REVIEW

nature chemical biology

Targeting and tinkering with interaction networks

Robert B Russell[1] & Patrick Aloy[2,3]

Biological interaction networks have been in the scientific limelight for nearly a decade. Increasingly, the concept of network biology and its various applications are becoming more commonplace in the community. Recent years have seen networks move from pretty pictures with limited application to solid concepts that are increasingly used to understand the fundamentals of biology. They are no longer merely results of postgenome analysis projects, but are now the starting point of many of the most exciting new scientific developments. We discuss here recent progress in identifying and understanding interaction networks, new tools that use them in predictive ways in exciting areas of biology, and how they have become the focus of many efforts to study, design and tinker with biological systems, with applications in biomedicine, bioengineering, ecology and beyond.

Before any attempt to study, perturb or redesign networks, it is important to understand their history and limitations. To this end, protein-protein interaction networks are a good representative, as they have been the subject of intense focus, and so far they contain the most data. Like the first results from genome sequencing projects[1], interaction networks available now should be considered as partial drafts of the complete story. Just as it is naïve to think that a partial set of expressed sequence tags constitutes enough to understand the complete genome of an organism, it is also naïve to think that a single high-throughput yeast two-hybrid screen of a proteome is enough to understand the complete interactome. Indeed, the interaction network that makes up an organism is possibly more complex than the genome that underlies it, given that it is the ultimate expression of the genome and is much more subject to temporal or environmental conditions than the underlying genomic recipe.

The two workhorses of protein interaction discovery are the two-hybrid system[2] and affinity purification or pull-down approaches[3], in their many flavors (for example, see refs. 4,5). Their popularity can be attributed to a coupling of straightforward genetic manipulations with relatively easy readouts to determine interactions, which makes them readily applicable to whole proteomes or very large numbers of proteins in a relatively short time. Both approaches have been applied to full genomes[6–9] and to significant fractions of the genomes for major model organisms[10,11] and for humans[12,13]. Although these first whole-organism interactome networks are far from comprehensive, they have revealed strengths and weaknesses of the approaches, and they have highlighted discrepancies. A recurrent inconsistency is the poor overlap observed between screens run by different groups on the same organism (for example, see refs. 7,14). The many efforts devoted to understanding this phenomenon have provided an excellent guide to establish a general framework for the quality standards that future experiments should follow.

The first issue to consider when comparing high-throughput interaction discovery experiments is the completeness of the screens. Terms such as 'comprehensive' or 'genome-wide' in the titles of large-scale studies can be somewhat misleading, and in fact the true coverage of both the genome and the interactome is often rather limited. Even those experiments that have targeted all open reading frames (ORFs) in an organism have shown ratios of homologous recombination lower than 85%, and of these only about 60% of the fusion proteins were successfully expressed[8], meaning that only about half of the total number of gene products are available for testing at the beginning of each screen. Private and public initiatives have already started to unify and provide collections of the ORFs from several model organisms in order to reduce the impact of this limitation (see refs. 15,16), and as a result the most recent studies in yeast cover about 80% of the genes[17].

Coverage of the interaction space is a different matter. Historically, the low overlap between individual experiments was attributed to a high number of false positives and false negatives in the screens[14]. For false positives, or interactions considered to be wrong, careful analyses identified the sources of most systematic errors, including self-activating baits in two-hybrid screens and "sticky" proteins in affinity purifications. The community is now aware of these effects, and there are now strategies to deal with them, such as regular auto-activation assays in yeast two-hybrid screens[18], probabilistic treatment of high-throughput affinity purification data[8] or new purification tags that permit more thorough washes to eliminate nonspecific interaction partners[19]. As a result the number of false interactions reported in the ongoing experiments is dropping[20]. An important point, however, is that most initiatives to chart the interactome are focused on the discovery of all interactions between macromolecules that are

[1]European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. [2]Institute for Research in Biomedicine and Barcelona Supercomputing Center, c/ Baldiri i Reixac 10-12, 08028 Barcelona, Spain. [3]Institució Catalana de Recerca i Estudis Avançats, Pg Lluís Companys 23, 08010 Barcelona, Spain. Correspondence should be addressed to R.B.R. (russell@embl.de) or P.A. (patrick.aloy@irbbarcelona.org).

Published online 20 October 2008; doi:10.1038/nchembio.119

biophysically possible, and not necessarily those that occur in biology. This is particularly relevant for the yeast two-hybrid system, where one can test interactions between proteins that might never see each other under physiological conditions because of different times of expression, different subcellular localizations or low copy numbers. The possibility of purely biophysical interactions raises a key issue about what constitutes a false positive: namely, whether biophysically real but nonphysiological interactions are false. The space of putative physiological conditions is clearly vast, and certain interactions only observed *in vitro* might yet occur under some not-yet-observed physiological state. Indeed, it has been proposed that some of such interactions might increase the robustness of cellular circuits acting as backups in the event of protein failure[21].

There have also been investigations to identify why large-scale studies miss many already known interactions (that is, false negatives). One possible explanation is that some proteins are inherently underrepresented in all large-scale screens[22], mainly owing to their biological properties. For instance, membrane and extracellular proteins often do not behave well in experiments relying on cytosolic or nuclear reporters, and many proteins aggregate when overexpressed. To address this, new methods are emerging to detect low-affinity interactions between receptors and ligands[23] or to test associations involving integral membrane proteins[24]. Current techniques are also not well suited to detect transient interactions or those that depend on post-translational modifications such as phosphorylation, but new strategies are also being developed to address this[25,26]. Recent research shows that many biological processes are mediated by ubiquitination, phosphorylation or other post-translational modification events, thus making it critical for new technical developments in interaction discovery to consider them. This unfortunately means a move from simple, easy-to-handle model organisms (for example, yeast) to more complex systems that permit proteins of interest to behave as they would under physiological conditions.

Assay sensitivity is also an issue. Not all methods are equally good at detecting all interactions. For instance, although the two main techniques above cover a wide range of affinities, the two-hybrid system is generally better at identifying weaker, more transient binary interactions, whereas the pull-down approach is generally better at recovering large-molecule machines with stable cores[8]. This observation can help explain many missing interactions from high-throughput screens, and more importantly it has highlighted the need for positive and negative reference ("gold standard") sets specific for each assay. Simply put, it is unfair to assess the accuracy of a method to discover large dedicated molecular machines on a reference of binary transient interactions, and vice versa. When specific benchmark sets are applied to evaluate the different interaction discovery methods (that is, binary interactions for two-hybrid screens and multiprotein complexes for affinity purification screens), results from large-scale screens have a similar quality to individual experiments, with the advantage of being free of any social bias[20]. The widely held notion that high-throughput interaction datasets are noisy and fragmentary might be a consequence of incorrect reference sets being used to assess their quality.

Finally, one has to consider the sensitivity and specificity of the sampling for any method applied on a large scale. Modern high-throughput interaction discovery experiments usually test millions of potential interactions, meaning there is a good chance that some will be missed, even when they fall into the category that the selected assay would normally detect. For instance, a single run of a two-hybrid screen detects roughly 40% of the interactions suitable for the technique, and one has to repeat the screen three times to raise this figure to 80%, and six times to get it to close to 100% (refs. 20,27).

How complete is the interactome today?

When considering all of the above, back-of-the-envelope calculations estimate some 20,000 binary protein-protein interactions in yeast, of which about 3,000 are currently known[20]. The picture is slightly more complete for macromolecular assemblies, where roughly 500 of an estimated 800 total core complexes are known[8], although owing to their dynamic nature, the characterization of all their components is still far from complete. The first datasets likely represent the 'low-hanging fruit', as many missing interactions and complexes are likely to be transient and/or dependent on physiological states not tested under the standard laboratory conditions normally used. This complete set will thus be much more difficult to uncover. With these numbers in mind, it is then easy to rationalize the low overlap observed—and often quoted as evidence of poor performance—between separate screens that have each identified fewer than 5% of the total number of interactions, and that have explored different regions of the total interaction space.

Finally, although proteins are important functional entities in a cell, one must not forget interactions involving other biomolecules. To this end, many new techniques are being developed to chart interactions involving proteins and nucleic acids[28–30], as well as small molecules[31] and other metabolites[32], and it is very likely that the derived networks will suffer from problems similar to those discussed for protein networks above. The quality standards developed for protein-protein interaction networks[33] should help to advance the understanding and interpretation of metabolic networks, where small molecules effectively link proteins together[34], or transcription regulatory networks, where transcription factors affect the expression of genes and their protein products[35].

Many past efforts were aimed at showing how well high-throughput interaction discovery experiments could reproduce small-scale biological results, in order to convince the community of their usefulness. This period is thankfully over, mainly owing to new benchmarks and standardization protocols that have established the framework for rapid developments. Interaction maps are now sufficiently reliable to uncover, for instance, unexpected cases of functional moonlighting, emerging properties of complex systems or the new rules that will drive biological research in the coming years. However, ten years of interactome research has also taught the community some tough lessons. Anybody working in the field has experienced the backlash from traditional molecular biologists, who are quick to condemn genome-scale analyses owing to a few missing interactions or false positives. As with much of science in general, people struggling to publish papers in the best journals are often guilty of overselling, and the resulting reactions can set back a field by years or even decades. But despite setbacks, the community needs to continue to think big in order to keep up with the grand promises of systems biology. Ambitious thinking, moreover, is in step with the many technological developments that make it possible to acquire large, diverse biological datasets with breakneck speed.

Exploiting networks

Early analyses of biological networks revealed a much higher connectivity than initially expected and established some basic principles that appear to hold over networks of many kinds. For instance, it is often noticed that they are 'scale free', meaning that most molecules in the network have only a few connections to others, with a small number of 'hubs' being highly connected, and that they are self-similar in that any part of the network is similar, in terms of connection statistics, to any other. Many are also 'small world', meaning that any two molecules can be connected by a limited number of intermediates.

There was an initial flurry of activity about these global, emerging network properties[36,37], almost to the point of an obsession, but so far these rather abstract concepts have had limited application in biology[38]. An exception to this was the correlation noted early on between the number of interactions that a protein makes and the tendency for the gene encoding this protein to be lethal when deleted[39], but even this observation has come under fire in recent years[20,40].

Nevertheless, the initial versions of interactome networks have provided very useful information. In the absence of comprehensive and accurate experimental networks, computational biology has been able to deliver rough models of cell processes, completing the maps with predictions and (perhaps more importantly) assessing the reliability of some others (see ref. 41 for a review). Probabilistic Bayesian network models (see ref. 42) deserve a special mention. Here the properties of a network are converted into a probability expressing the likelihood that one molecule affects another. These have many applications, and have been used, for instance, to greatly increase the accuracy in the prediction of phosphorylation events by treating the probability as an approximation for the proximity between a kinase and its potential substrate[26]. Bayesian networks have also been key to deciphering the first draft of the human B lymphocyte interactome from indirect expression data, which helped to identify deregulated interactions in specific pathologic or physiologic phenotypes, as well as causal lesions in several well-studied B cell malignancies[43]. Combined approaches have also been successfully used in cancer research. For instance, Pujana *et al.* used both predicted interactions and experimental yeast two-hybrid screens to model the network around BRCA1 and were able to identify novel genes associated with higher risk of breast cancer, in the process uncovering a genetic link with centrosome dysfunction[44]. The most exciting discovery, however, is that the current models are already accurate enough to allow global properties of the networks to emerge. A recent study illustrates how functional properties arising directly from the topology of networks can be used to identify new markers for breast cancer metastasis[45]. In this work, Chuang *et al.* applied a protein network–based approach to show that although genes known to be involved in breast cancer are not detected in analyses of differential expression, they play a central role in the network topology, interconnecting genes that are differentially expressed in people with metastasis. This showed that these subnetwork markers are more reproducible than individual marker genes selected without network information, and they are better predictors of metastatic tumors. It is clear that as the coverage, quality and variety of protein interaction data improve, the number of approaches exploiting emerging network properties will grow.

The success of network biology has rapidly diffused to other areas of biological research that have typically been more skeptical about computational models but that are now incorporating the analyses of networks in the planning of the experiments. The combination of interaction network models and co-complex experiments has recently permitted the rationalization of existing findings in proteomics as, for instance, substrate preferences in caspases responsible for triggering the apoptosis process: caspases do not indiscriminately cleave other proteins, but instead target a limited number of essential multiprotein complexes or functional pathways[19].

In evolutionary biology, network models are also illuminating certain old questions. It is reasonable to assume that interacting or functionally related protein families should share an evolutionary history, and thus their phylogenetic trees should be similar, although this has been hard to detect in the past[46]. Recently, it has been shown that these similarities are highly influenced by the number and type of partners connected to each protein, and thus to exploit these similarities in full, one must consider that the co-adaptation signal within a phylogenetic tree is influenced by its network of interactions[47]. There are also other interesting hypotheses with respect to speciation processes suggesting that what makes one organism different from another is not only their different genes (for example, chimps and humans share roughly 98.5% of their genes), but also the specific wiring of regulatory networks—mainly those involving transcription factors and their target genes[48]. However, the field where the topological analysis of networks has become a really hot topic is developmental biology. Here, the study of gene regulatory circuits in higher eukaryotes has revealed, for example, that the master regulators responsible for triggering cell differentiation are usually placed at the bottom of the networks, meaning that they are more often target hubs than primary regulators[35]. Conversely, for processes that require a faster response, such as cellular adaptation to an external change, the master regulators do not show any preferential positioning within the network. With great hopes put in stem cell research and its therapeutic applications, the correct identification of the small set of transcription factors that are able to regulate cell differentiation processes will certainly be extremely useful.

Targeting interactions

Protein networks, particularly when arranged into biological pathways, encapsulate sequential biological processes, and can then suggest rational points for intervention in order to develop therapies. There are now many dozens of examples of rational drug-design projects started with an eye on interactions downstream of more traditionally targeted receptors[49,50]. Comparatively new targets such as protein kinases or histone deacetylases are further into the cell than the earliest drug targets, which tended more toward surface receptors or channels.

Numerous studies of how proteins interact have also prompted the emergence of protein-protein interactions as targets for intervention. Although there has traditionally been skepticism about the possibility of targeting interactions with small molecules, much exciting recent research suggests that this is now possible[51,52]. Targeting interactions has certain advantages over more traditional targets such as enzyme active sites. Blocking an interaction often offers a more subtle, specific form of regulation that can avoid side effects due to off-target effects, or to total ablation of normal enzyme activity. A certain lack of specificity in kinase-targeting drugs, for instance, can help to explain differences in pharmacology, such as the fact that ostensibly similar compounds targeting BCR-ABL as treatments for leukemia differ markedly in the kinase and nonkinase targets they bind[53]. This has boosted research on allosteric inhibitors, or molecules that might target alternative sites, many of which are sites of protein-protein interactions[54].

Abbott Laboratories already has compounds in clinical development that inhibit interactions between several Bcl-2 family members, thus inducing the regression of solid tumors[55]. Other well-known examples of chemicals targeting an interaction include the cancer drug candidate nutlins[56] developed by Roche. These compounds act by blocking the interaction of the tumor suppressor protein p53 with its negative regulator MDM2 (which is overexpressed in many cancers), thus freeing p53 to go about its normal business of mediating cell death (**Fig. 1**). Chemicals have also been created to block the interaction between the translation initiation factors eIF4E and eIF4G (ref. 57) and to bind cytokines in order to disrupt receptor binding[58,59]. Discovery of new chemicals modulating interactions, or proteins inside of networks, has greatly progressed in recent years owing to new technologies. Many of these exploit chemical fragments libraries, which can be

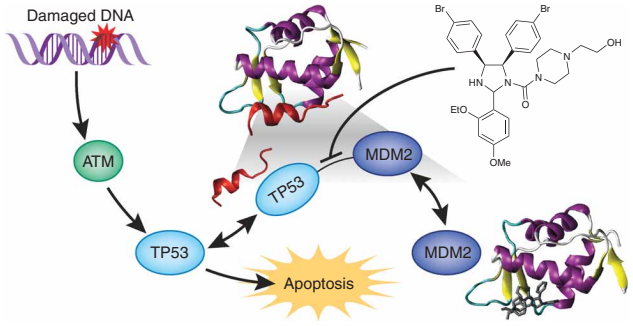

Figure 1 Targeting the interaction between MDM2 and p53. MDM2 normally binds to p53 (TP53) in the cell and prevents it from reacting to DNA damage. Several compounds have been developed that block this interaction, including the one shown. Blocking this interaction frees up more p53 and thus makes cells more responsive to DNA damage. Structure images rendered using Visual Molecular Dynamics (http://www.ks.uiuc.edu/Research/vmd/): peptide-bound structure, Protein Data Bank ID 1YCR; inhibitor-bound structure, Protein Data Bank ID 1RV1. The compound shown is from the second structure and is an imidazole derivative (PubChem CID 448419).

soaked into crystals (see ref. 60) or placed on a chemical microarray and screened for binding to a target molecule (see ref. 61).

Early skepticism about targeting interactions probably grew out of failed attempts to target large protein interfaces. However, the recent successes reveal a common trend—namely that most of the interactions involve a relatively small surface that accommodates the binding of a short peptide stretch in one protein by a globular domain in another. Many of the peptides, in turn, recur in multiple proteins according to a sequence pattern or linear motif that captures the general requirements for peptide binding. Various methods have been proposed to help to discover and validate new motifs (see refs. 62–64), and these are excellent starting points for additional attempts to target interactions rationally. Although it remains difficult to find drug-like molecules disrupting an interaction, it is likely that new technological developments will make this standard practice in the future.

Network-based therapeutics?

For almost a century, drug discovery was driven by the quest for magic bullets, which act by targeting one particular and critical point or step in a disease process and thus effect a cure with few other consequences. Though many drugs have been designed rationally in the past, the emerging picture is that drugs rarely bind specifically to a single target, thus challenging the magic bullet concept. Most rational drug discovery approaches focus on target-compound duets, and despite many success stories, this limited view has led to some expensive failures. Many promising drug candidates fail the last (and most expensive) clinical phases because of a poor understanding of the pathways involved in the mechanism of action, which can be due to an inappropriate choice of animal models that missed off-target effects in humans[65]. Resolving this is difficult, but to begin it is important to extend knowledge of the disease mechanism and consider the full biological context of a drug target and potential off-targets for a compound. In other words, one must integrate network biology and chemistry to identify putative secondary targets for a given compound or explore potential downstream effects of blocking the action of a key node in the biological network.

The first and most obvious practical application of interaction networks to drug discovery is to make more rational target selections. A detailed interaction map specific for a given pathology can indeed be very valuable—it can suggest potential points of intervention that might be selected based on (i) involvement in fewer pathways to avoid undesired effects in other routes[66] or on (ii) the topology of the interactome, opting for those strategic points that are vital to steer network traffic in one direction or another. Even some of the most specific marketed drugs target proteins that are not central in the physiological pathway relevant to the targeted disease; these drugs produce improvements only in a limited number of symptoms[65]. Interestingly, it has been noted that the wiring of interaction networks can change from a healthy to a diseased state, and charting such changes can also suggest excellent candidates for drug targets. For instance, several signaling pathways in the liver show a different functional wiring in the receptor-nucleus downstream routes when comparing normal hepatocytes with HepG2 (human hepatocellular liver carcinoma cell line) transformed cells, and such differences have already caught the attention of the pharmaceutical industry[67].

The accurate prediction of potential adverse reactions to compounds in early stages of drug development pipelines is one of the major challenges in the pharmaceutical industry, and network biology can also make important contributions here (**Fig. 2**). For instance, Pfizer has developed an *in vitro* testing strategy, based on Boolean models of hepatocyte death-survival pathways[68] and cell imaging, to predict drug-induced liver injury. This approach has been tested on 300 chemicals, including some known to cause rare idiosyncratic liver toxicity, with accuracies approaching 60% and with low rates of false positives[67]. Elsewhere, it is now possible to predict chemical toxicities, with increasing accuracy, by considering the proximity in a network of the target for a given compound and target proteins for other drugs known to cause some undesired side effects[66]. Most recently, it has been found that the set of common adverse reactions that two drugs share is an excellent descriptor for functional similarity and a means to classify the drug space[69].

Drug combinations have been used in the past to treat, for example, simultaneous symptoms of a disease. This is often effective, but it is also well established that many adverse reactions are known to occur when one drug has an effect on the metabolism of another (drug-drug interactions). An intriguing recent observation is that some combinations of drugs can exert therapeutically beneficial effects that have little to do with the effect of the compounds in isolation, which represents a powerful tool to explore new areas of action for already marketed compounds. CombinatoRx has tested several thousand combinations of off-patent drugs on a number of cell-based assays[70], and has uncovered surprises such as antitumor activity in a combination of the antipsychotic chlorpromazine with the antiprotozoal pentamidine. This highlights the fact that one rarely understands all effects of a compound on an organism, and much of this is down to interaction networks. This is perfectly illustrated in a study of the synergistic effects of these two compounds, which act at different and complementary parts of the cell cycle to halt tumor progression[71]. This work has been taken further in intriguing ways. For instance, it is possible to use the results of chemical combination screens to infer accurate biological relationships between the targets for separate chemicals[72]. There are also indications that suggest that by using a combination of drugs to target a disease, it is possible to reduce the dose of each individual compound, thereby avoiding many of the undesired off-target effects[73].

Another therapeutic field in which the study of networks will have a big impact is that of the selection of the best-suited model organism to study a given pathology. Many clinical studies fail because the underlying cell or animal models were poor predictors for the human

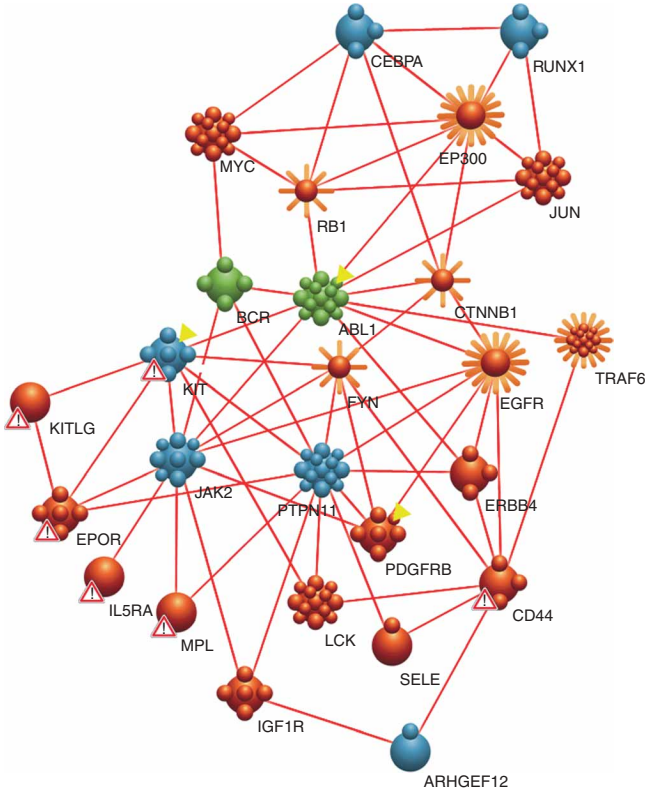

Figure 2 Leukemia disease network to predict undesired off-target effects. The disease network is built from known protein-protein interactions around genes associated with the onset of chronic (green) and acute (blue) forms of myeloid leukemia. Some of the nodes in the disease network are either primary or secondary targets (yellow triangles) of imatinib (Gleevec), a blockbuster drug used to treat people with Philadelphia chromosome–positive chronic myeloid leukemia (CML). The use of imatinib has several associated adverse effects, the most frequent of which is myelosupression. The primary target of imatinib, the ABL-BCR1 complex, is associated with aberrant clonal hematopoiesis in people with CML, and myelosupression is thus an expected therapeutic effect. Imatinib also inhibits KIT, the receptor for SCF (stem cell factor), which in turn affects other proteins in the network directly related to the formation of bone marrow tissue (alert signs). Proteins related to the molecular bases of other, more occasional, adverse effects detected in people taking imatinib (for example, nausea, diarrhea and dyspepsia) are also found very central in the disease network, although they are not depicted for clarity. The figure was drawn with AxPathBuilder (http://www.anaxomics.com/). Nodes are represented by different shapes depending on the number of interactions they make outside the depicted network, from the 6 interactions of IL5RA to the 347 of TRAF6.

disease[74]. The resolution of differences between established pathways and networks across species would thus lead to an increase in the predictive power. This is analogous to how gene prediction algorithms increased in accuracy upon the availability of additional genomes for organisms critically positioned in the evolutionary tree. Once networks of interest have been characterized in human and model organisms or cell types, it will be possible to compare them, in the same way one currently compares protein or DNA sequences[75], and select the systems that are closest to the disease state of interest. For instance, it is well documented that specific human brain cell cultures used to study Parkinson's disease are poor predictors of the final effects observed in clinical trials[76]. Although several methods to assess network similarity have been implemented[77–80], the field of comparative interactomics is still in its infancy[81,82], and will clearly develop in the future.

Tinkering with systems

One way to challenge the understanding of biological systems is to create an artificial one from scratch, applying the rules derived from those already known. Thirty years ago, researchers used the then limited understanding of the regulation of gene expression to start successfully designing and constructing pieces of DNA that would behave as predicted in living cells. This was an early demonstration that it is not always necessary to understand everything about a complex biological process in order to tinker with it.

A key concept in system design is modularity, or the notion that nature duplicates and reuses both parts and design principles again and again. It is apparent at nearly all levels, from the four bases making up the genetic code to the hierarchical organization of ecosystems. Proteins, the main nodes in most interaction networks, are also composed of modules or domains, which normally encode a specific function, such as catalytic activity or interactions with other proteins[83]. Nature duplicates and reuses these domains in a wide variety of contexts. It has also been proposed that protein complexes are modular in nature, with proteins or groups of proteins being reused for specific purposes[8].

Cell networks are also modular, meaning that the most complex structures can often be explained by combinations of a small number of network motifs. For instance, feedback or feedforward loops appear repeatedly throughout networks and are able to perform discrete functions almost independently[84]. These observations suggested the possibility to construct a toolkit of biological components for assembly into circuits to create new biological functions. Synthetic biology emerged as a combination of knowledge from diverse disciplines, such as molecular biology, engineering and mathematical modeling, to design and synthesize new metabolic circuits or signaling pathways that are not encoded in the original system. The approach is extremely powerful. Just as the development of synthetic chemistry in the nineteenth century allowed chemists to take control of chemical structures, the tools of synthetic biology should allow biologists to move from the discovery and analysis of existing networks to the *de novo* creation of new biological systems[85]. Indeed, several hundred modules that perform discrete functions, such as genetic inverters or protein reporters, are already available (http://parts.mit.edu/). Protein scaffolds can already be constructed for use as a platform onto which signaling pathways can be systematically assembled[86]—a technique similar to how chemists graft substituents onto a naturally occurring compound to alter its properties.

Although only in its infancy, synthetic biology already has a few success stories. One of the first big impacts is the incorporation of regulatory elements in classical metabolic engineering. Because engineered metabolic networks are often assembled from unrelated elements (even from different species) that have not been optimized through evolution, one of the challenges is to alter the kinetics of the individual components to make them function correctly in the new context[87]. Thus, incorporating the elements necessary for a desired metabolic route in a given organism, instead of introducing and expressing only one enzyme in isolation, has tremendously increased the number of successes, some of which have great socio-economic impact. For instance, synthetic pathways have been incorporated into yeast to significantly reduce the production costs of artemisinin, a compound naturally produced by *Artemisia annua* (sweet wormwood) that has proved to be very effective against multidrug-resistant strains of the malaria parasite *Plasmodium falciparum*[88]. Other biotechnological areas where current efforts at *de novo* design, or remodeling, of

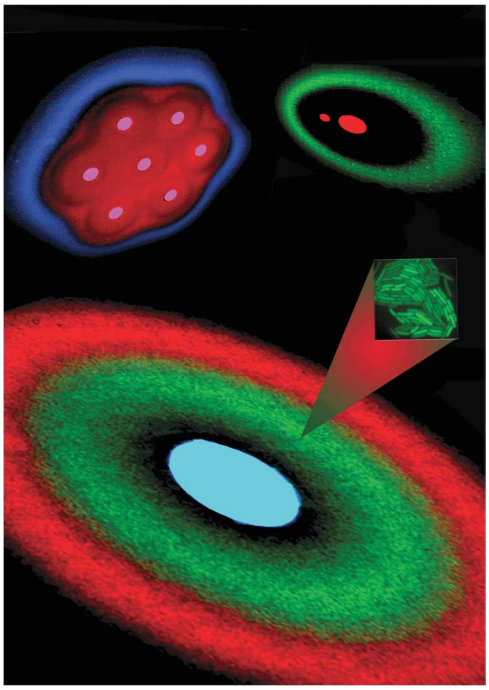

Figure 3 Ring-like patterns formed by programmed bacteria. Lawns of engineered *E. coli* "receiver" cells detect chemicals emitted by "senders" and produce different colored fluorescent proteins depending on their distance from the senders, as judged by chemical concentration[95]. Figure reprinted with permission.

synthetic pathways promise great steps forward are the production of biofuels[89,90] and the re-engineered oil-eating bacteria able to effectively deal with large oil spills[91] or other environmental contaminants[92]. These examples might sound like some of the first promises of biotechnological bacterial design, but the new systemic view, which considers the entire metabolic machinery and its control mechanisms rather than the introduction and overexpression of a foreign enzyme in an organism, is more likely to permit the true exploitation of bacteria as bioreactors to deliver the promised outcomes.

Synthetic biology is also a useful tool for studying the evolution of cell networks. So far, most experiments have been devoted to determining the importance of individual network nodes by knocking down or overexpressing a particular gene or a combination of them (see ref. 93). However, strategies exist to reveal systematically the importance of network edges by exploring the effect of adding new links to a biological circuit. In one study, gene regulatory networks in *Escherichia coli* were completely rewired to show surprisingly that 95% of new topologies are tolerated, with some even conferring a fitness advantage[94]. New links in such networks rarely appear to be a barrier for evolution.

The next level of complexity for synthetic biology is the power to tinker with multicellular systems, with an aim to build circuits able to achieve coordinated behaviors within a cellular population. However, these systems involve complex spatiotemporal dynamics regulated through cell-to-cell communication and intracellular signal processing systems, which makes them extremely challenging to model with available technology. Nevertheless, there have been attempts to engineer multicellular systems for programmed pattern formation—one of the hallmarks of coordinated cell behavior. For instance, Basu *et al.* were able to create ring-like patterns of differentiation based on chemical gradients of acylhomoserine lactone (a signaling molecule able to diffuse through the cell membrane) generated by "sender" cells and sensed through synthetic circuits by "receiver" cells[95] (**Fig. 3**). Other attempts to achieve coordinated actions in multicellular systems include the introduction into yeast of signaling mechanisms from *Arabidopsis thaliana*, and their coordination with endogenous responses[96]; and the coupling of gene and metabolic circuits in a tunable cell-cell communication system in *E. coli* to create an artificial sensor able to detect bacterial density[97]. Although the efficient design and implementation of synthetic multicellular communication systems are still a fantasy, early attempts have been very useful for improving the quantitative understanding of developmental processes (for example, pattern formation in *Drosophila melanogaster* embryos[98]). Moreover, they have inspired the creation of consortia to study the feasibility of much more ambitious projects, such as the redesign of pancreatic beta cells and programmed stem cell differentiation, which will certainly become a reality in the future.

Concluding remarks

Current interaction networks are still far from complete. However, numerous studies and new tools now mean that the community has a deeper understanding of the fraction of the interactome that is known and of how to complete it. It took 15 years from the development of DNA sequencing to complete the sequencing of the first yeast chromosome, and then nearly ten more until the human genome was fully sequenced. By analogy, the human interactome is probably some ten years from completion. This is a critical time for the community to rally support and make a convincing case for the utility of the final results.

The tools to study and perturb interaction networks are now sufficiently mature for them to be used systematically to enhance the understanding of biological processes, and more boldly to design new systems and extend existing systems. In addition, the understanding of networks themselves affords interesting opportunities to exploit them predictively. Overall, the hype about networks has changed into more down-to-earth (yet still exciting and innovative) applications that will clearly drive science forward. These developments, together with the emerging notions of network medicine, are probably what the community needs to push hard for interactome projects, ultimately to reap the many benefits that a more networked biology has to offer.

ACKNOWLEDGMENTS

We thank the Institute for Research in Biomedicine (IRB Barcelona) and the Banco Bilbao Vizcaya Argentaria Foundation (BBVA Foundation) for organizing the Barcelona Biomed Conference "Targeting and Tinkering with Interaction Networks." We also thank R. Weiss (Princeton University) for providing the images for **Figure 3** and A. Zanzoni (IRB Barcelona) for help with **Figure 1**. P.A. acknowledges the financial support received from the Spanish Ministerio de Educación y Ciencia through the grants PSE-010000-2007-1 and BIO2007-62426. We are both supported by the grant 3D-Repertoire from the European Commission under FP6 contract LSHG-CT-2005-512028.

Published online at http://www.nature.com/naturechemicalbiology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Oliver, S.G. *et al.* The complete DNA sequence of yeast chromosome III. *Nature* **357**, 38–46 (1992).
2. Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245–246 (1989).
3. Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**, 1030–1032 (1999).
4. Eyckerman, S. *et al.* Reverse MAPPIT: screening for protein-protein interaction modifiers in mammalian cells. *Nat. Methods* **2**, 427–433 (2005).
5. Paumi, C.M. *et al.* Mapping protein-protein interactions for the yeast ABC transporter Ycf1p by integrated split-ubiquitin membrane yeast two-hybrid analysis. *Mol. Cell* **26**, 15–25 (2007).

6. Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
7. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA* **98**, 4569–4574 (2001).
8. Gavin, A.C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
9. Krogan, N.J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
10. Giot, L. *et al.* A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736 (2003).
11. Li, S. *et al.* A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540–543 (2004).
12. Rual, J.F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).
13. Stelzl, U. *et al.* A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122**, 957–968 (2005).
14. Collins, S.R. *et al.* Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* **6**, 439–450 (2007).
15. Rual, J.F. *et al.* Human ORFeome version 1.1: a platform for reverse proteomics. *Genome Res.* **14**, 2128–2135 (2004).
16. Wei, C. *et al.* Closing in on the *C. elegans* ORFeome by cloning TWINSCAN predictions. *Genome Res.* **15**, 577–582 (2005).
17. Li, Q.R. *et al.* Revisiting the *Saccharomyces cerevisiae* predicted ORFeome. *Genome Res.* **18**, 1294–1303 (2008).
18. Vidalain, P.O., Boxem, M., Ge, H., Li, S. & Vidal, M. Increasing specificity in high-throughput yeast two-hybrid experiments. *Methods* **32**, 363–370 (2004).
19. Mahrus, S. *et al.* Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. *Cell* **134**, 866–876 (2008).
20. Yu, H. *et al.* High-quality binary protein interaction map of the yeast interactome network. *Science*, published online, doi:10.1126/science.1158684 (21 August 2008).
21. Stein, A. & Aloy, P. A molecular interpretation of genetic interactions in yeast. *FEBS Lett.* **582**, 1245–1250 (2008).
22. Pawson, T. & Linding, R. Synthetic modular systems–reverse engineering of signal transduction. *FEBS Lett.* **579**, 1808–1814 (2005).
23. Bushell, K.M., Sollner, C., Schuster-Boeckler, B., Bateman, A. & Wright, G.J. Large-scale screening for novel low-affinity extracellular protein interactions. *Genome Res.* **18**, 622–630 (2008).
24. Gisler, S.M. *et al.* Monitoring protein-protein interactions between the mammalian integral membrane transporters and PDZ-interacting partners using a modified split-ubiquitin membrane yeast two-hybrid system. *Mol. Cell Proteomics* **7**, 1362–1377 (2008).
25. Vazquez, M.E., Nitz, M., Stehn, J., Yaffe, M.B. & Imperiali, B. Fluorescent caged phosphoserine peptides as probes to investigate phosphorylation-dependent protein associations. *J. Am. Chem. Soc.* **125**, 10150–10151 (2003).
26. Linding, R. *et al.* Systematic discovery of *in vivo* phosphorylation networks. *Cell* **129**, 1415–1426 (2007).
27. Aloy, P. Shaping the future of interactome networks. *Genome Biol.* **8**, 316 (2007).
28. Berger, M.F. & Bulyk, M.L. Protein binding microarrays (PBMs) for rapid, high-throughput characterization of the sequence specificities of DNA binding proteins. *Methods Mol. Biol.* **338**, 245–260 (2006).
29. Warner, J.B. *et al.* Systematic identification of mammalian regulatory motifs' target genes and functions. *Nat. Methods* **5**, 347–353 (2008).
30. Mukhopadhyay, A., Deplancke, B., Walhout, A.J. & Tissenbaum, H.A. Chromatin immunoprecipitation (ChIP) coupled to detection by quantitative real-time PCR to study transcription factor binding to DNA in *Caenorhabditis elegans*. *Nat. Protoc.* **3**, 698–709 (2008).
31. Vetter, D. Chemical microarrays, fragment diversity, label-free imaging by plasmon resonance–a chemical genomics approach. *J. Cell. Biochem. Suppl.* **39**, 79–84 (2002).
32. Pearson, H. Meet the human metabolome. *Nature* **446**, 8 (2007).
33. Orchard, S. *et al.* The minimum information required for reporting a molecular interaction experiment (MIMIx). *Nat. Biotechnol.* **25**, 894–898 (2007).
34. Kanehisa, M. *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**, D480–D484 (2008).
35. Borneman, A.R. *et al.* Target hub proteins serve as master regulators of development in yeast. *Genes Dev.* **20**, 435–448 (2006).
36. Barabasi, A.L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
37. Albert, R., Jeong, H. & Barabasi, A.L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
38. Wolf, Y.I., Karev, G. & Koonin, E.V. Scale-free networks in biology: new insights into the fundamentals of evolution? *Bioessays* **24**, 105–109 (2002).
39. Jeong, H., Mason, S.P., Barabasi, A.L. & Oltvai, Z.N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
40. Pache, R., Babu, M. & Aloy, P. Exploiting gene deletion fitness effects in yeast to understand the modular architecture of protein complexes under different growth conditions. *PLoS Comput. Biol.* (in the press).
41. Aloy, P. & Russell, R.B. Structural systems biology: modelling protein interactions. *Nat. Rev. Mol. Cell Biol.* **7**, 188–197 (2006).
42. Jansen, R. *et al.* A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science* **302**, 449–453 (2003).
43. Mani, K.M. *et al.* A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. *Mol. Syst. Biol.* **4**, 169 (2008).
44. Pujana, M.A. *et al.* Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat. Genet.* **39**, 1338–1349 (2007).
45. Chuang, H.Y., Lee, E., Liu, Y.T., Lee, D. & Ideker, T. Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.* **3**, 140 (2007).
46. Pazos, F., Juan, D., Izarzugaza, J.M., Leon, E. & Valencia, A. Prediction of protein interaction based on similarity of phylogenetic trees. *Methods Mol. Biol.* **484**, 523–535 (2008).
47. Juan, D., Pazos, F. & Valencia, A. High-confidence prediction of global interactomes based on genome-wide coevolutionary networks. *Proc. Natl. Acad. Sci. USA* **105**, 934–939 (2008).
48. Borneman, A.R. *et al.* Divergence of transcription factor binding sites across related yeast species. *Science* **317**, 815–819 (2007).
49. Papageorgiou, A.C. & Wikman, L.E. Is JAK3 a new drug target for immunomodulation-based therapies? *Trends Pharmacol. Sci.* **25**, 558–562 (2004).
50. Johnson, Z. *et al.* Chemokine inhibition–why, when, where, which and how? *Biochem. Soc. Trans.* **32**, 366–377 (2004).
51. Wells, J.A. & McClendon, C.L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **450**, 1001–1009 (2007).
52. Neduva, V. & Russell, R.B. Peptides mediating interaction networks: new leads at last. *Curr. Opin. Biotechnol.* **17**, 465–471 (2006).
53. Rix, U. *et al.* Chemical proteomic profiles of the BCR-ABL inhibitors imatinib, nilotinib, and dasatinib reveal novel kinase and nonkinase targets. *Blood* **110**, 4055–4063 (2007).
54. Adrian, F.J. *et al.* Allosteric inhibitors of Bcr-abl-dependent cell proliferation. *Nat. Chem. Biol.* **2**, 95–102 (2006).
55. Oltersdorf, T. *et al.* An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* **435**, 677–681 (2005).
56. Vassilev, L.T. *et al.* *In vivo* activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**, 844–848 (2004).
57. Moerke, N.J. *et al.* Small-molecule inhibition of the interaction between the translation initiation factors eIF4E and eIF4G. *Cell* **128**, 257–267 (2007).
58. He, M.M. *et al.* Small-molecule inhibition of TNF-alpha. *Science* **310**, 1022–1025 (2005).
59. Braisted, A.C. *et al.* Discovery of a potent small molecule IL-2 inhibitor through fragment assembly. *J. Am. Chem. Soc.* **125**, 3714–3715 (2003).
60. Jhoti, H., Cleasby, A., Verdonk, M. & Williams, G. Fragment-based screening using X-ray crystallography and NMR spectroscopy. *Curr. Opin. Chem. Biol.* **11**, 485–493 (2007).
61. Neumann, T., Junker, H.D., Schmidt, K. & Sekul, R. SPR-based fragment screening: advantages and applications. *Curr. Top. Med. Chem.* **7**, 1630–1642 (2007).
62. Neduva, V. *et al.* Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol.* **3**, e405 (2005).
63. Edwards, R.J. *et al.* Bioinformatic discovery of novel bioactive peptides. *Nat. Chem. Biol.* **3**, 108–112 (2007).
64. Stein, A. & Aloy, P. Contextual specificity in peptide-mediated protein interactions. *PLoS ONE* **3**, e2524 (2008).
65. van der Greef, J. & McBurney, R.N. Innovation: rescuing drug discovery: *in vivo* systems pathology and systems pharmacology. *Nat. Rev. Drug Discov.* **4**, 961–967 (2005).
66. Apic, G., Ignjatovic, T., Boyer, S. & Russell, R.B. Illuminating drug discovery with biological pathways. *FEBS Lett.* **579**, 1872–1877 (2005).
67. Xu, J.J. *et al.* Cellular imaging predictions of clinical drug-induced liver injury. *Toxicol. Sci.* **105**, 97–105 (2008).
68. Saez-Rodriguez, J. *et al.* Flexible informatics for linking experimental data to mathematical models via DataRail. *Bioinformatics* **24**, 840–847 (2008).
69. Kuhn, M., von Mering, C., Campillos, M., Jensen, L.J. & Bork, P. STITCH: interaction networks of chemicals and proteins. *Nucleic Acids Res.* **36**, D684–D688 (2008).
70. Borisy, A.A. *et al.* Systematic discovery of multicomponent therapeutics. *Proc. Natl. Acad. Sci. USA* **100**, 7977–7982 (2003).
71. Lee, M.S. *et al.* The novel combination of chlorpromazine and pentamidine exerts synergistic antiproliferative effects through dual mitotic action. *Cancer Res.* **67**, 11359–11367 (2007).
72. Lehár, J. *et al.* Chemical combination effects predict connectivity in biological systems. *Mol. Syst. Biol.* **3**, 80 (2007).
73. Zimmermann, G.R., Lehar, J. & Keith, C.T. Multi-target therapeutics: when the whole is greater than the sum of the parts. *Drug Discov. Today* **12**, 34–42 (2007).
74. Lindsay, M.A. Finding new drug targets in the 21st century. *Drug Discov. Today* **10**, 1683–1687 (2005).
75. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
76. Matilla-Dueñas, A. The highly heterogeneous spinocerebellar ataxias: from genes to targets for therapeutic intervention. *Cerebellum* **7**, 97–100 (2008).
77. Suthram, S., Sittler, T. & Ideker, T. The *Plasmodium* protein network diverges from those of other eukaryotes. *Nature* **438**, 108–112 (2005).
78. Kelley, B.P. *et al.* Conserved pathways within bacteria and yeast as revealed by global protein network alignment. *Proc. Natl. Acad. Sci. USA* **100**, 11394–11399 (2003).
79. Koyutürk, M. *et al.* Pairwise alignment of protein interaction networks. *J. Comput. Biol.* **13**, 182–199 (2006).
80. Flannick, J., Novak, A., Srinivasan, B.S., McAdams, H.H. & Batzoglou, S. Graemlin: general and robust alignment of multiple large interaction networks. *Genome Res.* **16**, 1169–1181 (2006).

81. Sharan, R. & Ideker, T. Modeling cellular machinery through biological network comparison. *Nat. Biotechnol.* **24**, 427–433 (2006).
82. Kiemer, L. & Cesareni, G. Comparative interactomics: comparing apples and pears? *Trends Biotechnol.* **25**, 448–454 (2007).
83. Pawson, T. Organization of cell-regulatory systems through modular-protein-interaction domains. *Philos. Transact. A Math Phys. Eng. Sci.* **361**, 1251–1262 (2003).
84. Kashtan, N. & Alon, U. Spontaneous evolution of modularity and network motifs. *Proc. Natl. Acad. Sci. USA* **102**, 13773–13778 (2005).
85. Yeh, B.J. & Lim, W.A. Synthetic biology: lessons from the history of synthetic organic chemistry. *Nat. Chem. Biol.* **3**, 521–525 (2007).
86. Bashor, C.J., Helman, N.C., Yan, S. & Lim, W.A. Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. *Science* **319**, 1539–1543 (2008).
87. Yokobayashi, Y., Weiss, R. & Arnold, F.H. Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. USA* **99**, 16587–16591 (2002).
88. Ro, D.K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).
89. Keasling, J.D. & Chou, H. Metabolic engineering delivers next-generation biofuels. *Nat. Biotechnol.* **26**, 298–299 (2008).
90. Atsumi, S., Hanai, T. & Liao, J.C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86–89 (2008).
91. de Lorenzo, V. Blueprint of an oil-eating bacterium. *Nat. Biotechnol.* **24**, 952–953 (2006).
92. Gómez, M.J., Pazos, F., Guijarro, F.J., de Lorenzo, V. & Valencia, A. The environmental fate of organic pollutants through the global microbial metabolism. *Mol. Syst. Biol.* **3**, 114 (2007).
93. Steinmetz, L.M. *et al.* Systematic screen for human disease genes in yeast. *Nat. Genet.* **31**, 400–404 (2002).
94. Isalan, M. *et al.* Evolvability and hierarchy in rewired bacterial gene networks. *Nature* **452**, 840–845 (2008).
95. Basu, S., Gerchman, Y., Collins, C.H., Arnold, F.H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130–1134 (2005).
96. Chen, M.T. & Weiss, R. Artificial cell-cell communication in yeast *Saccharomyces cerevisiae* using signaling elements from *Arabidopsis thaliana*. *Nat. Biotechnol.* **23**, 1551–1555 (2005).
97. Bulter, T. *et al.* Design of artificial cell-cell communication using gene and metabolic networks. *Proc. Natl. Acad. Sci. USA* **101**, 2299–2304 (2004).
98. Isalan, M., Lemerle, C. & Serrano, L. Engineering gene networks to emulate *Drosophila* embryonic pattern formation. *PLoS Biol.* **3**, e64 (2005).