Research news
A road map of yeast interactions
Jonathan B Weitzman

Analysis of a yeast network that integrates five interaction datasets reveals the presence of large topological structures reflecting biological themes.

Open any atlas and you will find a variety of maps for each country or territory. These will include information about different features, such as geology, climate, population and so on. Integrating information from the different maps allows the reader to appreciate the landscape they are exploring. The same is true for cellular maps that chart genetic, protein or functional interactions within the cell. Now, in *Journal of Biology* [1], Frederick (Fritz) Roth from Harvard Medical School and colleagues from Toronto and Montreal describe key topological features of an enormous map of macromolecular interactions in yeast (see 'The bottom line' box for a summary of the work).

Integrating interaction maps
Cellular processes can be explored by investigating interactions between biological components. The complex system of the cell is a network of interconnections - proteins interact with other proteins or with DNA, and genes can interact functionally with one another. Large-scale projects have attempted to define the entire list of genetic components (the genome), their expression patterns (the transcriptome), their protein products (the proteome) and the interactions between

them (the interactome). A key challenge is to integrate these different maps so as to develop a conceptual model for dynamic cellular behavior.

Roth and colleagues have created an integrated network map that incorporates five different types of biological interaction data for the yeast

Saccharomyces cerevisiae [1]. Each **node** in their network represents a gene or its protein product (see the 'Background' box for further explanations and definitions). Genes can themselves be connected by sequence homology, or by mRNA expression correlations; their protein products can interact with

> **The bottom line**
>
> - An integrated network has been constructed by combining five different types of data from experiments in yeast that indicate protein or gene interactions.
>
> - The first unit of organization within the network is motifs of three or more genes, or proteins, connected via different types of interactions.
>
> - Analysis of the higher-order structure of the network reveals the presence of several network themes - topological clusters of connected network motifs that appear to have related biological functions.
>
> - A simplification of collections of themes into thematic maps provides insights into the fundamental architecture of cellular networks and the dynamic relationships between biological processes.
>
> - This approach provides a methodological framework for creating and characterizing networks that reflect the complexity of the molecular landscape of living cells.

<div style="border:1px solid #000">

Background

- Biological networks are made up of **nodes** (representing individual genes or their protein products) that are joined by **edges** (or links) which reflect a genetic, physical or functional interaction between two nodes.

- Interactions may be directly detected, for example by mapping **protein-protein interactions** using an approach such as the yeast two-hybrid assay or by mapping **protein-DNA interactions** using **chromatin-immunoprecipitation** (ChIP). Or they may be indirectly detected, for example on the basis of **co-expression** or genetic interactions.

- **Synthetic sick or lethal (SSL)** refers to a genetic interaction in which the combined mutation of two genes causes a phenotype (fitness reduction or death) that is more severe than either mutation alone.

- **Network motifs** are recurring interconnection patterns (or subgraphs) that are over-represented in biological networks compared to a randomized network.

- **Network themes** are enriched topological patterns that contain **clusters** of overlapping motifs. These higher-order themes represent genetic and regulatory interactions between complexes or between a transcriptional regulator and a complex.

- **Thematic maps** are simplified network graphs, in which theme structures are represented as the nodes, while the links represent inter-complex genetic interactions.

</div>

each other directly or may regulate the expression of other genes. Finally genes can also be linked genetically, if mutations in them cause **synthetic sick or lethal (SSL)** interactions. Roth's group combined data from sequence homology searches, **co-expression** microarray analysis, **protein-protein interaction** screens, genome-wide **chromatin immunoprecipitation** experiments and an SSL screen, to create a 'multi-color' integrated network, in which each color represents one type of interaction.

"Protein interaction mapping projects have emerged as an extremely powerful resource for understanding, and ultimately modeling, cell function

on a genome-wide scale," comments bioinformatics researcher Trey Ideker from the University of California, San Diego. "Although protein-protein interactions were some of the first to be measured at high-throughput, a variety of other interaction types are also being cataloged, such as genetic (synthetic-lethal) and **protein-DNA interactions**," he says, adding that the Roth study extends previous work by considering all of these different interaction types together. "The attempt to unify networks composed of heterologous components is certainly forward-looking," agrees Zoltan Oltvai from the University of Pittsburgh School of Medicine, Pennsylvania.

"In all five cases an interaction indicates a heightened chance of functional relationship," explains Roth. "These genes/proteins are more likely to have something to do with each other or to function together." He notes that several studies had reported a certain amount of overlap between different types of interaction, such as protein-protein and co-expression correlation or protein-protein interaction with phenotypic similarity. Roth was particularly interested in SSL genetic interactions and had begun collaborating with Charles Boone's laboratory at the University of Toronto, where work was underway to mutate pairs of genes in yeast to examine double-mutant phenotypes [2]. "This is a more abstract notion of interaction," notes Roth. "The protein products don't necessarily physically touch each other, but the presence of one gene can rescue the loss of the other." The Harvard group had already explored methods to predict SSL relationships and protein complexes, by combining multiple biological data types [3,4]. Roth was keen to improve methods for predicting interactions and function, and he wanted to explore the higher-order structure of an integrated network map (see the 'Behind the scenes' box for more of the rationale for the work).

Navigating towards motifs and themes

The yeast network produced by Roth and colleagues [1] contains 5,831 nodes (genes or proteins) linked together by a staggering 154,759 interactions ('**edges**' in network jargon). But building these networks is a lot easier than figuring out what they mean. To explore their map, Roth and colleagues were inspired by ideas from the field of network theory and the seminal work of Uri Alon at the Weizmann Institute of Science, Rehovot, Israel. Alon's group characterized the architecture of complex systems and defined basic network components called '**motifs**' [5,6]. "When Alon and

Behind the scenes

Journal of Biology asked Fritz Roth about the creation of the integrated yeast network and analysis of its topological features.

What motivated you to embark on the *S. cerevisiae* integrated network project?

The inspiration came from work by Uri Alon's group [5,6] that provided the idea of network motifs. We felt that these 'triangular' motifs might be signatures of a higher-order structure. We were also interested in synthetic-lethal genetic interactions and how these related to expression correlations or protein interactions and homology. Simple overlap analysis doesn't really tell the whole story, so we constructed the integrated yeast network, combining five different types of interaction, to see if we could distinguish between motifs and larger topological structures.

How long did the study take and what were the difficult steps you encountered?

In early 2003 we began collaborating with Charlie Boone's group to look at their synthetic lethal interaction data. One major hurdle was that in order to establish which motifs are enriched relative to random networks one has to generate randomized networks. This sounds simple, but is in fact a remarkably complicated question. We spent a long time arguing about what was the best way to randomize the graphs, about which network properties should be preserved and which randomized.

What was your initial reaction to the results and how were they received by others?

Our approach overlays multiple types of interaction and can characterize the properties of the network. Many of the motifs can be explained intuitively but some are less obvious. We were struck by how interconnected the motifs are and how we can understand relationships between genes and proteins. Everybody is particularly intrigued by the thematic maps. People have gotten most interested in the idea of drawing maps of redundant systems, where you have pairs of complexes with lots of genetic interactions between them.

What are the next steps?

Our chief interest is in predicting interactions and function. I think that this will get more exciting as we get more synthetic lethal interaction data. Right now we are limited by the roughly 4% of pairs of genes that have been tested for genetic interactions. It should also be feasible to do this in other organisms. We have partial protein interaction maps in worms, flies and humans, and I predict that we will find many of the same motifs. I would be shocked if we couldn't repeat this exercise in mammalian systems in the next two or three years.
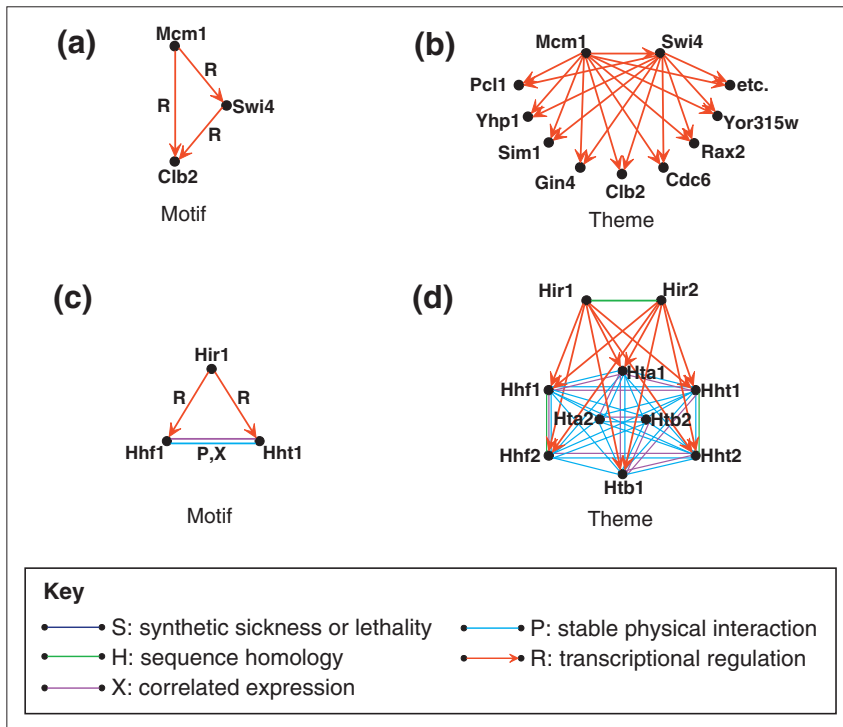
colleagues published the concept of elementary interaction patterns in cellular (and other) networks, it was important not only for our further understanding of network topology, but also because they could develop certain predictions regarding network behavior," explains Oltvai.

"Alon was the first to show that protein-protein interaction networks encode particular sub-circuits (motifs), such as feed-back and feed-forward loops," notes Ideker. These concepts were welcomed by researchers in the nascent field of systems biology, who construct complex network models. "Motif analysis is increasingly being used to understand the properties of integrated networks," comments Ernest Fraenkel from the Whitehead Institute in Cambridge, USA. "For example, network motifs were recently used systematically to assess the relationship between the transcription regulatory network and chromosomal organization in *Escherichia coli* and in budding yeast [7], yielding significant biological insight."

Roth and colleagues found many three-node 'triangle' motifs that were enriched within their network (see Figure 1a,b). They defined seven motif types in the yeast integrated network: transcriptional feed-forward (Figure 1a); co-pointing motifs, in which a gene is regulated by two related or interacting transcription factors (Figure 1c); regulonic motifs, in which co-regulation is accompanied by co-expression; protein complexes; SSL triangles; protein complexes with partially redundant members; and compensatory complexes/processes. They also identified some four-node motifs, but these are much more complex to identify and compute.

Both Alon's group and Oltvai's group (in collaboration with Barabási) had previously shown that motifs sometimes appear in **clusters** [5,8,9]. "We demonstrated that motifs mostly do not exist in isolation, but that they aggregate into larger structures and this is a natural consequence of the networks' global topological organization," notes Oltvai. Roth also found that most motifs were componenets of higher-order structures, and coined the term **'network themes'** to describe the recurrent examples of higher-order structures. Themes can be made up of

Figure 1
Examples of network motifs **(a,c)** and themes **(b,d)**. (a,b) A transcriptional feed-forward motif that occurs repeatedly in the control of the cell cycle. (c,d) Two targets of transcription that are regulated by co-expression, protein-protein interaction or homology during periodic histone gene expression. Images reproduced from [1].

of social cliques with a lot of interactions. Many of the same ideas apply." Roth adds that his group has previously used ideas that come straight out of communications theory to analyze protein interaction networks.

The motivation for computational modelling is to generate hypotheses that can then be tested experimentally. "In my view, one justification for looking at network motifs as interesting objects, aside from the fact that they form clusters, is that each motif (in transcription networks at least) can be assigned defined functions," comments Alon. "These functions can then be tested experimentally in living cells using measurements on motifs embedded inside the entire network." Indeed, laboratory results have supported many of the predictions made by Alon's group in fields as diverse as the *E.coli* flagellum and sporulation in *Bacillus subtilis*. Roth is keen to make further predictions about genetic links between the thematic groups in yeast.

Researchers agree that this approach will be enhanced by more data about genetic interactions. "I like the extensive analysis of multi-colored networks of diverse interactions," says Alon. "I think that the Roth paper is original and will have significant impact as we gain more and more data on integrated networks of interactions." Some experts in the field have raised questions about whether the different types of 'interactions' are all comparable. But analysis of these complex networks will indicate how reliable the links are, and how useful the concepts of motifs and themes are in predicting biologically relevant functions. The study by Roth and colleagues has laid down a methodology for large-scale integration of maps and multi-color network analysis. They are keen to see how similar approaches proceed in other organisms, and whether the general thematic maps are conserved. "I think that better use of topological patterns could help predict all sorts of interactions," concludes Roth.

multiple occurrences of the same motif (Figure 1b) or several different types of motif (Figure 1d).

"Roth shows that the types of molecular sub-circuits encoded by biology are exponentially richer than was previously thought. This complements work by others that is also directed at finding the commonality between networks of different types," says Ideker. A recent study of protein interactions from Ideker's group proposes a specific computational model of how physical and genetic interaction networks relate to each other to delineate redundant and/or synergistic molecular machinery [10]. "Roth's group goes beyond the motif analysis by providing a higher-level organizing principle," says Fraenkel. "The biological relevance of a network theme is often much clearer than the relevance of the underlying motifs. Network themes should also be less sensitive to the noise in individual data sources."

Complexes and cliques

The characterization of network themes led Roth and colleagues [1] to propose one further step: the construction of **thematic maps**, which chart a simplified landscape by showing only the larger structures and the links between them. He compares them to sub-graph structures in other complex networks. "For example, you could have social networks with certain groups of people, by whatever classification scheme that you wanted to impose, who were more likely to interact with each other. So, social networks have cliques just as protein networks have complexes. And there might be pairs of complexes that have a lot of synthetic-lethal interactions, just as there might be pairs

References

1. Zhang LV, King OD, Wong SL, Goldberg DS, Tong AHY, Lesage G, Andrews B, Bussey H, Boone C, Roth FP: **Motifs, themes and thematic maps of an integrated *Saccharomyces cerevisiae* interaction network.** *J Biol* 2005, **4:**6.
2. Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M *et al.*: **Global mapping of the yeast genetic interaction network.** *Science* 2004, **303:**808-813.
3. Wong SL, Zhang LV, Tong AH, Li Z, Goldberg DS, King OD, Lesage G, Vidal M, Andrews B, Bussey H *et al.*: **Combining biological networks to predict genetic interactions.** *Proc Natl Acad Sci USA* 2004, **101:**15682-15687.
4. Zhang LV, Wong SL, King OD, Roth FP: **Predicting co-complexed protein pairs using genomic and proteomic data integration.** *BMC Bioinformatics* 2004, **5:**38.
5. Shen-Orr SS, Milo R, Mangan S, Alon U: **Network motifs in the transcriptional regulation network of *Escherichia coli*.** *Nat Genet* 2002, **31:**64-68.
6. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U: **Network motifs: simple building blocks of complex networks.** *Science* 2002, **298:**824-827.
7. Hershberg R, Yeger-Lotem E, Margalit H: **Chromosomal organization is shaped by the transcription regulatory network.** *Trends Genet* 2005, **21:**138-142.
8. Dobrin R, Beg QK, Barabasi AL, Oltvai ZN: **Aggregation of topological motifs in the *Escherichia coli* transcriptional regulatory network.** *BMC Bioinformatics* 2004, **5:**10.
9. Vazquez A, Dobrin R, Sergi D, Eckmann JP, Oltvai ZN, Barabasi AL: **The topological relationship between the large-scale attributes and local interaction patterns of complex networks.** *Proc Natl Acad Sci USA* 2004, **101:**17940-51794.
10. Kelley R, Ideker T: **Systematic interpretation of genetic interactions using protein networks.** *Nat Biotechnol* 2005, **23:**561-566.

Jonathan B Weitzman is a scientist and science writer based in Paris, France.
E-mail: jonathanweitzman@hotmail.com