

Functional network motifs defined through integration of protein-protein and genetic interactions

Amruta Sahoo¹ and Sebastian Pechmann^{2*}

¹Département de Biochimie, Université de Montréal, QC, H3T 1J4, Canada

² 66125 Saarbrücken, Germany

*E-mail: sebastian@pechmannlab.net

Running Title: Functional network motifs

Keywords: Network motif, feedback regulation, protein interaction, genetic interaction

ABSTRACT

Cells are enticingly complex systems. The identification of feedback regulation is critically important for understanding this complexity. Network motifs defined as small graphlets that occur more frequently than expected by chance have revolutionized our understanding of feedback circuits in cellular networks. However, with their definition solely based on statistical over-representation, network motifs often lack biological context, which limits their usefulness. Here, we define functional network motifs (FNMs) through the systematic integration of genetic interaction data that directly informs on functional relationships between genes and encoded proteins. Occurring two orders of magnitude less frequently than conventional network motifs, we found FNMs significantly enriched in genes known to be functionally related. Moreover, our comprehensive analyses of FNMs in yeast showed that they are powerful at capturing both known and putative novel regulatory interactions, thus suggesting a promising strategy towards the systematic identification of feedback regulation in biological networks. Many FNMs appeared as excellent candidates for the prioritization of follow-up biochemical characterization, which is a recurring bottleneck in the targeting of complex diseases. More generally, our work highlights a fruitful avenue for integrating and harnessing genomic network data.

INTRODUCTION

Cells are enticingly complex systems (Wolf et al., 2018). Their phenotypic traits often depend on the concerted action of hundreds or even thousands of individual genes and encoded proteins (Lehner, 2013; Boyle et al., 2017). Rationalizing this interdependency is important both for understanding how cellular function is encoded in the genome (Watanabe et al., 2019) as well as the origins of complex diseases (Wray et al., 2018).

The large-scale mapping of protein-protein interactions (PPI) (Bork et al., 2004; Rual et al., 2005; Krogan et al., 2006; Tarassov et al., 2008; Yu et al., 2008) and genetic interactions (GI) (Schuldiner et al., 2005; Costanzo et al., 2016, 2019) has provided a foundation for understanding the functional organization of cellular networks (Barabasi, 2004). Most proteins function through direct contact, i.e. PPIs. Accordingly, clusters of high interconnectivity in PPI networks have revealed strong hierarchical modularity (Han et al., 2004) that ranges from protein complexes (Gavin et al., 2006) to sets of functionally related proteins (Ravasz et al., 2002; Huynen et al., 2003; von Mering et al., 2003). Significant efforts have been undertaken to identify such network modules from the network structure (Park and Bader, 2011) and link them to observable phenotypes (Wang et al., 2012; Vinayagam et al., 2014). In contrast, GIs inform on the functional relationships between genes by quantifying epistatic fitness effects of their deletion mutants (Mani et al., 2008). Positive GIs are often linked to dependencies within the same function while negative GIs are frequently observed between redundant and compensatory processes (Beltrao et al., 2010). However, many GIs have complex and non-trivial origins (Crona et al., 2017; Pirkl et al., 2017; Kuzmin et al., 2018) that await to be further characterized. Similar to PPIs, GI networks also display strong hierarchical modularity (Beltrao et al., 2010; Cornish and Markowitz, 2014; Fang et al., 2019) that confers robustness and versatility (Fortuna et al., 2017).

The smallest yet arguably most important recurring unit of modularity in biological networks are network motifs (Milo et al., 2004; Kashtan and Alon, 2005). Network motifs were initially defined as small graphlets of size $k \approx 3-6$ nodes that occur more frequently than expected by chance, thus assuming that their selection indicates importance (Milo et al., 2004). Enabled by efficient algorithms for their discovery (Kashtan et al., 2004; Schreiber and Schwöbbermeyer, 2005; Wernicke and Rasche, 2006; Przulj, 2006; Kashani et al., 2009), network motifs have dramatically improved our understanding of biological networks. Through their detailed characterization it has become clear that motifs often encode important feedback circuits with distinct functionality (Alon, 2007) such as feed-forward signaling (Goentoro et al., 2009), control of system states (Elowitz and Leibler, 2000), and coordination of decision making (Brandman et al., 2005). In complement, the large-scale analysis of motifs in networks has helped to connect network topology to function (Vázquez et al., 2004), for instance in transcription regulation (Jothi et al., 2009; Wang and Chen, 2010; Song et al., 2017; Teixeira et al., 2017), stress responses (Kim et al., 2012; Hahn et al., 2006), and development (Freeman, 2000).

Importantly, network motifs are primed to encode the evidently pervasive but so far largely elusive feedback and cross-talk pathways within cellular networks required to understand their complexity. However, the usefulness of network motifs to systematically identify regulatory interactions has so far been limited (Konagurthy and Lesk, 2008). Specifically, motifs are routinely found in very large numbers, which renders the selection of promising candidates for detailed follow-up characterization challenging. Moreover, despite statistical over-representation of their topology most individual network motifs are not evolutionarily conserved (Mazurie et al., 2005), suggesting that only a small fraction of the network motifs may indeed carry function. We therefore hypothesized that an improved definition of network motifs that takes into consideration their individual biological contexts may offer additional insights into the functioning and organization of biological networks.

Here, we defined functional network motifs (FNMs) based on integrating GIs as direct measure of functional relationships between genes into the discovery of motifs in the yeast PPI network. Remarkably, with occurrences of FNMs about two orders of magnitude less frequent than conventional network motifs, we found FNMs significantly enriched in genes known to be functionally related. Moreover, our comprehensive analyses of FNMs in yeast showed that they are powerful at re-identifying known regulatory interactions as well as discovering interesting new ones. Our work highlights a promising avenue to harness genomic network data towards systematically identifying regulatory interactions and further breaking down biological complexity.

RESULTS

The identification of regulatory interactions is critically important for understanding cellular complexity. Network motifs have made a foundational impact on our understanding of the regulation of biological networks. We here hypothesized that an expanded definition of network motifs based on both PPIs as record of protein activity as well as GIs that quantify functional relationships should improve their usefulness for biological discovery. To test this idea, we defined functional network motifs (FNMs) as graphlets in the yeast PPI network that were also enriched in GIs. Specifically, for a graphlet to be considered an FNM we required the presence of GIs between the source node and all nodes in the most distant layer, i.e. GIs that span the full graphlet, as well as overall in $\geq 50\%$ of all possible non-self edges (Figure 1A, S1A).

We next implemented an algorithm for the efficient and exhaustive enumeration of all motifs of size $k = 3, 4, 5, 6$ in the yeast PPI network (*see Methods*). Two deliberate constraints were imposed to reduce computational cost while increasing biological interpretability. Because the

most highly connected nodes in the yeast PPI are usually protein interaction hubs rather than specific regulators, we omitted the highest degree nodes from the network. Moreover, we only considered at most one protein complex subunit per motif as we were interested in links between functional units rather than the architecture of protein complexes.

Remarkably, we counted about two orders of magnitude fewer FNMs than conventional network motifs independent of motif size (Figure 1B). For comparison, we generated all conventional motifs in the yeast PPI and randomized PPI networks as well as FNMs upon randomization of the GI network (Figure S1B, *see Methods*). The obtained results followed expected trends. Illustrated for the most populated topologies, motifs occurred more frequently in the PPI network than in randomized PPI networks, thus could also be considered classical network motifs (Figure 1C). Similarly, FNMs were substantially more frequent than in randomized GI networks, suggesting that the selection of motifs based on high GI density is selective and meaningful (Figure 1C, S1C,D). Taken together, FNMs collectively identify the same topologies as conventional network motifs but their individual occurrences are substantially more rare.

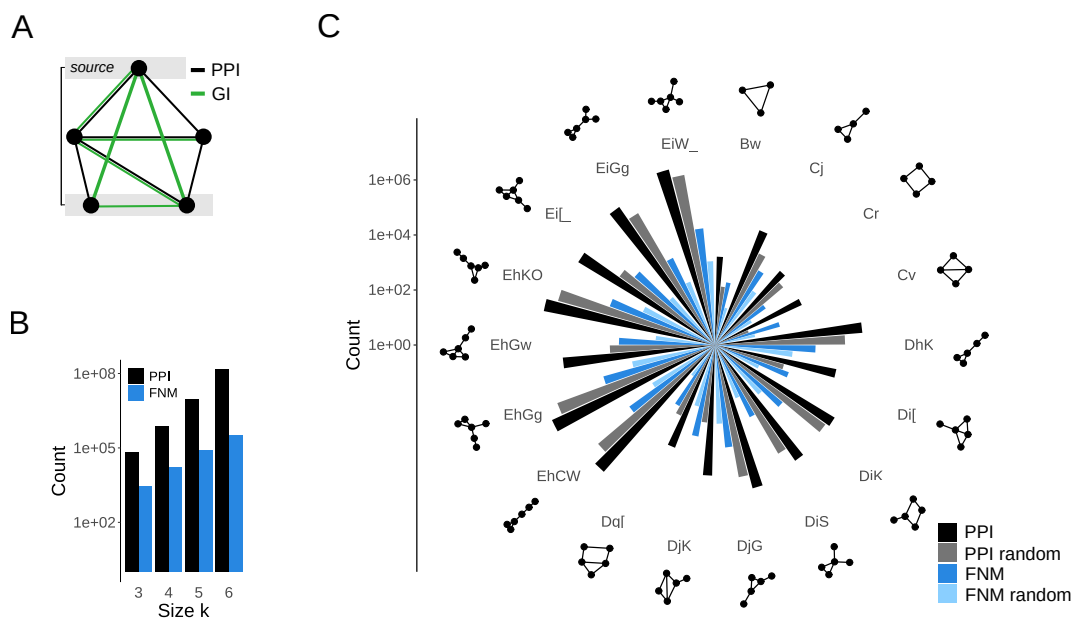


Figure 1 Definition of functional network motifs (FNMs). **(A)** Exemplary schematic of an FNM. Black edges represent protein-protein interactions (PPI) and define a graphlet. Green edges represent genetic interactions (GI). The source node and most distant layer in the motif are highlighted. A network motif becomes an FNM if GIs are present in at least 50% of all non-self edges, and between the source node and all nodes in the most distant layer. **(B)** Counts of conventional network motifs (PPI) and FNMs as function of the motif size k . The occurrences of FNMs are two orders of magnitude lower than that of classical network motifs. **(C)** Counts of motifs in the yeast protein and genetic networks for the most represented motif topologies. Shown are the total motif counts based only on the PPI network (PPI), for randomized PPI networks (PPI random), as well as for FNMs and FNMs computed from randomized GI network (FNM random). Motif topologies are indicated both graphically and in compressed graph6 format.

FNMs are enriched in functionally related genes

Naturally, a more selective definition of network motifs resulted in lower counts. To test whether our definition of FNMs was indeed meaningful, we evaluated if they preferentially contained functionally related and important genes. One intriguing example of functionally related genes is given by genetic suppressors, pairs of genes where the phenotypic effect of a mutation in one gene is compensated by the other. Importantly, while yeast suppressors genetically interact (van Leeuwen et al., 2016), a single GI alone would not bias a motif towards an FNM. Strikingly, we found a significant enrichment of suppressor interactions in FNMs compared to conventional PPI network motifs ($p \ll 10^{-16}$, Fisher's Exact Test, Figure 2A). Thus, FNMs were directly enriched in functional interactions.

Further support for our definition of FNMs was drawn from an analysis of essential genes and genes code for protein complex subunits. Essential genes are functionally indispensable and their deletion mutants are by definition not viable. However, GIs of essential genes can be measured for instance with DAmP alleles that perturb their expression, and are often pronounced. Similarly, protein complex subunits are important due to the functional importance of the complex as well as the high fitness cost of mutants in individual subunits that can destabilize the whole complex. We found FNMs to contain a higher fraction of both essential genes ($p \ll 10^{-16}$, Wilcoxon-Mann-Whitney (WMW) test, Figure 2B) and genes coding for protein complex subunits ($p \ll 10^{-16}$, WMW test, Figure 2B) than conventional PPI network motifs.

To estimate redundancy in the large number of FNMs, we next clustered them based on node overlap (*see Methods*). We found that 93.6%, i.e. the vast majority of FNMs concentrated in clusters of sizes $\approx 7 - 70$, while 26078 FNMs did not cluster (Figure 2C). Of note, very few genes were found responsible for this agglomeration of FNMs. While only 6.4% of FNMs did not cluster, they represented 87.7% of all genes and proteins found in FNMs (Figure 2C). The genes in clustered FNMs had a significantly higher representation in different motifs compared

to genes only found in non-clustered FNMs ($p = 5.3 \cdot 10^{-134}$, WMW test, Figure 2D) because they also had on average a significantly higher degree ($p = 2.3 \cdot 10^{-30}$, WMW test, Figure 2D). Importantly, the non-clustered FNMs retain their enrichment in suppressor interactions, essential genes and complex subunits. Thus, FNMs defined through integration of PPIs and GIs are indeed representing collections of important functional interactions. While clusters of FNMs share strong similarity to larger modules in the PPI network, the FNMs that do not cluster should be particularly interesting.

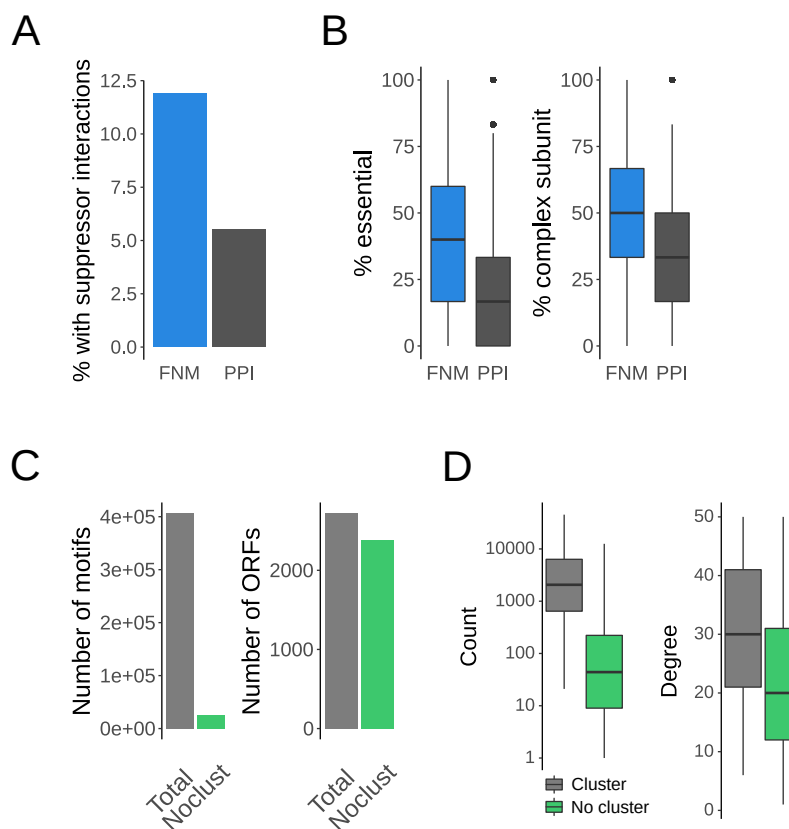


Figure 2 FNMs are enriched in functionally important and related genes. **(A)** Fractions of FNMs and classical network motifs (PPI) that contain suppressor interactions as proxy for functionally related

genes. **(B)** Fractions of essential genes and genes coding for protein complex subunits in FNMs and in classical network motifs (PPI). **(C)** Numbers of FNMs that do not cluster compared to all FNM, together with the number of genes (ORFs) in all FNMs and only in those that do not cluster. **(D)** Distributions of the number of motifs per gene for the genes in FNMs that cluster vs. those that do not cluster together with the network degree for the same groups of genes.

Genetic interactions in FNMs are orthogonal to protein interactions

Having established that FNMs indeed reflect known functional dependencies between genes, we next sought to better understand the composition of their genetic interactions. Overall, negative GIs were clearly more prevalent in FNMs (Figure 3A) but we could not identify any trends that link subsets of FNMs to predominantly positive or negative GIs. This was not unexpected as GIs are known to be complex and strongly context dependent (Markowitz et al., 2007; Kuzmin et al., 2018).

To test any consistency and correlation of GIs and PPIs, we generated a protein complex interaction network by considering all interactions contained in the non-clustering FNMs and mapping them onto joint nodes for the annotated yeast protein complexes. Protein complexes are a good test case because they form important functional units but may interact through multiple subunits. For each pairwise interaction between two complexes, we collected all GIs contained in the corresponding FNMs. While these included both repeated occurrences of the same GI edge from different FNMs and GIs between different complex subunits, almost all pairs of complexes had multiple different GIs. Remarkably, the consensus GIs between protein complexes extracted from the FNMs showed almost perfect consistency towards negative or positive interactions (Figure 3B).

The full yeast protein complex interaction map highlighted important functional connections as well as the orthogonal nature of PPIs and GIs (Figure 3C). Some of the most central protein machines required for genome maintenance such as DNA polymerases dominated the observed negative GIs. Complexes with positive GIs included examples of protein targeting and translocation such as the ESCRT complexes involved in endosomal sorting or the Vam3/Vam7 complex that functions in vacuolar trafficking. Only some complexes shared both PPIs and GIs. Notably, only 25.9% of protein complex pairs that interacted physically also interacted genetically (Figure 3D). In turn, 29.8% of edges between protein complexes were only PPIs, while 44.3% GIs (Figure 3D). Thus, the GIs captured by the FNMs were both consistent and largely orthogonal to the PPIs.

In the extreme, an FNM of size $k = 6$ can connect six different protein complexes, for instance complexes that function in DNA replication and repair (Figure 3E). In this exemplary FNM, SLD5 of the GINS complex that is important for DNA replication connected with both PPIs and negative GIs to DPB3 of the DNA polymerase epsilon subunit, DPB11 as scaffolding protein of the Dpb11p/Sld2p complex important for DNA replication initiation, and TOF1 of the replication pausing checkpoint complex. Further connections were observed from DPB11 to MMS4 of the Holiday-junction-resolvase that functions in recombination and DNA repair, as well as a PPI and positive GI between DBF4, which is the regulatory subunit of a kinase complex important for the initiation of DNA replication, and TOF1. The interplay between the regulation of DNA replication and repair is in general well established and this motif was just one example highlighting the potential of FNMs to identify strong candidates for regulatory interactions from genomic network data.

auxiliary roles such as sensing, regulation, or signalling. The betweenness centrality is an established network metric that quantifies the relative fraction of shortest paths that transverse through a node thus indicating its importance for instance for the flow of information across a network. Using the protein complex network generated from merging the FNM interactions of protein complex subunits into joint nodes, we therefore computed the betweenness centrality for all auxiliary nodes, i.e. non-complex nodes, as defined by the shortest paths between all pairs of protein complexes (Figure 4A).

To test whether the betweenness centrality of the auxiliary nodes identified by the FNMs differed, the corresponding betweenness centralities were also computed for all proteins in the yeast PPI network filtered at maximum degree $d_{max} < 50$, the network used to generate the FNMs. Moreover, we also computed as control betweenness centrality values for all auxiliary nodes in the full yeast PPI network. Betweenness centrality values were significantly higher for FNMs compared to the filtered PPI network ($p = 1.06 \cdot 10^{-6}$, WMW test, Figure 4B), and significantly higher for the filtered PPI network than for the full PPI network ($p = 9.24 \cdot 10^{-174}$, WMW test, Figure 4B). In part this observation scaled with the corresponding network sizes wherein the shortest paths between protein complexes concentrated to fewer nodes in smaller networks (Figure 4D). However, considering only contributing nodes with betweenness centrality values > 0 suggested that the differences in betweenness centralities for the filtered and full PPI networks was independent of the effective network size and rather a function of the presence of very highly connected network hubs (Figure 4D). The higher betweenness centrality for the FNM network must be the result of both a smaller network and the lack of hubs.

The investigation of the the auxiliary nodes with the highest betweenness centrality values in the FNM network lead to several proteins known for their regulatory or cross-talk roles (Figure 4E). BOL1 is a mitochondrial matrix protein and assembly factor involved in the transfer of [4Fe-4S] clusters from the ISA complex to client proteins (Uzarska et al., 2016). TOM71 is

assumed to aid in the targeting of proteins to mitochondria (Schlossmann et al., 1996). TAH1 is a HSP90 chaperone co-factor required for C/D small nucleolar ribonucleoprotein assembly (Jiménez et al., 2012). The mitochondrial thioredoxin peroxidase PRX1 is important for sensing and countering oxidative stress (Pedrajas et al., 2000). And the membrane transporter PDR15 contributes to the cross-talk between stress response and the pleiotropic drug resistance network for cellular detoxification (Wolfger et al., 2004).

To explore potentially novel regulatory interactions in a more controlled test scenario, we next identified all FNMs that connect two of the most important and best-characterized subnetworks in the cell, namely the set of transcription factors (TFs) forming the transcription regulatory network, and the yeast metabolic network. The metabolic network engages in extensive cross-talk to maintain overall energy homeostasis, predominantly through reporter metabolites that bind to regulatory proteins such as transcription factors or metabolic enzymes, but also mediated through PPIs (Grüning et al., 2010). Remarkably, we only found 11 FNMs or groups of interlinked FNMs that connect these large and important systems, each with distinct patterns of PPIs and GIs between TFs and metabolic enzymes (Figure 4E).

The FNMs included the well characterized response to glucose repression where the TF MIG1 directly inhibits the expression of the sucrose hydrolyzing enzyme SUC2. Both were additionally found to genetically interact with the pre-mRNA splicing machinery. The global regulator of respiratory gene expression HAP3 was linked to the enzyme ALG9 involved in N-linked glycosylation in the endoplasmic reticulum via two sensors of alkaline pH, RIM8 and RIM21, as well as USA1, the scaffold protein of the ubiquitin ligase HRD1. It is plausible to speculate about a feedback circuit that involves pH sensing and selective degradation. Similarly, the specialized proteasome and stress response TF RPN4 was found to interact not only with the core proteasome subunit PRE7, but via the E3 ligases UBR1 and UBR2 as well as the E2 RAD6 also with the metabolic enzyme DFR1 involved in tetrahydrofolate biosynthe-

sis within respiratory metabolism. RAD6 is known to form complexes with UBR1 and UBR2 respectively with different substrate specificities, thus may well contribute to the coordination of RPN4 and DFR1 protein levels. As such, this FNM appeared as an exceptionally strong candidate for contributing to the coordination of transcription regulation and metabolism under stress. In summary, the FNMs have revealed very promising candidates for putative regulatory interactions in low enough numbers that render them immediately amenable for biochemical characterization.

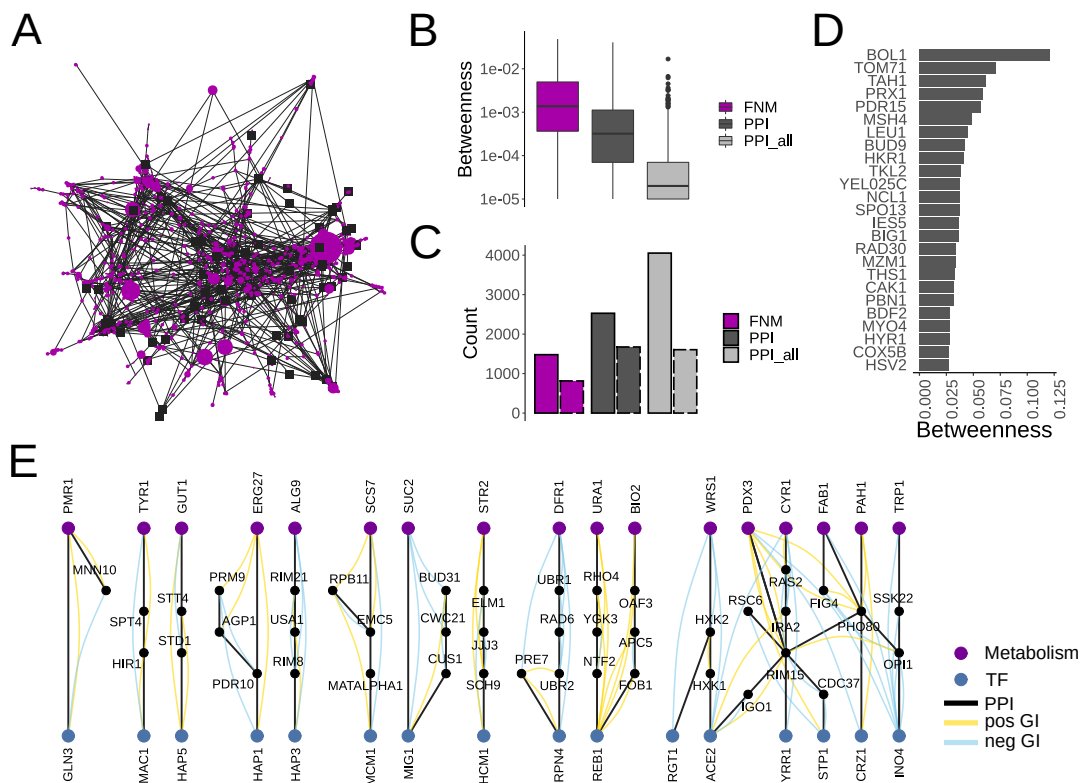


Figure 4 FNMs identify known and novel feedback and cross-talk interactions. **(A)** Yeast protein complex interaction network based on the interactions in the FNMs. Protein complex subunits are merged into joint nodes (black squares). Auxiliary, non-complex, nodes (purple circles) are scaled based on their

betweenness centrality. **(B)** Distributions of the betweenness centrality of the auxiliary nodes connecting protein complexes. Shown are the values for the network derived from the FNMs, from the PPI network with $d_{max} < 50$ as used to derive the FNMs, and from the full PPI network (PPI_all). **(C)** Network sizes of the FNM, PPI and PPI_all networks. Also shown are the effective network sizes that only include auxiliary nodes with non-zero betweenness centralities (dashed borders). **(D)** Ranking of the auxiliary nodes with the highest betweenness centrality in the FNM network. **(E)** All FNMs that connect the yeast consensus transcription factors to the yeast metabolic network.

Transcriptionally co-regulated FNMs are rare but important

Finally, cellular systems are highly dynamic and constantly rewire their networks (Pe'er et al., 2001), complexes (de Lichtenberg et al., 2005), and also network motifs (Prill et al., 2005; Doyle and Csete, 2005) under changing conditions. For instance, dynamically adapting motifs temporally coordinate the global transcription regulation of metabolism (Chechik et al., 2008). Without knowledge of the network states under perturbation, the inference of dynamically changing network motifs and their activity under different conditions is challenging and at most approximate. However, the identification of strongly co-regulated motifs offered an indication of which motifs retain their activity under changing conditions.

Specifically, by analyzing the transcriptional responses to environmental perturbations we found that co-regulated FNM were rare (Figure 5A). In fact, less than 1% of FNMs were what can be considered very strongly co-regulated, and about 5% as strongly co-regulated (Figure 5B). Moreover, while smaller motifs of size $k = 3$ naturally had higher fraction of strongly co-regulated FNMs, co-regulation was else not biased by motif size (Figure 5C).

Some of the co-regulated FNMs were interlinked and clustered into well-characterized mod-

ules, for example the core RNA splicing machinery that was found systematically co-regulated in response to several stress conditions (Figure 5D). But the small individual FNMs highlighted the more interesting connections. For instance, the nuclear and exosome-associated RNA binding protein MPP6, involved in pre-mRNA surveillance, linked through PPIs to both AIR2, an RNA-binding protein of the TRAMP nuclear RNA surveillance complex, and the non-catalytic exosome core component RRP4 (Figure 5E). Moreover, RRP4 was interacting through negative GIs with both MPP6 and AIR2, underlining substantial cross-talk between RNA surveillance and quality control (Lisbeth-Carolina et al., 2020).

Last, many motifs likely alter their activity under changing conditions. For instance, the previously identified FNM of the stress response TF RPN4 that is linked by an E2/E3 ligase system to the metabolic enzyme DFR1 showed a distinct pattern of expression changes. RPN4 as well as the direct TF target PRE7 were strongly induced under heat stress (Figure 5F). In turn, the RPN4-facing E3 UBR2 was initially mildly down-regulated, but overall the expression of the E2 and E3 ligases did not change while DFR1 was strongly down-regulated to no longer detected (Figure 5F). It is plausible to speculate that this FNM contributes to the coupling of transcription and protein levels in adaptation to heat stress by shifting the protein degradation activity to the RAD6-UBR1 complex for the degradation of DFR1. Of note, while UBR1 is normally nuclear, it readily becomes cytosolic under stress (Breker et al., 2013) where it could target DFR1 and other proteins to maintain cellular homeostasis. FNMs are no definitive blueprint of cellular regulation, but offer a most promising glimpse into putative regulatory interactions from already available genomic data that warrant further investigation.

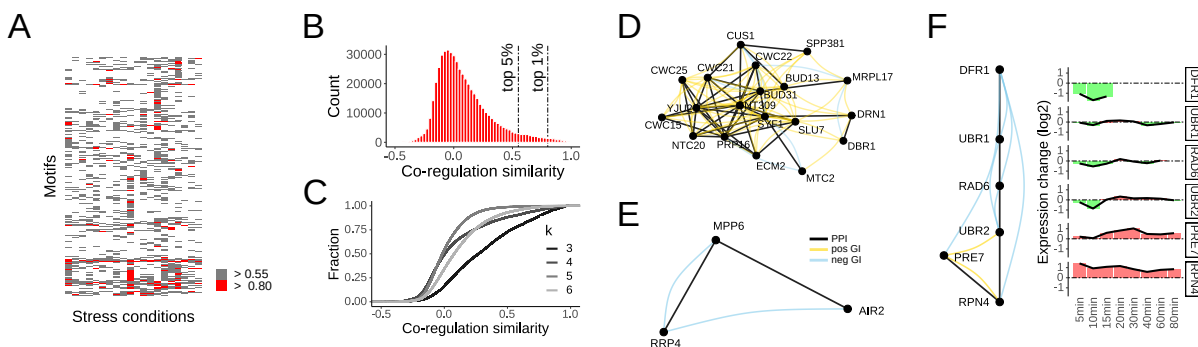


Figure 5 Transcriptional co-regulation of FNMs in response to environmental perturbations. **(A)** Heatmap of co-regulation expressed through average pairwise cosine similarity of the expression profiles of the individual nodes in the motif. The top 5% and 1% are color coded. Rows of motifs with no co-regulation in at least the top 5% are not shown. **(B)** Distribution of the FNM co-regulation scores. The top 5% and 1% are indicated in the far right tail of the distribution. **(C)** Cumulative distribution curves of the co-regulation scores as function of motif size k . **(D)** An exemplary cluster of motifs that is strongly transcriptionally co-regulated during heat stress identifies the core splicing machinery. **(E)** An exemplary FNM that is strongly co-regulated during heat stress highlights extensive cross-talk between RNA surveillance and quality control. **(F)** An FNM that connects the transcription regulatory network to the metabolic network may indicate a regulatory function of the E2-E3 complexes formed by RAD6, UBR1 and UBR2 in coupling transcript and protein levels during heat stress.

DISCUSSION

Network motifs have revolutionized our understanding of feedback circuits and the organization of cellular networks. However, solely defined through statistical over-representation they often lack biological context. Here, we have defined functional network motifs, or FNMs, through the integration of genetic interaction data that directly informs on functional relationships between genes and encoded proteins. By capturing known and novel regulatory interactions, FNMs have been found a very promising strategy towards the systematic identification of feedback and cross-talk in cellular regulation. Importantly, the description as 'functional' is intended as categorical rather than definite. Any notion of biological function itself is strongly context-dependent and subject to debate (Keeling et al., 2019). More opportunities await for further improving any definition of FNMs.

Ultimately, causation can only be established through targeted perturbations. Large-scale perturbation studies (McIsaac et al., 2012; Hackett et al., 2016; Caldera et al., 2019; Hackett et al., 2020) offer a promising outlook. However, even in absence of control perturbations, additional information, especially the directionality of interactions (Vinayagam et al., 2014), can be very informative as it allows to explore functional parameter ranges through the modeling of motif dynamics. As more genomic data are generated under non-standard laboratory conditions, it will certainly become easier to dissect the intricacies of cellular regulation.

Already now, FNMs offer many interesting insights and a strategy through the motif-based integration of omics data that is readily extendable to additional data types. Support for this rationale is given by the GIs observed in the FNMs. Similar to other genome-scale measurables, GIs are inherently noisy. Consequently, GIs are routinely analyzed through correlations of their genomic interaction profiles rather than as individual interactions (VanderSluis et al., 2018). Through small clusters of GIs between functionally related genes, the FNMs achieve more

confidence than would be offered by single GIs but are less generic than genome-wide profiles.

To this end, regulatory events in the cell extend far beyond physical interactions between macromolecules as captured by PPI networks. Signaling can involve feedback through temporal (Harrigan et al., 2018) and spatial organization (Santos et al., 2012), the dynamic activation and deactivation through modification (Hirano et al., 2016), and even the folding of proteins and RNA (Rutherford and Zuker, 1994). Moreover, some systems are truly determined by collective behavior that can only be understood at a systems level. For instance, the cellular protein folding capacity balances the proteome with available quality control pathways (Draceni and Pechmann, 2019) wherein the competition for the shared folding capacity is governed through differential interaction specificities (Pechmann, 2000). Central to many aspects of cell integrity, protein homeostasis is a particularly important example where integrative approaches have been very promising with a lot of open opportunity (Rizzolo et al., 2017).

An orthogonal approach to the discovery of critical regulators follows recent advances in control and systems theory to identify the controlling nodes of networks from their topology (Liu et al., 2011; Zanudo Tejada et al., 2017) and even dynamics (Baggio et al., 2021). Such avenues currently remain challenging for genome-scale networks, predominantly due to strong assumptions as well as insufficient and incomplete data, but will most certainly continue to gain in importance. Even stronger poised will be the synthesis of these ideas from genomics, systems biology, and control. Due to their simplicity, FNMs can immediately contribute to the systematic identification of cross-talk and regulatory interactions. Iterative cycles of improvement will continue to refine the prioritization of genes for follow-up characterization (Kuang et al., 2020), decipher complex regulatory logic (Buchler et al., 2003), and support the re-engineering of critical protein sense-response systems (Glasgow et al., 2020).

METHODS

Data and code availability

Project data and computer code generated for this study are available at

<https://www.github.com/pechmannlab/FNM>.

Data sources

The *S. cerevisiae* protein interaction network was obtained from BIOGRID release 3.5.185 (Oughtred et al., 2019) and filtered for physical protein interactions. Yeast genetic interactions were retrieved from (Costanzo et al., 2016). The top and bottom 5% of all pairwise interaction scores were considered significant and used to build a genetic interaction network. The set of curated suppressor interactions from (van Leeuwen et al., 2016) was used as benchmark example of functionally related genes. We further used the annotations of yeast protein complexes from the MIPS database (Guldener et al., 2006), the consensus transcription factors from YEASTRACT (Teixeira et al., 2017; Monteiro et al., 2020), and the metabolic network model Yeast8 (Lu et al., 2019). Co-expression in response to environmental perturbations was assessed with the expression data from (Gasch et al., 2000).

Enumeration of FNMs

Network motifs were generated through exhaustive enumeration of graphlets in the *S. cerevisiae* protein-protein interaction (PPI) network. Our approach followed the algorithm developed by (Kashani et al., 2009) wherein each motif class of numbers of nodes per layer, defined through the integer compositions of the motif size k , was enumerated by standard depth-first-search. We considered a graphlet, i.e. a motif in the PPI network, as a 'functional network motif' (FNM) if at least 50% of all possible non-self genetic interaction edges within the graphlet were present, and the source node had direct genetic interactions with all nodes in the most distant layer. All

motifs of sizes $k = 3, 4, 5, 6$ were generated.

Two deliberate constraints served to improve both computational tractability and biological interpretability. Proteins that interact with hundreds or thousands of other proteins play important roles in the organization of cellular networks but rarely exhibit strong specificity and selectivity. To identify functional network motifs that may be involved in specific feedback circuits rather than network hubs, we only considered proteins of degree $d_{max} < 50$, while thresholds of $d_{max} < 25$ and $d_{max} < 100$ were also tested. Importantly, the excluding of the most highly connected network hubs made the exhaustive motif enumeration in the yeast PPI computationally feasible up to a motif size of $k = 6$. Moreover, we only considered interactions between protein complexes but not between subunits of the same protein complexes. The rationale for this was that we wanted to identify putative feedback links that connect biological function rather than understand the architecture of protein complexes that often act as functional unit. Network randomizations were performed with the R package *BiRewire* (Iorio et al., 2016). Motifs were generated for randomized PPI networks, and randomized genetic interaction networks. 30 randomizations were found to yield sufficiently converging reference values (Figure S1B).

Analysis of FNMs

Motifs were clustered by iteratively identifying and merging maximally overlapping individual motifs. A complex-based motif interaction network was generated by considering all yeast genes/proteins from the FNMs and merging protein complex subunit genes into joint nodes. Genes encoding proteins that are part of more than one protein complex were omitted. Transcriptional co-regulation of genes from the same motif was quantified through their average pairwise cosine similarity between the time-course expression changes in response to environmental perturbations (Gasch et al., 2000). All visualization was done in R.

ACKNOWLEDGMENTS

This research was in part supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada (06504-2016) and the Canada Research Chair in Computational Systems Biology.

DISCLOSURE DECLARATION

The authors declare no competing interests.

References

- Alon U. 2007. Network motifs: theory and experimental approaches. *Nat Rev Genet* **8**(6):450–61.
- Baggio G, Bassett DS, and Pasqualetti F. 2021. Data-driven control of complex networks. *Nat Comms* **12**(1):1429.
- Barabasi A. 2004. NETWORK BIOLOGY: UNDERSTANDING THE CELL'S FUNCTIONAL ORGANIZATION. *Net Rev Genet* **5**:101–114.
- Beltrao P, Cagney G, and Krogan NJ. 2010. Quantitative genetic interactions reveal biological modularity. *Cell* **141**(5):739–745.
- Bork P, Jensen LJ, von Mering C, Ramani AK, Lee I, and Marcotte EM. 2004. Protein interaction networks from yeast to human. *Curr Opin Struct Biol* **14**(3):292–299.
- Boyle EA, Li YI, and Pritchard JK. 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**(7):1177–1186.
- Brandman O, Ferrell JJ, Li R, and T. M. 2005. Interlinked fast and slow positive feedback loops drive reliable cell decisions. *Science* **310**(5747):496–8.
- Breker M, Gymrek M, Moldavski O, and Schuldiner M. 2013. LoQAtE Localization and Quantitation ATlas of the yeast proteome. A new tool for multiparametric dissection of single-protein behaviour in response to biological perturbations in yeast. *Nucl Acids Res* **42**(D1):D726–D730.
- Buchler NE, Gerland U, and Hwa T. 2003. On schemes of combinatorial transcription logic. *Proc Acad Natl Sci* **100**(9):5136–5141.

- Caldera M, Müller F, Kaltenbrunner I, Licciardello MP, Lardeau CH, Kubicek S, and Menche J. 2019. Mapping the perturbome network of cellular perturbations. *Nat Comms* **10**:5140.
- Chechik G, Oh E, Rando O, Weissman J, Regev A, and Koller D. 2008. Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. *Nat Biotechnol* **26**(11):1251–9.
- Cornish AJ and Markowitz F. 2014. SANTA: Quantifying the Functional Content of Molecular Networks. *PLoS Comput Biol* **10**(9):e1003808.
- Costanzo M, Kuzmin E, van Leeuwen J, Mair B, Moffat J, Boone C, and Andrews B. 2019. Global Genetic Network and the Genotype-to-Phenotype Relationship. *Cell* **177**:85–97.
- Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, Wang W, Usaj M, Hanchard J, Lee SD, *et al.*. 2016. A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**(6306):aaf1420–1.
- Crona K, Gavryushkin A, Greene D, and Beerenwinkel N. 2017. Inferring genetic interactions from comparative fitness data. *eLife* **6**:e28629.
- de Lichtenberg U, Jensen LJ, Brunak S, and Bork P. 2005. Dynamic complex formation during the yeast cell cycle. *Science* **307**(5710):724–727.
- Doyle J and Csete M. 2005. Motifs, Control, and Stability. *PLoS Biol* **3**(11):e392.
- Draceni Y and Pechmann S. 2019. Pervasive convergent evolution and extreme phenotypes define chaperone requirements of protein homeostasis. *Proc Natl Acad Sci* **116**:20009–20014.
- Elowitz M and Leibler S. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**(6767):335–8.

- Fang G, Wang W, Paunic V, Heydari H, Costanzo M, Liu X, Liu X, VanderSluis B, Oatley B, Steinbach M, *et al.*. 2019. Discovering genetic interactions bridging pathways in genome-wide association studies. *Nat Comms* **10**:4272.
- Fortuna MA, Zaman L, Ofria C, and Wagner A. 2017. The genotype-phenotype map of an evolving digital organism. *PLoS Comput Biol* **13**(2):e1005414.
- Freeman M. 2000. Feedback control of intercellular signalling in development. *Nature* **408**(6810):313–9.
- Gasch A, Spellman P, Kao C, Carmel-Harel O, Eisen M, Storz G, Botstein D, and Brown P. 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* **11**(12):4241–57.
- Gavin A, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen L, Bastuck S, Dümpelfeld B, *et al.*. 2006. Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**(7084):631–636.
- Glasgow AA, Huang YM, Mandell DJ, Thompson M, Ritterson R, Loshbaugh AL, Pellegrino J, Krivacic C, Pache RA, Barlow KA, *et al.*. 2020. Computational design of a modular protein sense-response system. *Science* **366**(6468):1024–1028.
- Goentoro L, Shoval O, Kirschner MW, and Alon U. 2009. The Incoherent Feedforward Loop Can Provide Fold-Change Detection in Gene Regulation. *Mol Cell* **36**:894–899.
- Grüning NM, Lehrach H, and Ralser M. 2010. Regulatory crosstalk of the metabolic network. *Trends Biochem Sci* **35**(4):220–227.
- Güldener U, Münsterkötter M, Oesterheld M, Pagel P, Ruepp A, Mewes HW, and Stümflen V. 2006. MPact: the MIPS protein interaction resource on yeast. *Nucl Acids Res* **34**:D436–41.

Hackett SR, Baltz EA, Coram M, Wranik BJ, Kim G, Baker A, Fan M, Hendrickson DG, Berndl M, and McIsaac RS, *et al.*. 2020. Learning causal networks using inducible transcription factors and transcriptome-wide time series. *Mol Syst Biol* **16**:e9174.

Hackett SR, Zanutelli VRT, Xu W, Goya J, Park JO, Perlman DH, Gibney PA, Botstein D, Storey JD, and Rabinowitz JD, *et al.*. 2016. Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* **354**(6311):aaf2786.

Hahn JS, Neef DW, and Thiele DJ. 2006. A stress regulatory network for co-ordinated activation of proteasome expression mediated by yeast heat shock transcription factor. *Mol Microbiol* **60**(1):240–251.

Han JDJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJM, Cusick ME, Roth FP, *et al.*. 2004. Evidence for Dynamically Organized Modularity in the Yeast Protein-Protein Interaction Network. *Nature* **430**(6995):88–93.

Harrigan P, Madhani HD, and El-Samad H. 2018. Real-Time Genetic Compensation Defines the Dynamic Demands of Feedback Control. *Cell* **175**(3):877–886.

Hirano A, Fu YH, and Ptacek LJ. 2016. The intricate dance of post-translational modifications in the rhythm of life. *Nat Struct Mol Biol* **23**(12):1053–1060.

Huynen MA, Snel B, von Mering C, and Bork P. 2003. Function prediction and protein networks. *Curr Opin Cell Biol* **15**(2):191–198.

Iorio F, Bernardo-Faura M, Gobbi A, Cokelaer T, Jurman G, and Saez-Rodriguez J. 2016. Efficient randomization of biological networks while preserving functional characterization of individual nodes. *BMC Bioinformatics* **17**:542.

- Jiménez B, Ugwu F, Zhao R, Ortí L, Makhnevych T, Pineda-Lucena A, and Houry WA. 2012. Structure of minimal tetratricopeptide repeat domain protein Tah1 reveals mechanism of its interaction with Pih1 and Hsp90. *J Biol Chem* **287**(8):5698–709.
- Jothi R, Balaji S, Wuster A, Grochow JA, Gsponer J, Przytycka TM, Aravind L, and Babu MM. 2009. Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture. *Mol Syst Biol* **5**:294.
- Kashani ZRM, Ahrabian H, Elahi E, Nowzari-Dalini A, Ansari ES, Asadi S, Mohammadi S, Schreiber F, and Masoudi-Nejad A. 2009. Kavosh: a new algorithm for finding network motifs. *BMC Bioinformatics* **10**:318.
- Kashtan N and Alon U. 2005. Spontaneous evolution of modularity and network motifs. *Proc Natl Acad Sci* **102**(39):13773–13778.
- Kashtan N, Itzkovitz S, Milo R, and Alon U. 2004. Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics* **20**(11):1746–1758.
- Keeling DM, Garza P, Nartey CM, and Carvunis AR. 2019. The meanings of 'function' in biology and the problematic case of de novo gene emergence. *eLife* **8**:e47014.
- Kim D, Kim MS, and Cho KH. 2012. The core regulation module of stress-responsive regulatory networks in yeast. *Nucl Acids Res* **40**(18):8793–8802.
- Konagurthy AS and Lesk AM. 2008. On the origin of distribution patterns of motifs in biological networks. *BMC Syst Biol* **2**:73.
- Krogan N, Cagney G, Yu H, Zhong G, Guo x, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis A, et al.. 2006. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**(7084):637–634.

- Kuang D, Truty R, Weile J, Johnson B, Nykamp K, Araya C, Nussbaum RL, and Roth FP. 2020. Prioritizing genes for systematic variant effect mapping. *Bioinformatics* **36**(22-23):5448–5455.
- Kuzmin E, VanderSluis B, Wan W, Tan G, Deshpande R, Chen Y, Usaj M, Balint A, Usaj MM, van Leeuwen J, *et al.*. 2018. Systematic analysis of complex genetic interactions. *Science* **360**:eaao1729.
- Lehner B. 2013. Genotype to phenotype: lessons from model organisms for human genetics. *Nat Rev Genet* **14**(3):168–178.
- Lisbeth-Carolina A, Paul B, Reiter T, Gendron L, Rajan AAN, Montpetit R, Trahan C, Pechmann S, Oeffinger M, and Montpetit B, *et al.*. 2020. Altered rRNA processing disrupts nuclear RNA homeostasis via competition for the poly(A)-binding protein Nab2. *Nucl Acids Res* **48**(20):11675–11694.
- Liu YY, Slotine JJ, and Barabási AL. 2011. Controllability of complex networks. *Nature* **473**(7346):167–73.
- Lu H, Li F, Sanchez B, Zhu Z, Li G, Domenzain I, Marcisaukas S, Anton P, D L, Lieven C, *et al.*. 2019. A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat Commun* **10**(1):3586.
- Mani R, St Onge RP, Hartman 4th JL, Giaever G, and Roth FP. 2008. Defining Genetic Interactions. *Proc Natl Acad Sci* **105**(9):3461–6.
- Markowetz F, Kostka D, Troyanskaya OG, and Spang R. 2007. Nested effects models for high-dimensional phenotyping screens. *Bioinformatics* **23**:i305–i312.

- Mazurie A, Bottani S, and Vergassola M. 2005. An evolutionary and functional assessment of regulatory network motifs. *Genome Biol* **6**:R35.
- McIsaac RS, Petti AA, Bussemaker HJ, and Botstein D. 2012. Perturbation-based analysis and modeling of combinatorial regulation in the yeast sulfur assimilation pathway. *Mol Biol Cell* **23**:2993–3007.
- Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, Ayzenshtat I, Sheffer M, and Alon U. 2004. Superfamilies of Evolved and Designed Networks. *Science* **303**(5663):1538–42.
- Monteiro PT, Oliveira J, Pais P, Antunes M, Palma M, Cavalheiro M, Galocha M, Godinho CP, Martins LC, Bourbon N, *et al.*. 2020. YEASTRACT+: a portal for cross-species comparative genomics of transcription regulation in yeasts. *Nucl Acids Res* **48**:D642–D649.
- Oughtred R, Stark C, Breitkreutz BJ, Rust J, Boucher L, Chang C, Kolas N, O’Donnell L, Leung G, McAdam R, *et al.*. 2019. The bioGRID interaction database: 2019 update. *Nucl Acids Res* **47**(D1):D529–D541.
- Park Y and Bader JS. 2011. Resolving the structure of interactomes with hierarchical agglomerative clustering. *BMC Bioinformatics* **12**:S44.
- Pechmann S. 2000. Programmed trade-offs in protein folding networks. *Structure* **28**:1361–1375.
- Pedrajas J, Miranda-Vizueté A, Javanmardy N, Gustafsson J, and Spyrou G. 2000. Mitochondria of *Saccharomyces cerevisiae* contain one conserved cysteine type peroxiredoxin with thioredoxin peroxidase activity. *J Biol Chem* **275**(21):16296–301.
- Pe’er D, Regev A, Elidan G, and Friedman N. 2001. Inferring subnetworks from perturbed expression profiles. *Bioinformatics* **17**:S215–S224.

- Pirkl M, Diekmann M, van der Wees M, Beerenwinkel N, Fröhlich H, and Markowetz F. 2017. Inferring modulators of genetic interactions with epistatic nested effect models. *PLoS Comput Biol* **13**(4):e1005496.
- Prill RJ, Iglesias PA, and Levchenko A. 2005. Dynamic Properties of Network Motifs Contribute to Biological Network Organization. *PLoS Biol* **3**(11):e343.
- Przulj N. 2006. Biological network comparison using graphlet degree distribution. *Bioinformatics* **23**:e177–e183.
- Ravasz E, Somera A, Mongru D, Oltvai Z, and Barabasi A. 2002. Hierarchical Organization of Modularity in Metabolic Networks. *Science* **297**:1551–1555.
- Rizzolo K, Huen J, Kumar A, Phanse S, Vlasblom J, Kakihara Y, Zeineddine HA, Minic Z, Snidder J, Wang W, *et al.*. 2017. Features of the Chaperone Cellular Network Revealed through Systematic Interaction Mapping. *Cell Rep* **20**:2735–2748.
- Rual J, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz G, Gibbons F, Dreze M, Ayivi-Guedehoussou N, *et al.*. 2005. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**:1173–1178.
- Rutherford S and Zuker C. 1994. Protein folding and the regulation of signaling pathways. *Cell* **79**(7):1129–1132.
- Santos SD, Wollman R, Meyer T, and Ferrell Jr JE. 2012. Spatial Positive Feedback at the Onset of Mitosis. *Cell* **149**(7):1500–1513.
- Schlossmann J, Lill R, Neupert W, and Court D. 1996. Tom71, a novel homologue of the mitochondrial preprotein receptor Tom70. *J Biol Chem* **271**(30):17890–5.

- Schreiber F and Schwöbbermeyer. 2005. MAVisto: a tool for the exploration of network motifs. *Bioinformatics* **21**(17):3572–3574.
- Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, Punna T, Ihmels J, Andrews B, Boone C, Greenblatt JF, *et al.*. 2005. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**(3):507–519.
- Song Q, Grene R, Heath LS, and Li S. 2017. Identification of regulatory modules in genome scale transcription regulatory networks. *BMC Systems Biology* **11**:140.
- Tarassov K, Messier V, Landry CR, Radinovic S, Serna Molina MM, Shames I, Malitskaya Y, Vogel J, Bussey H, and Michnick SW, *et al.*. 2008. An in Vivo Map of the Yeast Protein Interactome. *Science* **320**(5882):1465–1470.
- Teixeira MC, Monteiro PT, Palma M, Costa C, Godinho CP, Pais P, Cavalheiro M, Antunes M, Lemos A, Pedreira T, *et al.*. 2017. YEASTRACT: an upgraded database for the analysis of transcription regulatory networks in *Saccharomyces cerevisiae*. *Nucl Acids Res* **46**:D348–D353.
- Uzarska MA, Veronica N, Weiler BD, Spantgar F, Ciofi-Baffoni S, Saviello MR, Leonardo G, Mühlhoff U, Banci L, and Lill R, *et al.*. 2016. Mitochondrial Bol1 and Bol3 function as assembly factors for specific iron-sulfur proteins. *eLife* **5**:e16673.
- van Leeuwen J, Pons C, Mellor JC, Yamaguchi TN, Friesen H, Koschwanez J, Usaj MM, Pechlaner M, Takar M, Usaj M, *et al.*. 2016. Exploring genetic suppression interactions on a global scale. *Science* **354**(6312):aag0839.
- VanderSluis B, Costanzo M, Billmann M, Ward HN, Myers CL, Andrews BJ, and Boone C. 2018. Integrating genetic and protein-protein interaction networks maps a functional wiring diagram of a cell. *Curr Opin Microbiol* **45**:170–179.

- Vázquez A, Dobrin R, Sergi D, Eckmann JP, Oltvai ZN, and Barabási AL. 2004. The topological relationship between the large-scale attributes and local interaction patterns of complex networks. *Proc Acad Natl Sci* **101**(52):17940–17945.
- Vinayagam A, Zirin J, Roesel C, Hu Y, Yilmazel B, Samsonova AA, Neumüller RA, Mohr SE, and Perrimon N. 2014. Integrating protein-protein interaction networks with phenotypes reveals signs of interactions. *Nat Methods* **11**(1):94–99.
- von Mering C, Zdobnov EM, Tsoka S, Ciccarelli FD, Pereira-Leal JB, Ouzounis CA, and Bork P. 2003. Genome evolution reveals biochemical networks and functional modules. *Proc Natl Acad Sci* **100**(26):15428–15433.
- Wang X, Castro MA, Mulder KW, and Markowetz F. 2012. Posterior Association Networks and Functional Modules Inferred from Rich Phenotypes of Gene Perturbations. *PLoS Comput Biol* **8**(6):e1002566.
- Wang YC and Chen BS. 2010. Integrated cellular network of transcription regulations and protein-protein interactions. *BMC Systems Biology* **4**:20.
- Watanabe K, Stringer S, Frei O, Mirkov MU, de Leeuw C, Polderman TJC, van der Sluis S, Andreassen OA, Neale BM, and Posthuma D, *et al.*. 2019. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* **51**(9):1339–1348.
- Wernicke S and Rasche F. 2006. FANMOD: a tool for fast network motif detection. *Bioinformatics* **22**(9):1152–1153.
- Wolf YI, Katsnelson MI, and Koonin EV. 2018. Physical foundations of biological complexity. *Proc Acad Natl Sci* **115**(37):E8678–E8687.

Wolfger H, Mamnun YM, and Kuchler K. 2004. The yeast Pdr15p ATP-binding cassette (ABC) protein is a general stress response factor implicated in cellular detoxification. *J Biol Chem* **279**(12):11593–9.

Wray NR, Wijmenga C, Sullivan PF, Yang J, and Visscher PM. 2018. Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. *Cell* **173**(7):1573–1580.

Yu H, Braun P, Yildirim M, 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**(5898):104–110.

Zanudo Tejada JG, Yang G, and Albert R. 2017. Structure-based control of complex networks with nonlinear dynamics. *Proc Acad Natl Sci* **114**(28):7234–7239.

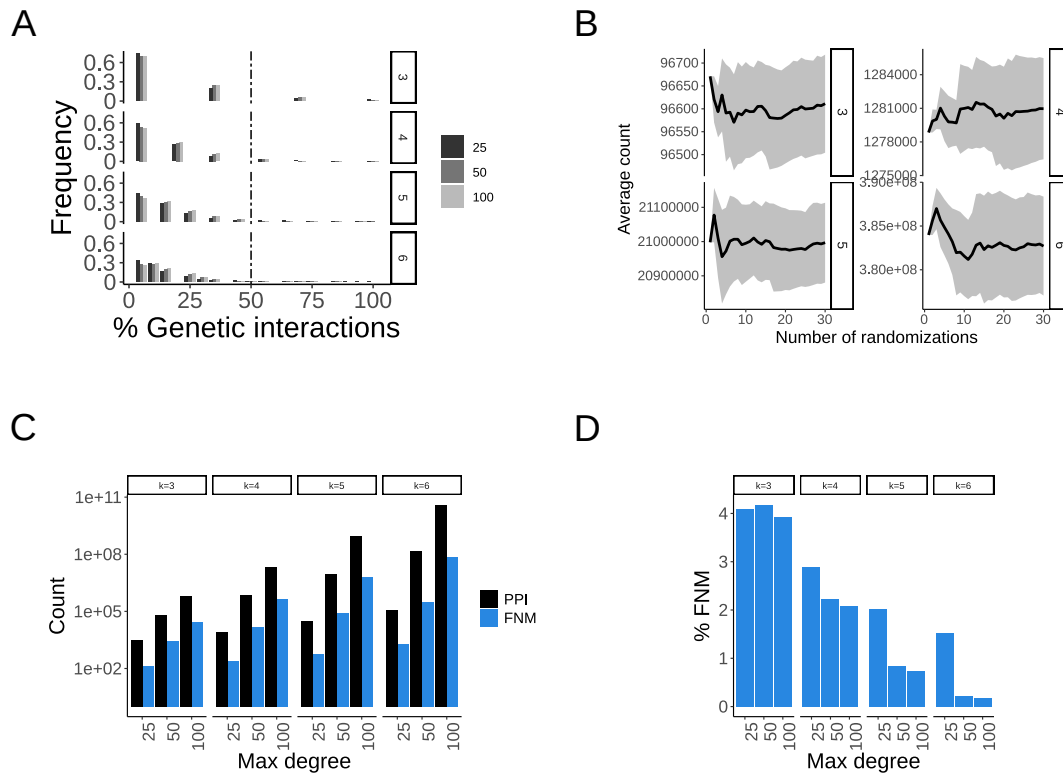


Figure S1 Definition and analysis of FNMs. **(A)** Fractions of GIs observed in classical network motifs as function of motif size k and the maximum degree of the initial PPI network. **(B)** Expected counts of classical network motifs obtained from randomized networks as function of the number of randomizations and motif size k . Shown are mean (black line) and standard deviation (grey shaded area). Randomizations are computationally costly and 30 randomizations appear sufficient for stable estimates. **(C)** Counts of FNMs and classical network motifs as function of motif size k and maximum degree in the PPI network. **(D)** Percentages of FNMs relative to classical network motifs as function of motif size k and maximum degree in the PPI network.