Analyzing Network Models to Make Discoveries About Biological Mechanisms

William Bechtel
Department of Philosophy
University of California, San Diego

Abstract

Systems biology provides alternatives to the strategies to developing mechanistic explanations traditionally pursued in cell and molecular biology and much discussed in accounts of mechanistic explanation. Rather than starting by identifying a mechanism for a given phenomenon and decomposing it, systems biologists often start by developing cell-wide networks of detected connections between proteins or genes and construe clusters of highly interactive components as potential mechanisms. Using inference strategies such as 'guilt-by-association', researchers advance hypotheses about functions performed of these mechanisms. I examine several examples of research on budding yeast, first on what are taken to be enduring networks and subsequently on networks that change as cells perform different activities or respond to different external conditions.

I. Introduction

The explanations biologists offer often take the form of accounts of mechanisms responsible for phenomena to be explained. These explanations identify a phenomenon with a particular mechanism, characterize the parts and operations that constitute the mechanism, represent their organization, and show that this mechanism is capable of generating the phenomenon when situated in its normal environment (Bechtel & Abrahamsen, 2005; Machamer, Darden, & Craver, 2000). Accounts of the development of mechanistic explanations in biology emphasize the empirical research that links a phenomenon with a mechanism and strategies for decomposing mechanisms to discover their component parts and operations as well as the reasoning strategies that build upon this information (Bechtel & Richardson, 1993/2010; Craver & Darden, 2013). Since this approach is well exemplified in research in cell and molecular biology of the 19th and 20th century, I refer to it as the 'traditional mechanistic approach'. My aim in this paper is to contrast this traditional approach with a relatively new approach that has emerged with the advent of systems biology. Systems biologists are developing large databases of information about gene interactions, protein-protein interactions and protein-gene interactions in cells.[1] This data is collected from automated versions of more traditional

---

[1] Over the past 15 years numerous large databases based on a variety of interactions between genes or between molecules in cells have become available. A regularly updated compilation of molecular biology databases is maintained at https://www.oxfordjournals.org/our_journals/nar/database/c/. It currently includes 685 databases. Starting with supplementary issues in April 1991 and May 1992 and a regular

experimental investigations that manipulate components of cells. What is distinctive is the way in which this data is represented and analyzed. To make sense of such data and to invoke them in explanations of biological phenomena, many systems biologists represent data in networks and apply analysis tools to these networks.[2] These network representations take a far more system-wide perspective on cells than more traditional mechanistic approaches. However, network representations are not ends in themselves; rather they provide a means for developing new mechanistic explanations and adding to existing ones.

My main goal in this paper is to characterize the new network-based approaches to mechanistic explanation and show how they have been employed in systems biology to develop new hypotheses about mechanisms. Elaborating a bit on traditional mechanism approaches will facilitate drawing the contrast. Traditional approaches begin by characterizing a single phenomenon and subsequent inquiry is directed at understanding that phenomenon. Research proceeds by identifying a mechanism responsible for the phenomenon and taking it apart (conceptually or experimentally) to identify its component parts (structures) and the operations (functions) they perform. As illustrated in the case of the discovery of cell mechanisms (Bechtel, 2006), different research strategies enable structural and functional decompositions. Ultimately both decompositions need to be integrated by localizing operations in the parts of mechanisms. Research on oxidative phosphorylation between 1940 and 1970 provides an illustrative example. By the 1940s oxidative phosphorylation was known to involve the catabolism of pyruvate to carbon dioxide and water while capturing energy in the phosphate bonds of ATP. Classical biochemical techniques that used heat or poisons to inhibit individual enzymes revealed the reactions of the Krebs cycle and the electron transport chain and associated phosphorylation with specific reactions in the electron transport chain. Cell fractionation techniques demonstrated that the identified enzymes were located in the mitochondrion while electron microscopy revealed the structure of the mitochondrion, especially its protruding internal membranes. Through a variety of experimental approaches, the steps of electron transport were linked to the internal membrane and the crucial processes of phosphorylation were localized to small protrusions on it. Building on Mitchell's (1961) chemiosmotic hypothesis, researchers developed a mechanistic account whereby protons released during oxidative reactions in the electron transport chain accumulate between the mitochondrial membranes, creating a gradient over the membrane that provides the energy for synthesizing ATP.

Application of this traditional mechanistic approach required first delineating a phenomenon, which often itself required experimentation. What are now regarded as

---

issue in July, 1993, the journal *Nucleic Acids Research* has regularly reviewed databases. Starting in 1996, the journal identified its first issue of each year a Database Issue.

[2] Many of these studies use tools such as Cytoscape (Shannon, Markiel, Ozier et al., 2003; Su, Morris, Demchak et al., 2014) and VisANT (Hu, Chang, Wang et al., 2013), which can take data from databases and rend it into a network representation. These tools provide resources for visualizing networks in different ways and to perform operations such as identifying clusters.

central phenomena in cell life—the replication of cells through division, the synthesis of proteins, and the recycling of failed cell components—were only identified through a history of experimental research. In some cases, the accounts of the phenomena were further developed in conjunction with the decomposition of the associated mechanisms into constituents. But the legacy of discovering new cell phenomena indicates that only some mechanisms have been discovered to date. Moreover, the process of decomposition into parts and operations typically begins with just a few of the parts and major operations that can be explored with experimental interventions. Even with what are regarded as well-worked out mechanistic accounts, ongoing research continues to identify additional parts and operations. An indication of just how incomplete accounts of mechanisms are is that at the beginning of the twenty-first century, even in the best studied model organism, budding yeast or *Saccharomyces cerevisiae*, the roles played by only about one-third of the proteins had been determined, either directly through experimental investigations or indirectly on the basis of sequence homology to proteins in other species where experiments had revealed function. Much of the yeast cell remains *terra incognita*.

Instead of starting locally to delineate and explain a specific phenomenon, network research starts from a cell-wide perspective. Large throughput procedures generate massive amounts of data about interactions between large numbers of genes or proteins in individual cells. The results of these experimental investigations are typically collected into publically accessible databases that strive to include interactions between all genes or proteins of an organism. Although actual databases fall short of these aspirations, they include vastly more genes and proteins than have been the focus of traditional mechanistic research. Moreover, this process is not guided by prior characterization of a phenomenon or an identification of a responsible mechanism. Rather, information about phenomena and mechanisms is developed from the analysis of the interaction data. Further, it is often different scientists, many of whom are trained in bioinformatics, that develop mechanistic accounts from these databases. To do so, these researchers represent the data they extract from databases in networks, making a number of choices along the way about how to lay out the genes or proteins as nodes and connections between them as edges. They then employ a variety of analysis procedures on these networks to find patterns within them. One important class of analytic techniques aims to identify nodes that cluster together in terms of the increased density of connections between those nodes (there are a variety of algorithms for identifying clusters, and each produces somewhat different results). Researchers interpret increased connectivity in a cluster (also referred to as a 'module', 'complex', or 'community') as indicating that the entities represented by those nodes work together to perform an activity in the cell—that is, that they constitute a mechanism. This process can be applied at different scales, often yielding clusters within clusters.

Since the network researchers did not set out to explain a specified phenomenon, once they have identified a cluster and construed it as a mechanism, they must figure out to what phenomenon it contributes. In some cases clusters include nodes for proteins that on the basis of previous research, often as compiled in Gene Ontology (Ashburner, Ball, Blake et al., 2000), had been recognized as parts of a mechanism responsible for a specific phenomenon. Employing an inference strategy known as 'guilt-by-association', researchers infer that the other nodes in the cluster contribute to the same phenomenon. In this way,

network research advances hypotheses about new parts of mechanisms and what these parts are doing that can spur further research into operations within mechanisms. In many instances, the clusters or modules do not contain nodes identified with any known mechanism. The researchers view these clusters as representing new mechanisms. The hypothesis that such a cluster constitutes a mechanism motivates an investigation (using more traditional mechanistic approaches) into the phenomenon for which it is responsible.

The approach I have just outlined is a potent heuristic or hypothesis generator about mechanisms that functions differently than the heuristics of decomposition and localization (Bechtel and Richardson, 1993/2010) or forward and backward chaining (Craver and Darden, 2013). It is far more data-driven than these traditional strategies. Instead of relying on experiments guided by a hypothesis about a phenomenon or a mechanism, network research starts with broad-scale interventions that identify interactions between large numbers of entities and takes advantage of the fact that some of them cluster together. Once the clusters are identified, the interpretation of the nodes as representing potential parts is straightforward. The edges, on the other hand, do not represent specific operations but simply interactions between the parts. The location of an edge between clustered nodes may provide clues about the operation a part performs, but filling in these mechanistic details requires more traditional mechanistic studies. The potency of the network approach results from starting from a cell-wide perspective. This results in identifying many entities that traditional mechanistic research never associated with the mechanism and clusters that do not correspond to any previously characterized mechanisms. Network research thus pursues different heuristic strategies than traditional mechanistic research and these offer novel insights into biological mechanisms.

Although the strategies for discovering mechanisms are different, the mechanisms sought are much like those characterized by mechanistic philosophers of science—they consist of entities or parts (e.g., enzymes) performing activities or operations (catalyzing reactions) organized in a particular manner. Moreover, the hypotheses put forward must be tested in the same manner as those advanced in more traditional mechanistic research—e.g., by experimentally demonstrating that changing the part or operation alters the phenomenon and that the part or operation is altered appropriately when the mechanism is functioning. The differences between traditional mechanistic and network approaches are ones of emphasis. As noted above, network research itself typically does not provide a detailed account of the operations performed by the parts—that remains to be filled in by more traditional modes of research. On the other hand, it provides insight into aspects of mechanisms often missed in traditional mechanistic approaches, especially how individual mechanisms are integrated into larger systems, ultimately cells and organisms. Once it has identified a mechanism, traditional mechanistic research focuses on its components and only occasionally recognizes how parts of a given mechanism are often connected to parts of other mechanisms. Identifying interactions, whether within or between mechanisms, is a strength of network approaches, with the result that it emphasizes more than traditional approaches the interactions between components of different mechanisms, treating the whole organism as much more interconnected than did traditional mechanistic accounts.

Even though systems biologists analyzing networks approach mechanisms from a cell-wide perspective, they employ an analog to the contrast between structural and functional decomposition. Sometimes clustering is done to reveal how components are composed into structural units (e.g., how proteins are bound to other proteins or to genes).[3] In other cases it is done to reveal how components contribute functionally to the activities of the cell (e.g., cell replication or procuring energy). I will illustrate how these strategies figure in the analyses researchers perform on networks and will focus in particular in how they are integrated in a way that parallels the role played by localization, the linking of operations with parts, in traditional mechanistic research.

My focus will be on the actual inference strategies researchers have developed to analyze networks in terms of mechanisms. I will present several examples from network research over the last fifteen years. In the next section, I will focus on inferences about mechanisms derived from constructing what are treated as enduring networks, networks representing protein-protein interactions or gene interactions that exist in the cell independently of the conditions the cell confronts. The first example involves a network based on structural relations between proteins, but subsequent examples will include interactions between genes that address their functional contribution. In section three I turn to attempts to use network analyses to understand how network organization in the cell changes as cells undergo mutations or engage in different activities either initiated endogenously (e.g., the cell cycle) or exogenously (responding to environmental stress). Once again, I will start with networks emphasizing structural relations before bringing in studies employing functional information about gene interactions.

Research of the type I am describing has been pursued on a number of model organisms, but budding yeast (*Saccharomyces cerevisiae*) has been the most studied. Accordingly, I limit myself to examples of research conducted on this model organism.

2. Analyzing Networks under Static Conditions

Traditional mechanistic research assumed mechanisms were enduring entities that performed their functions whenever appropriate conditions arose. Philosophical accounts of mechanistic explanation likewise have treated mechanisms as static entities. A similar perspective was adopted in the initial studies using networks to study cellular mechanisms—the networks, and the clusters found in them, were viewed as representing enduring conditions within the cell. I will focus on three exemplar studies, published between 2003 and 2008, that adopted this perspective and generated new hypotheses about mechanisms in yeast.

Bechtel and Richardson (1993/2010) note that in traditional mechanistic research investigators sometimes begin by decomposing a mechanism structurally and build up to a

---

[3] Although on some accounts, the interaction of a protein with another or with DNA might count as functional, biologists characterize them as structural (often researchers also use the term 'physical') since they result in the construction of a complex. The term 'functional' is then reserved for research focused on what the component contributes to cell function.

functional characterization, and other times begin by characterizing a mechanism functionally, and subsequently link the functions to structural components of the mechanism. A comparable distinction is found in network research. Protein-protein interactions are intended to identify complexes of proteins that constitute structures in the cell. In section 2a I consider an example in which researchers built their network based on structural information (protein-protein interactions) and then seek information about how components in it contribute to cell life. In section 2b I consider two examples in which researchers included functional interactions (gene interactions) in the construction of the network itself.

2a. Networks constructed from structural relations between proteins

Proteins are major structural components of biological cells, important in part because of their functional role as enzymes catalyzing chemical reactions. Their functional role has long been linked to their complex three-dimensional structure. In classical biochemical approaches to biological phenomena, researchers focused on individual enzymes as catalysts. They were linked into pathways through the metabolites on which each of them operated. But it has increasingly become clear that proteins function in biological mechanisms as complexes, often semi-stable ones. Even when they do not form stable complexes, interactions between proteins enable one protein to alter the behavior of others. Knowing which proteins can form complexes with other proteins provides evidence about the working parts in functioning cells.

In the late twentieth century researchers developed two techniques that enabled mass screening for protein interactions (complex formation). The first, yeast two-hybrid screening, involves separating the domains of a transcription factor and binding one domain to one protein and the other either to another protein (or a library of cDNA fragments from multiple proteins). If the proteins are able to bind, the domains of the transcription factors are reunited and activate transcription of a reporter gene (Fields & Song, 1989; Young, 1998). The second, affinity purification followed by mass spectrometry, first segregates a tagged protein together with its interaction partners and then uses mass spectrometry to identify them (Rigaut, Shevchenko, Rutz et al., 1999). Many proteins interact with multiple partners. Thus, one can construct an interconnected network by treating individual proteins as nodes and detected interactions between proteins as edges.

Spirin and Mirny (2003) provide an early example of the basic strategy for using protein-protein interactions to identify mechanisms. Drawing upon the Munich Information Center for Protein Sequences (MIPS) database of protein-protein interactions, they constructed a network that treated proteins as nodes and interactions as edges. The resulting network consisted of 3,992 nodes and 6,500 edges.[4] Using the criterion that to count as a cluster, nodes must be totally or very highly interconnected, Spirin and Mirny identified more than

---

[4] It is important to recognize that the databases upon which networks are being constructed are themselves growing as new experimental research is conducted. A subsequent attempt to build a network from the same database may generate somewhat different results.

fifty clusters with from four to thirty-five nodes. Three example clusters are shown against a portion of the background network in Figure 1. Just looking at the network, it is notable that there are a large number of connections outside of the clusters and that nodes in a cluster are also connected to many other nodes, some in other clusters. What makes clusters distinctive is that the nodes in them are highly connected to each other, suggesting that they form structural complexes or modules. Individual proteins in the complexes may be able to bind to other proteins, but these bindings are viewed as less central to the structure of the cell.

These clusters are candidate mechanisms and the proteins in them candidate parts of these mechanisms. Although the clusters are identified on structural criteria, the researchers also sought to determine how these mechanisms contribute functionally to the cell. Spirin and Mirny thus applied the functional annotations provided in MIPS to the nodes in each module and used this information to propose a function for each mechanism. Thus, they identified the nodes in the SAGA_TFIID transcription factor cluster (shown in red) as involved in transcription regulation, those in the anaphase-promoting complex (blue) as involved in cell-cycle/cell-fate control, and those in the TRAPP complex (yellow) as involved in protein transport.[5] They thus treated the clusters as cell mechanisms.

---

[5] Although Spirin and Mirny analyzed the network as if it were static, they do note that many of the clusters consist of proteins that are synthesized at different times. As such, they cannot form a complex. They proposed that these proteins form *functional modules* that operate in a coordinated manner over time, e.g., in a signaling pathway or being synthesized successively to advance a cell through the cell cycle.
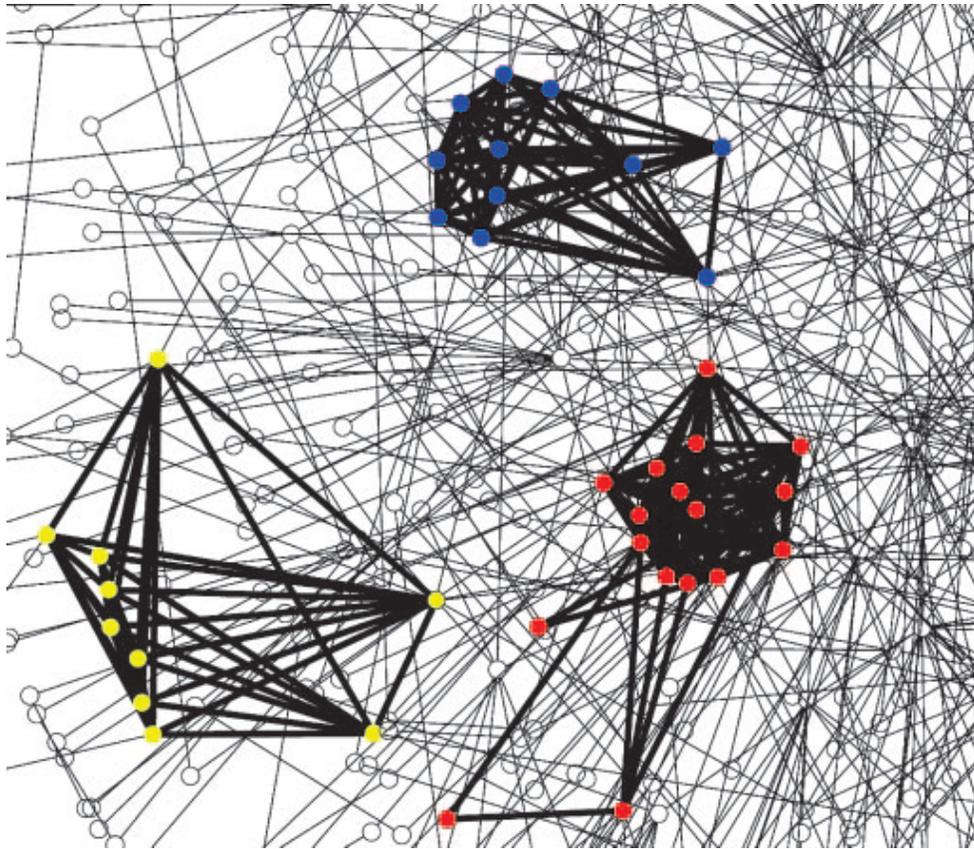
Figure 1. A fragment of the protein-protein interaction network used by Spirin and Mirny (2003). Colored nodes and the connections between them constitute modules that are interpreted as complexes: the SAGA/TFIID complex (red), the anaphase-promoting complex (blue), and the TRAPP complex (yellow). Reprinted from Spirin, V., & Mirny, L. A., 'Protein Complexes and Functional Modules in Molecular Networks'. *Proceedings of the National Academy of Sciences of the United States of America, 100*, 12123-12128, Copyright (2003) National Academy of Sciences.

The inference from nodes forming a cluster to nodes representing parts of a mechanism with particular functions was grounded on the fact that some of the parts were already recognized as parts of a mechanism that performed a cell function. Since they began with all nodes that clustered together, Spirin and Mirny identified many new parts of thirteen known mechanisms. For example, because they clustered together, they identified two small ribosomal subunits as members of the Lsm splicing mechanism. They employed the guilt-by-association strategy to propose that these new components contributed to the same cell activity. Not all clusters could be linked to known mechanisms in this manner. Spirin and Mirny report finding eight previously uncharacterized complexes and seven uncharacterized functional modules (see note 5) that could function as mechanisms. One complex consists of six proteins, only one of which, YIP1, had previously been characterized—as a Golgi membrane protein. Since the others share a degree of homology with other membrane proteins, Spirin and Mirny propose that this complex is a mechanism involved in membrane construction and maintenance.

2b. Networks constructed using information about functional interactions

I turn now to network approaches that start with functional data about genes, specifically, when genes interact in the generation of a phenotype. A simple form of interaction occurs when either of two genes can be knocked out without killing the organism, but knocking out both together is lethal. This is known as 'synthetic lethality'. In its simplest form, synthetic lethality results when the two gene products are involved in alternative ways to achieve a required function. In that situation, when one is knocked out the product from the other still performs the function. A related condition, 'synthetic sickness' arises when each knockout affects colony growth but the effect of knocking out both differs from the multiplicative effect that would be expected if the two were independent. Neither synthetic lethality nor synthetic sickness reveals what the function of a given gene is, but either effect indicates that two genes are performing the same function or closely related ones.

One advantage of focusing on gene-interactions rather than protein-protein interactions is that they can occur both between genes that code for proteins that interact to form a complex and also between genes whose proteins reside in different complexes. They thus provide insight into how the individual complexes are organized into larger mechanisms. To explore this, Kelley and Ideker (2005) constructed both a structural network (based not just on protein-protein interactions but also protein-DNA interactions and interactions via shared metabolites) and a functional network based on synthetic-lethal or synthetic-sick interactions. Kelley and Ideker identified any densely connected set of proteins as pathways (modules). Their analysis focused on functional (genetic) interactions between the structurally identified pathways; they refer to these sets of pathways as 'models'. Figure 2A shows three such models, with the rectangles in each representing pathways and the circles within them genes/proteins belonging to that pathway. Blue lines between circles indicate structural relations between the proteins and red arrows indicate functional (genetic) interactions between the respective genes. In Figure 2A some of the red arrows are contained within the rectangles representing pathways, but most cross between two pathways belonging to the same model. The arrows between pathways, Kelley and Ideker propose, reflect that the different pathways perform redundant or complementary functions. Whereas the pathways represent low-level mechanisms, the models correspond to larger-scale mechanisms composed from the pathways.

A.   B. Between pathway   Within Pathway
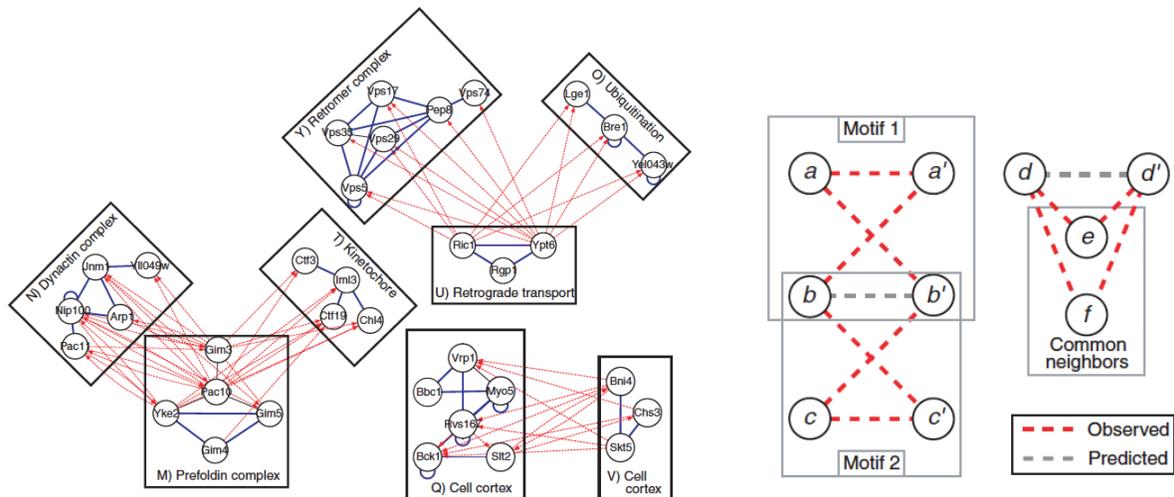
Figure 2. A. Three models of interconnected pathways in which blue edges indicate protein-protein interactions while red arrows indicate genetic-interactions. B. Two examples, one between pathways and one within a pathway, of using incomplete motifs to propose additional edges between nodes. Reprinted by permission from Macmillan Publishers Ltd: *Nature Biotechnology*, Kelley, R and Ideker, T., 'Systematic Interpretation of Genetic Interactions using Protein Networks', copyright 2005.

By building the network using information about synthetic lethals, Kelley and Ideker included functional relations in the network itself. But just the fact that two genes produce a synthetic lethal does not reveal what functions are involved. Thus, like Spirin and Mirny, Kelley and Ideker appeal to what was already known (e.g., in GO) about the individual genes/proteins in their network, for example, that proteins in pathway M constitute the prefoldin complex, which promotes the folding of α- and β-tubulin into functional microtubules. These microtubules are important for the function of the proteins in the other two complexes to which they are linked: proteins in the dynactin complex are involved in translocating the spindle and other molecules along microtubules while those in the kinetochore complex anchor chromosomes to spindle microtubules during metaphase. These annotations allow Kelley and Ideker to interpret the edges between pathways generated from the gene interaction data mechanistically.

Kelley and Ideker's goal was not just to show that a mechanistic analysis could be recovered from these networks but also to show that by incorporating functional information about synthetic lethals in their network they could develop new functional hypotheses about what the parts of these mechanisms do. To do this, they selected from their network structural pathways in which annotations were assigned to less than one hundred genes overall and in which the percentage of the proteins assigned common annotations was statistically significant ($p < .05$) when compared to random networks. Employing guilt-by-association they predicted 973 annotations for 343 proteins in these models. This was four times as many as could have been predicted from the structural network alone (i.e., using the strategy exhibited in the Spirin and Mirny study discussed above). The difference was due to including functional synthetic lethal information in the network itself and not just in the annotation process. As an illustrative example, the

authors predicted that Yll049w in pathway N (shown on the left middle of Figure 2A), which binds Jnm1, and is regulated by common elements in pathway M, is a dynactin protein required for spindle partitioning in anaphase. This prediction would have been difficult without correlating the gene interaction with protein interaction information since each protein engages in more than a dozen other interactions in the genetic network.

Both the protein-protein interaction and gene interaction databases on which Kelley and Ideker's network analysis is based are known to be incomplete. This presents an interesting challenge: can the network analysis itself predict new functional, gene interactions? To show how it might, Kelley and Ideker invoked Alon's account of motifs (Milo, Shen-Orr, Itzkovitz et al., 2002). Motifs are subnetworks (two to four nodes) with a specific pattern of connectivity that occurs in a larger network far more frequently than would be expected by chance. In their between-pathway models, Kelley and Ideker found many four-node 'complete bipartite motifs', in which the first two nodes are connected in all possible ways to the second two nodes. Kelley and Ideker then took a step beyond Alon, interpreting other sub-networks in which only three of the four connections were realized as incomplete motifs and predicted that the fourth connection should occur as well (Figure 2B, left). Kelley and Ideker could not directly test the correctness of this inference procedure, but they offered indirect evidence of its reliability by withholding data about one-fifth of the interactions and constructing a network from the reduced data set. They then identified incomplete motifs in these networks and were able to predict eight-seven per cent of the connections about which information had been withheld.[6] This novel strategy for inferring new connections suggests a way of extending the ability of network analysis to advance new mechanistic hypotheses. Not only might researchers be able to use the structural and functional connections identified in databases to identify previously unrecognized structural and functional interactions in biological mechanisms, but they may also be able to predict new structural and functional relations not yet observed in the data.

A subsequent study by Bandyopadhyay, Kelley, Krogan et al. (2008) expanded upon the Kelley and Ideker study and revealed even more potential for networks built from both functional and structural interaction data to generate new hypotheses about mechanisms. They utilized data about both synthetic lethality and synthetic sickness, and in the case of synthetic sickness differentiated to 'positive' or 'alleviating' effects in which the reduction in colony growth is less than the product of the individual knockouts and 'negative' or 'aggravating' effects if it is greater than the product.[7]

---

[6] Kelley and Ideker found a different sub-network pattern in within-pathway models built from structural information: two nodes that share connections to common neighbors are also connected (Figure 2B, right). They performed the same sort of analysis of deleting edges and predicting them from the remaining network. In this case the accuracy in the validation test was far less (thirty-eight per cent) but still much greater than chance.

[7] Segrè, DeLuna, Church et al. (2005) found that interactions within the same cluster tended to be either all positive or all negative, and introduced the concept of 'monochromicity' that fostered merging clusters into a hierarchy that minimized clustering together positive and negative interactions.

Rather than relying on gene interaction databases, the researchers took advantage of rapid screening techniques to develop epistatic miniarray profiles (E-MAPs) of interactions among all pairs of genes in a group. Thus, for each pair of genes they determined whether it generated synthetic lethality or aggravated or alleviated synthetic sickness. They integrated these functional results with structural relations based on protein-protein interactions, assigning those genes/proteins that exhibited both strong genetic interactions and structural relations to a single complex and those with strong genetic but weak physical interactions to separate but functionally related modules. As shown in Figure 3, with this procedure they identified ninety-one distinct complexes (one, the co-chaperone prefoldin complex consisting of seven proteins that has twenty-five links to other modules, is not shown to enhance clarity). The color in which a complex is shown indicates whether the interactions within it are primarily alleviating or aggravating while the color of the edges indicates whether the interconnections between complexes are primarily alleviating or aggravating. Ten of these modules had not been picked out by any other techniques and since they have no common name, are indicated by just the names of the constituent proteins (for example, VPS64,HCS1 on the bottom left). These are new candidate mechanisms. The procedure also identified eight-four new members of modules corresponding to known mechanisms—new parts of previously identified mechanisms.
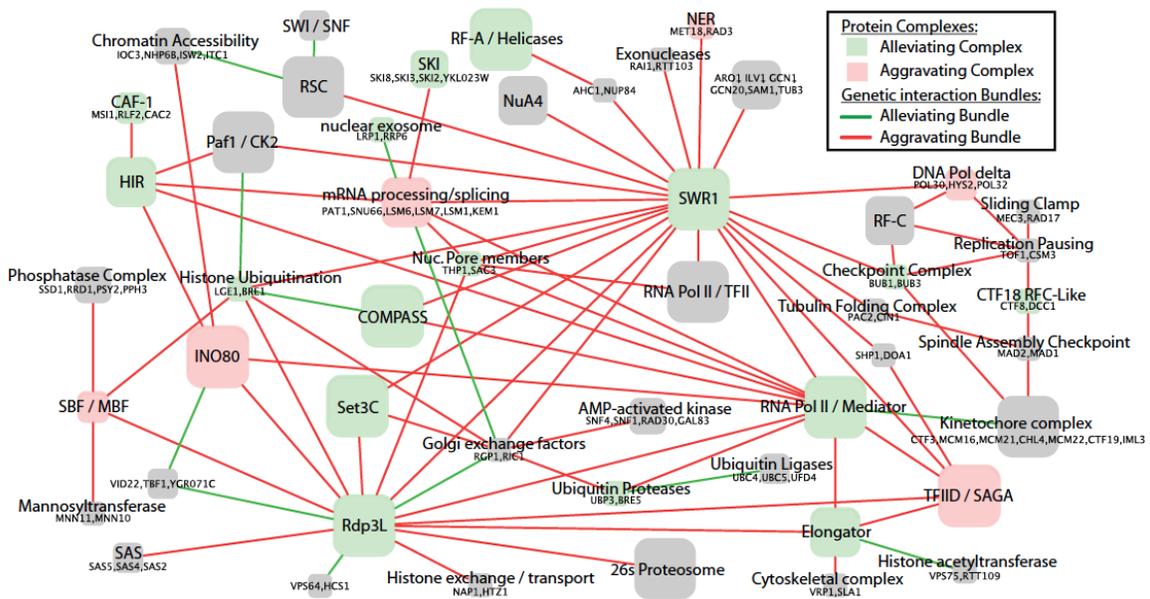
Figure 3. Network of protein complexes involved in chromosome biology. The size of nodes indicates the number of proteins in the complex whereas the color indicates whether the interactions within it are predominately alleviating or aggravating. Likewise, the color of the edges indicates whether the interactions between modules are predominately alleviating or aggravating. Reprinted from Bandyopadhyay, S., Kelley, R. M., Krogan, N. J., & Ideker, T. (2008). 'Functional Maps of Protein Complexes from Quantitative Genetic Interaction Data. *Plos Computational Biology, 4*(4): e1000065.

The studies I have described in this section show how researchers are using protein-protein interaction and gene interaction networks (individually or together) to advance new mechanistic hypotheses. Clusters or modules that do not correspond to previously identified mechanisms are new candidate mechanisms. Genes or proteins that cluster with those in known mechanisms are new candidate parts. Using guilt-by-association, researchers advanced new hypotheses about the function of mechanisms and their parts. Although confirming evidence for some of these has been generated, many of these hypotheses have not yet been tested. For my purposes of showing how network analyses can contribute to making discoveries about mechanisms, whether hypotheses have been confirmed is less important than that network research is being used to construct such hypotheses.

3. Analyzing Changes in Networks Across Conditions

A long-standing, if implicit, assumption in cellular research is that the mechanisms that are identified in research are enduring entities. Their parts, operations, and organization remain the same. They may behave differently on different occasions due to different inputs or different control processes, but the mechanisms are not themselves fundamentally altered. This assumption carried over to the network research I discussed in the previous section. The databases from which the network analyses are constructed do not take into account different conditions under which a cell might function. Rather, they extract information about protein-protein interactions or gene interactions from yeast cells grown on a rich medium under normal conditions. While this strategy has been highly successful, biologists have long recognized that cells are highly dynamic entities that function differently under different conditions. For example, different genes are expressed under different conditions, and accordingly different proteins are operative. Systems biologists are employing network analytic tools to characterize the differences in cell function at different times. If interacting nodes constitute mechanisms, revealing different interactions at different times can reveal that some mechanisms are differentially active under specific conditions, or even come into existence only in those conditions. These variations in mechanisms are missed if one does not look at specific conditions separately but averages over all.[8] In this section I describe two forms of network analysis that are revealing mechanisms that are differentially active or come into existence when cells experience different conditions.

---

[8] Once one recognizes variability in the mechanisms that exist in cells in different circumstances, one may wonder why researchers pursue static networks. One answer is that if static networks offer only a time-limited perspective on cell mechanisms, the mechanisms they identify do exist at those times. Another is that there are practical limits to how many conditions can be investigated in a given study as well as limits in the ability to represent and interpret the results.

3a. Active subnetworks

To investigate dynamic changes in cells, Ideker, Ozier, Schwikowski et al. (2002) proposed comparing gene expression and protein interactions under different conditions[9] to identify 'active subnetworks' (sometimes referred to as 'active modules', 'network hotspots', or 'responsive subnetworks'). They define active subnetworks as "connected sets of genes with unexpectedly high levels of differential expression" when the biological system is confronted with a specific circumstance. Since the cell exhibits different phenomena when it functions in different conditions, it is plausible to view the subnetworks active under specific circumstances as corresponding to the mechanisms responsible for these distinctive phenomena.

A study by Luscombe, Babu, Yu et al. (2004) illustrates the approach to discovering active sub-networks or mechanisms. These researchers contrasted the networks active when yeast cells were in conditions requiring one of five different activities—cell cycle, sporulation, diauxic shift, DNA damage, and stress response. Unlike most philosophical accounts of mechanistic explanations, which emphasize the generation of a product (e.g., the synthesis of a protein), the mechanisms Luscombe et al. investigated were those that controlled more basic mechanisms. Thus, they focused on transcription factors that regulate gene expression, yielding the proteins that figure in the sort of mechanisms characterized in philosophical accounts. They represented the network in which transcription factors regulate the expression of target genes in a circular plot (Figure 4, top left). Transcription factors are shown at the top of the plot, grouped by the number of conditions in which they are active (indicated by the color of the band around them and the number inside the band). Target genes are shown around the bottom, and the band around them indicates the number of conditions in which they are expressed. An arrow from a transcription factor to a target gene indicates that the transcription factor regulates a given target gene, with the color of the arrow indicating the number of conditions in which the target gene is expressed. Transcription factors can also operate on other transcription factors, so some arrows terminate in transcription factors. This plot makes it clear that over half of the target genes are only active in one condition.[10] Since the target genes that are expressed in each condition are the constituents of mechanisms active in that condition, these results indicate different mechanisms operative in yeast cells in these different conditions.

---

[9] Ideker et al.'s approach to identifying mechanisms is a version of the subtraction methodology familiar in other fields. For example, to identify the brain activity involved in performing a given mental operation, researchers will often subtract the activity found when a participant is performing a related task that does not require the operation. One important point to note is that just because a given mechanism is not differentially active in a given condition does not mean it is not performing important roles in that condition. Network analysts recognize this, but characterize the operations that are not differentially involved as 'housekeeping'.

[10] Only sixty-six interactions are active in four or more conditions. The researchers termed these 'hot links' and identify them as mostly regulating housekeeping functions.

On the right side of Figure 4B Luscombe et al. plot the connections between transcription factors and genes active in individual conditions and distinguish two clusters. They group cell cycle and sporulation as *endogenous* conditions since both involve the cell progressing through multiple stages under internal control. They term the three other cell conditions (diauxic shift, DNA damage, and stress response) 'exogenous' since they involve cellular responses to external stimuli.

To analyze how the networks are structured in each condition, Luscombe et al. employ a number of standard measures used in graph theory to characterize networks. I focus on two. Path length is the number of edges that must be traversed on the shortest path between two nodes. If mean path length is short, then signals are transmitted rapidly through the whole network. The clustering coefficient measures how interconnected a local set of nodes is. If it is high, then nodes are highly connected to their neighbors. Interconnected nodes form clusters (of the sort identified in the previously discussed studies) that can work together to perform a common activity.
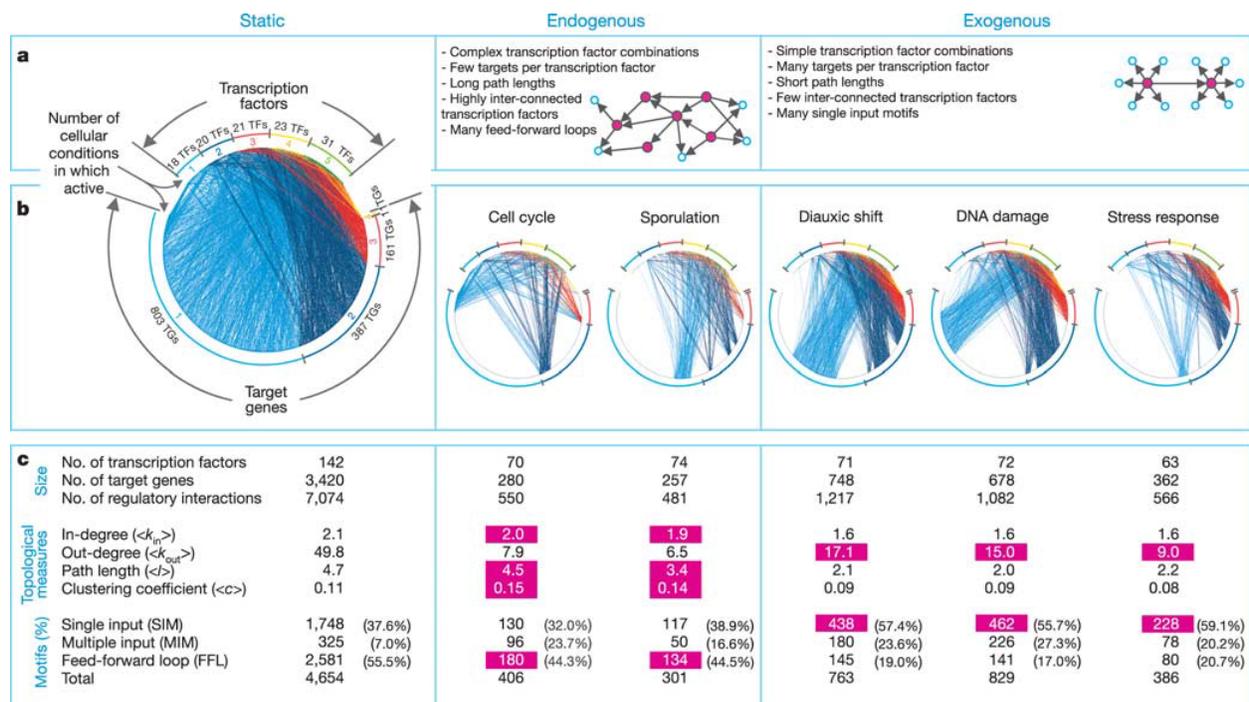
Figure 4. Changes in the network of transcription factors—target gene network under two endogenous conditions and three exogenous conditions. The static network is shown in the upper left. A. Contrast in network structure under endogenous and exogenous conditions. B. Network projections active in each condition. C. Network statistic. Reprinted by permission from Macmillan Publishers Ltd: *Nature,* Luscombe, N. M., Babu, M. M., Yu, H., Snyder, M., Teichmann, S. A., & Gerstein, M., 'Genomic Analysis of Regulatory Network Dynamics Reveals Large Topological Changes', copyright 2004.

The endogenous and exogenous conditions exhibit very different profiles on these network measures, as indicated by the highlighting of the values in the middle rows of Figure 4C.

Luscombe et al. relate these to the mechanisms functioning in the different conditions. They propose that the short path length in the exogenous condition indicates rapid response across the network while the low clustering suggests little coordination between components in generating a response. In contrast, the authors propose that the 'long paths in the multi-stage, endogenous conditions suggest slower action arising from the formation of regulatory chains to control intermediate phases' (p. 309). The larger clustering in endogenous conditions indicates that the nodes constitute modules that collaborate in performing activities. Overall, Luscombe et al. hypothesize that 'sub-networks [mechanisms] have evolved to produce rapid, large-scale responses in exogenous states, and carefully coordinated processes in endogenous conditions' (p. 309).

In addition to identifying different overall organization of mechanisms operative in the two types of conditions, Luscombe et al. determined that they exhibited different local organization in terms of sub-graphs or motifs. Of particular interest is their observation that feed-forward loops, in which an input node has both direct effects on an output unit and effects mediated by an intermediate, are twice as frequent in mechanisms operative in the endogenous condition. Alon and his colleagues showed that feed forward loops in which all connections are excitatory and in which input on both pathways is required for the output are able to function as persistence detectors in which the output only is generated when the input condition persists (Milo et al., 2002). This is important in mechanisms in which it would be problematic to commence an operation if the requirements for its successful completion were uncertain.

Having used their network representation both to identify different types of control mechanisms operative in different cell conditions and to reveal relevant features of their organization, Luscombe et al. also analyze their network to reveal how transcription factors exercise their control. They zeroed in on one of the endogenous control conditions in their study, the cell cycle, identified genes differentially expressed in specific phases of the cell cycle (early G1, late G1, S, G2, and M) and determined what transcription factors regulated each of these genes. They differentiated two classes of transcription factors: those that are approximately equally active in all phases, and those that are primarily active in just one phase. The saturation of the colors in Figure 5A indicates how active each of the transcription factors listed across the top is in each phase. The phase-specific transcription factors each have a dark row indicating they are most active in that phase. There are no dark rows for the ubiquitous transcription factors. Focusing on those most active in a specific phase, they showed in Figure 5B that these tend to regulate transcription factors active in subsequent phases (for those active in late phases, subsequent phases are those at the beginning of a subsequent cycle). This activation pattern presumably explains cells progress through the stages of the cell cycle. As Figure 5C illustrates, the ubiquitous transcription factors tend to regulate other transcription factors that are phase specific. This analysis of transcription factors provides perspective on how particular cell mechanisms are regulated.

(a) phase-specific and ubiquitous transcription factors
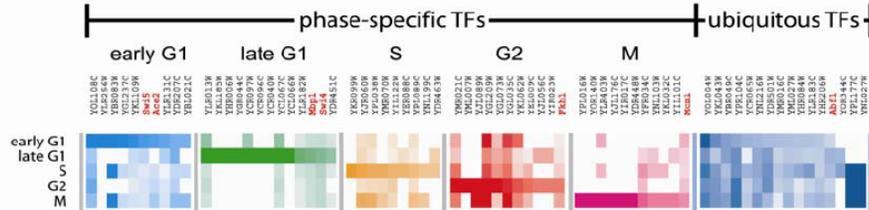
(b) serial inter-regulation
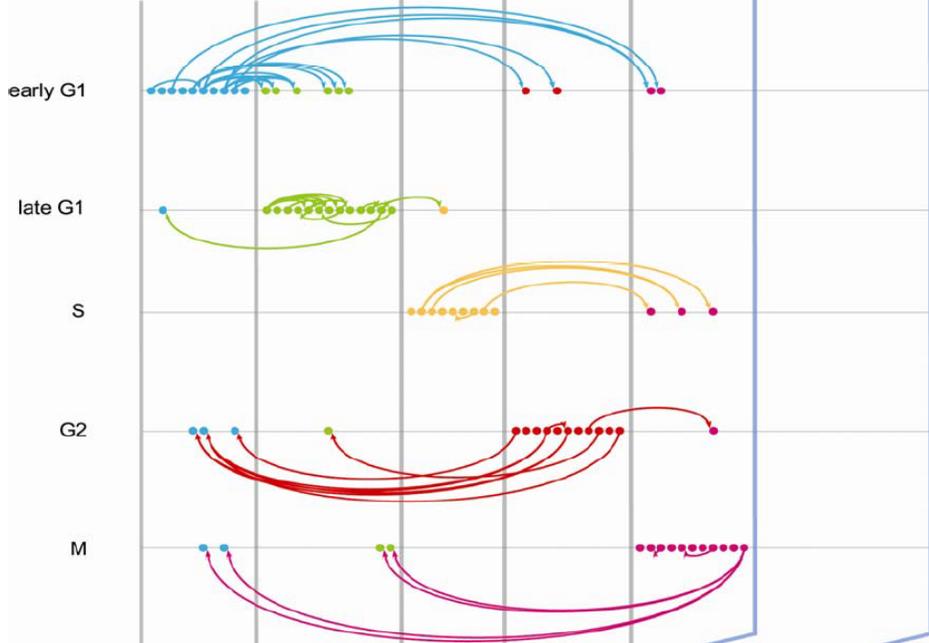
(c) parallel inter-regulation
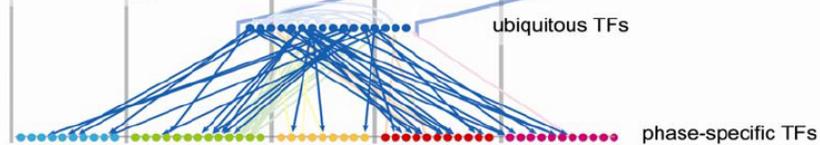
Figure 5. A. Genes clustered according to the phase of the cell cycle in which they are active—the darkness of the shading indicates how active each gene is during each phase. B. Projections of regulation from those active in one phase, mostly to those in a subsequent phase. C. Projections from transcription factors active in all phases to those active in specific phases. Reprinted by permission from Macmillan Publishers Ltd: Nature, Luscombe, N. M., Babu, M. M., Yu, H., Snyder, M., Teichmann, S. A., & Gerstein, M., 'Genomic Analysis of Regulatory Network Dynamics Reveals Large Topological Changes', copyright 2004.

The Luscombe et al. study on which I have focused exemplifies how network analyses can be used to identify active subnetworks and thus control mechanisms that are active or even come into existence when the cell confronts specific conditions. By focusing on transcription factors that regulate the transcription of target genes and showing that

different networks are involved in specific cell conditions, Luscombe et al. provided a means to identify active control and target mechanisms. By applying network measures to the networks active in different conditions, they were able to reveal organizational differences between control mechanisms operative in different types of conditions and even to zero in on the temporal dynamics through which control is exercised.

3b. Differential network biology

In a second example I consider how network analyses such as those discussed in section 2 above can be extended to reveal mechanisms operative or even existing only in specific cell conditions. Ideker and Krogan (2012) term this approach 'differential network biology'. In section 2b I introduced epistatic miniarray profiles (E-MAPs) of gene interactions. Differential network biology subtracts the E-MAP for a cell in an altered state from one for a cell in the default state to yield a differential E-MAP (dE-MAP). A dE-MAP thus focuses attention on those genes that interact differently in the two conditions. As the cases below illustrate, the contrast reveals genes/proteins whose interactions do not appear significant in either condition treated alone but may constitute important mechanisms in the cell.

Bandyopadhyay, Mehta, Kuo et al. (2010) introduced the dE-MAP strategy in a study in which they compared yeast colonies grown under two conditions, an unperturbed condition in a rich medium and a perturbed condition resulting from adding to the medium methyl methanesulfonate (MMS), a DNA-alkylating agent that creates base modifications in DNA. When creating the differential map from the E-maps constructed in each condition, Bandyopadhyay et al. identified 873 highly significant interactions, sixty-two per cent of which were not detectable in either static E-MAP. The researchers developed several new hypotheses about the mechanisms involved in DNA damage response, but of particular interest was the finding that the connections that showed up as changed in the differential network were more likely to connect modules (identified using a variety of criteria) than nodes within modules. The researchers conclude 'known protein complexes tend to be stable across conditions—it is the genetic interactions between these modules that are reprogrammed in response to perturbation.' Put in the language of mechanisms, the components of mechanisms may exist in both conditions, but they are assembled into different mechanisms on different occasions.

Guénolé, Srivas, Vreeken et al. (2013) followed up on this finding in order -'to understand how functional interconnections between pathways are formed and altered in response to various genotoxic insults.' They deployed three different DNA damaging agents: MMS; Camptothecin (CPT), a topoisomerase I inhibitor that causes double-strand breaks in DNA; and Zeocin (ZEO), a DNA-intercalating agent that causes single-strand lesions. They used the dE-MAP strategy first to identify 'specific repair mechanisms' activated in the DNA damage response process. This involved creating yeast strains which each had a deletion of one of fifty-five representative genes known to figure in DNA damage repair pathways. They crossed these with approximately 2000 strains involving deletions of genes involved either in DNA damage repair or in a related process—cell-cycle regulation, chromatin organization, replication, transcription, and protein transport (included so as to investigate the interaction of DNA damage repair with these other cell processes). From each of these

four E-MAPs they generated networks. While these networks shared a number of edges (interpreted as involved in basic cellular processes), they also exhibited substantial differences—almost one-half of the positive interactions and one-third of the negative interactions were unique to one of the treated networks. By subtracting the network in each of the three damage conditions from the unperturbed case from the network, they generated three differential networks. As shown in Figure 6A, most interactions only showed up in one of the differential networks.[11]
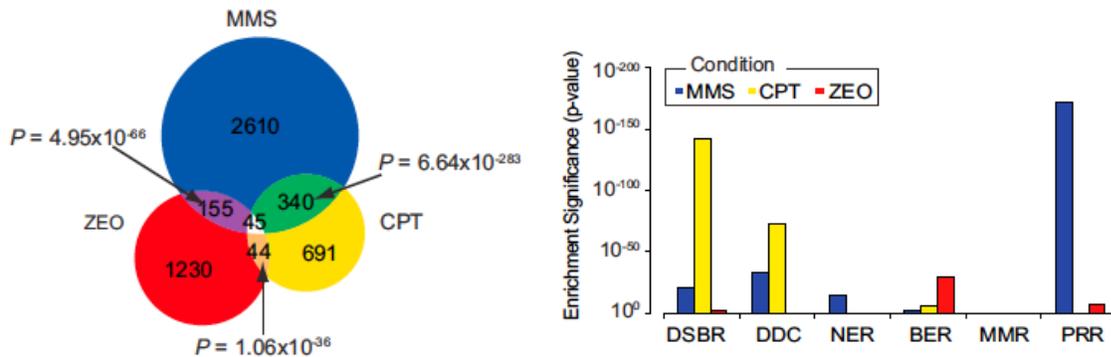
Figure 6. Analysis of the differential networks identify by Guénolé et al. The labels MMS, ZEO, and CPT refer to the differential networks in which the static network generated in the unperturbed condition was subtracted from the static network in the designated perturbed condition. A. Most of the connections identified in the differential networks were unique to one of the differential networks. B. The *p*-value for the enrichment for genes in different pathways for the three differential networks. Reprinted from *Molecular Cell*, Vol 49, Guénolé, A., Srivas, R., Vreeken, K., Wang, Z. Z., Wang, S., Krogan, N. J., Ideker, T., van Attikum, H., 'Dissection of DNA Damage Responses Using Multiconditional Genetic Interaction Maps', Figure 2c, Copyright 2010, with permission from Elsevier.

To determine whether the interactions in the three differential networks reflected known mechanisms, Guénolé et al. investigated whether they were specifically enriched for genes in the six major DNA damage repair pathways. They found that the individual networks were differentially enriched (Figure 6b) and that the enrichment in each network made sense given the action of the specific damage agents. For example, CPT is known to act by stabilizing DNA topoisomerase 1-DNA complexes, resulting in double stranded breaks (DSB) during S-phase. Fittingly, the CPT differential network was greatly enriched for DSB repair. In contrast, the ZEO network was less enriched for DSB repair. Instead, it was enriched for genes employed in base excision repair (BER) and postreplication repair (PBR). This fits with the fact that at the concentrations used ZEO generates abasic sites

---

[11] Guénolé et al. investigted the forty-five genes involved in all differential networks. Several of them were highly conserved genes involved in DNA damage repair pathways. The authors interpreted as showing a 'a nonredundant role for these factors in multiple DDR mechanisms, including the resection of DSBs and subsequent Mec1-dependent activation of the DDC.'

rather than DNA strand breaks. Accordingly, Guénolé et al. conclude they were detecting repair mechanisms suited for the particular condition in the cell.

To determine whether their approach could generate new information about these DNA-damage repair mechanisms, Guénolé et al. investigated the hubs (highly connected nodes) in the differential networks. They determined that the hubs were more sensitive to DNA damage but also took on more diverse functions. They focused specifically on SAE2, a hub found in the CPT network and known to encode the homolog of the human endonuclease CtIP that functions in processing double-strand breaks into 3' single-stranded tails. They found that the majority of *SAE2*'s interactions were induced by CPT and that *SAE2* interacted with numerous DNA repair genes. An unanticipated finding was that *SAE2* has negative interactions with genes belonging to the PP4 complex that are needed to dephosphorylate Rad53, the major checkpoint kinase, and to allow recovery from the arrest of the cell-cycle. The researchers proposed that SAE2 works in parallel with PP4, a proposal they confirmed by examining the effects of mutating just one of *SAE2* and *PPH3* (a component of the PP4 complex) or of mutating both. The authors conclude: 'this example illustrates the power of our differential network analysis in identifying connections between different DDR factors'—that is, in finding connections between components in a mechanism.

A final step in Guénolé et al.'s analysis was to integrate the differential genetic interaction data with structural (protein-protein) interaction data into a global map. Much like the network analyses I discussed in the section 2, modules are identified as sharing both structural and functional (genetic) interactions. These modules are shown as nodes in Figure 7 while the edges reflect funcitonal (genetic) interactions between modules. The edges derived from the three differential networks are overlaid in color, with those from just one differential network shown in red, yellow, and blue. The few edges derived from two differential networks are shown in purple, green, and orange. Not only does this figure make clear how the various repair mechanisms are coordinated, it also reveals prevously unknown relations, such as that between RTT109 and Polδ, a repair specific polymerase involved in translesion synthesis. The researchers then investigated the mechanistic connection in more detail by acetylating HEK56, which is required for recruiting Polδ to the site at which DNA replication is stalled.

Figure 7. Functional interactions between modules, shown as nodes. Colors indicate the dE-MAPs in which genetic interactions are found between modules. Those shown in blue, yellow, and red only occur in one dE-MAP, suggesting that the mechanism involving the different modules only forms in those conditions. Reprinted from *Molecular Cell*, Vol 49, Guénolé, A., Srivas, R., Vreeken, K., Wang, Z. Z., Wang, S., Krogan, N. J., Ideker, T., van Attikum, H., 'Dissection of DNA Damage Responses using Multiconditional Genetic Interaction Maps', Figure 6A, Copyright 2010, with permission from Elsevier.

The two examples of differential network biology that I have presented illustrate how static representations of functional (gene) interactions (E-MAPs) developed for yeast grown in different conditions can be combined to reveal mechanisms that become operative only when the cell is grown in specific conditions. Not only did the studies reveal mechanisms already known to be operative in these specific conditions, but also previously unknown mechanisms. Moreover, the Guénolé et al. study makes clear that these mechanisms result from establishing connections between modules already present in the cell. Together with the Bandyopadhyay et al. (2010) example discussed earlier, these studies show how network analyses enable researchers to advance beyond a static view of cell mechanisms to one that sees mechanisms as dynamically changing given conditions in the cell.

4. Conclusion

Traditional strategies for developing mechanistic explanations start with a mechanism identified as responsible for a phenomenon and invoke experimental manipulations to identify its parts and operations. For example, demonstrating that lesioning or poisoning a part inhibits a phenomenon is interpreted as showing that the part inhibited is a working part of the mechanism. Likewise, demonstrating that activity in a component increases when the phenomenon is exhibited is again taken to show that the part figures in the operation of the mechanism. My focus has been to describe a different strategy for developing hypotheses about biological mechanisms and their components that has been developed and deployed within systems biology. It takes advantage of the large databases characterizing structural and functional interactions between components of cells, especially genes and proteins, from which network representations are constructed. These networks are analyzed to identify clusters of highly connected nodes, which sometimes can only be detected when data from yeast grown in different conditions are compared. The interconnections between these nodes suggest that the entities being represented interact in the manner of a mechanism. In the examples I have presented, this has resulted in the identification of new mechanisms and new parts of mechanisms beyond those that have been discovered through more traditional mechanistic strategies. Invoking the guilt-by-association strategy, researchers hypothesize that genes or proteins not previously associated with a specific cell activity are involved in the same cell phenomenon as other nodes that correspond to genes or proteins whose role is already known. That is, they are new constituents of the same mechanism.

In articulating how network analyses facilitate making new discoveries about mechanisms I am not suggesting that these strategies operate totally independently of more traditional mechanistic research strategies. First, the data from which networks are discovered result from (automated) versions of more traditional molecular experiments. Second, interpretation of clusters in networks relies directly on annotations, often derived from ontologies such as GO. The hierarchical organization of these ontologies is intended to encapsulate curated mechanistic knowledge about the structures in which proteins appear, the chemical reactions in which they are involved, and the cellular phenomena in which they figure.[12] Finally, the hypotheses generated from the network analyses are further investigated using more traditional mechanistic interventions. The network strategy may generate a hypothesis that a given gene or protein is involved in a given cell function, but to confirm or falsify this inference, researchers must intervene directly on the gene or protein and determine whether it has the predicted effects. Moreover, such intervention is needed

---

[12] More recently, Dutkowski, Kramer, Surma et al. (2013) have taken advantage of the cell-wide nature of network representations to develop a Network Extracted Ontology (NeXO). From NeXO they have made mechanistic discoveries that have been accepted into GO. NeXO also includes large numbers of genes that reside outside the scope of the curated knowledge on which GO is based. Nonetheless, NeXO is also dependent on GO. The researchers constrained themselves to developing an ontology of roughly the same size as GO and devoted major efforts to align NeXO with GO to take advantage of the annotations already in GO.

to work out the details of how it contributes to the phenomenon (i.e., what operation it performs in the mechanism).

While acknowledging major respects in which network reasoning about mechanisms is reliant on more traditional mechanistic approaches, my goal has been to show how it provides new tools for developing and expanding mechanistic explanations. The reasoning begins not with trying to explain a particular phenomenon but with the databases storing data about protein and gene interactions and uses clusters in the networks constructed from this data to make inferences about mechanisms. These steps are not part of the traditional mechanistic explanations, but as I have tried to show, provide additional strategies for constructing and amending mechanistic explanations. Network approaches also points to ways the traditional conception of mechanisms needs to be expanded and amended. As a result of adopting a cell-wide focus, network approaches reveal that the various mechanisms of cells are far more interconnected than traditionally assumed. Moreover, as the examples in section 3 showed, biological mechanisms may not be a stable as traditionally conceived. The constituents of cells are organized into different clusters, corresponding to different mechanisms, when cells confront different conditions. The clusters identified in networks, whether representing static or differential conditions, though, are still treated by researchers as representing mechanisms. Network analysis has introduced new avenues for generating hypotheses about mechanisms beyond those pursued in traditional mechanistic research, and has thereby extended the ability of biologists to make discoveries about biological mechanisms.

Acknowledgements

References

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: tool for the unification of biology. *Nature Genetics, 25*, 25-29.

Bandyopadhyay, S., Kelley, R. M., Krogan, N. J., & Ideker, T. (2008). Functional maps of protein complexes from quantitative genetic interaction data. *Plos Computational Biology, 4*, e1000065.

Bandyopadhyay, S., Mehta, M., Kuo, D., Sung, M. K., Chuang, R., Jaehnig, E. J., Bodenmiller, B., Licon, K., Copeland, W., Shales, M., Fiedler, D., Dutkowski, J., Guenole, A., van Attikum, H., Shokat, K. M., Kolodner, R. D., Huh, W. K., Aebersold, R., Keogh, M. C., Krogan, N. J.,

& Ideker, T. (2010). Rewiring of genetic networks in response to DNA damage. *Science, 330*, 1385-1389.

Bechtel, W. (2006). *Discovering cell mechanisms: The creation of modern cell biology*. Cambridge: Cambridge University Press.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 421-441.

Bechtel, W., & Richardson, R. C. (1993/2010). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Cambridge, MA: MIT Press. 1993 edition published by Princeton University Press.

Craver, C. F., & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. Chicago: University of Chicago Press.

Dutkowski, J., Kramer, M., Surma, M. A., Balakrishnan, R., Cherry, J. M., Krogan, N. J., & Ideker, T. (2013). A gene ontology inferred from molecular networks. *Nature Biotechnology, 31*, 38-45.

Fields, S., & Song, O. K. (1989). A Novel Genetic System to Detect Protein Protein Interactions. *Nature, 340*, 245-246.

Guénolé, A., Srivas, R., Vreeken, K., Wang, Z. Z., Wang, S. Y., Krogan, N. J., Ideker, T., & van Attikum, H. (2013). Dissection of DNA Damage Responses Using Multiconditional Genetic Interaction Maps. *Molecular Cell, 49*, 346-358.

Hu, Z., Chang, Y.-C., Wang, Y., Huang, C.-L., Liu, Y., Tian, F., Granger, B., & DeLisi, C. (2013). VisANT 4.0: Integrative network platform to connect genes, drugs, diseases and therapies. *Nucleic Acids Research*.

Ideker, T., & Krogan, Nevan J. (2012). Differential network biology. *Molecular Systems Biology, 8*, 565.

Ideker, T., Ozier, O., Schwikowski, B., & Siegel, A. F. (2002). Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics, 18 Suppl 1*, S233-240.

Kelley, R. M., & Ideker, T. (2005). Systematic interpretation of genetic interactions using protein networks. *Nature Biotechnology, 23*, 561-566.

Luscombe, N. M., Babu, M. M., Yu, H., Snyder, M., Teichmann, S. A., & Gerstein, M. (2004). Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature, 431*, 308-312.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1-25.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network Motifs: Simple Building Blocks of Complex Networks. *Science, 298*, 824-827.

Mitchell, P. (1961). Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature, 191*, 144-148.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., & Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnology, 17*, 1030-1032.

Segrè, D., DeLuna, A., Church, G. M., & Kishony, R. (2005). Modular epistasis in yeast metabolism. *Nature Genetics, 37*, 77-83.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for

integrated models of biomolecular interaction networks. *Genome Res, 13*, 2498-2504.

Spirin, V., & Mirny, L. A. (2003). Protein complexes and functional modules in molecular networks. *Proceedings of the National Academy of Sciences of the United States of America, 100*, 12123-12128.

Su, G., Morris, J. H., Demchak, B., & Bader, G. D. (2014). Biological network exploration with cytoscape 3. *Curr Protoc Bioinformatics, 47*, 8 13 11-18 13 24.

Young, K. H. (1998). Yeast two-hybrid: So many interactions, (in) so little time … *Biology of Reproduction, 58*, 302-311.