Systematic Functional Annotation and Visualization of Biological Networks

Anastasia Baryshnikova

Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, 08544, USA

Correspondence should be addressed to: abarysh@princeton.edu

ABSTRACT

Large-scale biological networks map functional relationships between most genes in the genome and can potentially uncover high level organizing principles governing cellular functions. Despite the availability of an incredible wealth of network data, our current understanding of their functional organization is very limited and essentially opaque to biologists. To facilitate the discovery and the interpretation of network functional structure, I developed a systematic quantitative approach to determine which functions are represented in a network, which parts of the network they are associated with and how they are related to one another. This method, named Spatial Analysis of Functional Enrichment (SAFE), detects network regions that are statistically overrepresented for a functional group or a quantitative phenotype of interest, and provides a visual representation of their relative positioning within the network. This visual representation is remarkably intuitive to biologists and maps the connection between a network and a set of functions with unprecedented clarity. By systematically annotating the genetic interaction similarity network from *Saccharomyces cerevisiae* with Gene Ontology (GO) biological process terms, SAFE proved to be accurate and robust to several sources of variation. In addition to GO, SAFE successfully annotated the network with other types of functional information, including quantitative chemical genomics and genetic interaction data. The union of these annotations produced a global view of the yeast cellular response to chemical treatment, which recapitulated known modes-of-action of chemical compounds and identified a potentially novel mechanism of resistance to the anti-cancer drug bortezomib. Finally, SAFE annotated the yeast protein-protein interaction network and revealed that, despite the lack of any visible topological structure on a global scale, the network presents a clear functional structure. These results demonstrate that SAFE is a powerful new tool for systematically annotating biological networks and investigating the global wiring diagram of the cell.

Keywords: yeast, networks, functional annotation, genetic interactions, protein-protein interactions, chemical genomics, bortezomib, cell map

INTRODUCTION

Understanding the functional organization of living cells is essential for predicting their behavior in normal and diseased conditions and designing effective therapeutic strategies to control them. Despite rapid advancements in technology development and data collection, our current knowledge about functional organization is limited to individual molecular pathways and a few well-characterized processes, but does not extend to the entire cell.

A unique opportunity to tackle this problem comes from budding yeast *Saccharomyces cerevisiae* which benefits from the availability of extensive molecular interaction networks that map physical, biochemical and phenotypic relationships between nearly all genes in the genome (Costanzo et al., 2010; Gavin et al., 2006; Krogan et al., 2006; Mo et al., 2009; Tarassov et al., 2008; Yu et al., 2008; Zhu et al., 2007). Since each network likely captures a different aspect of a gene-gene relationship, it is ultimately the integration of all networks that will provide the most comprehensive view of the functional organization of the yeast cell. However, to devise good strategies for integrating networks, we must first understand their individual functional structure, i.e. we must determine which biological functions are represented in each network, which parts of the network they are associated with and how they are related to one another. This process, analogous to understanding the functional structure of a newly sequenced genome, can be referred to as functional annotation of biological networks.

Large-scale biological networks are famously complex and often compared to disordered masses of tangled interactions ("hairballs") that are nearly impenetrable to biologists. However, early attempts to visualize and annotate networks, carried out through largely manual analysis, unveiled their potential for advancing our knowledge about the cell (Costanzo et al., 2010). For example, close examination of the yeast genetic interaction network revealed that genes naturally assembled into 13 major functional groups whose relative positioning within the network was consistent with their functional relatedness (Costanzo et al., 2010). To assess these trends quantitatively and gain insight into other molecular networks through similar annotations, we need to approach the question systematically using rigorous statistical methods and reproducible workflows.

Systematic annotation of a biological network requires three basic steps. First, we must obtain a comprehensive map of the network showing all of its nodes and their connections to one another. This map can be produced using a network layout algorithm which embeds the network in two- or three-dimensional space and positions all nodes based on their connectivity (Kobourov, 2012). Second, we need to secure multiple independent sources of functional information that would characterize each node relative to all other nodes in a variety of biological contexts. Such functional resources are readily available in yeast, thanks to the development of numerous high-throughput phenotypic assays and a long history of literature curation (Botstein and Fink, 2011). Finally, we must implement automated statistical procedures to overlay functional information onto the network map and locate functionally coherent regions. While functional subnetworks have been discovered on a case-by-case basis (Khatri et al., 2012; Mitra et al., 2013), no method currently exists to identify them exhaustively. Furthermore, no method is currently equipped to visualize functional regions within the network and generate a

comprehensive functional map that is both quantitative and intuitive to biologists, which is an essential requirement for its correct interpretation.

A few existing approaches could potentially be stretched to achieve a comprehensive annotation of a biological network. However, these methods would require extensive manipulations and/or laborious post-processing, and thus are unsuitable for automated analysis. For example, numerous network algorithms have been developed to search for sets of interconnected nodes that share a common phenotype or a consistent response across experimental conditions (Ideker et al., 2002; Mitra et al., 2013). Despite their potential for annotating networks, the main purpose of these algorithms is typically to evaluate experimental datasets and identify the most promising candidate genes, using networks only as independent supporting evidence. As a result, substantial modifications would be required to redirect these approaches towards comprehensive annotation tasks. Similarly, network clustering algorithms could potentially be used to identify sets of densely connected nodes that correspond to known, as well as novel, functional modules (Newman, 2006). However, clustering acts by partitioning the network into discrete and, in some cases, overlapping subnetworks, which must be annotated separately and assembled back together to provide a global functional view of the network. In addition, clustering disregards sets of nodes that are loosely connected, which would cause many sparse yet functionally coherent network regions to remain unannotated. More recently, a spatial statistics-based approach has been proposed to identify functional groups that co-cluster within the network more closely than expected by random chance (Cornish and Markowetz, 2014). While significant co-clustering does indicate a strong association between the functional group and the network, it cannot reveal in which part of the network this association occurs. As a result, the localization of functional groups cannot be mapped relative to one another and the global functional coverage of the network cannot be resolved.

To address these and other limitations of the existing methods, I developed an automated approach to annotate biological networks and generate quantitative and intuitive functional maps. This method, named Spatial Analysis of Functional Enrichment (SAFE), visualizes the network in two-dimensional space and detects network regions that are statistically overrepresented for a functional group or a quantitative phenotype of interest, such as a Gene Ontology term or a growth response to a chemical treatment. By mapping a function (or a phenotype) to a specific part of the network, SAFE provides statistical evidence, as well as an intuitive visual representation, for the relative positioning of different functions within the network. Using SAFE, I performed an automated functional analysis of the yeast genetic interaction network and showed that an iterative annotation of the network with multiple independent functional standards provides novel insight into the yeast response to chemical treatment. Furthermore, I showed that SAFE can detect functional structure within extremely dense and visually challenging biological networks, such as the extensive yeast protein-protein interaction network. These results indicate that SAFE is a powerful tool for annotating biological networks in a systematic and unbiased way and a unique framework for investigating the global functional organization of the cell.

RESULTS

Description of the SAFE method

SAFE annotates a biological network by calculating and visually representing local enrichment for a set of functional attributes (**Figure 1**; Experimental Procedures). To achieve this, SAFE first generates a two-dimensional map of the network by applying the Kamada-Kawai layout algorithm (Kamada and Kawai, 1989) or adopting a network embedding computed by a third-party software (e.g., Cytoscape (Shannon et al., 2003)) (**Figure 1A**). In the Kamada-Kawai layout, connected nodes attract each other, whereas disconnected nodes are pushed further apart. As a result, the position of every node is determined by an equilibrium of forces that reflect network topology.

For every node on the map, SAFE identifies a local neighborhood, defined as the set of nodes located within a chosen distance from it (**Figure 1B**). Several distance metrics can be used (Experimental Procedures), including, for example, a map-based weighted shortest path length. Using this metric, the distance between two nodes equals to the smallest number of steps required to reach one node from the other, multiplied by the physical length of each step on the map (**Figure 1B**).

For every neighborhood, SAFE calculates a set of quantitative scores, each corresponding to the sum of the neighbors' values for one of the functional attributes under examination (**Figure 1C**; Experimental Procedures). Attributes may include binary annotations to pre-defined functional groups, such as Gene Ontology (GO) terms, or quantitative phenotypes derived from systematic genomic experiments. To estimate the significance of a neighborhood's score for a given attribute, SAFE computes a hypergeometric p-value (for binary annotations) or an empirical p-value (for quantitative annotations) based on 1,000 network randomizations that preserve network topology but reshuffle attribute assignments (**Figure 1C**). A log enrichment score ($-\log_{10} p$) is then assigned to the original node that is the center of the neighborhood (**Figure 1D**; Experimental Procedures).

The enrichment scores for any given attribute across all nodes in the network define the attribute's enrichment landscape (**Figure 1E**). Size, shape and ruggedness (i.e., number of peaks and valleys) of an enrichment landscape characterize the distribution of the attribute throughout the network and measure the strength of their association (see Discussion).

Assigning different colors to different attributes enables SAFE to combine their enrichment landscapes into a composite functional map, where peaks and valleys of color intensity indicate regions of high and low enrichment for specific attributes (**Figure 1F**). Regions enriched for multiple attributes are represented with mixed colors, derived by combining the attribute-specific colors proportionally to their relative enrichment in the region (**Figure 1F**). This approach results in a global view of all functional attributes and an accurate quantitative representation of the network's functional enrichment landscape.

If several attributes are closely related and map to the same area of the network (as occurs, for example, for a GO term and its descendants), SAFE can combine them into functional regions based on the similarity of their enrichment landscapes. To facilitate the analysis of the resulting functional landscape, all attributes within a region are

represented with the same color and labeled by a concise tag list, composed of the five most frequent words among the attributes' denominations (Experimental Procedures).

SAFE annotates the yeast genetic interaction similarity network with Gene Ontology

To test the SAFE method, I annotated the yeast genetic interaction similarity network (Costanzo et al., 2010) using Gene Ontology biological process terms as attributes (**Figure 2**). A genetic interaction is a phenotypic relationship between two genes occurring when the fitness of a double mutant deviates from an expectation based on the fitness values of the two single mutants (Baryshnikova et al., 2013). Genetic interactions are strongly associated with genes sharing a common biological function or acting within the same biological process: these genes often genetically interact with each other and share similar genetic interaction partners throughout the genome (Baryshnikova et al., 2013). Similarity of genetic interaction profiles is a particularly robust measure of functional relationship and has enabled the construction of a functional network connecting ~75% of all genes in the yeast genome (Baryshnikova et al., 2010b; Costanzo et al., 2010). A high-confidence version of this genetic interaction similarity network, consisting of 2,838 genes and 10,016 edges (**Figure 2A**), was visualized and manually annotated in its original study (Costanzo et al., 2010) and therefore provides a good test case for SAFE.

By mapping local enrichment for 4,373 GO biological process terms, each associated with at least one yeast gene, SAFE revealed that different GO terms performed differently at annotating the genetic interaction similarity network. The vast majority of GO terms (84%) were small or sparsely distributed throughout the network and were only enriched within the neighborhoods of 10 genes or less (**Figure 2B**). A relatively small subset of GO terms (4%) showed complex enrichment landscapes with two or more peaks of enrichment in different regions of the network (**Figure 2C**). The presence of multiple peaks suggests that these GO terms comprise several functionally distinct subgroups of genes that localize separately within the network and may be better described by more specific annotations. Indeed, each of the remaining 12% of GO terms presented a single peak of enrichment within a specific region of the network (**Figure 2D**) and cumulatively covered all multi-regional enrichment landscapes. This set of 506 region-specific GO terms (12% of 4,373) was therefore adopted for annotating the genetic interaction similarity network. Interestingly, while region-specific GO terms were generally smaller than multi-regional terms (p-value = 10^{-49}, rank-sum test), their size distributions were largely overlapping (**Supplementary Figure 1**), suggesting that term size is not the sole responsible for landscape differences (see Discussion).

Since many of the 506 region-specific GO terms appeared to be functionally related and showed overlapping enrichment landscapes, their contributions towards annotating the genetic interaction similarity network were partially redundant. To minimize redundancy and simplify the annotation process, SAFE measured all pair-wise similarities among the 506 enrichment landscapes and identified 19 distinct enrichment patterns, each corresponding to a group of GO terms, that can be referred to as functional regions (Experimental Procedures). These regions, represented by different colors and

labeled with their individual tag lists, formed a comprehensive GO-based functional annotation map of the genetic interaction similarity network (**Figure 2E**).

By comparing the automated functional map, produced by SAFE, to the manual annotation, carried out in (Costanzo et al., 2010), I found that the outcomes of the two approaches were highly consistent (**Figure 2E-F**). Specifically, all manually annotated regions significantly overlapped at least one of the 19 SAFE regions (all p-values $< 2 \times 10^{-4}$, Fisher's exact test; **Figure 2F**). Furthermore, the GO terms associated with each region were in agreement with the manually assigned label (**Figure 2F**). For example, the region overlapping with SAFE region *r* was manually annotated to ribosome and translation-associated processes. Indeed, SAFE recognized that the region is enriched for 19 closely related GO terms, including ribosome biogenesis (GO:0042254) and cytoplasmic translation (GO:0002181). The complete list of GO terms associated with each functional region is reported in **Supplementary Data 1**.

Interestingly, several network regions that have been previously unified under a single functional label were recognized as separate regions by the automated SAFE analysis (**Figure 2E-F**). For example, SAFE partitioned mitosis and chromosome segregation into two separate regions, one enriched primarily for mitosis and cell cycle-related functions (region *k*) and the other enriched for microtubule-based processes and cytoskeleton organization (region *l*). In addition, SAFE refined the localization of metabolism-related genes (regions *b* and *c*) and recognized three previously unannotated network regions that were likely missed during manual analysis due to their reduced size and specific localization (regions *d*, *i* and *p*). These regions corresponded to the vacuolar H+-ATPase protein complex (region *d*), the tRNA wobble base modification pathway (region *i*) and the pheromone-induced cell cycle arrest factor (FAR complex, region *p*).

SAFE is robust against stochastic variations in network layout and differences in distance measures

To annotate the yeast genetic interaction similarity network, SAFE defined local neighborhoods using a map-based distance metric which may be subject to variability due to the non-deterministic nature of the layout algorithm. Specifically, two nodes were in the same neighborhood if the shortest path between them, defined as the minimum distance one had to travel to get from one node to the other, was shorter than a chosen threshold (**Figure 1B**; Experimental Procedures). The lengths of the edges composing the paths were determined by the Kamada-Kawai layout algorithm which positions each node based on the balance of attractive and repulsive forces exerted, respectively, by nodes that are connected and disconnected from it (Kamada and Kawai, 1989). Given that the initial node positions are seeded randomly, the layout is non-deterministic and produces a different network map at every instance (**Figure 3A-B, i**). Since node positioning may affect local neighborhoods and their enrichments, it is important to assess how much these stochastic variations in network layout affect the functional annotations produced by SAFE.

To address this question directly, I generated 10 different maps of the genetic interaction similarity network by repeatedly applying the Kamada-Kawai layout (**Figure 3A-B, i**). I used SAFE to annotate each network map with GO biological process terms

and compared the enrichment landscapes of each individual term across all 45 map pairs ($10 \times 9 / 2 = 45$) (**Figure 3A-B, ii**). I found that GO terms enriched within the neighborhoods of at least 10 genes showed highly similar enrichment landscapes in any two network maps (median Spearman's rank correlation $\rho = 0.82$; **Figure 3C**). This indicated that, despite differences in absolute node positioning between maps, the neighborhoods of the individual genes remained largely unchanged and tended to be enriched at the same level for the same GO terms. Importantly, the number of enriched neighborhoods per GO term was also consistent: of all GO terms that had at least 10 enriched neighborhoods in one map, 83% made the same threshold in at least 5 maps and 67% in all 10 (**Supplementary Figure 2**).

In addition to being robust to stochastic variations in network layout, SAFE can also, in theory, be completely independent from them by defining local neighborhoods via a map-independent distance metric. For example, in the genetic interaction similarity network, distance between two genes can be measured using the correlation-weighted shortest path length where each step equals to 1 minus the correlation of the genetic interaction profiles of the two connected genes (**Figure 3D, i**). This distance metric is map-independent because the quantitative correlation values were not employed by the layout algorithm to generate the network map (Experimental Procedures). Instead, the network was binarized and every edge, connecting two genes with correlated genetic interaction profiles (Pearson correlation coefficient $PCC > 0.2$), had the same weight.

To compare the outcomes of map-dependent and map-independent distance metrics in SAFE, I re-annotated the genetic interaction similarity network with GO biological process terms using the correlation-weighted shortest path length. For each individual GO term, I compared the enrichment landscapes produced by the two distance metrics (**Figure 3A, ii**; **Figure 3D, ii**) and found them to be very similar (median Spearman's rank correlation $\rho = 0.70$; **Figure 3E**). This indicates that the local neighborhoods defined with or without network visualization tend to be enriched for the same GO terms to a similar degree and produce equally accurate representations of the functional relationships between genes.

SAFE annotates networks with chemical genomic data and recapitulates known drug modes-of-action

I showed that an automated GO-based annotation of a biological network can reveal which biological processes are associated with the network and which parts of the network they are associated with (**Figure 2**). However, annotating a network with a single type of functional information, such as Gene Ontology, is unlikely to provide a full picture of the network's functional organization due to the inherent noise and limitations of each individual functional standard. A more robust strategy is to use multiple independent sources of functional data and apply them iteratively to annotate the same network. Such an approach could not only provide a more accurate description of different network regions, but has also the potential to uncover novel relationships between functional standards.

In yeast, a rich source of systematic functional information is provided by chemical genomics, a powerful technology for characterizing gene functions and identifying the

molecular mechanisms of chemical compounds (Ho et al., 2011; Roemer et al., 2012). In a chemical genomic screen, a genome-wide collection of yeast mutants is grown in the presence of a chemical compound and the relative fitness of each mutant is measured with respect to an untreated control (Giaever et al., 2002). Identifying mutants that are particularly resistant or sensitive to a given compound is instrumental for mapping out pathways that mediate the compound's toxicity or are required to protect the cell against its detrimental effects (Hillenmeyer et al., 2008; Hoepfner et al., 2014; Lee et al., 2014; Parsons et al., 2004; Parsons et al., 2006). I hypothesized that SAFE could facilitate the identification of these pathways by annotating the genetic interaction similarity network with chemical genomic data, in addition to GO biological process terms. Specifically, I predicted that SAFE could identify network regions that are enriched for genes that confer sensitivity or resistance to a given compound and determine whether those same regions are enriched for GO biological process terms that would be indicative of the compound's mode-of-action.

To test this prediction, I used one of the most recent chemical genomics datasets in yeast, which measured quantitative fitness scores for ~5,000 yeast homozygous deletion mutants upon exposure to 132 chemical compounds with known modes-of-action (Hoepfner et al., 2014). For any given compound, low or high fitness scores corresponded to mutants that grew slower or faster compared to their corresponding untreated controls and the median of the population (Hoepfner et al., 2014). As a result, fitness scores measured the relative sensitivity or resistance of mutants to the chemicals (Hoepfner et al., 2014). Using the genetic interaction similarity network and 132 sets of fitness scores, SAFE generated 132 compound-specific fitness enrichment landscapes that mapped local enrichment for genes associated with chemical sensitivity or resistance phenotypes (**Figure 4A-B**).

I examined the compound-specific fitness enrichment landscapes (**Figure 4A-B**) in the context of the GO biological process map (**Figure 2E**) and found them to be highly consistent with the compounds' known modes-of-action. For example, mutants sensitive to doxorubicin, a DNA intercalator that prevents DNA replication by blocking the progression of topoisomerase II (Fornari et al., 1994; Momparler et al., 1976; Tacar et al., 2013), were specifically enriched in the network region that was also enriched for GO terms describing DNA replication and repair processes (**Figure 4A**). Similarly, regions enriched for ribosome- and mitochondria-related GO terms were also overrepresented for mutants sensitive to verrucarin A, a protein synthesis inhibitor (Hernandez and Cannon, 1982) with a reported toxicity towards mitochondria (Schappert and Khachatourians, 1986) (**Figure 4B**).

By combining all sensitivity and all resistance enrichment landscapes together, SAFE generated two global maps of chemical genomic annotations for the genetic interaction similarity network (**Figure 4C-D**). These maps grouped chemical compounds based on the similarity of their cellular effects and captured general trends associated with the yeast cell response to chemical treatment (**Figure 4C-D**).

For example, several studies reported that mutations affecting vesicle-mediated transport sensitize the cell to a wide range of chemical compounds (Hillenmeyer et al., 2008; Parsons et al., 2004; Zakrzewska et al., 2010). Vesicle-mediated transport controls the localization of membrane proteins, including drug transporters and signaling

receptors which are expected to play an important role in regulating the efflux/influx of substances from/into the cell. Consistent with these observations, the sensitivity to 79 of 132 compounds (60%) was linked to mutations in endocytic and vesicle-mediated transport pathways (**Figure 4C**).

Similarly, it has been reported that chemical treatment often triggers the downregulation of ribosomal genes as part of the environmental stress response (ESR) program (Holcik and Sonenberg, 2005; Warner, 1999). This program enables the cell to cope with exogenous perturbations by activating several protective mechanisms, including the deceleration of protein synthesis and the delay of cell cycle progression (Causton et al., 2001; Gasch et al., 2000; Gasch and Werner-Washburne, 2002; Zakrzewska et al., 2010). Indeed, we observed that resistance to 44 of 132 compounds (33%) was increased upon mutation of genes involved in ribosome biogenesis and translation (**Figure 4D**).

SAFE provides novel insights into the molecular basis of resistance to bortezomib

By systematically examining a biological network and mapping local enrichment for sensitivity/resistance to chemical compounds, as well as for GO term annotations, SAFE provides a comprehensive view of the cellular response to chemical treatment. In doing so, SAFE not only displays known modes-of-action of chemical compounds in a simple and intuitive visual form, but can also uncover novel response patterns that may extend our understanding of the molecular mechanisms of chemical activity. To illustrate this potential application of SAFE, I examined the fitness enrichment landscape of bortezomib, a proteasome inhibitor approved for treating multiple myeloma and mantle cell lymphoma, and undergoing clinical trials for several other types of cancer (Chen et al., 2011) (**Figure 5**).

By blocking the proteasome's proteolytic activity, bortezomib is thought to prevent the degradation of pro-apoptotic factors and indirectly promote programmed cell death in tumor cells (Chen et al., 2011). In addition, bortezomib's anti-cancer effect is enhanced in combination with several histone deacetylase inhibitors (Canestraro et al., 2010; Deleu et al., 2009; Feng et al., 2007; Yu et al., 2003), suggesting that transcriptional regulation may also play a role in mediating bortezomib's activity. Consistent with these observations in human cells, SAFE showed that similar processes are likely involved in exerting bortezomib's effect in yeast: genes showing sensitivity to the drug were specifically overrepresented within regions of the genetic interaction similarity network that were enriched for proteasome-mediated protein degradation, cell cycle control and transcriptional regulation functions (**Figure 5A**).

Interestingly, SAFE also showed that resistance to bortezomib was strongly associated with the network region enriched for secretion and vesicle-mediated transport GO terms (**Figure 5A**). Although several reports have suggested that proteasome inactivation may cause endoplasmic reticulum (ER) stress, due to a higher abundance of misfolded proteins within the ER (Lee et al., 2003; Obeng et al., 2006), a protective effect of mutations in ER or other secretory genes against bortezomib toxicity has not been anticipated.

In order to confirm this potentially novel finding, I asked whether, similarly to other drugs (Costanzo et al., 2010; Parsons et al., 2004), the effects of bortezomib treatment could be recapitulated via genetic inactivation of its molecular target, i.e. the proteasome. Specifically, I hypothesized that mutations providing a relative growth advantage in the presence of bortezomib should also display a relatively high fitness when combined with proteasome mutants and would therefore result in positive genetic interactions.

To test this hypothesis, I obtained quantitative negative and positive genetic interactions involving members of the core and the regulatory particles of the proteasome (Experimental Procedures, **Supplementary Data 2**) and used them to annotate the genetic interaction similarity network with SAFE. For every gene encoding a subunit of the proteasomal machinery, SAFE calculated a genetic interaction enrichment landscape which mapped local enrichment for the gene's negative and positive genetic interaction partners throughout the network (**Figure 5B**). I found that at least 7 of the 13 subunits of the proteasomal regulatory particle that have been tested showed significant enrichment for positive genetic interactions within the region enriched for secretion-related GO terms (**Supplementary Figure 3**). In particular, *rpt4-145*, a temperature-sensitive mutant in an essential ATPase of the regulatory subunit that preferentially contributes to ER-associated protein degradation (ERAD) (Lipson et al., 2008), presented a genetic interaction enrichment landscape remarkably similar to bortezomib's fitness enrichment landscape (**Figure 5B**).

These findings support the hypothesis that mutations in secretory functions may partially compensate for proteasome inactivity and alleviate the detrimental effects of bortezomib treatment, although the precise mechanisms of this alleviation remain to be uncovered. Considered that resistance to bortezomib is a common complication in treating multiple myeloma and other cancers (Murray et al., 2014), understanding its molecular mechanisms is critical for designing targeted therapeutic approaches and effective drug combination treatments. The identification of secretory pathways as a potential focus of these new therapies illustrates the power of SAFE to uncover novel biological responses by comprehensively annotating biological networks with multiple independent functional standards.

SAFE reveals functional structure in super-dense biological networks

By iteratively incorporating GO biological process, chemical genomic and genetic interaction annotations, SAFE produced accurate and informative functional maps of the genetic interaction similarity network. However, annotating the genetic network may seem like a relatively easy challenge because of the network's highly organized topological structure where large modules of densely connected—and thus likely functionally related—nodes are readily distinguishable from one another, even by the naked eye (**Figure 2A**). In contrast, most other biological networks in yeast and other organisms present more complex global topologies, characterized by higher interaction density and lower node modularity (Arabidopsis Interactome Mapping, 2011; Giot et al., 2003; Jeong et al., 2001; Nayak et al., 2009; Rual et al., 2005; Yu et al., 2008). As a result, resolving the functional organization of these networks may be much more difficult than suggested by the analysis of genetic interactions.

To assess whether SAFE can potentially annotate more complex biological networks, I first tested whether the functional annotation of the genetic interaction similarity network was dependent upon network density. Given that genes in the network were originally connected if the similarity of their genetic interaction profiles exceeded an arbitrary correlation threshold ($PCC > 0.2$), I progressively reduced the stringency of the threshold ($PCC > 0.18$, $PCC > 0.16$ and $PCC > 0.14$) and generated three additional versions of genetic interaction similarity network, containing 1.4, 2.1 and 3.4 times more edges than the original (**Figure 6A, i**). As expected, the maps of these networks, produced by the Kamada-Kawai layout algorithm, showed rapidly decreasing levels of visible structure and were nearly impossible to explore by eye (**Figure 6A, i**).

I used SAFE to annotate the three network maps with GO biological process terms and found that local functional enrichment could be easily detected in spite of the increasing number of edges (**Figure 6A, ii**). Furthermore, the enrichment landscapes for each individual GO term were similar, albeit to a gradually decreasing extent, to the original landscape derived from the lowest density map (median Spearman's rank correlations $\rho = 0.75$, $\rho = 0.68$ and $\rho = 0.61$; **Figure 6B**). Thus, even a ~3-fold difference in network density appeared to have a limited impact on the GO term enrichments in most gene neighborhoods.

The ability to map functional enrichment in a dense genetic interaction similarity network reflects SAFE's sensitivity to coherent functional structure. However, the existence of a functional structure itself may be a distinctive property of the genetic interaction similarity network. To test whether that is the case, I examined the extensive yeast protein-protein interaction network which maps discrete physical binding events between all ~6,000 proteins in the cell (Chatr-Aryamontri et al., 2015) (Experimental Procedures). While protein-protein interactions are known to occur more frequently between members of the same protein complex and other functionally related proteins (Cusick et al., 2005), a genome-wide map of this network, produced by the Kamada-Kawai layout algorithm and covering 5,699 nodes connected by 78,406 edges, showed no visible topological structure on a global scale (**Figure 6C, i**). To search for possible underlying functional structure, I used SAFE to annotate the network with GO biological process terms. Strikingly, SAFE revealed 21 separate functional network regions, each associated with a distinctive GO term enrichment landscape and a unique list of enriched GO terms (**Figure 6C, ii; Supplementary Data 3**). The resulting functional map of the protein-protein interaction network strongly suggests that physical binding, just like genetic interaction profile similarity, can define high level protein communities that correspond to broad-scale biological processes and may reveal the global functional organization of a cell.

DISCUSSION

Here I described the development and validation of SAFE, the first automated method for annotating biological networks and exploring their functional organization. Given a network and a visual map of its connectivity, SAFE locates all network regions enriched for one or more functional attributes, such as Gene Ontology terms or

quantitative gene phenotypes. In doing so, SAFE addresses three fundamental questions that help to elucidate the network's functional organization. First, are any regions of this network specifically associated with a given function or phenotype? Second, where in the network are these regions localized? And third, how does their localization compare to that of other functions or phenotypes? By answering these questions, SAFE builds a functional map of the network and enables the investigation of inter-process relationships within the cell.

The functional mapping of network domains in a rapid and systematic manner is a distinctive feature of SAFE that encourages deep exploration of biological networks through iterative annotation with multiple complementary functional standards. In this study, I showed that the iterative annotation of the yeast genetic interaction similarity network with GO biological process, chemical genomics and genetic interaction data provides a comprehensive view of the cellular response to chemical perturbation, linking genes and their functions to bioactive compounds and enabling the prediction of novel mechanisms of drug resistance (**Figure 4-5**).

Another distinctive feature of SAFE is that, through the power of visualization, it can improve our understanding of functional standards. For example, by annotating the genetic interaction similarity network with GO biological process terms, SAFE showed that some GO terms are enriched within a single network region, whereas others are multi-regional (**Figure 2A-C**). Although multi-regional GO terms tended to be generally larger than region-specific ones, term size alone could not fully explain the difference between the two groups (**Supplementary Figure 1**). An interesting possibility is that region-specific GO terms share a similar level of functional specificity, which is defined by the topology of the genetic interaction similarity network. If that is the case, these terms could be used to delineate a cross section of the GO hierarchy and form a flat subset of GO annotations, analogous to GO slim (Ashburner et al., 2000; Consortium). Flat annotation resources have been instrumental for many genomic analyses thanks to their smaller size and lower redundancy (Costanzo et al., 2010; Myers et al., 2006). SAFE may offer the possibility to generate data-driven network-specific flat annotation standards which would enable more targeted functional analyses and further promote the exploration of high throughput data.

A third feature that enables SAFE to annotate biological networks in an accurate and comprehensive manner is its robustness against incomplete or noisy data. Validation results showed that SAFE is largely insensitive to several types of input variation, including stochastic effects in network layout, different distance measures and overall density of the network (**Figures 3, 6**). In addition, we also expect false positive and false negative errors in functional annotation standards to have little impact on the outcome of SAFE. Wrong or missing annotations will likely scatter randomly throughout the network and fail to significantly alter the enrichments of local neighborhoods. For example, it has been reported that chemical genomics data may be affected by an experimental bias whereby deletion mutants constructed in the same laboratory share common secondary mutations and show similar growth phenotypes in response to a wide range of chemical compounds (Hoepfner et al., 2014). SAFE would not recognize these phenotypes as a reliable biological signal because genes mutated in the same laboratory do not

preferentially interact with one another or localize to the same region of the genetic interaction similarity network (**Supplementary Figure 4**).

In addition to being robust to noise, SAFE is also sensitive to true biological signal, as shown by its ability to detect local functional enrichment in super-dense genetic interaction similarity and protein-protein interaction networks (**Figure 6**). This feature enables SAFE to be applied to a wide range of biological networks, especially those that have been considered inaccessible to biologists because of their complexity and lack of a clear topological structure. With the availability of SAFE, global visualization and examination of these networks no longer requires radical edge trimming via arbitrary cutoffs, thus potentially preserving valuable biological information.

Finally, SAFE provides a unique opportunity to exhaustively examine the functional landscape of a network because it explores all local neighborhoods and does not focus exclusively on densely connected clusters. A neighborhood is a set of nodes located within a short distance from a source node, irrespective of the total number of connections between them. As a result, SAFE does not miss functional coherence within sparse network regions (**Figure 2D**) and is capable of annotating networks with virtually no variation in interaction density, such as a lattice. In this respect, SAFE is conceptually different from other approaches for network analysis, such as clustering, and provides a more comprehensive overview of a network's functional content.

Mapping the functional organization of biological networks is a complex task that requires accurate statistical analysis and informative data representations. While SAFE addresses some of the main challenges associated with this task, the method is still far from complete and presents vast opportunities for future development.

For example, the interpretation of the functional maps produced by SAFE would greatly benefit from a better understanding of network layout algorithms. Data-driven network layouts, such as the Kamada-Kawai algorithm (Kamada and Kawai, 1989), are unsupervised methods for organizing nodes based on their connectivity. SAFE relies on layouts to identify local neighborhoods and visually map their functional enrichment. However, very little is known about how a particular layout should be chosen. Despite their great potential for uncovering hidden patterns within the data, layouts are typically used to generate esthetically pleasant network visualizations (Agapito et al., 2013) and are rarely the basis of any systematic network analysis. As a result, we have limited experience in evaluating network layouts and a poor understanding of their relative performance in the context of different networks.

So far, the decision to use the Kamada-Kawai layout to represent and annotate the genetic interaction similarity network was founded on theoretical considerations: the basic working principles of the algorithm are in agreement with our understanding of genetic interactions and produce an easily interpretable outcome (Costanzo et al., 2010). While that decision has proved to be very successful, SAFE offers a unique opportunity to quantitatively support it or to identify a potentially better approach for annotating the network. To achieve this, SAFE must be systematically applied to annotate and evaluate different layouts for the same network using a common set of functional attributes. The ultimate outcome of this analysis would be the identification of an optimal layout for

each individual network type and the preparation of a common ground for comparing networks to each another.

Quantitative comparison of biological networks is indeed a major goal of systems biology (Sharan and Ideker, 2006). Without a deep understanding of how genes, pathways and processes are connected in different kinds of molecular networks, we have little hope of developing successful strategies for integrating networks into a single comprehensive model of a living cell. By mapping enrichment for the same functional attributes in different networks (**Figure 2**, **Figure 6**), SAFE can make an important contribution to this goal. However, rigorous statistical approaches must be implemented to compare SAFE functional annotation maps across networks and draw meaningful conclusions about differences and similarities between interaction types.

In summary, SAFE represents an important step towards advanced understanding of biological networks. Unlike most other methods for network analysis, SAFE provides a global perspective onto the functional organization of a network by mapping statistical associations between functional groups and network regions. SAFE shows that, despite previous lack of attention, network visualization coupled with robust enrichment analysis is a valid strategy for studying molecular networks and gaining insight into the biological systems they represent. As both the number and the extent of molecular networks grows exponentially over time, understanding network functional organization with methods like SAFE is becoming of primary importance for building a unified model of a living cell.

ACKNOWLEDGEMENTS

I am deeply grateful to Dmitriy Gorenshteyn for his invaluable input on writing this manuscript. Also, I would like to thank David Botstein, Michael Costanzo, Charles Boone and Brenda Andrews for critically reading the manuscript and providing useful suggestions. The Boone and Andrews labs also generously provided the proteasome genetic interaction data. This work was supported in its entirety by the Lewis-Sigler fellowship at Princeton University.

FIGURE LEGENDS

Figure 1. Spatial Analysis of Functional Enrichment (SAFE). (**A**) Given a biological network, as an edge list or an adjacency matrix, SAFE generates a two-dimensional map of the network by applying the Kamada-Kawai layout algorithm. Alternatively, SAFE can import a network and its visual map directly from a Cytoscape session. (**B**) For each node *X* on the network map, SAFE defines a local neighborhood by identifying all other nodes (*A*-*H*) located within a chosen distance from it. By default, distance between two nodes is measured by the shortest physical path between them, where the length of a physical path equals to the total length of the edges that must be traversed to reach one node from the other (Experimental Procedures). Other distance measures are also available. (**C**) For each neighborhood, SAFE calculates a quantitative

score that equals to the sum of the neighbors' values for any given functional attribute, such as, for example, a yes/no annotation to a Gene Ontology term. This score is then compared to the scores for the same neighborhood across 1,000 network randomizations which reshuffle node labels (Experimental Procedures). (**D**) To measure the deviation of the neighborhood score from random expectation, SAFE calculates an empirical p-value and assigns it as a –log10 enrichment score to the original node *X*. (**E**) The enrichment scores for any given functional attribute across all nodes in the network define the attribute's enrichment landscape, which illustrates the distribution of the attribute throughout the network and measures the strength of their association. Node opacity is used to represent variations in enrichment scores. (**F**) Finally, SAFE constructs a composite functional enrichment map of the network by assigning different colors to different attributes and summing their enrichment landscapes. For every node, the final color corresponds to the weighted sum of all of its enrichments, where the weights are the enrichment scores for every attribute. As a result, multifunctional nodes are automatically represented by mixed colors.

Figure 2. Annotating the yeast genetic interaction similarity network with Gene Ontology (GO) biological process terms. (**A**) A map of the yeast genetic interaction similarity network, containing 2,838 nodes and 10,016 edges, was originally constructed in Costanzo et al. (Costanzo et al., 2010) by applying the Kamada-Kawai layout algorithm in Cytoscape (Shannon et al., 2003). (**B-D**) The analysis of 4,373 GO terms with SAFE showed that different GO terms present different enrichment landscapes. (**B**) Most GO terms (84%) were only enriched within the neighborhoods of 10 genes or less and thus could not be used for annotating the network. (**C**) 4% of GO terms showed multi-regional enrichment landscapes where two or more distinct regions of the network were associated with peaks of enrichment. (**D**) 12% of GO terms presented region-specific enrichment landscapes where a single region of the network was associated with a peak of enrichment. All together, these region-specific GO terms covered all of the multi-regional enrichment landscapes. (**E**) The composite functional enrichment map of the yeast genetic interaction similarity network was constructed by combining all region-specific GO terms and by grouping them into 19 functional regions based on the similarity of their enrichment landscapes. Different colors represent different regions. The label of each region is a tag list derived from the top five most common words in the denominations of the GO terms associated with the region. (**F**) The systematic comparison of the automated (E) and the manual (Costanzo et al., 2010) annotations of the genetic interaction similarity network. The significance of overlap between each of the 19 SAFE region (E) and each of the 13 manually defined regions (Costanzo et al., 2010) was measured by Fisher's exact test. The resulting p-values were –log10 transformed and represented as a heatmap. All of the manually defined regions overlap at least one of the SAFE regions. Four additional regions (labeled *b*, *d*, *h* and *p*) were only detected by the automated analysis.

Figure 3. SAFE is robust against stochastic variations in network layout and differences in distance measures. (**A**) The original map of the yeast genetic interaction similarity network (i) and the enrichment landscape of a region-specific GO term (ii). (**B**) A different map of the network, produced by re-applying the Kamada-Kawai network layout (i) and re-annotating the network with the same GO term (ii). (**C**) Distribution of enrichment landscape similarities of region-specific GO terms across 10 versions of the

genetic interaction similarity map. Ten network maps were generated by repeatedly applying the Kamada-Kawai layout algorithm. Each map was re-annotated with GO biological process terms, resulting in 10 enrichment landscapes for every GO term. A Spearman's rank correlation coefficient was computed for every pair of enrichment landscapes, associated with the same GO term, that had at least 10 significant enrichment scores. The resulting correlation values for each GO term were averaged and plotted as a distribution. The median of the distribution is indicated. (**D**) Using correlation-weighted shortest path length as a map-independent inter-node distance metric. Here, the length of every step of the path equals to 1 minus the correlation of the genetic interaction profiles of the two connected genes. The quantitative correlation values were not used by the layout algorithm to define the node positions. As a result, this metric is independent from the default edge length-based shortest path length. (**E**) The original map (A) was re-annotated with GO biological process terms using the correlation-weighted shortest path length distance measure. The enrichment landscapes produced by this metric were compared to those produced by the default edge length-based shortest path length and the distribution of the pair-wise Spearman's rank correlation coefficient was plotted. The median of the distribution is indicated.

Figure 4. SAFE annotates networks with chemical genomic data and recapitulates known drug modes-of-action. (**A-B**) The fitness enrichment landscapes map local enrichment for genes associated with sensitivity or resistance to specific chemical compounds and are consistent with the compounds' known modes-of-action. The outlines indicate the location of the 19 functional regions as defined by the GO biological process annotation in Figure 2D. (**A**) Fitness enrichment landscape of doxorubicin, a DNA intercalator that blocks the progression of topoisomerase II and prevents DNA replication. (**B)** Fitness enrichment landscape of verrucarin A, a protein synthesis inhibitor with a known toxicity towards mitochondria. (**C-D**) The composite sensitivity and resistance landscapes of all 132 chemical compounds. Both region-specific and multi-regional chemical compounds were included. Colors were assigned to compounds based on the similarity of their sensitivity (C) or resistance (D) enrichment landscapes, such that compounds with similar landscapes were associated with similar colors. All 132 sensitivity and resistance landscapes were then summed together. For every node, the final color corresponds to the weighted sum of all of its enrichments, where the weights are the enrichment scores for every attribute. As a result, multifunctional nodes are automatically represented by mixed colors.

Figure 5. SAFE provides novel insights into the molecular basis of resistance to bortezomib. (**A**) The fitness enrichment landscape of the proteasome inhibitor bortezomib shows that sensitivity to the drug is associated with mutations in proteasome-mediated protein degradation, cell cycle control and transcriptional regulation functions. In contrast, resistance to bortezomib is specific to genes involved in secretion and vesicle-mediated transport processes. (**B**) The genetic interaction enrichment landscape of *rpt4-145*, an essential member of the regulatory particle of the proteasome, is highly consistent with the fitness landscape of bortezomib and supports the role of secretory mutants in compensating for the effects of proteasome inactivation.

Figure 6. SAFE annotates dense genetic and protein-protein interaction networks. (**A**) **i.** Three additional versions of the genetic interaction similarity network were

generated by progressively reducing the stringency of the correlation threshold (*PCC* > 0.18, *PCC* > 0.16 and *PCC* > 0.14). Compared to the original network (*PCC* > 0.2), this resulted in 1.4, 2.1 and 3.4 times higher network density, respectively. Maps of the networks were generated by applying the Kamada-Kawai layout algorithm in Cytoscape. **ii.** Despite the rapidly decreasing levels of visible structure, SAFE detected functional enrichment in all three network maps. (**B**) Distribution of enrichment landscape similarities of region-specific GO terms across all network densities. For every GO term, a Spearman's rank correlation coefficient was computed between the original enrichment landscape, derived from the PCC > 0.2 network, and the enrichment landscapes derived from denser networks. The medians of the distributions are indicated. (**C**) **i.** A global *S. cerevisiae* protein-protein interaction network, containing 5,699 nodes and 78,406 edges, was constructed from all protein pairs sharing a physical interaction type in the BioGRID database (Chatr-Aryamontri et al., 2015) (Experimental Procedures). A map of the network was generated by applying the Kamada-Kawai layout algorithm in Cytoscape. **ii.** In a manner analogous to the genetic interaction similarity map (Figure 2), the protein-protein interaction map was annotated with SAFE using 4,373 GO biological process terms. All region-specific GO terms were combined into 21 functional regions based on the similarity of their enrichment landscapes and produced the composite functional enrichment map of the yeast protein-protein interaction network. Different colors represent different regions. The label of each region is a tag list derived from the top five most common words in the denominations of the GO terms associated with the region.

EXPERIMENTAL PROCEDURES

The current version of SAFE and its future updates are openly accessible as MATLAB code at https://bitbucket.org/abarysh/safe. I strongly encourage users to subscribe as "watchers" to the Bitbucket repository in order to receive prompt notifications about code improvements and bug fixes.

The basic flow of the algorithm, including key inputs and outputs, are described below. A more detailed description, including examples, is provided in the README file of the Bitbucket repository.

The SAFE algorithm

Software requirements

MATLAB, Bioinformatics Toolbox, Statistics and Machine Learning Toolbox

Input data

1) A network $G=(V,E)$ that consists of a set of nodes V and a set of undirected edges $E \subseteq V \times V$. Edge weights $w_e \mid e \in E$ can be binary ($w_e \in \{0,1\}$) or quantitative ($w_e \in R$).

2) A functional annotation standard in the form of a matrix $A = L \times F$, where L are node labels and F are functional node attributes. The matrix A can contain binary or quantitative functional attributes. In case of binary attributes (e.g., Gene Ontology annotations), the entry a_{ij} in matrix A equals 1 if the i-th node is annotated to the j-th functional attribute, and 0 otherwise. In case of quantitative attributes (e.g., a chemical genomics dataset), the entry a_{ij} in matrix A equals to the quantitative measurement associated with the i-th node in the j-th experimental condition.

Algorithm

1) Load the network.

2) Optional. Create a new network map by applying the Kamada-Kawai network layout algorithm implemented as part of the MatlabBGL toolbox (Gleich, 2009).

3) Load the functional annotation standard.

4) For every node v_i in network G, define a local neighborhood as the set of nodes U_i that can be reached from v_i by traveling no more than a distance d. Distance between any two nodes is measured as an unweighted, a map-based (default) or an edge weight-based shortest path length between them.

    a. In the unweighted case, traversing the edge between nodes v_i and v_k has a constant cost $C_{ik} = 1$.

    b. In the map-based case, traversing the edge between nodes v_i and v_k has the cost of the Euclidean distance between the nodes on the network map: $C_{ik} = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2}$, where x and y are the spatial coordinates of nodes v_i and v_k.

    c. In the edge weight-based case, traversing the edge between nodes v_i and v_k has the cost of the original edge weight w_{ik} (see Input data, above).

    By default, the neighborhood distance threshold d equals to the 0.5th-percentile of all pair-wise node distances in the network.

5) Calculate the enrichment of the neighborhood U_i for each functional attribute $F_j \in F$ in the annotation standard A. Two ways of calculating enrichment are used, depending on whether the attributes are binary or quantitative.

    a. For binary attributes, the enrichment is calculated by performing a one-tailed Fisher's exact test. The significance of the enrichment is determined by the probability p_{ij} that a random overlap between U_i and F_j will fall in the interval $[S_{ij}\ \infty)$:

$$p_{ij} = 1 - \sum_{k=0}^{S_{ij}} \frac{\binom{F_j}{S_{ij}}\binom{N - F_j}{U_i - k}}{\binom{N}{U_i}}$$

where:

N = number of nodes in the network

F_j = number of nodes in the network annotated to functional attribute F_j

U_i = number of nodes in neighborhood U_i

S_{ij} = number of nodes in neighborhood U_i annotated to functional attribute F_j

b. For quantitative attributes, the enrichment is calculated using an empirical test. A neighborhood score S_{ij} is computed by summing the quantitative attribute values F_j for all nodes in U_i. The score S_{ij} is then compared to the mean μ and the standard deviation σ of 1,000 random neighborhood scores obtained by reshuffling the network labels and, consequently, the functional attribute values. The significance of the enrichment is determined as the probability that a single observation from a normal distribution with mean μ and standard deviation σ will fall in the interval $(-\infty\ S_{ij}]$ or $[S_{ij}\ \infty)$, depending on whether the highest or the lowest scores are of interest.

6) Convert neighborhood significance p-values p_{ij} into neighborhood enrichment scores O_{ij}, normalized to a range from 0 to 1, by computing:

$$O_{ij} = \frac{\min\left\{-\log_{10} p_{ij}, -\log_{10} P\right\}}{-\log_{10} P} = \left[0,1\right]$$

where $P = 10^{-16}$ and corresponds to the smallest p-value that Matlab can calculate in a Fisher's exact test.

A neighborhood U_i is considered significantly enriched for the functional attribute F_j if:

$$O_{ij} > \frac{-\log_{10}\left(0.05/F\right)}{-\log_{10} P}$$

where F is the number of functional attributes in the annotation standard.

The enrichment landscape of the functional attribute F_j is defined as the vector of O_j values for all nodes i in the network.

7) Optional. Restrict the analysis to region-specific functional attributes using one of the three available methods.

a. Hartigan's dip test of unimodality (Hartigan and Hartigan, 1985): defines a functional attribute as region-specific if the distribution of pair-wise distances between its significantly enriched neighborhoods is unimodal.

b. Radius-based method (default): defines a functional attribute as region-specific if at least 65% of significantly enriched neighborhoods are within a distance of $2d$, where d is the radius of the neighborhood (see #4 above).

c. Subtractive clustering-based method (Chiu, 1994; Yager and Filev, 1994): defines a functional attribute as region-specific if there is at most one cluster center among the significantly enriched neighborhoods.

8) Optional. Group functional attributes into functional regions based on the similarity of their enrichment landscapes. Distance between functional attributes F_j and F_k is computed using one minus Jaccard (default) or Pearson correlation coefficients between their corresponding enrichment landscapes O_j and O_k. Given an $F \times F$ distance matrix, functional regions are formed using agglomerative hierarchical clustering with average linkage. Clusters are defined by a distance threshold, set by default at 75% of the cluster tree height.

9) Assign a random RGB color to each functional region. All functional attributes in the same region are assigned the same color.

10) Determine the color RGB_i of every node v_i in the network G by computing a weighted average of the colors associated with the functional attributes for which the neighborhood U_i is significantly enriched. The weights correspond to the square of the neighborhood enrichment scores O_{ij} for each functional attribute F_j.

$$RGB_i = \frac{\sum_{j=1}^{F} RGB_j \times O_{ij}^2 \times \hat{O}_{ij}}{F}$$

where $\hat{O}_{ij}$ equals 1 if the neighborhood is significantly enriched for functional attribute F_j and 0 otherwise.

11) Output the results.

a. Plot all nodes according to their position on the network and their computed color.

b. Generate an automatic label for each functional region by identifying the five most frequent words in the denominations of the functional attributes that belong to that region.

c. Print all data, including the list of functional regions, their automatically generated labels and the complete list of functional attributes enriched within each region, into text files.

12) Save all data and settings into a SAFE session file.

Annotation of the yeast genetic interaction similarity network

The yeast genetic interaction similarity network was constructed as described in Costanzo et al. (Costanzo et al., 2010). Briefly, similarity of genetic interaction profiles for all pairs of 1,712 query genes and all pairs of 3,885 array genes was computed using Pearson correlation coefficients (PCC). Correlation values for gene pairs that have been tested both as queries and as arrays were averaged. Gene pairs presenting similarities greater than PCC = 0.2 were connected in a network and visualized using the edge-weighted spring-embedded layout in Cytoscape 2.8 (Shannon et al., 2003). Despite the

choice of an edge-weighted layout algorithm, it was later discovered that a bug in Cytoscape 2.8 caused edge weights to be ignored. The bug is now fixed in the current Cytoscape 3.0 version.

The Gene Ontology (GO) biological process data and the yeast gene association files were downloaded from www.geneontology.org on August 19, 2014 (Ashburner et al., 2000). Annotations were propagated from child to parent terms, such that a gene was associated with a GO term if it was directly annotated to the term or any of its descendants.

The chemical genomics dataset from Hoepfner et al. (Hoepfner et al., 2014) was downloaded from the Dryad digital repository (Hoepfner D, 2013) on December 3, 2013. Only homozygous profiling (HOP) data for 132 chemical compounds with known modes of action (as listed in the Supplementary Table 1 of that publication) were used.

The quantitative genetic interaction data involving 13 members of the proteasome regulatory subunit were obtained and processed as described previously (Baryshnikova et al., 2010a; Baryshnikova et al., 2010b). Specifically, genome-wide synthetic genetic array (SGA) experiments were conducted to construct double mutants involving a deletion or a temperature-sensitive allele of a proteasome member (*RPN1*, *RPN5*, *RPN6*, *RPN7*, *RPN10*, *RPN11*, *RPN12*, *RPT1*, *RPT2*, *RPT3*, *RPT4*, *RPT6* and *SEM1*) and the deletion mutants of all non-essential genes in the yeast genome. Quantitative double mutant fitness values were measured using colony size and compared to the fitness values of the two corresponding single mutants to produce a genetic interaction score that quantifies negative and positive deviations of the observed double mutant fitness from the expected combination of the two singles. The genetic interaction scores for these 13 SGA screens are available as **Supplementary Data 2**.

Annotation of the yeast protein-protein interaction network

A complete set of yeast protein-protein interactions was downloaded from BioGRID on April 26, 2015. The dataset was filtered to include only "physical" interaction types and exclude "biochemical activity" and "protein-RNA" experiments. The network was visualized using the spring-embedded layout in Cytoscape 3.2 (Shannon et al., 2003).

REFERENCES

Agapito, G., Guzzi, P.H., and Cannataro, M. (2013). Visualization of protein interaction networks: problems and solutions. BMC Bioinformatics *14 Suppl 1*, S1.

Arabidopsis Interactome Mapping, C. (2011). Evidence for network evolution in an Arabidopsis interactome map. Science *333*, 601-607.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T.*, et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25-29.

Baryshnikova, A., Costanzo, M., Dixon, S., Vizeacoumar, F.J., Myers, C.L., Andrews, B., and Boone, C. (2010a). Synthetic genetic array (SGA) analysis in Saccharomyces cerevisiae and Schizosaccharomyces pombe. Methods Enzymol *470*, 145-179.

Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.Y., Ou, J., San Luis, B.J., Bandyopadhyay, S.*, et al.* (2010b). Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. Nat Methods *7*, 1017-1024.

Baryshnikova, A., Costanzo, M., Myers, C.L., Andrews, B., and Boone, C. (2013). Genetic interaction networks: toward an understanding of heritability. Annual review of genomics and human genetics *14*, 111-133.

Botstein, D., and Fink, G.R. (2011). Yeast: an experimental organism for 21st Century biology. Genetics *189*, 695-704.

Canestraro, M., Galimberti, S., Savli, H., Palumbo, G.A., Tibullo, D., Nagy, B., Guerrini, F., Piaggi, S., Cine, N., Metelli, M.R.*, et al.* (2010). Synergistic antiproliferative effect of arsenic trioxide combined with bortezomib in HL60 cell line and primary blasts from patients affected by myeloproliferative disorders. Cancer genetics and cytogenetics *199*, 110-120.

Causton, H.C., Ren, B., Koh, S.S., Harbison, C.T., Kanin, E., Jennings, E.G., Lee, T.I., True, H.L., Lander, E.S., and Young, R.A. (2001). Remodeling of yeast genome expression in response to environmental changes. Mol Biol Cell *12*, 323-337.

Chatr-Aryamontri, A., Breitkreutz, B.J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L.*, et al.* (2015). The BioGRID interaction database: 2015 update. Nucleic Acids Res *43*, D470-478.

Chen, D., Frezza, M., Schmitt, S., Kanwar, J., and Dou, Q.P. (2011). Bortezomib as the first proteasome inhibitor anticancer drug: current status and future perspectives. Current cancer drug targets *11*, 239-253.

Chiu, S. (1994). Fuzzy model identification based on cluster estimation. Journal of Intelligent and Fuzzy Systems *2*.

Consortium, T.G.O. GO Slim and Subset Guide.

Cornish, A.J., and Markowetz, F. (2014). SANTA: quantifying the functional content of molecular networks. PLoS Comput Biol *10*, e1003808.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S.*, et al.* (2010). The genetic landscape of a cell. Science *327*, 425-431.

Cusick, M.E., Klitgord, N., Vidal, M., and Hill, D.E. (2005). Interactome: gateway into systems biology. Hum Mol Genet *14 Spec No. 2*, R171-181.

Deleu, S., Lemaire, M., Arts, J., Menu, E., Van Valckenborgh, E., Vande Broek, I., De Raeve, H., Coulton, L., Van Camp, B., Croucher, P.*, et al.* (2009). Bortezomib alone or in combination with the histone deacetylase inhibitor JNJ-26481585: effect on myeloma bone disease in the 5T2MM murine model of myeloma. Cancer research *69*, 5307-5311.

Feng, R., Oton, A., Mapara, M.Y., Anderson, G., Belani, C., and Lentzsch, S. (2007). The histone deacetylase inhibitor, PXD101, potentiates bortezomib-induced anti-multiple myeloma effect by induction of oxidative stress and DNA damage. British journal of haematology *139*, 385-397.

Fornari, F.A., Randolph, J.K., Yalowich, J.C., Ritke, M.K., and Gewirtz, D.A. (1994). Interference by doxorubicin with DNA unwinding in MCF-7 breast tumor cells. Molecular pharmacology *45*, 649-656.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell *11*, 4241-4257.

Gasch, A.P., and Werner-Washburne, M. (2002). The genomics of yeast responses to environmental stress and starvation. Functional & integrative genomics *2*, 181-192.

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B*., et al.* (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631-636.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B*., et al.* (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature *418*, 387-391.

Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y.L., Ooi, C.E., Godwin, B., Vitols, E*., et al.* (2003). A protein interaction map of Drosophila melanogaster. Science *302*, 1727-1736.

Gleich, D.F. (2009). Models and Algorithms for PageRank Sensitivity (Stanford University).

Hartigan, J.A., and Hartigan, P.M. (1985). The Dip Test of Unimodality. The Annals of Statistics *13*, 70-84.

Hernandez, F., and Cannon, M. (1982). Inhibition of protein synthesis in Saccharomyces cerevisiae by the 12,13-epoxytrichothecenes trichodermol, diacetoxyscirpenol and verrucarin A. Reversibility of the effects. The Journal of antibiotics *35*, 875-881.

Hillenmeyer, M.E., Fung, E., Wildenhain, J., Pierce, S.E., Hoon, S., Lee, W., Proctor, M., St Onge, R.P., Tyers, M., Koller, D*., et al.* (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science *320*, 362-365.

Ho, C.H., Piotrowski, J., Dixon, S.J., Baryshnikova, A., Costanzo, M., and Boone, C. (2011). Combining functional genomics and chemical biology to identify targets of bioactive compounds. Curr Opin Chem Biol *15*, 66-78.

Hoepfner, D., Helliwell, S.B., Sadlish, H., Schuierer, S., Filipuzzi, I., Brachat, S., Bhullar, B., Plikat, U., Abraham, Y., Altorfer, M*., et al.* (2014). High-resolution chemical dissection of a model eukaryote reveals targets, pathways and gene functions. Microbiological research *169*, 107-120.

Hoepfner D, H.S., Sadlish H, Schuierer S, Filipuzzi I, Brachat S, Bhullar B, Plikat U, Abraham Y, Altorfer M, Aust T, Baeriswyl L, Cerino R, Chang L, Estoppey D, Eichenberger J, Frederiksen M, Hartmann N, Hohendahl A, Knapp B, Krastel P, Melin

N, Nigsch F, Oakeley EJ, Petitjean V, Petersen F, Riedl R, Schmitt EK, Staedtler F, Studer C, Tallarico JA, Wetzel S, Fishman MC, Porter JA, Movva NR (2013). Data from: High-resolution chemical dissection of a model eukaryote reveals targets, pathways and gene functions (Dryad Digital Repository).

Holcik, M., and Sonenberg, N. (2005). Translational control in stress and apoptosis. Nature reviews Molecular cell biology *6*, 318-327.

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

Jeong, H., Mason, S.P., Barabasi, A.L., and Oltvai, Z.N. (2001). Lethality and centrality in protein networks. Nature *411*, 41-42.

Kamada, T., and Kawai, S. (1989). An algorithm for drawing general undirected graphs. Information Processing Letters *31*, 7-15.

Khatri, P., Sirota, M., and Butte, A.J. (2012). Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol *8*, e1002375.

Kobourov, S.G. (2012). Spring Embedders and Force Directed Graph Drawing Algorithms. arXiv:12013011v1.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P.*, et al.* (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature *440*, 637-643.

Lee, A.H., Iwakoshi, N.N., Anderson, K.C., and Glimcher, L.H. (2003). Proteasome inhibitors disrupt the unfolded protein response in myeloma cells. Proc Natl Acad Sci U S A *100*, 9946-9951.

Lee, A.Y., St Onge, R.P., Proctor, M.J., Wallace, I.M., Nile, A.H., Spagnuolo, P.A., Jitkova, Y., Gronda, M., Wu, Y., Kim, M.K.*, et al.* (2014). Mapping the cellular response to small molecules using chemogenomic fitness signatures. Science *344*, 208-211.

Lipson, C., Alalouf, G., Bajorek, M., Rabinovich, E., Atir-Lande, A., Glickman, M., and Bar-Nun, S. (2008). A proteasomal ATPase contributes to dislocation of endoplasmic reticulum-associated degradation (ERAD) substrates. J Biol Chem *283*, 7166-7175.

Mitra, K., Carvunis, A.R., Ramesh, S.K., and Ideker, T. (2013). Integrative approaches for finding modular structure in biological networks. Nat Rev Genet *14*, 719-732.

Mo, M.L., Palsson, B.O., and Herrgard, M.J. (2009). Connecting extracellular metabolomic measurements to intracellular flux states in yeast. BMC Syst Biol *3*, 37.

Momparler, R.L., Karon, M., Siegel, S.E., and Avila, F. (1976). Effect of adriamycin on DNA, RNA, and protein synthesis in cell-free systems and intact cells. Cancer research *36*, 2891-2895.

Murray, M.Y., Auger, M.J., and Bowles, K.M. (2014). Overcoming bortezomib resistance in multiple myeloma. Biochem Soc Trans *42*, 804-808.

Myers, C.L., Barrett, D.R., Hibbs, M.A., Huttenhower, C., and Troyanskaya, O.G. (2006). Finding function: evaluation methods for functional genomic data. BMC Genomics *7*, 187.

Nayak, R.R., Kearns, M., Spielman, R.S., and Cheung, V.G. (2009). Coexpression network based on natural variation in human gene expression reveals gene interactions and functions. Genome Res *19*, 1953-1962.

Newman, M.E. (2006). Modularity and community structure in networks. Proc Natl Acad Sci U S A *103*, 8577-8582.

Obeng, E.A., Carlson, L.M., Gutman, D.M., Harrington, W.J., Jr., Lee, K.P., and Boise, L.H. (2006). Proteasome inhibitors induce a terminal unfolded protein response in multiple myeloma cells. Blood *107*, 4907-4916.

Parsons, A.B., Brost, R.L., Ding, H., Li, Z., Zhang, C., Sheikh, B., Brown, G.W., Kane, P.M., Hughes, T.R., and Boone, C. (2004). Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. Nat Biotechnol *22*, 62-69.

Parsons, A.B., Lopez, A., Givoni, I.E., Williams, D.E., Gray, C.A., Porter, J., Chua, G., Sopko, R., Brost, R.L., Ho, C.H.*, et al.* (2006). Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. Cell *126*, 611-625.

Roemer, T., Davies, J., Giaever, G., and Nislow, C. (2012). Bugs, drugs and chemical genomics. Nat Chem Biol *8*, 46-56.

Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N.*, et al.* (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature *437*, 1173-1178.

Schappert, K.T., and Khachatourians, G.G. (1986). Effects of T-2 toxin on induction of petite mutants and mitochondrial function in Saccharomyces cerevisiae. Curr Genet *10*, 671-676.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.

Sharan, R., and Ideker, T. (2006). Modeling cellular machinery through biological network comparison. Nat Biotechnol *24*, 427-433.

Tacar, O., Sriamornsak, P., and Dass, C.R. (2013). Doxorubicin: an update on anticancer molecular action, toxicity and novel drug delivery systems. The Journal of pharmacy and pharmacology *65*, 157-170.

Tarassov, K., Messier, V., Landry, C.R., Radinovic, S., Serna Molina, M.M., Shames, I., Malitskaya, Y., Vogel, J., Bussey, H., and Michnick, S.W. (2008). An in vivo map of the yeast protein interactome. Science *320*, 1465-1470.

Warner, J.R. (1999). The economics of ribosome biosynthesis in yeast. Trends in biochemical sciences *24*, 437-440.

Yager, R., and Filev, D. (1994). Generation of fuzzy rules by mountain clustering. Journal of Intelligent and Fuzzy Systems *2*, 209-219.

Yu, C., Rahmani, M., Conrad, D., Subler, M., Dent, P., and Grant, S. (2003). The proteasome inhibitor bortezomib interacts synergistically with histone deacetylase inhibitors to induce apoptosis in Bcr/Abl+ cells sensitive and resistant to STI571. Blood *102*, 3765-3774.

Yu, H., Braun, P., Yildirim, M.A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N*., et al.* (2008). High-quality binary protein interaction map of the yeast interactome network. Science *322*, 104-110.

Zakrzewska, A., Boorsma, A., Beek, A.T., Hageman, J.A., Westerhuis, J.A., Hellingwerf, K.J., Brul, S., Klis, F.M., and Smits, G.J. (2010). Comparative analysis of transcriptome and fitness profiles reveals general and condition-specific cellular functions involved in adaptation to environmental change in Saccharomyces cerevisiae. Omics : a journal of integrative biology *14*, 603-614.

Zhu, X., Gerstein, M., and Snyder, M. (2007). Getting connected: analysis and principles of biological networks. Genes Dev *21*, 1010-1024.

Figure 1

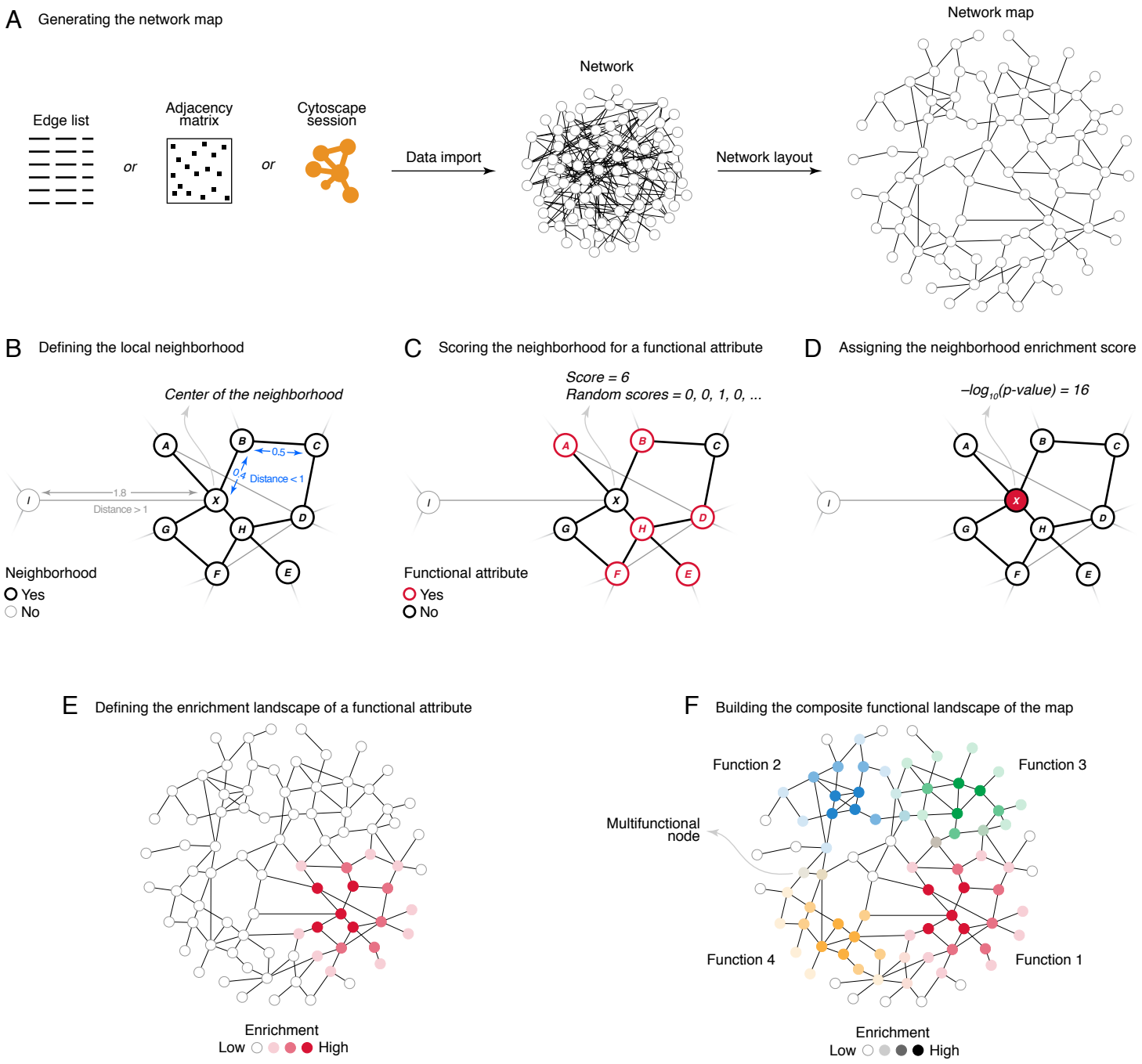
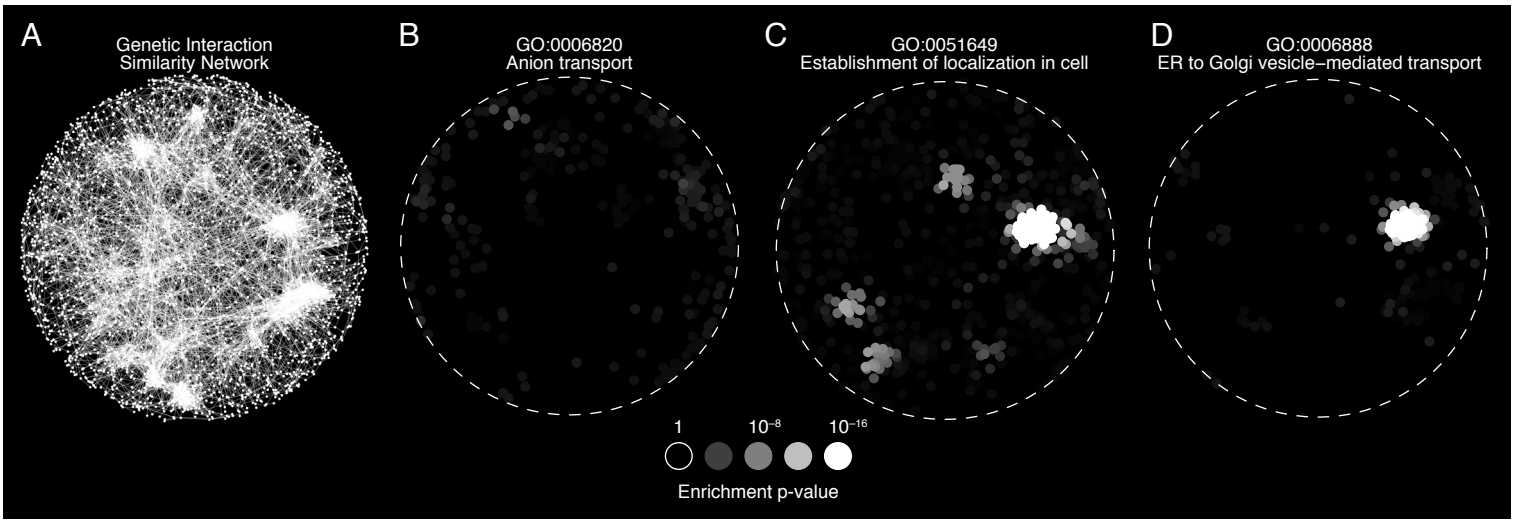
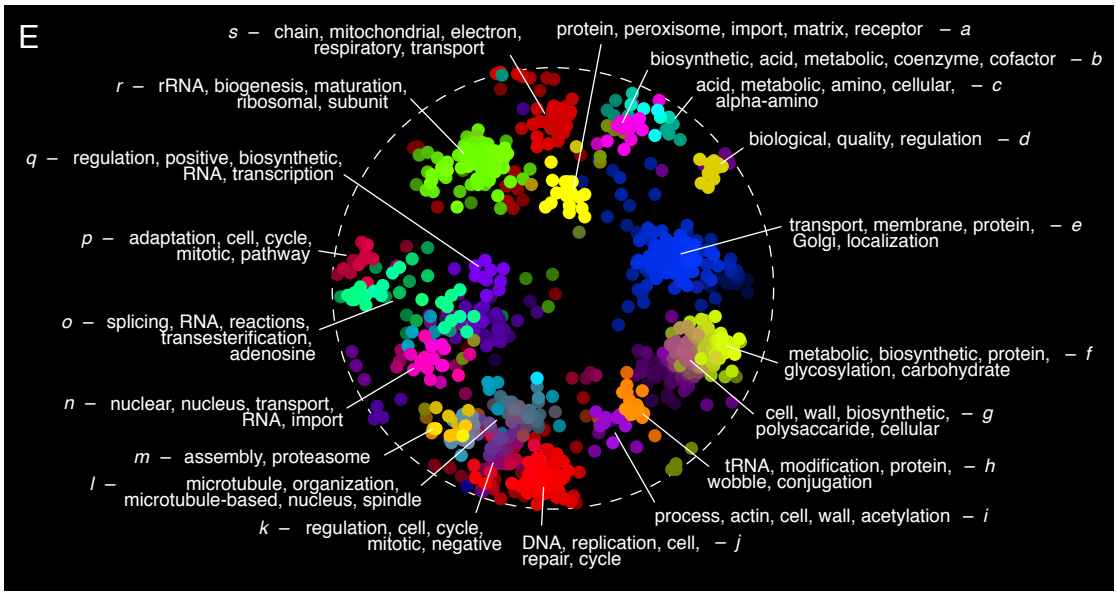

A
Generating the network map
Edge list
or
Adjacency matrix
or
Cytoscape session
Data import
Network
Network layout
Network map
B
Defining the local neighborhood
Center of the neighborhood
1.8
Distance > 1
0.5
0.4
Distance < 1
Neighborhood
Yes
No
C
Scoring the neighborhood for a functional attribute
Score = 6
Random scores = 0, 0, 1, 0, ...
Functional attribute
Yes
No
D
Assigning the neighborhood enrichment score
–log10(p-value) = 16
E
Defining the enrichment landscape of a functional attribute
Enrichment
Low
High
F
Building the composite functional landscape of the map
Function 2
Function 3
Multifunctional node
Function 4
Function 1
Enrichment
Low
High

Figure 2


A
Genetic Interaction
Similarity Network
B
GO:0006820
Anion transport
C
GO:0051649
Establishment of localization in cell
D
GO:0006888
ER to Golgi vesicle–mediated transport
1
10⁻⁸
10⁻¹⁶
Enrichment p-value


E
s – chain, mitochondrial, electron, respiratory, transport
protein, peroxisome, import, matrix, receptor – a
biosynthetic, acid, metabolic, coenzyme, cofactor – b
r – rRNA, biogenesis, maturation, ribosomal, subunit
acid, metabolic, amino, cellular, alpha-amino – c
biological, quality, regulation – d
q – regulation, positive, biosynthetic, RNA, transcription
transport, membrane, protein, Golgi, localization – e
p – adaptation, cell, cycle, mitotic, pathway
o – splicing, RNA, reactions, transesterification, adenosine
metabolic, biosynthetic, protein, glycosylation, carbohydrate – f
n – nuclear, nucleus, transport, RNA, import
cell, wall, biosynthetic, polysaccaride, cellular – g
m – assembly, proteasome
tRNA, modification, protein, wobble, conjugation – h
l – microtubule, organization, microtubule-based, nucleus, spindle
process, actin, cell, wall, acetylation – i
k – regulation, cell, cycle, mitotic, negative
DNA, replication, cell, repair, cycle – j

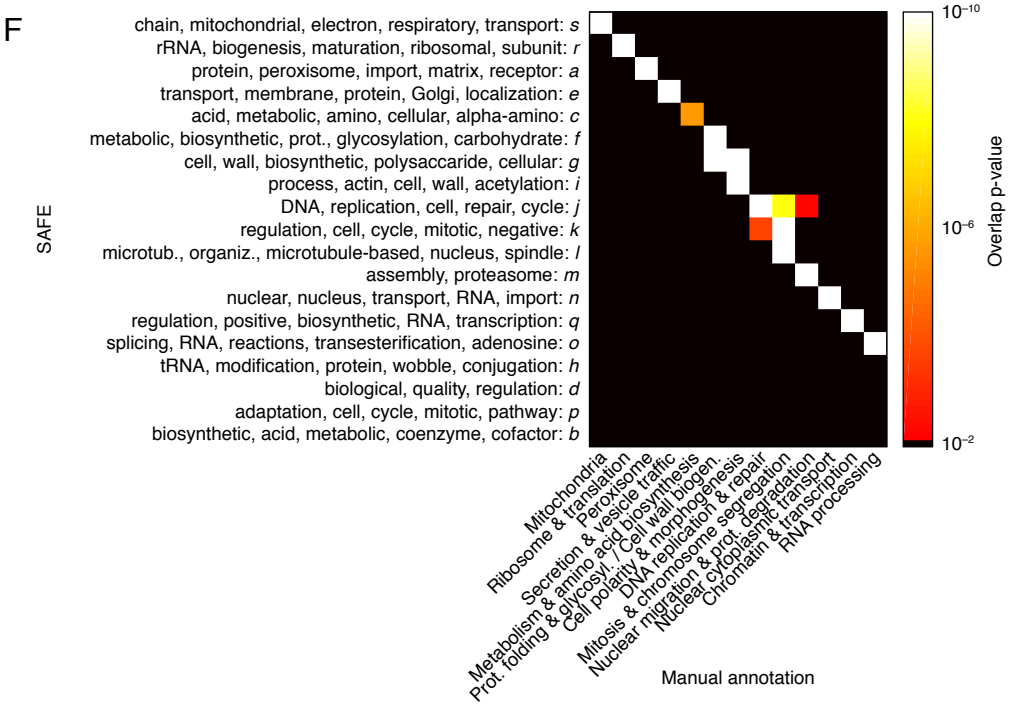

F
chain, mitochondrial, electron, respiratory, transport: s
rRNA, biogenesis, maturation, ribosomal, subunit: r
protein, peroxisome, import, matrix, receptor: a
transport, membrane, protein, Golgi, localization: e
acid, metabolic, amino, cellular, alpha-amino: c
metabolic, biosynthetic, prot., glycosylation, carbohydrate: f
cell, wall, biosynthetic, polysaccaride, cellular: g
process, actin, cell, wall, acetylation: i
DNA, replication, cell, repair, cycle: j
regulation, cell, cycle, mitotic, negative: k
microtub., organiz., microtubule-based, nucleus, spindle: l
assembly, proteasome: m
nuclear, nucleus, transport, RNA, import: n
regulation, positive, biosynthetic, RNA, transcription: q
splicing, RNA, reactions, transesterification, adenosine: o
tRNA, modification, protein, wobble, conjugation: h
biological, quality, regulation: d
adaptation, cell, cycle, mitotic, pathway: p
biosynthetic, acid, metabolic, coenzyme, cofactor: b
SAFE
Mitochondria
Ribosome & translation
Peroxisome
Secretion & vesicle traffic
Metabolism & amino acid biosynthesis
Prot. folding & glycosyl. / Cell wall biogen.
Cell polarity & morphogenesis
DNA replication & repair
Mitosis & chromosome segregation
Nuclear migration & prot. degradation
Nuclear cytoplasmic transport
Chromatin & transcription
RNA processing
Manual annotation
10⁻¹⁰
10⁻⁶
10⁻²
Overlap p-value

Figure 3

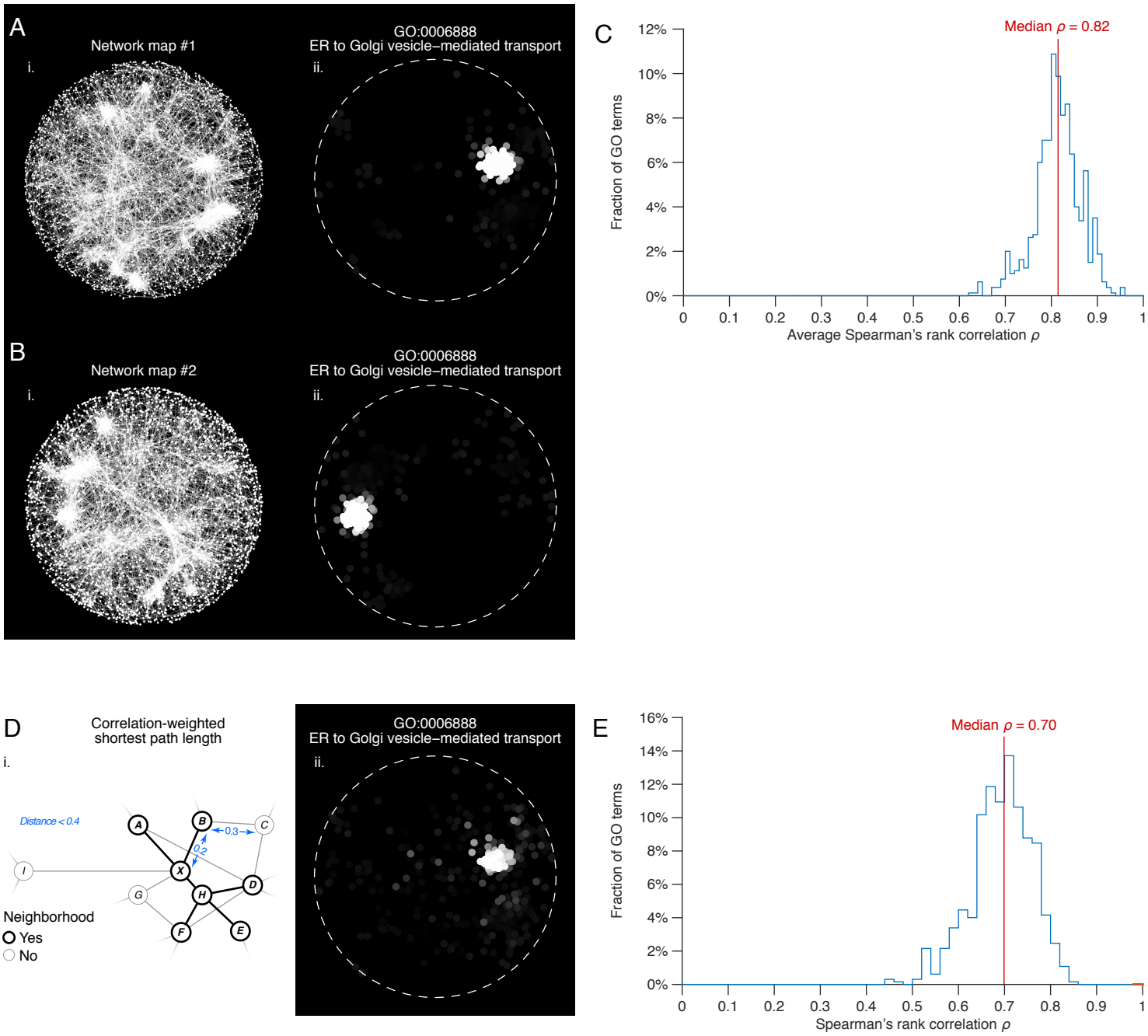

A
Network map #1
i.
GO:0006888
ER to Golgi vesicle−mediated transport
ii.
B
Network map #2
i.
GO:0006888
ER to Golgi vesicle−mediated transport
ii.
C
Median ρ = 0.82
Fraction of GO terms
12%
10%
8%
6%
4%
2%
0%
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Average Spearman's rank correlation ρ
D
Correlation-weighted
shortest path length
i.
Distance < 0.4
A
B
C
I
X
G
H
D
F
E
0.3
0.2
Neighborhood
Yes
No
GO:0006888
ER to Golgi vesicle−mediated transport
ii.
E
Median ρ = 0.70
Fraction of GO terms
16%
14%
12%
10%
8%
6%
4%
2%
0%
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Spearman's rank correlation ρ

Figure 4

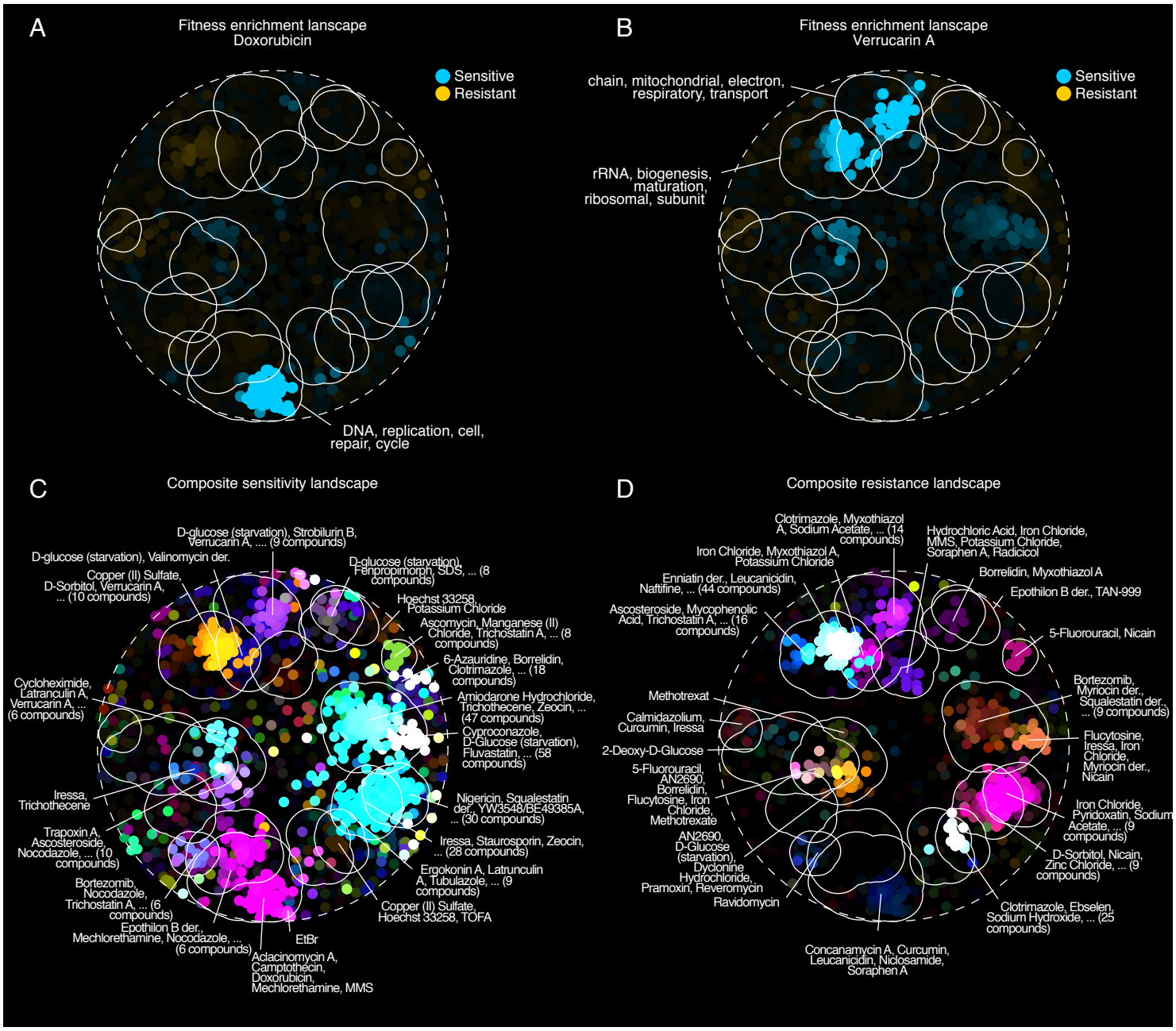

A
Fitness enrichment lanscape
Doxorubicin
Sensitive
Resistant
DNA, replication, cell, repair, cycle
B
Fitness enrichment lanscape
Verrucarin A
Sensitive
Resistant
chain, mitochondrial, electron, respiratory, transport
rRNA, biogenesis, maturation, ribosomal, subunit
C
Composite sensitivity landscape
D-glucose (starvation), Strobilurin B, Verrucarin A, ... (9 compounds)
D-glucose (starvation), Valinomycin der.
Copper (II) Sulfate, D-Sorbitol, Verrucarin A, ... (10 compounds)
D-glucose (starvation), Fenpropimorph, SDS, ... (8 compounds)
Hoechst 33258, Potassium Chloride
Ascomycin, Manganese (II) Chloride, Trichostatin A, ... (8 compounds)
6-Azauridine, Borrelidin, Clotrimazole, ... (18 compounds)
Cycloheximide, Latranculin A, Verrucarin A, ... (6 compounds)
Amiodarone Hydrochloride, Trichothecene, Zeocin, ... (47 compounds)
Cyproconazole, D-Glucose (starvation), Fluvastatin, ... (58 compounds)
Iressa, Trichothecene
Nigericin, Squalestatin der., YW3548/BE49385A, ... (30 compounds)
Trapoxin A, Ascosteroside, Nocodazole, ... (10 compounds)
Iressa, Staurosporin, Zeocin, ... (28 compounds)
Ergokonin A, Latrunculin A, Tubulazole, ... (9 compounds)
Bortezomib, Nocodazole, Trichostatin A, ... (6 compounds)
Copper (II) Sulfate, Hoechst 33258, TOFA
Epothilon B der., Mechlorethamine, Nocodazole, ... (6 compounds)
EtBr
Aclacinomycin A, Camptothecin, Doxorubicin, Mechlorethamine, MMS
D
Composite resistance landscape
Clotrimazole, Myxothiazol A, Sodium Acetate, ... (14 compounds)
Hydrochloric Acid, Iron Chloride, MMS, Potassium Chloride, Soraphen A, Radicicol
Iron Chloride, Myxothiazol A, Potassium Chloride
Borrelidin, Myxothiazol A
Enniatin der., Leucanicidin, Naftifine, ... (44 compounds)
Epothilon B der., TAN-999
Ascosteroside, Mycophenolic Acid, Trichostatin A, ... (16 compounds)
5-Fluorouracil, Nicain
Bortezomib, Myriocin der., Squalestatin der., ... (9 compounds)
Methotrexat
Calmidazolium, Curcumin, Iressa
Flucytosine, Iressa, Iron Chloride, Myriocin der., Nicain
2-Deoxy-D-Glucose
5-Fluorouracil, AN2690, Borrelidin, Flucytosine, Iron Chloride, Methotrexate
Iron Chloride, Pyridoxatin, Sodium Acetate, ... (9 compounds)
AN2690, D-Glucose (starvation), Dyclonine Hydrochloride, Pramoxin, Reveromycin
D-Sorbitol, Nicain, Zinc Chloride, ... (9 compounds)
Ravidomycin
Clotrimazole, Ebselen, Sodium Hydroxide, ... (25 compounds)
Concanamycin A, Curcumin, Leucanicidin, Niclosamide, Soraphen A

Figure 5

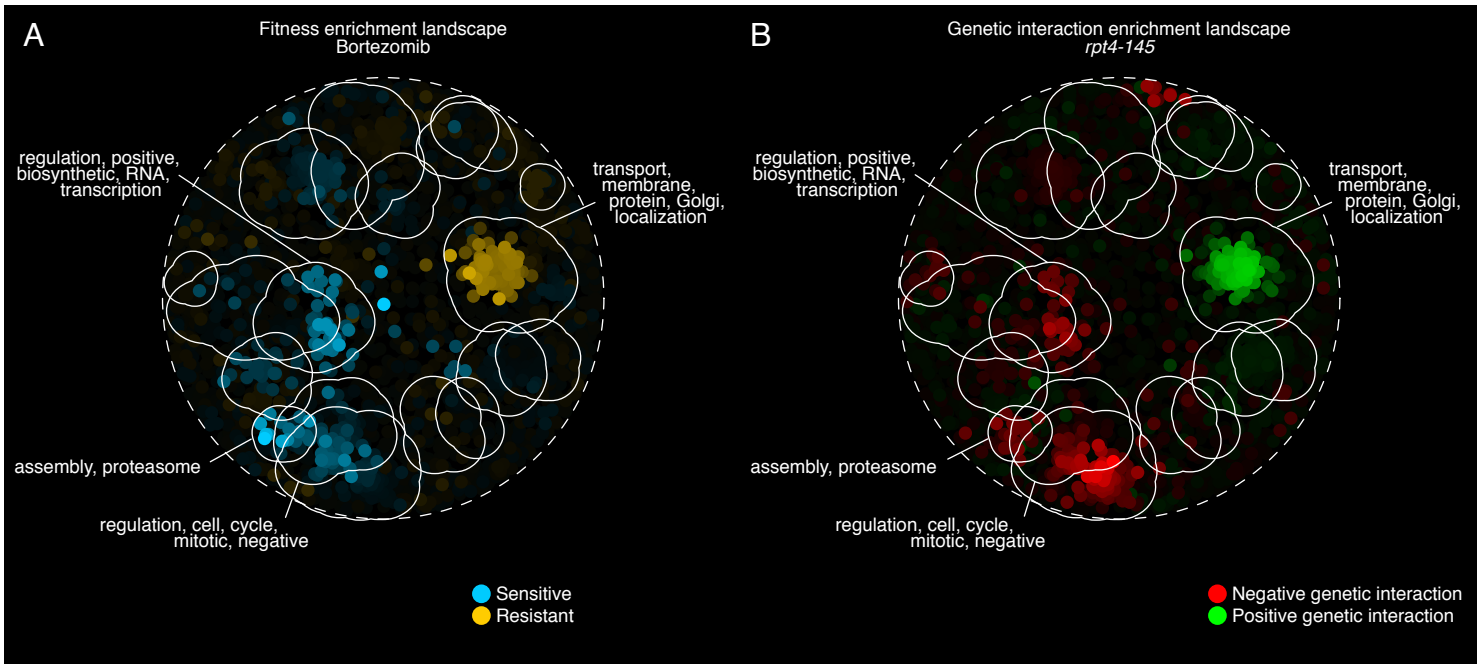
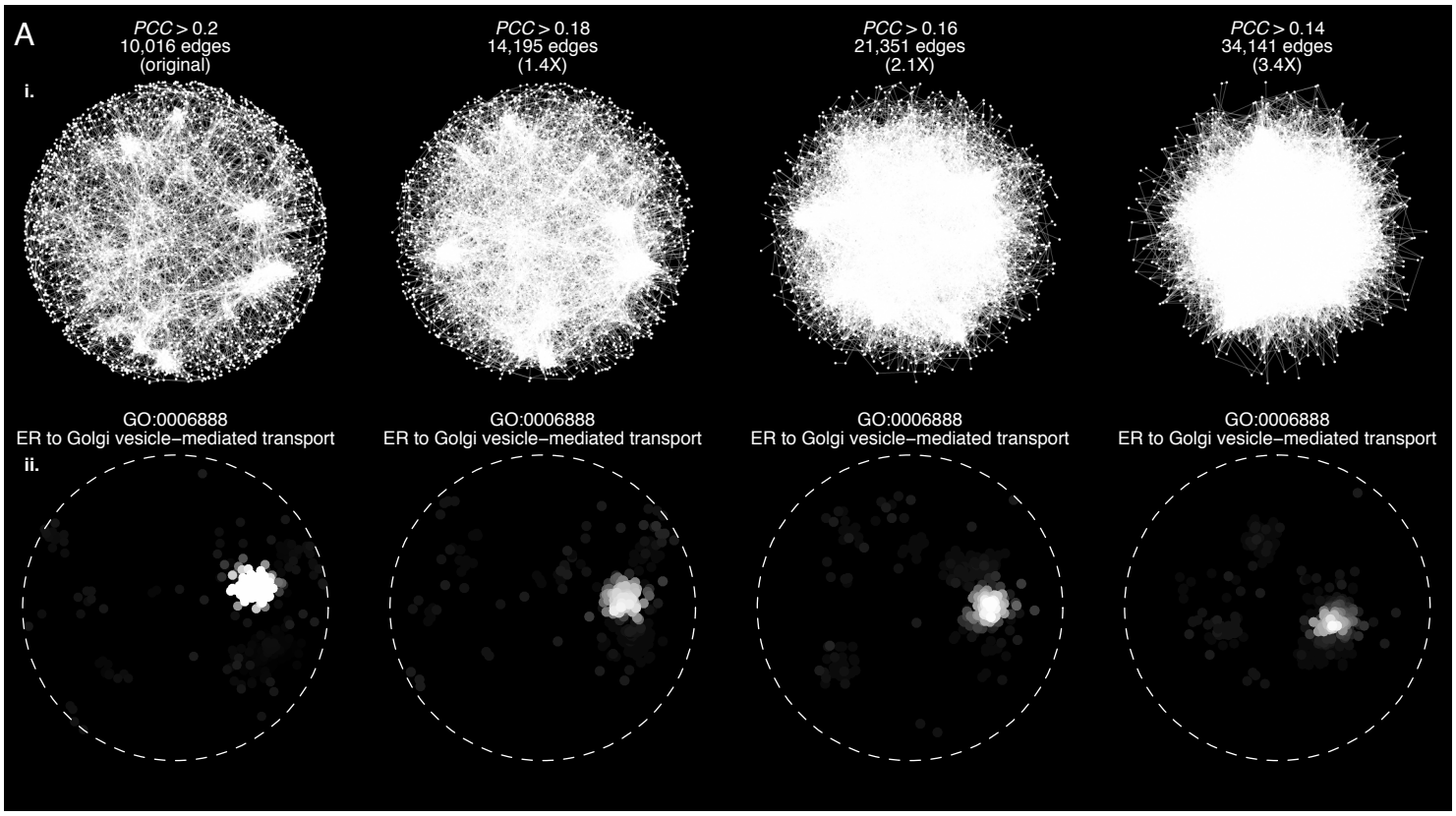
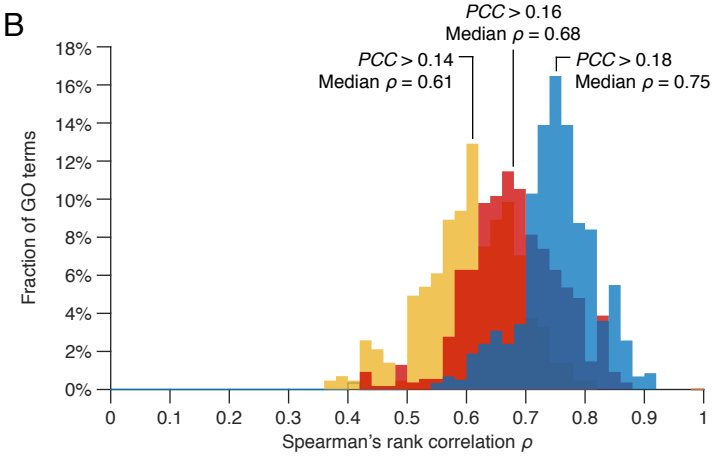
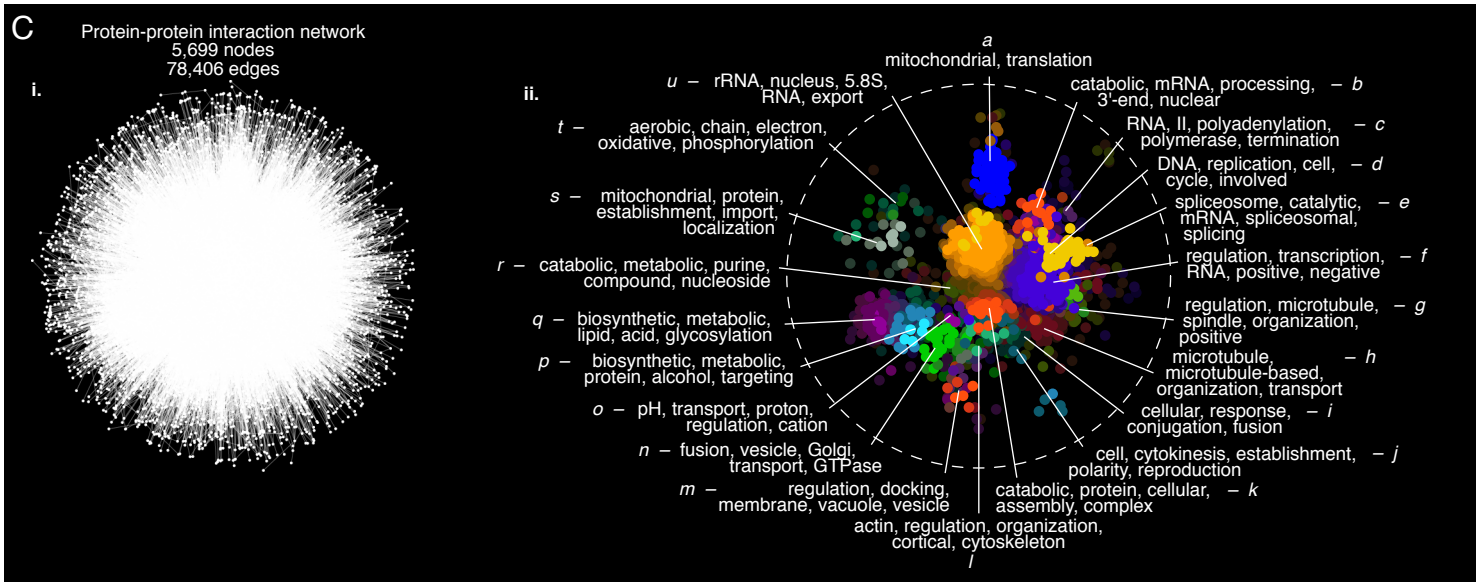

A
Fitness enrichment landscape
Bortezomib
regulation, positive, biosynthetic, RNA, transcription
transport, membrane, protein, Golgi, localization
assembly, proteasome
regulation, cell, cycle, mitotic, negative
Sensitive
Resistant
B
Genetic interaction enrichment landscape
rpt4-145
regulation, positive, biosynthetic, RNA, transcription
transport, membrane, protein, Golgi, localization
assembly, proteasome
regulation, cell, cycle, mitotic, negative
Negative genetic interaction
Positive genetic interaction

Figure 6


A
i.
PCC > 0.2
10,016 edges
(original)
PCC > 0.18
14,195 edges
(1.4X)
PCC > 0.16
21,351 edges
(2.1X)
PCC > 0.14
34,141 edges
(3.4X)
ii.
GO:0006888
ER to Golgi vesicle–mediated transport
GO:0006888
ER to Golgi vesicle–mediated transport
GO:0006888
ER to Golgi vesicle–mediated transport
GO:0006888
ER to Golgi vesicle–mediated transport


B
PCC > 0.16
Median ρ = 0.68
PCC > 0.14
Median ρ = 0.61
PCC > 0.18
Median ρ = 0.75
Fraction of GO terms
18%
16%
14%
12%
10%
8%
6%
4%
2%
0%
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Spearman's rank correlation ρ


C
Protein-protein interaction network
5,699 nodes
78,406 edges
i.
ii.
a
mitochondrial, translation
b – catabolic, mRNA, processing, 3'-end, nuclear
c – RNA, II, polyadenylation, polymerase, termination
d – DNA, replication, cell, cycle, involved
e – spliceosome, catalytic, mRNA, spliceosomal, splicing
f – regulation, transcription, RNA, positive, negative
g – regulation, microtubule, spindle, organization, positive
h – microtubule, microtubule-based, organization, transport
i – cellular, response, conjugation, fusion
j – cell, cytokinesis, establishment, polarity, reproduction
k – catabolic, protein, cellular, assembly, complex
l
actin, regulation, organization, cortical, cytoskeleton
m – regulation, docking, membrane, vacuole, vesicle
n – fusion, vesicle, Golgi, transport, GTPase
o – pH, transport, proton, regulation, cation
p – biosynthetic, metabolic, protein, alcohol, targeting
q – biosynthetic, metabolic, lipid, acid, glycosylation
r – catabolic, metabolic, purine, compound, nucleoside
s – mitochondrial, protein, establishment, import, localization
t – aerobic, chain, electron, oxidative, phosphorylation
u – rRNA, nucleus, 5.8S, RNA, export