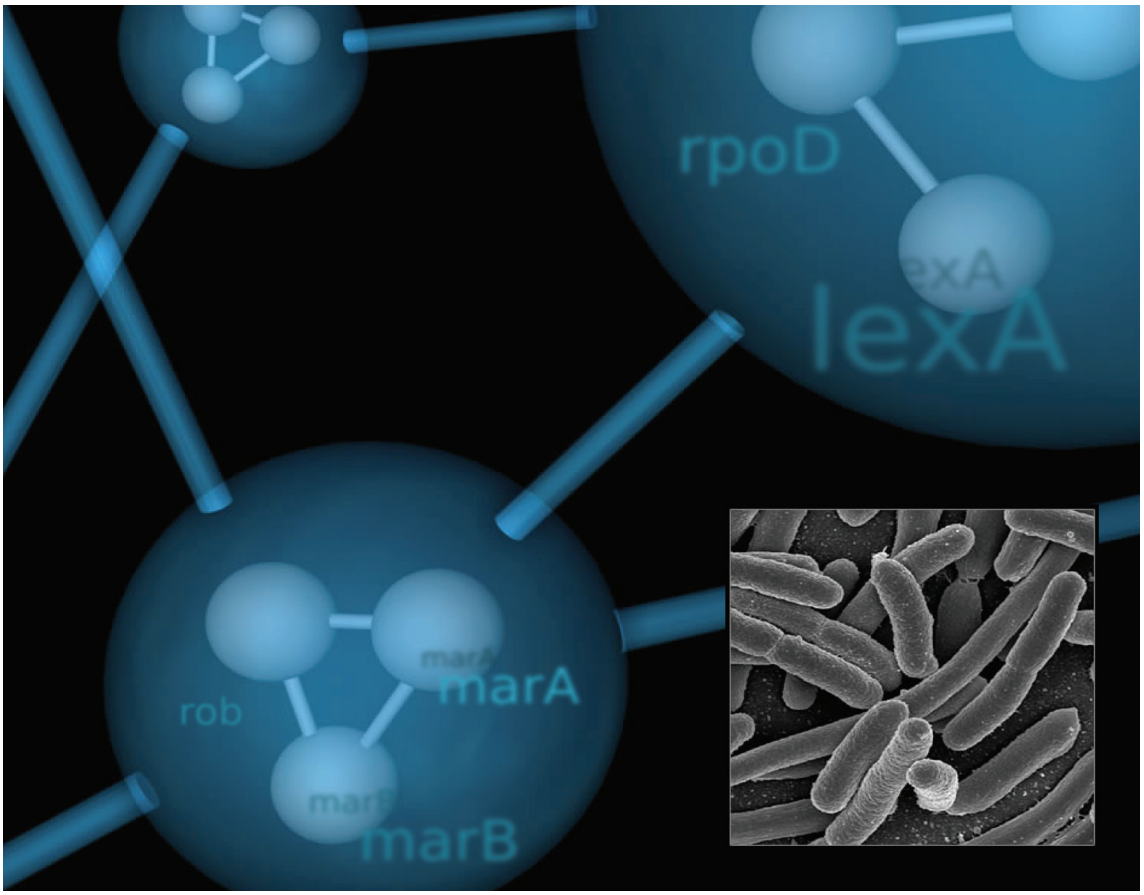
Molecular BioSystems

This article was published as part of the

Computational and Systems Biology themed issue

Please take a look at the full table of contents to access the other papers in this issue.



Explorations in topology–delving underneath the surface of genetic interaction maps†

Michal Breker[a] and Maya Schuldiner[b]

Received 7th April 2009, Accepted 10th August 2009
First published as an Advance Article on the web 8th September 2009
DOI: 10.1039/b907076c

High throughput assays, as well as advances in computational approaches, have recently allowed the acquisition of vast amounts of genetic interaction (GI) data in several organisms. Since GIs are a functional measure that reports on the effect of a mutation in one gene on the phenotype of a mutation in another, they can serve as a powerful tool to study both the function of individual genes and the wiring of biological networks. Therefore, these data hold much promise for advancing our understanding of cellular systems. In this review we focus on the methodologies currently available for using and interpreting large datasets of GIs for functional gene groups (GI maps), and elaborate on the challenges ahead. In addition, we discuss potential applications for the study of evolution and disease mechanisms, and highlight the need for comprehensive integrative analysis to extract the wealth of information found in these maps.

Introduction

The ability to translate the genetic code of an organism to a set of conceptual frameworks that underlie the function of a cellular system is the holy grail of systems biology. In order to attain such a holistic understanding, we must acquire information on many aspects of cellular organization. This should include comprehensive data on individual proteins (such as structure, function, localization, amount and modifications), their physical association, and the interconnectivity between cellular functions. This has become feasible during the last decade with the emergence of high throughput and computational tools. These allow for the acquisition of systematic data on various cellular features, which can then be analyzed to predict the function of single molecules. Moreover, they can serve to characterize patterns and designs that would not have been visible from the combination of small-scale experiments. Merging different types of data acquired systematically should offer us a way to view the multi dimensional architecture of the cell.

Methods for measuring genetic interactions in yeast

One type of data that is used to study the relationship between proteins is their genetic interactions (GIs). A GI is defined as the effect of a mutation in one gene on the phenotypic severity of a mutation in a second gene. This is measured by combining pairs of mutations and quantifying their phenotypes relative to those of the single mutants. GIs can be either aggravating, where the double-mutant displays a more severe phenotype than expected by the phenotype of each single mutant; or alleviating, where the double-mutant has a less severe phenotype than expected. The presence of a GI therefore implies that the processes in which these proteins partake are somehow connected to each other. The direction and extent of these deviations report on the form of functional dependence between the proteins in this specific cellular context. Since GIs are not dependent on the presence of a physical association they provide a powerful tool for ascertaining functional relations.

In recent years, the yeast *Saccharomyces cerevisiae* has been extensively used as a model to understand eukaryotic cells through systematic approaches. Along these lines, two major tools were developed in yeast to allow acquisition of GIs for a large number of gene pairs. The first is the Synthetic Genetic Array (SGA),[1,2] in which a yeast deletion strain of interest (query) can be crossed into a library of deletion strains. Following sporulation and selection, haploid double-mutant colonies arise and are scored for colony size. Double-mutants that display a colony size that is much smaller than expected are regarded as having an aggravating interaction (synthetic sick or in the extreme case-synthetic lethal). The second tool is Diploid based Synthetic Lethality Analysis with Microarrays (dSLAM),[3,4] where deletion strains are created in batch on the background of a query strain of choice. The resulting double-mutant strains then compete for growth and their relative fitness can be determined after several generations by their frequencies in the culture mixture. As each deletion strain carries a short unique genetic barcode,[5] identification of each strain can be performed by hybridization of genomic DNA to a microarray. Strains that have disappeared from the batch culture are again regarded as having extremely aggravating GIs. Analysis of the first sets of GI data demonstrated them as capturing a wealth of information on many aspects of protein roles within the cell. This led to a hope to obtain a comprehensive dataset comprised of the entire repertoire of GIs for all genes in the genome. Since such a

[a] *Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: michal.breker@weizmann.ac.il; Fax: +972 8 9346 373; Tel: +972 8 9346 517*
[b] *Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: maya.schuldiner@weizmann.ac.il; Fax: +972 8 9346 373; Tel: +972 8 9346 346*
† This article is part of a *Molecular BioSystems* themed issue on Computational and Systems Biology.

dataset would not be restricted to pre-chosen groups of genes, it would allow for unexpected biological observations to arise.[6] However, current methodology still cannot support this amount of double-mutant creation even in a simple genome like that of yeast.

Creation of genetic interaction maps in yeast

As would be expected, it has been shown that most random combinations of gene pairs do not display GIs.[2] Thus, in order to maximize the information content of GI experiments an SGA based methodology was created, called E-MAPs (Epistasis Mini-Array Profiles). This approach was developed on the notion that by selecting functionally related subsets of genes, it would be feasible to create smaller arrays of double-mutants while still retrieving the bulk of important GIs for that group of proteins.[7] The first E-MAP focused on ~400 proteins all localized to the early secretory pathway (and thus expected to be functionally related) and measured the GIs between them (~80 000 double-mutant combinations).[7] This method demonstrated that the lion's share of interactions could be obtained through the creation of smaller and more manageably sized GI maps therefore making it possible to obtain near complete coverage of whole genome GIs in the near future by carefully choosing gene subsets. More importantly, this approach was novel in using computational tools[8] to move from a binary measure of GIs (either "no interaction" or "synthetic sick/lethal") to a quantitative one which results in a score for the entire spectrum of GIs ranging from negative (aggravating or synthetic sick/lethal) all the way through to a positive score (alleviating or buffering/suppressing interaction) (Fig. 1). This paradigm shift allowed a unique tool for in-depth analysis of these connected subsets of proteins as it revealed an entire realm of interactions that were previously not measured. In addition, E-MAPs integrated essential genes into the analysis by utilizing a new method to construct hypomorphic alleles of essential proteins in high throughput, termed DAmP (decreased abundance by mRNA perturbation).[7,9] Additional methods for integrating essential proteins into GI maps have also been published and include the use of repressible promoters (Tet-off)[10] and temperature sensitive[11] alleles. The ability of E-MAPs to combine essential proteins alongside quantitative scores dramatically enriched the ability of these maps to shed light on the underlying cellular structure. In fact it was shown that analysis of the data found within them allowed the prediction of protein functions, the arrangement of proteins into complexes or pathways and the delineation of functional dependencies between different protein modules.[7]

Improving data acquisition

Despite the high data content of E-MAPs, there are still many ways by which they can be made even more informative:

(a) Improving the choice of proteins in each map: this could be done by relying even more on computational tools to integrate genomic data (such as sub-cellular localization, Protein–Protein Interactions (PPI), transcriptional profiles, homology to studied organisms[12] *etc*...) to efficiently choose the gene groups to be included in the E-MAP. The second relies on experimentally finding a functional subgroup of proteins.[6] One possible way of doing that is to systematically screen for all single mutant strains for those displaying a phenotype of choice (such as misexpression of a gene[13]). Creating a GI map for such a phenotypically related gene groups should enrich the set in interactions since they all affect a process of interest.[14] Moreover, the emerging map would be able to give clues as to the workings of that specific cellular function. This approach should also be useful in more complex organisms where genomic data is less available. Optimally, both methods should be integrated to enable us to capture both inter and intra process related information.

(b) Increasing the variability of GI measurements: most GI maps to date have focused on a single condition or a comparison of two.[15] However many interactions may be context-dependent. Therefore future efforts should aim to measure the same double-mutant strains under a variety of different growth conditions. This should allow us to understand how cells respond to cellular stimuli and how dynamic the protein networks are under those conditions.[16,17] Variability could also be uncovered by the use of multiple alleles (such as hypomorphs, specific splice forms, point mutations and gain of function alleles) instead of only complete deletions for each protein. Previous use of alleles in GI maps has demonstrated that various alleles can reveal different functional properties of a protein.[18] Moreover, use of downregulated alleles instead of complete knock-outs could help alleviate some of the background phenotypes that often occur when an entire protein is eliminated from the genome.

(c) Refining the scoring: methods for improving the quality and sensitivity of GI data by more accurate measurement systems are prevailing.[9,15] Using these it should be possible to quantify more interactions, as well as more accurately define them in terms of directionality and severity.[14,19] Quantitative scoring of interactions has many implications on the amount of information that can be gleaned from GI maps, therefore it is essential to consider the method of scoring used. The major theoretical question is which neutrality function to use. A neutrality function is the description of the double-mutant phenotype that represents no interaction and is necessary to define the extent of all other found GIs. Two options have been proposed; the first is a multiplicative function, which predicts that the double-mutant fitness (the exponential growth rate of the mutant strain relative to the WT) is the product of the corresponding single-mutant fitness values; while the second is a minimum neutrality function, which defines non-interacting mutations when the double-mutant yields the fitness of the less-fit single mutant.[20] Since current GI maps are utilizing additional phenotypes other than fitness, it is becoming more challenging to define an optimal neutrality function and this should be given much thought as it may affect the calculated assignment of GIs. This would be especially important in attempts to join together data from different experimental systems. Although most GI data to date is being calculated on the basis of a multiplicative neutrality function, it should be taken into consideration, that using similar raw data one could potentially obtain different views of interaction networks depending on the type of mathematical model used to compute the presence of GIs.[20]

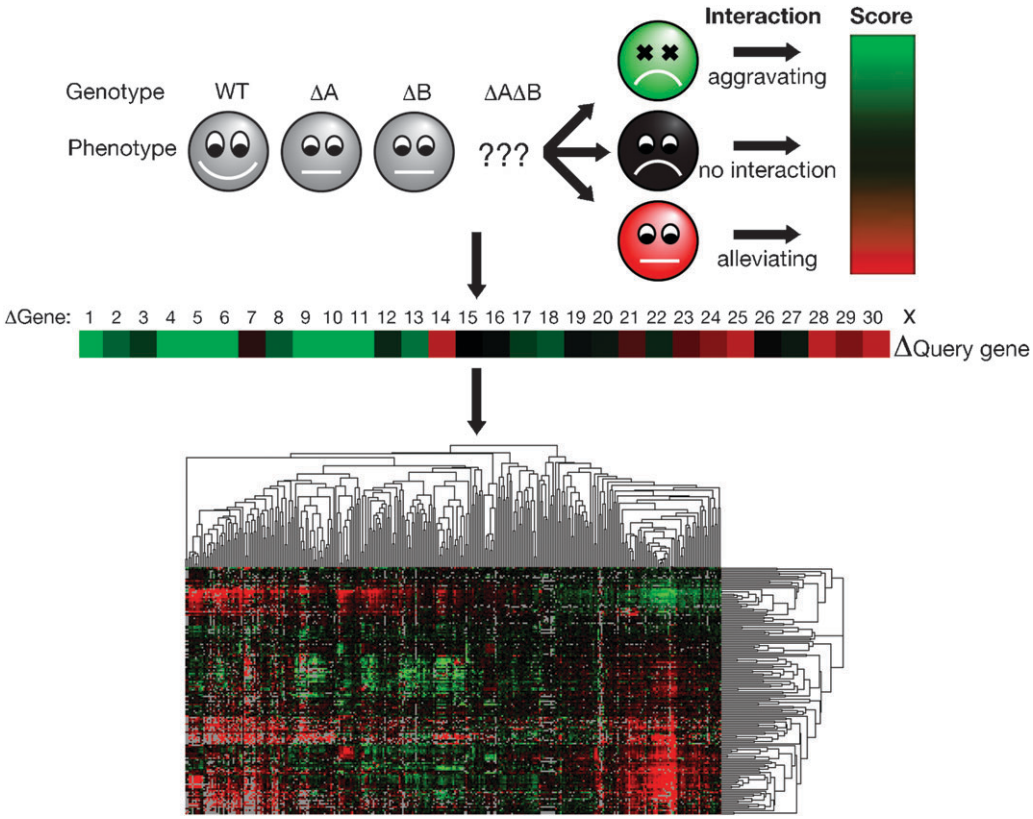


Fig. 1 Creating Genetic Interaction (GI) maps. In order to generate GI maps it is first necessary to create double mutant strains and measure their phenotype. Using mathematical models (such as the multiplicative model presented here—see text for others), it is possible to calculate the GI between each pair of genes by its deviation from the expected double mutant phenotype (top row). This assignment gives rise to a GI pattern for each gene deletion, which serves as its functional "finger print" (middle row). These functional patterns can then be hierarchically clustered according to statistically significant similarity (bottom row). The emerging GI map can be analyzed on many levels.

(d) Extending the analysis to higher combinatorial gene perturbations such as triple or quadruple knockouts.[2] This may be extremely useful in mammalian systems where large groups of homologous proteins exist and also where combinatorial addition of RNAi molecules should be easier to implement.

From method creation to extensive utilization

Although many improvements can still be made, it is now well established that quantitative measures of GIs between dense subsets of genes contain an enormous amount of information regarding many aspects of cellular organization. Therefore, there is an ongoing effort to create them for as many gene subsets as possible. To date, E-MAPs have been created for the secretory pathway,[7] chromosome biology,[3,18] RNA processing,[21] phosphorylation networks,[22] endosomal functions (T. Walther, personal communications) and mitochondria (J. Weissman, personal communications) with more constantly being created. Recently this methodology has started spreading to additional single cell organisms such as *Schisosaccharomyces pombe*[23–25] and *Escherichia coli*.[26,27]

Moving from single cells (either vertebrate cells in culture or microorganisms) to complex multi-cellular organisms is not a trivial task. This is because growth rate is not a good phenotype to assay and because any other phenotype measured is a result of a complex process of cellular differentiation and communication contributed to by an enormous number of variable cellular networks. Current GI interaction maps in whole organisms such as *C. elegans*[28] have used fitness as a readout, however this is not applicable in higher vertebrates and for gene sets whose function does not impact fitness. Moreover they may be hard to interpret functionally due to the myriad of individual affects that contribute to it. To this end we are far from being able to truly understand the meaning of any single GI in a whole animal, though often, specific GIs extracted from experimental model systems can be used to infer functional dependency.

It has recently been shown that GI maps can be made generally applicable by the use of fluorescent reporters instead of growth rate to measure interactions.[14] This was done by using a green fluorescent protein (GFP) under the regulation of a stress induced enhancer (specifically, the Endoplasmic Reticulum Unfolded Protein Response). The level of GFP was then used as a measure of protein misfolding in the secretory pathway in either single or double mutants and as a basis for calculating GIs that does not require growth rate measurements for fitness calculations.[14] Using such fluorescent reporters alongside silencing techniques (such as si/shRNA) to create double perturbations, it should now be possible to generate GI maps in higher organisms. Indeed, some ground breaking studies have all ready been published in

Drosophila melanogaster[12,29] and *Caenorhabditis elegans*[28,30,31] and there are efforts to set up such protocols in human cell lines (A. Kimchi, J. Weissman, personal communications). This rapid accumulation of GI data sets is now shifting the bottleneck from data generation to data analysis.

Data analysis

Creating functional predictions based on hierarchical clustering methods

One of the big challenges of high-throughput GI maps is how to translate the quantitative trait of a GI to a biological understanding. Analyses of GI maps have shown that each gene can have a large number of interactions.[2,7] While we do not yet know of a systematic way for explaining every GI observed, it is possible to utilize the spectrum of GIs for each query gene as a predictive tool for analyzing its molecular function. This is because the distinct pattern of interactions displayed by each gene represents the effect of its deletion on the cellular environment and thus can be used as its compound phenotype. Deletions of two proteins functioning in the same complex or pathway would be expected to have a similar effect on the cell and therefore would generate similar patterns of GIs. Such similarities are easily revealed by the use of two-dimensional hierarchical clustering of the GI patterns. From these similarities it is possible to predict functional roles of proteins.[8] One striking example is that of the clustering of the glycosylation genes in an E-MAP of the early secretory pathway.[7] The clustering of the glycosylation genes not only put all of the genes in the pathway together (Fig. 2, top), but also differentiated between two sub-clusters (Fig. 2, bottom right) corresponding to the distinct functions of these proteins (Fig. 2, bottom left).[8]

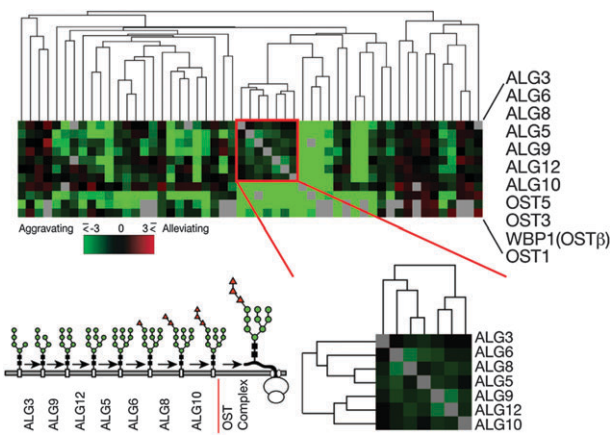


Fig. 2 Using hierarchical clustering methods to predict protein function. An example of the power of two-dimensional hierarchical clustering to uncover functional modules using GI data. Shown is a cluster of 7 genes from an E-MAP of the early secretory pathway (top). All 7 genes are involved in the endoplasmic reticulum glycosylation pathway. Based on the lack of GIs between several of the genes (bottom right), the algorithm could distinguish two subsets of the proteins—the ALG genes involved in the linear pathway (bottom left) for creation of the glycan chain and the OST genes involved in transfer of the chains to the nascent polypeptide.

Examples of the power of hierarchical clustering to reveal modules are ample and can be found in every GI map created to date. When tight clusters contain uncharacterized or poorly characterized proteins alongside well known ones, it is possible to predict their function as was demonstrated by the finding that Rtt109 is the founder of a novel class of histone acetyltransferases that acetylate lysine 56 on histone H3 in an Asf1-dependent manner;[18] and the discovery that Sem1 functions in three distinct complexes which link mRNA export, splicing and the proteasome.[21] When all proteins in the cluster are unknown, it is still possible to define new complexes, however uncovering their function can be more labor intensive. Examples include the characterization of the GET complex[7], and only later the discovery of its role in insertion of tail anchored proteins into membranes;[14,32] and the discovery of the novel EMC complex, a 6 member transmembrane complex in the ER[14] whose function is still unknown. Carefully studying the specific interactors of proteins can also give insight into an unknown proteins function even in the absence of a coherent cluster, as was shown by the identification of Phs1 as the very long-chain fatty acid dehydratase;[7,33] and the finding that SWR-C-mediated incorporation of Htz1 into chromatin is dependant on the function of Glc7/Bud14.[22] Although predictions are plentiful, they still require manual sifting through the large number of current GI maps and depend upon a basal knowledge of protein function for a substantial number of the proteins in the map. In addition, they work mostly for proteins that share some of their functions or that are in a coherent complex or pathway. To extract functional predictions for proteins in an automated manner, and mainly to reveal key properties in the molecular networking topology, higher-level analysis methods are necessary.

Defining protein modules

In an attempt to move from a two-dimensional view of GI maps to the real world of multi dimensional depictions it is necessary to go beyond hierarchical clustering. The first major use of GI data to recreate a model of the higher level organization of proteins into functional modules came from the study of metabolic networks.[34] By using theoretical data that contained both positive and negative measures of GIs, it was shown that proteins that function together in a coherent group share the same type of GIs (either positive or negative) towards other protein-groups. This property was called monochromaticity and could be uncovered using the PRISM algorithm, which defines protein modules by the tendency of proteins to be monochromatic towards other proteins.[34] The idea of monochromatic interactions later served as the basis for many (if not all) of the computational approaches to define networks (see below).

Since early GI maps did not contain quantitative measures, the first algorithms for analyzing non theoretical/measured data could only use the presence of aggravating interactions (synthetic sick/lethal) for analysis. However, based on the notion of monochromaticity, it was still possible to define functionally coherent groups of proteins based on a Congruence Score.[35] This score organizes genes as being in a single

complex or pathway, if they share synthetic lethal partners.[35] For example, identification of synthetic lethal interactions between mitotic exit network (MEN) complex and components of the Sin3/Rpd3 histone deacetylase complex allowed the definition of two protein modules. Moreover, the presence of an aggravating GI between them suggested a novel function for Sin3/Rpd3 in the promotion of mitotic exit redundantly to MEN.[35] Later, similar algorithms were created for quantitative measures (both positive and negative) of GIs. One such is the COP score which measures the extent of similarity (correlation) between the GI patterns of two deletion strains (as in a Congruence Score) but also uses the GI between them to predict their functional relatedness.[8] In a COP score, groups of proteins defined as functioning in a coherent manner have both a high correlation between their GI patterns and positive/buffering interactions between them.[8] Using the additional measure of the GIs between genes can help uncover complexes of proteins even in cases where the GI pattern is not identical (which may be due to both biological diversity and the inherent noise in measurement of GIs). Such methods for uncovering complexes and pathways give rise to many more functional predictions and override some of the problems associated with clustering alone.

Hierarchical clustering of entire profiles results in the assignment of each gene into only one cluster, thus losing the ability to uncover the pleiotropic functions that proteins often have. To overcome this over-simplification it is possible not to focus on the entire spectrum of GIs for a given deletion strain, but on recurring ''signatures'', meaning short repeated motifs in the GI profiles. Each signature may arise due to one aspect of the query genes function and would therefore be shared with other proteins that have a similar role. The concept of signature algorithms, initially demonstrated for analysis of transcriptional profiles,[36] was used for GI maps by construction of the Local Coherence Detection (LCD) algorithm.[37] This algorithm permits assignment of individual genes to multiple clusters, thus uncovering more than one function. An example of the power of this method comes from their analysis of Spf1. The hierarchical clustering of Spf1 with Sec66, suggested its involvement in the import of proteins into the ER.[38] However, the LCD algorithm also associated Spf1, as was experimentally proven, with calcium homeostasis, glycosylation, ER quality control, lipid biosynthesis and protein translocation.[37] An additional algorithm that builds on non conventional clustering (CHAMP), works by optimizing clusters of proteins which share GI patterns such that the predicted modules interact to form a network similar to those expected in a cellular context (N. Friedman, personal communications).

Novel approaches for organizing proteins in a given GI map into modules are to use data driven methods that are model independent. Recently a complexity metric has been shown to be useful in analyzing published datasets in the absence of any prior knowledge as to the function of genes within them.[39] This should prove extremely useful in analysis of GI maps in higher organisms or less characterized ones, where information on individual protein's functions may be more sparse than in the yeast.

Integration of proteomic information

In recent years there has been a flood of PPI data which comes from several sources, including affinity purification followed by mass spectrometry,[40,41] two hybrid methodologies[42,43] and bi-molecular complementation.[44] Since GIs report on a phenotypic consequence for the loss of two proteins which are often related by non-physical means (such as protein modification, transfer of substrate or parallel pathways), the overlap between GI and PPI data can be rare.[2,45] Thus, merging these two vast datasets is essential for reconstructing the most precise view of molecular networks (for a review see ref. 46).

Merging these two datasets has several roles. In cases where GIs exist for proteins all ready annotated as being within a complex, they can allude to the function of the complex. Additionally, GIs can be used to elucidate the structure of the complex.[47] This can be by: (a) differentiating between complexes with a single coherent function *versus* those in which different members may have additional functions. This can be done by use of algorithms such as the COP score (see above).[7,8] (b) Assigning complex modularity. For example study of the large mediator complex (made from more than twenty proteins and four functionally distinct subunits) showed that GIs for this complex could define the functional dependency between the head, tail, middle and CDK8 submodules as well as their interactions with other proteins outside of the complex.[48] In cases where alleviating GIs exist for proteins with a similar pattern of GIs that do not display PPIs this could be a method for predicting a complex that could not be seen by PPI detection methods due to technical issues (such as: transient interactions, posttranslational modifications[22] and transmembrane complexes that are more difficult to pull down for technical reasons).

Reciprocally, the finding that two proteins share a PPI can help explain the presence of a GI between two previously unconnected proteins. An example for how this can be done systematically comes from a novel software tool termed ANAT[49] (Advanced Network Analysis Tool). This analytical platform allows the user to plug in a group of proteins (for example those with unexplained GIs) and receive a model of the physical connectivity between them. The program provides a selection of algorithms[50] for predicting the most probable network of physical interactions that connects the proteins by using PPI data from previously published datasets. Since it allows the uncovering of paths that are indirect it allows the interpretation of GIs and creation of functional predictions based on the measured GIs.

To tap into the important information found by integrating these two datasets, several algorithms have been published. By integrating PPI data from affinity purification alongside GI data from a chromosome biology GI map,[51,52] these methods enabled not only the high-precision identification of new protein complexes, but also the addition of previously unknown proteins to known complexes and the characterization of their interconnectivity. These combined analyses also created new insights as to the nature of cellular complexes. For example, it was shown that aggravating interactions amongst complex members usually occur between non-essential subunits of essential complexes.[51]

The requirement for PPI data limits such approaches to protein sets and organisms where such data exists, and may miss many functional pathways that are not mediated by protein complexes (*e.g.*, metabolic pathways). However an enormous variety of high throughput/genomic/proteomic datasets exist in most model organisms such as transcriptional,[53,54] translational,[55] DNA-protein interaction data[56,57] and phenotypic data.[16,17] It would therefore be important to create algorithms that integrate a wide variety of data with GI maps.[58] Such integration could be performed in a stepwise fashion, by first defining protein complexes and then predicting their function based on the additional systematic data (Fig. 3) or by novel algorithms for merging all the different data to build a conceptual framework of the relevant proteins, not only in terms of molecular function and physical connection, but rather as a deep view into their cellular context. Indeed initial efforts to do that have combined GI data alongside PPI data, protein–DNA interactions, and either metabolic networks[59] or sequence homology and expression correlation.[28,47] Such integration methods in yeast, have only been performed to date on a relatively small set of binary synthetic sick/lethal interactions and should be much more powerful when utilized to study the current available datasets with quantitative measures of both negative and positive interactions as well as a much larger repertoire of deletion strains.

An additional type of data that has been used in conjuncture with GI data is the growth rate of single mutant strains under a variety of stress conditions. When this data was combined with GI data, it was shown that GI data, although recorded under only a single condition, could report on a broad spectrum of cellular conditions.[52] An interesting observation that arises using such stress data is that as many as a third of the aggravating interactions in current GI maps can be attributed to one gene buffering the effect of a cellular stress induced by the perturbation of another gene (N. Friedman, personal communications).

Initial analysis of GI data focused on making biological predictions on the function of proteins or their organization into modules. These findings, however, only represent the tip of the iceberg, revealing a glimpse of how much information can still be extracted. It is possible that with more sophisticated analysis techniques GI maps hold the information to create a model of the cellular environment. Retrieving this information requires going beyond simple analysis into complex computational methods and more profound analytical tools in the search of the biological meaning "encoded" in each pixel of these maps.

Genetic interactions and evolutionary constraints

One of the unique consequences of sexual reproduction is the creation of novel combinations of gene alleles, which may also be mutations, that did not exist in either parental strain. Understanding how these novel combinations act together to affect the phenotype (*i.e.*, what are the GIs between the different alleles/mutations), may allow better modeling of the evolutionary forces driving the emergence of this reproductive

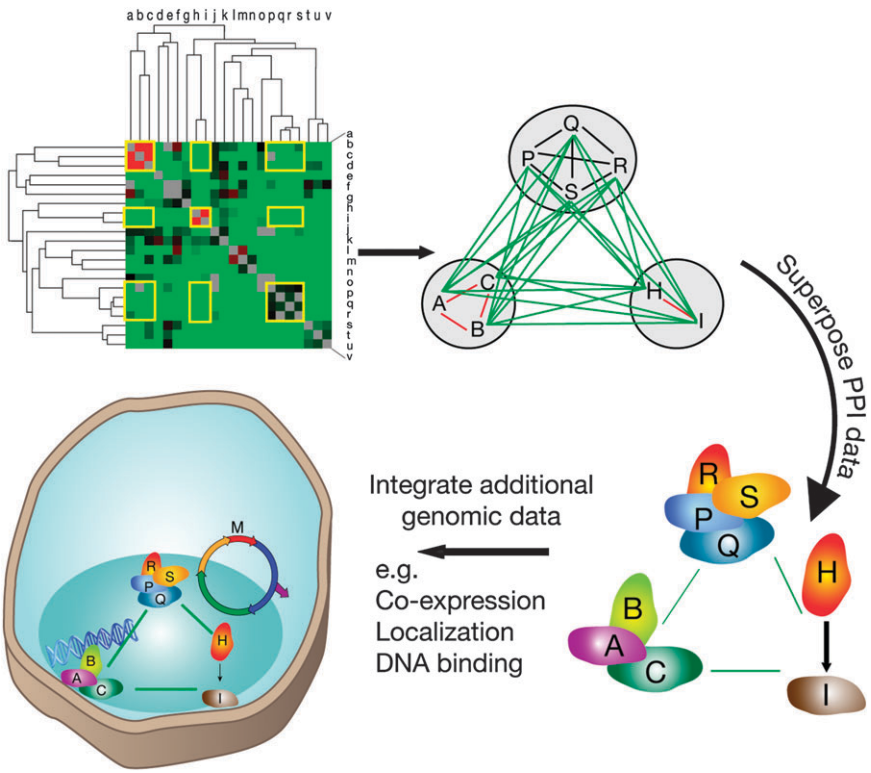


Fig. 3 Integrating various data sources to create a holistic view of cell architecture. Local areas of "monochromaticity" in GI patterns of clustered genes (top left) can be translated into protein networks (top middle). By integrating these networks with available PPI data it is possible to create a model of pathway and complex architecture and their interconnections (bottom right). Finally, by integrating all other available genomic/proteomic scale data it should be possible to obtain a model representing the true dynamics of the network in a given cell type (bottom left).

mode.[60] Simlarly, GI data in specific organisms could enlarge our ability to estimate their robustness (how many changes can occur without affecting phenotypic consequences) or their evolvability (their ability to support complex changes which occur by multiple genetic perturbations).[61–63] For example: networks of proteins that exhibit buffering interactions could allow for genetic variability while maintaining phenotypic robustness. This would support complex changes in genotype sometimes necessary for dramatic evolution of phenotypes.[62] In contrast, aggravating GIs could help "purge" the genome of deleterious mutations more rapidly than by random drift alone. Studies aimed at understanding the evolution of genomes often try to estimate the rate of mutation for an individual gene. However, the ideas above rely on the notion that certain gene pairs or alleles must evolve in parallel in order to sustain viability of an organism. Allowing for such cases we can find non-stereotypic evolution such as that of "hub" proteins that display GIs with many other proteins. These have a more cardinal effect on fitness than less connected genes[2,64] and have changed very little through evolution,[2,64,65] making them even more conserved than essential proteins.[66,67] Thus GIs can be used to start understanding the evolutionary constrains for evolving complex networks.

GIs can also be used to define the evolutionary relations between proteins. For example by comparing GI patterns it is possible to uncover the extent of functional redundancy between duplicated genes (paralogs). Such an analysis in yeast found that complete redundancy is extremely rare.[66] However, in the few cases where true or even partial redundancy is present it is easily detected by the overlapping pattern of GIs.[66] Another use of GIs is in the definition of orthologs (homologous genes in different organisms that share the same function). Often such proteins are hard to determine based on sequence or even structural homology alone due to extensive divergence between distantly related organisms or the presence of multiple homologs by sequence. However, true orthologs should have a similarity in their GI patterns or share their location within the hierarchy of GI networks.

More globally, influences of engineering on biology have pushed towards the characterization of general "design principles".[68] Using GI data, this could be achieved by comparing GI patterns from various organisms to reveal recurring network motifs.[69] Such motifs could create insights into the considerations for maximizing fitness by using specific solutions to perturbation. For example, comparing between a chromosome biology E-MAP in *S. pombe* and *S. cerevisiae* revealed that GIs within complexes as well as the genetic profiles of physically associated pairs of genes were more conserved than GIs between complexes. This suggests that the interdependancy between functional modules is more easily re-wired than the re-creation of a protein complex.[23]

Using genetic interactions to study disease mechanisms

GIs can be used for more than just studying the underlying principles of a normal cell. Using GIs in cell culture we can start to probe multi-genic phenotypes/diseases such as cancer,[64,70] viral resistance and even cystic fibrosis (CF)[71] (which is now acknowledged as being multi-genic). These cellular states all exhibit phenotypes that rely on multiple factors that confound simple genotype–phenotype relationships. In these cellular contexts, a single mutation can be tested for its GI pattern against the entire genome by using siRNA/shRNA libraries. The resulting "hits" would be good candidates to act as "modifiers" of the phenotype of the query mutation. Such screens are performed routinely to date and should soon become widespread enough to allow rapid accumulation of data for many genes and gene-subsets. It would be interesting to compare results from such a screen to more traditional methods that have, in the past, uncovered functional relationships between disease genes in human subjects (such as association studies or mapping of quantitative trait loci (QTLs)). After uncovering modifier genes by any given method, their function can be studied by using combinatorial silencing to create GI maps. The resulting GI maps would shed light on the way by which the network of proteins, found in the initial screen, function together or in parallel to promote or harness the phenotype of interest. Both of these avenues should contribute to our understanding of disease progression, which is an essential path to the development of targeted drugs. Moreover, knowledge regarding the genetic gene pairs that aggravate or alleviate the phenotype of the query cell can suggest appropriate drug combinations that may inhibit or rescue the cellular systems. For example, selecting drugs that will synergistically affect cell viability *via* their affect on cell targets that are themselves synthetically lethal can make the required dose of each drug lower, thus improving efficiency of treatment relative to negative side effects[64] (Fig. 4).

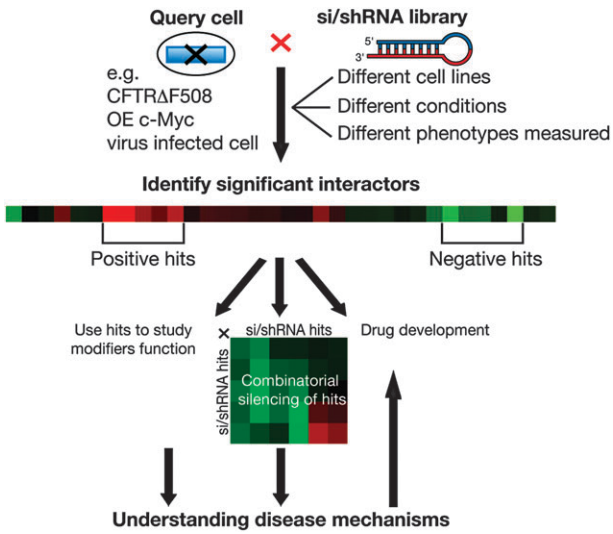


Fig. 4 Using systematic screens followed by GI maps to study disease mechanisms. A possible procedure to utilize high throughput silencing screens alongside GI maps to elucidate the biological mechanism underlying multi-genic disorders. Using current silencing libraries it is possible to start with a query cell and find all of the genes that act as "modifiers" of its phenotype (by either aggravating or alleviating a measured phenotype). The resulting hits could then be studied independently for their role in modifying the disease or as drug targets or silenced in combination to create a GI map. The construction of such a map should shed light on the role of these proteins in the cell and their role in modifying the phenotype, again contributing to drug development.

Perspective

Less than a decade after the invention of methodologies for creating large, systematic maps of GIs we are entering an era where such datasets are widespread. It is now time to create analytical tools to extract the immense amount of information they contain about the structure and function of cellular networks. In addition, we still lack robust algorithms to integrate GI data with other types of high throughput data. This is not a trivial problem as each type of biological data is assigned a different scoring method with matching units, and holds a different biological meaning. Moreover, even within the same type of data, such as GIs, datasets could differ based on the conditions in which they were measured or the phenotype followed and, in higher organisms, in which cell type and in which developmental stage they were assayed. For this reason, integration and representation into true understanding of the underlying system remains the next frontier.

Acknowledgements

We would like to thank Nir Friedman, Jonathan Weissman, Adi Kimchi and Tobias Walther for sharing unpublished data; Martin Jonikas and Ariel Jaimovich for input on the figures; Nir Friedman, Ruth Rinott, Adi Kimchi, Einat Zalckvar, Nir Yosef, Martin Jonikas, Oren Schuldiner and members of the Schuldiner lab for stimulating discussions and critical reading of the manuscript. MS is funded by the Human Frontiers Science Program (HFSP) Career Development Award, and the Israeli Science Foundation (ISF) Morasha grant.

Notes and references

1 A. H. Tong, M. Evangelista, A. B. Parsons, H. Xu, G. D. Bader, N. Page, M. Robinson, S. Raghibizadeh, C. W. Hogue, H. Bussey, B. Andrews, M. Tyers and C. Boone, *Science*, 2001, **294**, 2364–2368.
2 A. H. Tong, G. Lesage, G. D. Bader, H. Ding, H. Xu, X. Xin, J. Young, G. F. Berriz, R. L. Brost, M. Chang, Y. Chen, X. Cheng, G. Chua, H. Friesen and D. S. Goldberg, *et al.*, *Science*, 2004, **303**, 808–813.
3 X. Pan, P. Ye, D. S. Yuan, X. Wang, J. S. Bader and J. D. Boeke, *Cell*, 2006, **124**, 1069–1081.
4 X. Pan, D. S. Yuan, D. Xiang, X. Wang, S. Sookhai-Mahadeo, J. S. Bader, P. Hieter, F. Spencer and J. D. Boeke, *Mol. Cell*, 2004, **16**, 487–496.
5 G. Giaever, A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. Andre, A. P. Arkin, A. Astromoff, M. El-Bakkoury, R. Bangham and R. Benito, *et al.*, *Nature*, 2002, **418**, 387–391.
6 E. Conibear, *Cell*, 2005, **123**, 366–368.
7 M. Schuldiner, S. R. Collins, N. J. Thompson, V. Denic, A. Bhamidipati, T. Punna, J. Ihmels, B. Andrews, C. Boone, J. F. Greenblatt, J. S. Weissman and N. J. Krogan, *Cell*, 2005, **123**, 507–519.
8 S. R. Collins, M. Schuldiner, N. J. Krogan and J. S. Weissman, *GenomeBiology*, 2006, **7**, R63.
9 D. K. Breslow, D. M. Cameron, S. R. Collins, M. Schuldiner, J. Stewart-Ornstein, H. W. Newman, S. Braun, H. D. Madhani, N. J. Krogan and J. S. Weissman, *Nat. Methods*, 2008, **5**, 711–718.
10 S. Mnaimneh, A. P. Davierwala, J. Haynes, J. Moffat, W. T. Peng, W. Zhang, X. Yang, J. Pootoolal, G. Chua, A. Lopez, M. Trochesset, D. Morse, N. J. Krogan, S. L. Hiley and Z. Li, *et al.*, *Cell*, 2004, **118**, 31–44.
11 R. J. Dohmen and A. Varshavsky, *Methods Enzymol.*, 2005, **399**, 799–822.
12 S. Aerts, S. Vilain, S. Hu, L. C. Tranchevent, R. Barriot, J. Yan, Y. Moreau, B. A. Hassan and X. J. Quan, *PLoS Genet.*, 2009, **5**, e1000351.
13 A. P. Capaldi, T. Kaplan, Y. Liu, N. Habib, A. Regev, N. Friedman and E. K. O'Shea, *Nat. Genet.*, 2008, **40**, 1300–1306.
14 M. C. Jonikas, S. R. Collins, V. Denic, E. Oh, E. M. Quan, V. Schmid, J. Weibezahn, B. Schwappach, P. Walter, J. S. Weissman and M. Schuldiner, *Science*, 2009, **323**, 1693–1697.
15 R. P. St Onge, R. Mani, J. Oh, M. Proctor, E. Fung, R. W. Davis, C. Nislow, F. P. Roth and G. Giaever, *Nat. Genet.*, 2007, **39**, 199–206.
16 A. M. Deutschbauer, D. F. Jaramillo, M. Proctor, J. Kumm, M. E. Hillenmeyer, R. W. Davis, C. Nislow and G. Giaever, *Genetics*, 2005, **169**, 1915–1925.
17 M. E. Hillenmeyer, E. Fung, J. Wildenhain, S. E. Pierce, S. Hoon, W. Lee, M. Proctor, R. P. St Onge, M. Tyers, D. Koller, R. B. Altman, R. W. Davis, C. Nislow and G. Giaever, *Science*, 2008, **320**, 362–365.
18 S. R. Collins, K. M. Miller, N. L. Maas, A. Roguev, J. Fillingham, C. S. Chu, M. Schuldiner, M. Gebbia, J. Recht, M. Shales, H. Ding, H. Xu, J. Han, K. Ingvarsdottir and B. Cheng, *et al.*, *Nature*, 2007, **446**, 806–810.
19 B. L. Drees, V. Thorsson, G. W. Carter, A. W. Rives, M. Z. Raymond, I. Avila-Campillo, P. Shannon and T. Galitski, *GenomeBiology*, 2005, **6**, R38.
20 R. Mani, R. P. St Onge, J. L. t. Hartman, G. Giaever and F. P. Roth, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 3461–3466.
21 G. M. Wilmes, M. Bergkessel, S. Bandyopadhyay, M. Shales, H. Braberg, G. Cagney, S. R. Collins, G. B. Whitworth, T. L. Kress, J. S. Weissman, T. Ideker, C. Guthrie and N. J. Krogan, *Mol. Cell*, 2008, **32**, 735–746.
22 D. Fiedler, H. Braberg, M. Mehta, G. Chechik, G. Cagney, P. Mukherjee, A. C. Silva, M. Shales, S. R. Collins, S. van Wageningen, P. Kemmeren, F. C. Holstege, J. S. Weissman, M. C. Keogh and D. Koller, *et al.*, *Cell*, 2009, **136**, 952–963.
23 A. Roguev, S. Bandyopadhyay, M. Zofall, K. Zhang, T. Fischer, S. R. Collins, H. Qu, M. Shales, H. O. Park, J. Hayles, K. L. Hoe, D. U. Kim, T. Ideker, S. I. Grewal and J. S. Weissman, *et al.*, *Science*, 2008, **322**, 405–410.
24 S. J. Dixon, Y. Fedyshyn, J. L. Koh, T. S. Prasad, C. Chahwan, G. Chua, K. Toufighi, A. Baryshnikova, J. Hayles, K. L. Hoe, D. U. Kim, H. O. Park, C. L. Myers, A. Pandey and D. Durocher, *et al.*, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 16653–16658.
25 A. Roguev, M. Wiren, J. S. Weissman and N. J. Krogan, *Nat. Methods*, 2007, **4**, 861–866.
26 G. Butland, M. Babu, J. J. Diaz-Mejia, F. Bohdana, S. Phanse, B. Gold, W. Yang, J. Li, A. G. Gagarinova, O. Pogoutse, H. Mori, B. L. Wanner, H. Lo, J. Wasniewski and C. Christopolous, *et al.*, *Nat. Methods*, 2008, **5**, 789–795.
27 A. Typas, R. J. Nichols, D. A. Siegele, M. Shales, S. R. Collins, B. Lim, H. Braberg, N. Yamamoto, R. Takeuchi, B. L. Wanner, H. Mori, J. S. Weissman, N. J. Krogan and C. A. Gross, *Nat. Methods*, 2008, **5**, 781–787.
28 A. B. Byrne, M. T. Weirauch, V. Wong, M. Koeva, S. J. Dixon, J. M. Stuart and P. J. Roy, *J. Biol.*, 2007, **6**, 8.
29 C. Bakal, R. Linding, F. Llense, E. Heffern, E. Martin-Blanco, T. Pawson and N. Perrimon, *Science*, 2008, **322**, 453–456.
30 A. Fortunato, *Genomics*, 2009, **93**, 392–396.
31 B. Lehner, C. Crombie, J. Tischler, A. Fortunato and A. G. Fraser, *Nat. Genet.*, 2006, **38**, 896–903.
32 M. Schuldiner, J. Metz, V. Schmid, V. Denic, M. Rakwalska, H. D. Schmitt, B. Schwappach and J. S. Weissman, *Cell*, 2008, **134**, 634–645.
33 V. Denic and J. S. Weissman, *Cell*, 2007, **130**, 663–677.
34 D. Segre, A. Deluna, G. M. Church and R. Kishony, *Nat. Genet.*, 2005, **37**, 77–83.
35 P. Ye, B. D. Peyser, X. Pan, J. D. Boeke, F. A. Spencer and J. S. Bader, *Mol. Syst. Biol.*, 2005, **1**, E1.
36 J. Ihmels, G. Friedlander, S. Bergmann, O. Sarig, Y. Ziv and N. Barkai, *Nat. Genet.*, 2002, **31**, 370–377.
37 S. Pu, K. Ronen, J. Vlasblom, J. Greenblatt and S. J. Wodak, *Bioinformatics*, 2008, **24**, 2376–2383.

38 D. J. Tipper and C. A. Harley, *Mol. Biol. Cell*, 2002, **13**, 1158–1174.
39 G. W. Carter, D. J. Galas and T. Galitski, *PLoS Comput. Biol.*, 2009, **5**, e1000347.
40 A. C. Gavin, M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon, C. M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic and H. Ruffner, *et al.*, *Nature*, 2002, **415**, 141–147.
41 Y. Ho, A. Gruhler, A. Heilbut, G. D. Bader, L. Moore, S. L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff and J. Shewnarane, *et al.*, *Nature*, 2002, **415**, 180–183.
42 T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori and Y. Sakaki, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 4569–4574.
43 P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover and T. Kalbfleisch, *et al.*, *Nature*, 2000, **403**, 623–627.
44 K. Tarassov, V. Messier, C. R. Landry, S. Radinovic, M. M. Serna Molina, I. Shames, Y. Malitskaya, J. Vogel, H. Bussey and S. W. Michnick, *Science*, 2008, **320**, 1465–1470.
45 S. L. Wong, L. V. Zhang and F. P. Roth, *Trends Genet.*, 2005, **21**, 424–427.
46 A. Beyer, S. Bandyopadhyay and T. Ideker, *Nat. Rev. Genet.*, 2007, **8**, 699–710.
47 L. V. Zhang, O. D. King, S. L. Wong, D. S. Goldberg, A. H. Tong, G. Lesage, B. Andrews, H. Bussey, C. Boone and F. P. Roth, *J. Biol.*, 2005, **4**, 6.
48 S. R. Collins, P. Kemmeren, X. C. Zhao, J. F. Greenblatt, F. Spencer, F. C. Holstege, J. S. Weissman and N. J. Krogan, *Mol. Cell Proteomics*, 2007, **6**, 439–450.
49 http://www.cs.tau.ac.il/~bnet/.
50 N. Yosef, L. Ungar, E. Zalckvar, A. Kimchi, M. Kupiec, E. Ruppin and R. Sharan, *Molecular Systems Biology*, 2009, **5**, 248.
51 S. Bandyopadhyay, R. Kelley, N. J. Krogan and T. Ideker, *PLoS Comput. Biol.*, 2008, **4**, e1000065.
52 I. Ulitsky, T. Shlomi, M. Kupiec and R. Shamir, *Mol. Syst. Biol.*, 2008, **4**, 209.
53 M. B. Eisen, P. T. Spellman, P. O. Brown and D. Botstein, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 14863–14868.
54 T. R. Hughes, M. J. Marton, A. R. Jones, C. J. Roberts, R. Stoughton, C. D. Armour, H. A. Bennett, E. Coffey, H. Dai, Y. D. He, M. J. Kidd, A. M. King, M. R. Meyer, D. Slade and P. Y. Lum, *et al.*, *Cell*, 2000, **102**, 109–126.
55 N. T. Ingolia, S. Ghaemmaghami, J. R. Newman and J. S. Weissman, *Science*, 2009, **324**, 218–223.
56 T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon and B. Ren, *et al.*, *Science*, 2002, **298**, 799–804.
57 C. T. Harbison, D. B. Gordon, T. I. Lee, N. J. Rinaldi, K. D. Macisaac, T. W. Danford, N. M. Hannett, J. B. Tagne, D. B. Reynolds, J. Yoo, E. G. Jennings, J. Zeitlinger, D. K. Pokholok, M. Kellis and P. A. Rolfe, *et al.*, *Nature*, 2004, **431**, 99–104.
58 A. R. Joyce and B. O. Palsson, *Nat. Rev. Mol. Cell Biol.*, 2006, **7**, 198–210.
59 R. Kelley and T. Ideker, *Nat. Biotechnol.*, 2005, **23**, 561–566.
60 S. F. Elena and R. E. Lenski, *Nature*, 1997, **390**, 395–398.
61 A. Wagner, *Proc. R. Soc. London, Ser. B*, 2008, **275**, 91–100.
62 J. L. Hartman, B. Garvik and L. Hartwell, *Science*, 2001, **291**, 1001–1004.
63 S. L. Rutherford and S. Lindquist, *Nature*, 1998, **396**, 336–342.
64 C. Boone, H. Bussey and B. J. Andrews, *Nat. Rev. Genet.*, 2007, **8**, 437–449.
65 A. P. Davierwala, J. Haynes, Z. Li, R. L. Brost, M. D. Robinson, L. Yu, S. Mnaimneh, H. Ding, H. Zhu, Y. Chen, X. Cheng, G. W. Brown, C. Boone, B. J. Andrews and T. R. Hughes, *Nat. Genet.*, 2005, **37**, 1147–1152.
66 J. Ihmels, S. R. Collins, M. Schuldiner, N. J. Krogan and J. S. Weissman, *Molecular Systems Biology*, 2007, **3**, 86.
67 D. Waxman and J. R. Peck, *Science*, 1998, **279**, 1210–1213.
68 L. H. Hartwell, J. J. Hopfield, S. Leibler and A. W. Murray, *Nature*, 1999, **402**, C47–52.
69 R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon, *Science*, 2002, **298**, 824–827.
70 P. Demant, *Nat. Rev. Genet.*, 2003, **4**, 721–734.
71 J. R. Riordan, *Annu. Rev. Biochem.*, 2008, **77**, 701–726.