guide the next round of cell modifications. These methods are truly systemic approaches to network reconstruction and flux determination and allow the generation of high resolution flux estimates of parallel pathways with many branching points and reversible reactions. A recent example where these methods have been applied to the complete identification of a metabolic network while satisfying all redundant balances can be found in (Klapa et al., 2003). In a variation of these methods, fluxes can be determined such as to satisfy a maximum growth criterion. The approach, known as Flux Balance Analysis (FBA), has yielded maximum growth flux estimates for a variety of organisms whose pathways were reconstructed from genomic information (Ibarra et al., 2002).

Metabolic Engineering

Metabolic Engineering is known as the field of study aiming at the improvement of strains using modern genetic tools. Strains are modified by introducing specific transport, conversion or deregulation changes that yield flux redistribution and product yield improvement (Stephanopoulos et al., 1998). As these modifications rely to a significant extent on modern methods from molecular

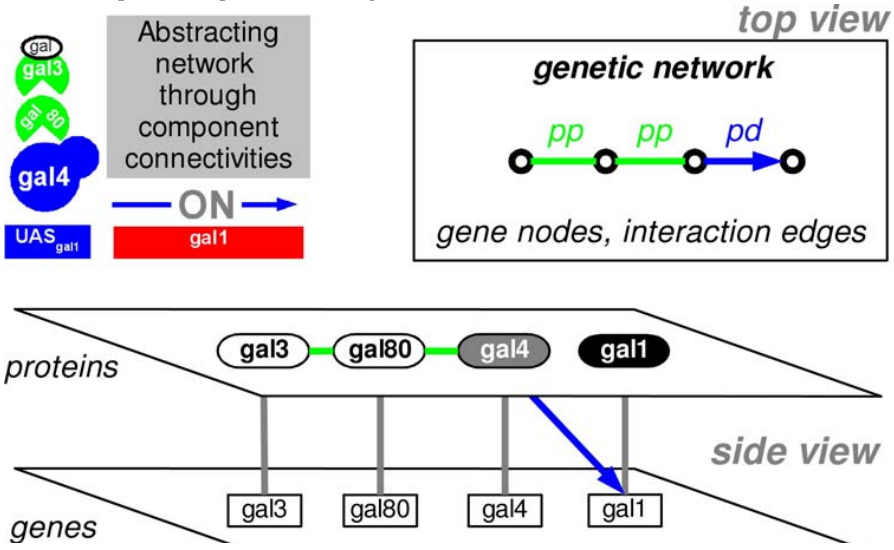

Figure 1: Abstracting networks from pathways using component connectivities. The top view integrates gene, transcript, and protein levels and represents protein-protein interactions with an undirected edge and protein-DNA transcriptional relationships with a directed edge. The networks lump the interactions into these two categories, but do not describe the specific nature of each connection. For instance, from the presence of a connection alone, GAL80 may not be identified as a GAL4 transcription repressant deactivated by galactose-GAL3 complex binding. We also clarify that interaction screens give only putative network connections, not the underlying mechanisms described here for the galactose system.

biology, a central question that naturally arises is: what is the real difference between genetic engineering and metabolic engineering? We submit that the main difference is that metabolic engineering is concerned with the entire metabolic system, in contrast to the specific focus of genetic engineering with the over-expression of a particular gene. In other words, while genetic engineering can be successful in getting a gene overexpressed or otherwise modified, this may have very little impact on the final goal of altering cell physiology. By examining the properties of the metabolic network in its entirety, metabolic engineering attempts to identify targets for amplification as well as assess rationally the effect of such changes to the properties of the overall network. As such, metabolic engineering is a progenitor of functional genomics and systems biology in the sense that it represents the first organized effort to reconstruct and modify pathways making use of genomic tools and guided by the information of post-genomic developments (Stephanopoulos and Vallino, 1991).

Continuing and Expanding the Systems Paradigm in Post-Genome Research

In a genetic circuit board representation of a cell, a low-level model replete with kinetic parameters describes the precise wiring joining the components. With a perfect genetic circuit board description, a model can propagate an input signal through the board and determines the output.

Although we now know the genetic components for many organisms and can screen possible connections among components, we rarely understand the precise wiring. And even if knowledge of the wiring were available we would additionally need to worry about the nature and timing of the signals being exchanged by the various components. Figure 1 shows how connectivity networks may be abstracted from a genetic switch. In this case, we understand the protein signals and transcription activator mechanisms involved, but more generally a genetic interaction screen gives only the more abstract network representation without providing any more detail. But how does one go about modeling? And if we were missing (needed) components such as previously unreported genes, how would we ever know? And how would their absence affect our modeling effort? Also, should we accept this lack of knowledge as a given in our effort? Or should we continue to strive and further hone our skills in gene discovery, gene and protein

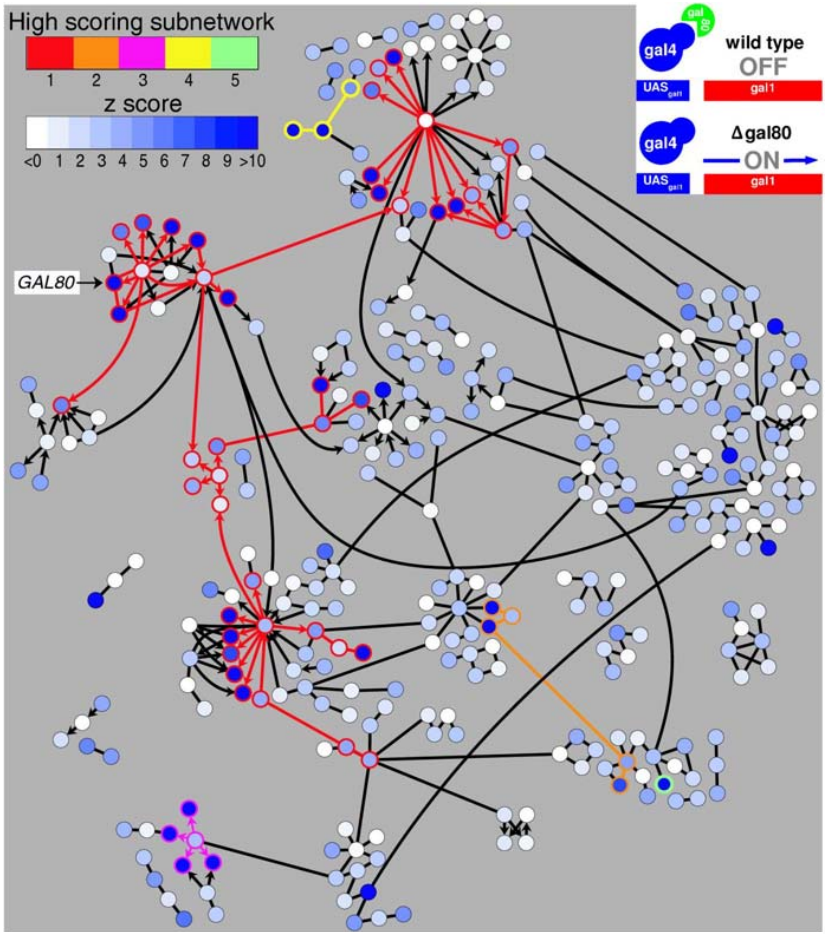

Figure 2: Determining genetic pathways activated by GAL80 gene deletion\For a set of protein-protein (pp) signaling and protein-DNA (pd) regulatory interactions among genes involved in the galactose pathway, differential gene expression after GAL80 gene deletion reveals the corresponding activated subnetworks. Even without galactose present, the removal of the GAL80 triggers cellular pathways through genetic elimination of GAL4 transcription factor represent. Node color indicates differential expression statistical significance for the particular gene, while node outline color and interaction edges between nodes indicates activated subnetworks. Significance of differential expression does not distinguish between upregulation and downregulation states; thus, both GAL80 (here, eliminated) and GAL1 (here, upregulated) will possess high confidence in differentially activity. (courtesy Trey Ideker, UCSD)

annotation, gene expression analysis, protein-protein interaction elucidation etc.?

Clearly, it is unavoidable that initially we will have to adopt a high-level cellular view devoid of detailed mechanistic knowledge. We know the genetic components, and pair-wise screens give us putative connections. Also properly applied data mining methods can uncover previously unsuspected gene relationships and consequently a pathway's members. In general, transcriptome data, proteome data, RNA interference data mass spectrometry data, and so forth can provide a good starting point that permits us to construct putative genetic interaction networks. Often, putative genetic interaction networks consist of protein-protein and protein-DNA interactions gleaned from these methods. Understanding the methods by which we determine and refine the network databases facilitates our appropriate use of the data.

Protein-protein interactions provide diverse cellular functionalities including the fulfillment of signaling roles. For instance, a cascade of signaling reactions accompanies the binding of an extracellular stimulant and initiates a shift in gene expression programs in response. For both yeast two hybrid and mass spectrometry (MS) of purified protein complexes, the large set of potential interactions makes the high-throughput screening methods prone to false positives (Aebersold and Mann, 2003). Subsequent analyses refine this data by integrating other weaker, indirect interaction indicators as well such as coexpression. Regulatory interactions control gene transcription. After a cell senses an environmental perturbation through a signaling pathway, a shift in transcriptional program occurs chiefly via trans-regulatory proteins binding to cis-regulatory upstream activation sequence DNA. While we often infer regulation through gene expression patterns, chromatin immunoprecipitation offers the leading direct, high-throughput experimental assay of gene regulation in yeast and mammals (Lee et al., 2001). Just as with protein-protein interactions, merging relevant data sets upgrades data quality.

Using these putative biomolecular networks, one notable analysis in yeast by Ideker and colleagues demonstrated that, for galactose pathway perturbations, the networks connected many of the differentially active genes (Ideker et al., 2001). Effectively, this result reinforces the intuitive notion that genes in a biomolecular pathway are upregulated and downregulated together in response to a perturbation. Going a step further, Ideker's Cytoscape software platform and the accompanying ActiveModules plugin identify the biomolecular pathways most affected by sets of perturbations. These types of abstracted models process global data and output potentially interesting relationships that may be investigated with lower-level, more detailed modeling and experiments.

Currently, Cytoscape represents the best software platform for visualizing biomolecular relationship networks and searching those networks for active subnetworks. A Cytoscape analysis requires inputting connectivities as well as differential state data for each component. Typically, protein-protein and protein-DNA interactions among genes are integrated with differential gene expression data. The component connectivities distinguish among the unidirectional nature of protein-protein interactions and the directional nature of protein-DNA interactions. Given this molecular relationship map and differential states, Cytoscape then allows searching for differentially active subnetworks mediating the perturbed state. Searching for active subnetworks requires a methodology for scoring a given subnetwork before subsequent comparison with other subnetworks. Scoring a subnetwork first requires scoring individual components. For a gene, high confidence in differential expression contributes to the activity of any subnetwork containing this gene. Global optimization techniques such as simulated annealing techniques allow robust determination of the most active subnetworks. Figure 2 shows one such active subnetwork search (Ideker et al., 2002).

Typically, a lack of robust interaction data for higher organisms, such as mammals, limits searching for response-mediating pathways. Among other difficulties, the sheer magnitude of the possible interactions among tens of thousands of genes prevents systematic interaction screens for such systems. However, experimental efforts have documented many genetic relationships for specific pathways in numerous publications over the years. Ingenuity Systems and the Stanford Genome Technology Center recently made available the results of exhaustive literature searches that document known genetic interactions. Furthermore, the reasonably characterized central metabolism in human and mouse parallel yeast metabolic pathways. These genetic and metabolic

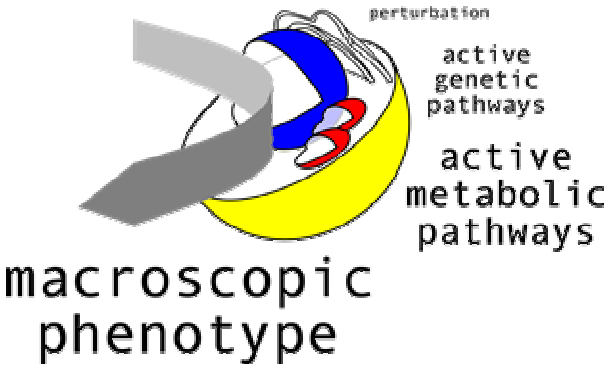

Figure 3: Bridging genetic measurements and macroscopic phenotype with metabolic pathways. For a given genetic or environmental perturbation, protein signaling and transcriptional regulators orchestrate metabolic changes that manifest a new macroscopic phenotype such as growth rate, disease, or product production. In this way, understanding the metabolic changes leading to the new macroscopic phenotype elucidates changes occurring at the genetic level.

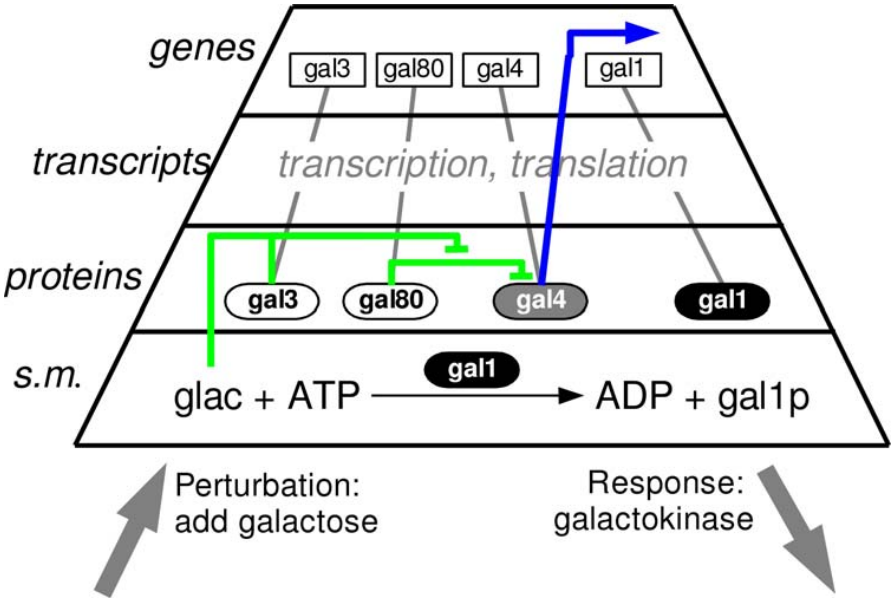

Figure 4: Hierarchy for perturbation and response. The introduction of intracellular galactose at the small molecule level initiates genetic pathways resulting in galactose phosphorylation at the small molecule level in preparation for further metabolic processing. Changes at the small molecule level typically occur on the order of seconds, while changes in gene expression program typically occur on the order of minutes. Thus, although an immediate responses may be mediated strictly at the small molecule level, the longer term responses stem from a change in gene expression program. Also, macromolecule perturbations such as peptide hormone introduction initiate changes beginning at the macromolecule rather than small molecule level.

networks offer the opportunity to apply the pathway analysis techniques for yeast to human and mouse transcriptional data.

A high-level screen may reveal a potentially interesting pathway (subnetwork) that activates in response to a stress condition. As any researcher with gene clustering results can attest, the following question arises: How do you gain mechanistic insight simply based on a set of interesting genes? Unlike gene clusters however, previously identified genetic interactions link the screened active networks and provide a launching pad for gaining mechanistic insight. We then hope that subsequent specific, low-level exploration of the differentially active subnetworks will provide such mechanistic insights.

Realistic v. Unrealistic Expectations

The wide scope of both data and analysis tools justifiably leads to ambitious expectations for systems biology. After so many years of studying genes in isolation or just small groups of genes at a time, a deluge of information gives us snapshots of a dynamic world where cells divide, organisms develop, tumors take root, etc. No one can deny the opportunities that present themselves, but one must also be mindful that the problem which we set out to address is several orders of magnitude larger than those with which we are experienced. Consequently, it is important that we temper our expectation of immediate results. We would like to suggest that researchers bear in mind the following points:

Despite the wealth of available genomic data there are still lots of genes of whose presence we are unaware and which are involved in important interactions. A commonly repeated mis-statement is that the genomic effort and the analysis that followed have generated all the data that would be necessary to effectively practice systems biology. However, as recent results have demonstrated this could not be further than the truth: genomic maps are continuously updated by discarding or adding (occasionally substantial amounts of) genes, new important protein interactions are elucidated for pathways that were considered well understood, etc. (Berns et al., 2004; Dornan et al., 2004)

Constructing biomolecular networks for new systems demands significant resources and expertise. Biomolecular networks incorporate a multitude of relationships connecting several types of components. At the genome-scale, constructing interaction maps requires extremely large experimental investments and subsequent analysis and curation. For instance, global protein-protein interaction maps exist for only a handful of model species and represent the work of large collaborative efforts. Even reconstructing well-studied and well-documented networks such as metabolic pathways in a genome context requires years of curation for a lone researcher. Thus, a do-it-yourself approach for a new biomolecular network is not a possibility.

Even when you work with a biomolecular network database, do not expect a full system picture. In the post-genomic era, the effort to uncover the structure and function of genetic regulatory networks gave birth to many databases each of which attempts to distill the most salient

features from incomplete and at times flawed knowledge. For yeast, databases document over 80,000 putative

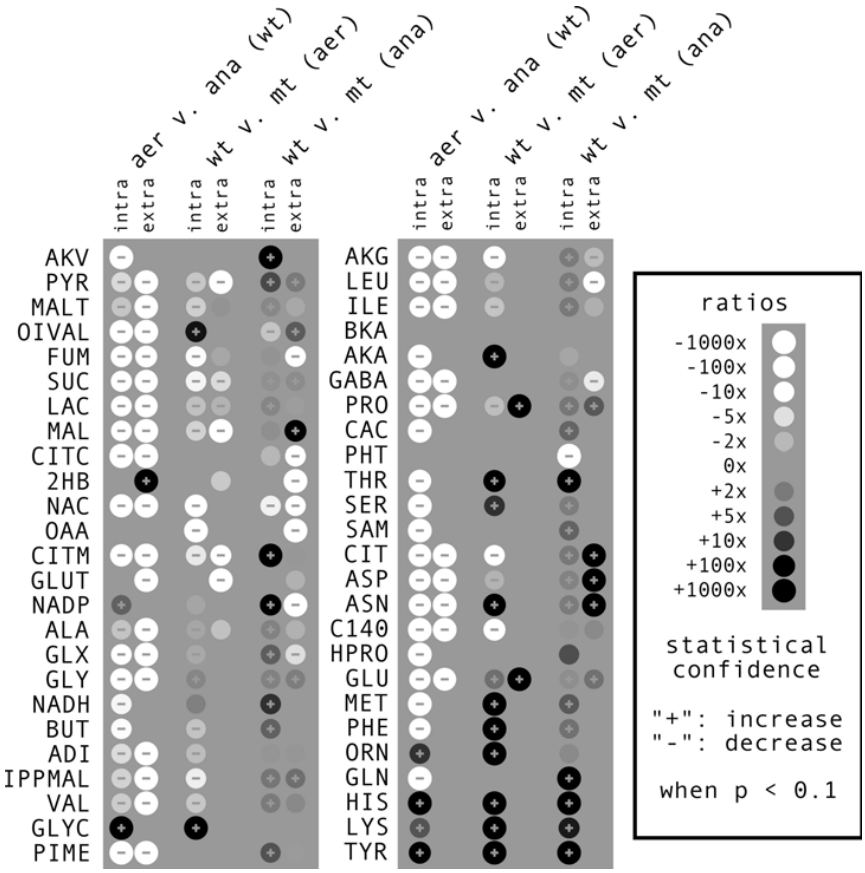

Figure 5: Metabolic states Just as differential transcription changes may be determined with gene expression arrays, differential metabolite levels may be determined with high-throughput assays. Here, this table compares metabolite levels for wild type and mutant yeast strains under aerobic and anaerobic conditions.

protein-protein interactions for the yeast two-hybrid (Y2H) protein-protein screen (Aebersold and Mann, 2003). Of these interactions, only a few thousand agree between the individual screens among labs. Clearly, "accepted" interactions vary significantly across databases and over time even under the best-case scenario. Of course, we must note that yeast databases benefit from heavy experimentation and scrutiny as well as a relatively smaller interaction space explored (pairwise interactions among approximately 6000 genes). In other organisms with lower levels of direct interaction experimentation and scrutiny (e.g., E. coli) or much larger interaction spaces (e.g., mouse and human), databases represent even less of a full picture. Furthermore, many databases do not distinguish among direct and indirect interactions, especially where direct interaction screens are not feasible: for instance, a human genetic interaction in the Ingenuity database will often be an observation in the literature that "gene X affects gene Y."

Making the necessary measurements for your system demands significant resources and expertise. Presently, the only broadly available tool for measuring gene expression is the DNA chip. Moreover, conducting a transcriptional experiment requires training, and large-scale studies naturally incur significant costs. Going further and measuring protein levels, protein states, regulatory elements, and metabolites requires complex and specialized equipment and each can consume one or more full graduate student careers. Consequently, systems biology necessitates the creation of partnerships and the collaboration among cross-disciplinary faculty members. No one group or department possesses all the technology or computational tools necessary for a full analysis. Biologists, engineers, chemists, physicists, mathematicians and computer scientists must learn to speak one another's language and to work together.

It is unlikely that a single/complex microarray experiment will elucidate interactions you want to know. Large amounts of data and the presence of noise aside, many relevant interactions simply do not result in large, direct transcriptional changes. The practitioner must be mindful of the fact that transcript levels do not correlate well with protein levels, and protein levels do not correlate well with activity level.

Patience is advised as the more complex hypotheses derived from systems approaches are disproportionately more demanding to validate. For a perturbation, differential gene expression and network searching reveals active putative biomolecular networks. Often, the networks span dozens, if not hundreds, of genes. The larger active networks provide the statistical support necessary for signal recognition in the midst of the noise within the individual gene measurements and the noise within the imperfect putative networks. Figure 2 shows that even for a small network, a single gene removal causes a cascade of activation and deactivation among many genes. While we can demonstrate cause and

eventual effect, verifying the specific mechanism for each gene's activation/deactivation becomes a herculean task.

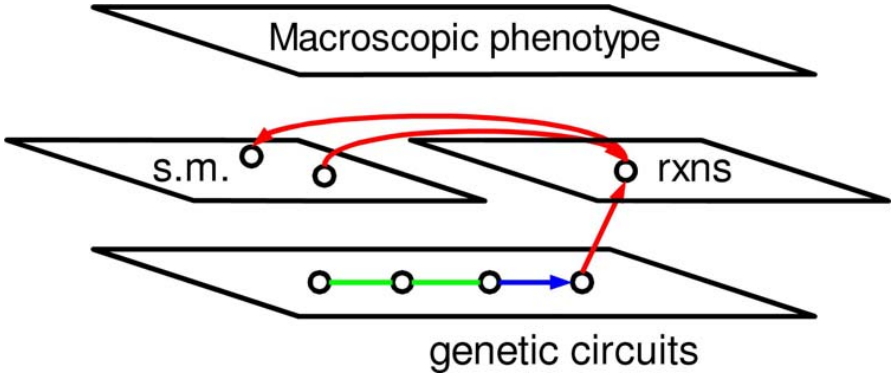

Figure 6: Connecting biomolecular networks Because macroscopic phenotype adjustment occurs through metabolic pathways, activated genetic pathways must be viewed in small molecule context to provide insight into the macroscopic phenotype. Here, reaction nodes connect small molecule (s.m.) nodes to enzymes at the genetic level.

Vision

Ultimately, systems biology hopes to unravel the complex underpinnings of macroscopic phenomena like cell division, animal development, tumor growth etc. through the integration and analysis of data of different modalities. With time, we refine our genetic network as well as our ability to accurately monitor transcripts and proteins. Also, the current state of the art enables us to more accurately determine specific genetic interactions active under any genetic or environmental perturbation.

However, we still need to link the genetic interactions to the macroscopic phenomena that we observe. For instance, we may discover that under a certain condition a transcription factor causes activation of a previously unknown pathway. In the best case scenario where we can determine and model the specific binding involved, we can only postulate on reasons why the pathway became activated and how the pathway influences the observed macroscopic phenotype.

Metabolism links the genetic interactions with macroscopic phenotype. The responses propagated through protein-protein and protein-DNA interactions are ultimately incorporated as adjustments to metabolic pathways to fulfill the new cellular demands. Unfortunately, transcript and protein levels and states do not correlate with metabolic activity. Figure 3 shows a cartoon propagation of a cellular perturbation manifested macroscopically through active metabolic pathways. Upon traumatic injury, the human body responds by initiating a shift in gene expression which propels the body into a hypermetabolic state. Likewise, an oxygen deficiency in yeast initiates a shift in gene expression which maximizes energy production by diverting metabolic fluxes from the TCA cycle to produce ethanol. For eukaryotic organisms, signaling and regulatory pathways at the genetic level alter metabolic state by interconverting small molecules in such a way as to fulfill the new cellular demands through tight enzymatic control. In the galactose utilization example, perturbations at the small molecule level propagate through signaling and regulatory pathways at the genetic level to generate responses at the small molecule level (see Figure 4).

The cellular response hierarchy requires bridging activated genetic networks to cellular phenotype by incorporating a small molecule perspective into biomolecular networks. Since genetic networks indirectly mediate metabolic adjustments, active genetic networks infer responses whereas metabolic measurements directly assay response mechanism. Documented metabolic pathways and the known genetic relationships provide the ability to integrate genetic interaction networks with metabolism. The availability of genetic interaction and metabolic networks, gene expression states, and metabolic levels and reaction fluxes provide an opportunity to conduct integrated pathway analyses and further discern genetic interactions with direct metabolic measurements.

This metabolomic information is an experimentally accessible feature of the cell that reveals important and extensive phenotypic information. As we mention above, it is difficult to understand how these pathways mediate a macroscopic phenotype. Generally, it is not possible to draw accurate metabolic conclusions using only genetic data. For instance, an increase in transcriptional activity does not necessarily imply increase in protein levels; moreover, once translated a protein may or may not be enzymatically active.

The shortcomings of the current state-of-the-art metabolic analysis result not from a lack of investigator interest but from the difficulty of measuring metabolic states. Diverse chemical properties make broad, simultaneous measurement of metabolites difficult. Indeed, unlike transcriptional states, high-throughput metabolic state measurements have only recently begun to mature. Metabolic analysis techniques that use gas chromatograph mass spectrometry (GCMS) measure metabolite levels and reaction fluxes for central carbon metabolism with increasing throughput. (Villas-Boas 2003) Figure 5 displays sample results of metabolite levels differing among genetic and environmental perturbations. Incorporating this metabolic state data into

a tool like Cytoscape requires expanding the genetic interaction network to include metabolism and associating differential state data with each metabolic node -- see Figure 6. For the characterized yeast metabolic network, interfacing reactions with genes and small molecules to those reactions using the built-in small molecule and reaction node capabilities is straightforward. For associating differential metabolic state data with each node, differential metabolite levels and reaction fluxes may be employed.

With metabolic state data supplementing genetic pathway activities, we may anchor our analysis and address basic questions. How do the changes in genetic state cause metabolic changes? How do the metabolic changes account for our macroscopic phenotype observations? The metabolic state data provide the direct mechanistic insight necessary to tackle these questions when combined with well-defined perturbations across genetic and environmental perturbations.

References

Aebersold, R. and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, 422, 198.

Berns, K., et al. (2004). A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature*, 428, 431.

Dornan, D., et al. (2004). The ubiquitin ligase COP1 is a critical negative regulator of p53. *Nature*, 429, 86.

Femenia, J. (2004). Signaling Networks. PhD Thesis, MIT.

Ideker, T., et al. (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292, 929.

Ideker, T., Ozier, O., Schwikowski, B., and Siegel, A. F. (2002). Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, 18 Suppl 1, 233.

Kacser, H.rand Burns, J. A. (1973). The control of flux. *Symp Soc Exp Biol*, 27, 65.

Klapa, M. I., Aon, J. C., and Stephanopoulos, G. (2003). Systematic quantification of complex metabolic flux networks using stable isotopes and mass spectrometry.. *Eur J Biochem*, 270, 3525.

Koffas, M. A., Jung, G. Y., and Stephanopoulos, G. (2003). Engineering metabolism and product formation in *Corynebacterium glutamicum* by coordinated gene overexpression. *Metab Eng*, 5, 32.

Lee, T. I. et al. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, 298, 799.

Ibarra, R. U., Edwards, J. S., and Palsson, B. E. (2002). Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, 420, 186.

Stephanopoulos, G. and Vallino, J. J. (1991). Network rigidity and metabolic engineering in metabolite overproduction.. *Science* 252, 1675.

Stephanopoulos, G., Aristidou, A., and Nielsen, J. (1998). *Metabolic Engineering*, Academic Press.

Stephanopoulos, G. (1991). Metabolic fluxes and metabolic engineering. *Metab Eng*, 1, 1.

Villas-Boas, S. G., Delicado, D. G., Akesson, M., and Nielsen, J. (2003). Simultaneous analysis of amino and nonamino organic acids as methyl chloroformate derivatives using gas chromatography-mass spectrometry *Anal Biochem*, 322, 134.