Molecular Omics

ROYAL SOCIETY
OF CHEMISTRY

RESEARCH ARTICLE
Mohan Babu, Walid A. Houry *et al.*
Systems analysis of the genetic interaction network of yeast molecular
chaperones

Indexed in
Medline!

Systems analysis of the genetic interaction network of yeast molecular chaperones†

Kamran Rizzolo,‡[a] Ashwani Kumar,‡[b] Yoshito Kakihara,§[a] Sadhna Phanse,[c] Zoran Minic,[c] Jamie Snider,[d] Igor Stagljar,[ade] Sandra Zilles,[b] Mohan Babu ORCID *[c] and Walid A. Houry ORCID *[af]

Molecular chaperones are typically promiscuous interacting proteins that function globally in the cell to maintain protein homeostasis. Recently, we had carried out experiments that elucidated a comprehensive interaction network for the core 67 chaperones and 15 cochaperones in the budding yeast *Saccharomyces cerevisiae* [Rizzolo *et al.*, *Cell Rep.*, 2017, **20**, 2735–2748]. Here, the genetic (*i.e.* epistatic) interaction network obtained for chaperones was further analyzed, revealing that the global topological parameters of the resulting network have a more central role in mediating interactions in comparison to the rest of the proteins in the cell. Most notably, we observed Hsp10, Hsp70 Ssz1 chaperone, and Hsp90 cochaperone Cdc37 to be the main drivers of the network architecture. Systematic analysis on the physicochemical properties for all chaperone interactors further revealed the presence of preferential domains and folds that are highly interactive with chaperones such as the WD40 repeat domain. Further analysis with established cellular complexes revealed the involvement of R2TP chaperone in quaternary structure formation. Our results thus provide a global overview of the chaperone network properties in yeast, expanding our understanding of their functional diversity and their role in protein homeostasis.

Introduction

Molecular chaperones are a highly interactive group of proteins that are primarily involved in cellular stress responses.[1,2] They are typically tightly regulated to rapidly sense physiological stresses that eventually engage their chaperone function, for example, in protein folding, holding, and maturation of substrate clients in different cellular compartments.[2] Chaperones that are conserved across many organisms are classified into families based on their functional similarities. In the budding

yeast *Saccharomyces cerevisiae*, there are a total of 67 chaperones: 2 Hsp90s, 14 Hsp70s, 22 Hsp40s, 8 CCTs, 1 Hsp60, 1 Hsp10, 6 prefoldins, 5 ATPases associated with diverse cellular activities (AAA+), 7 small heat shock proteins (sHsps), and 1 calnexin. In addition, the Hsp70 and Hsp90 families contain 4 and 11 partner proteins, respectively, termed cochaperones. Here, we refer to chaperones and cochaperones as CCos. Many biochemical reductionist studies have been performed to understand individual CCo function and their interaction with clients, yet such studies do not provide the global view of the functions performed by the CCo network.

Recently, we published a global systematic network of the CCos in yeast using two types of high-throughput data:[3] (1) genetic interactions (GIs; when mutations in two or more genes combine to generate an unexpected phenotype) obtained from synthetic genetic array (SGA) technology,[4] and (2) protein–protein interactions (PPIs; physical interactions between two or more proteins) obtained from tandem-affinity purification followed by mass spectrometry meta-analysis from four different large-scale studies.[1,5–7] In that work, 22 443 GIs were obtained with 13 704 negative GIs (*i.e.* double mutants with a more severe fitness defect than the expected multiplicative effect of combining the individual mutants, with the extreme case being synthetic lethality) and 8739 positive GIs (*i.e.* double mutants with a less severe defect in fitness than expected). In the case of PPIs, we compiled a total of 43 020 interactions.[3]

[a] *Department of Biochemistry, University of Toronto, Toronto, Ontario M5G 1M1, Canada. E-mail: walid.houry@utoronto.ca*

[b] *Department of Computer Science, University of Regina, Regina, Saskatchewan S4S 0A2, Canada*

[c] *Department of Biochemistry, Research and Innovation Centre, University of Regina, Regina, Saskatchewan S4S 0A2, Canada. E-mail: mohan.babu@uregina.ca*

[d] *The Donnelly Centre, University of Toronto, Toronto, Ontario M5S 3E1, Canada*

[e] *Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 3E1, Canada*

[f] *Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7mo00142h

‡ Co-first authors.

§ Present address: Division of Dental Pharmacology, Niigata University Graduate School of Medical and Dental Sciences, 2-5274 Gakkocho-dori, Chuo-ku, Japan.

Network concepts are typically used to understand the properties of biological networks in the context of single protein–protein or multi-protein relationships[8–10] and to gain insights into important biochemical functions[11] based on the network modularity. These can be generally achieved by analyzing the network's topological properties from the large amount of biological data that are being accumulated by virtue of large-scale experiments and computational predictions. Since biological networks are known to be scale-free and unevenly organized, their topological properties are very powerful in deciphering functional modules[12] and in establishing global relationships between the groups of interacting components in the network. For example, in the case of chaperones, topological studies performed using PPI networks revealed their central role as highly-connected proteins (*i.e.* hubs) in the cell.[13,14] Such is the case of Hsp90 in tumor cells where its connectivity with a highly rewired proteome has made it an important drug target for cancer treatments.[15] Similarly, the quantitative proteomic analysis of Hsp70 chaperones *SSA1* and *SSB1* deletion mutant strains showed that the cellular protein concentrations are mostly unchanged as a result of the chaperone network structure.[16]

Despite the aforesaid studies revealing topological roles of individual or subgroups of CCos in a cellular network, a more complete analysis encompassing all CCos has not yet been explored. Here, we have addressed this by analyzing the CCo GI network to provide a statistical description of its topology and to identify key CCo and non-CCo proteins in the network. This analysis is important given that more than double the deciphered interactions in yeast come from GIs as opposed to physical interaction studies.[17] Our findings revealed that CCos are critical to shaping the whole yeast genome GI network and have a pronounced association with protein complexes throughout the cell likely to modulate their assemblies.

Materials and methods

Topological properties of the whole genome and CCo GI networks

To characterize the topological properties of the interaction network, the following parameters were calculated:

Degree centrality (DC) is defined as the fraction of connections a node v has in the network.

Betweenness centrality (BC) of a node v is the summation of the portion of all-pairs shortest paths that go through v.

$$C_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

where V is the group of nodes, $\sigma(s,t)$ is the number of shortest (s,t)-paths, and $\sigma(s,t|v)$ is the total paths passing through some node v other than s and t.

The *closeness centrality* (CC) for a node u in the network is defined as the reciprocal of the average shortest path to u over all other $(n-1)$ nodes.

$$C(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)}$$

where n is the number of vertices that can be reached by u whereas $d(v,u)$ is the shortest path between u and v. A higher closeness value produces a higher centrality value of a node in the network.

The *eigenvector centrality* measure is defined as follows. Let $A = (a_{i,j})$ be the adjacency matrix of the network. The eigenvector centrality e_v of node v is:

$$e_v = \frac{1}{\lambda} \sum_k a_{k,v} e_k$$

where eigen value $\lambda \neq 0$ is a constant. In matrix form:

$$\lambda e = eA$$

The *normalized centrality* parameter is used to obtain an unbiased combined centrality score of each node. In Fig. 2, we calculated a normalized centrality score (NCS) by using degree, betweenness and closeness centrality measures. NCS of a node v is:

$$NCS_v = \frac{((DC_v/DC_{max}) + (BC_v/BC_{max}) + (CC_v/CC_{max}))}{3}$$

where DC_v, BC_v, and CC_v are degree, betweenness, and closeness centrality values of node v, respectively, whereas max represents the highest score of a centrality measure in the network.

As a control for each centrality score, we bootstrapped (1000 times) a random sample set of non-CCos of the same size as the CCo set and performed Wilcoxon–Mann–Whitney test to calculate a P-value for each iteration compared to the CCo set. The distribution of P-values between the randomly selected non-CCo set and the CCos for each centrality score are plotted in Fig. S1 (ESI†). To reduce the influence of outliers in the P-value distributions, we highlight the median values for each plot as they better reflect the overall P-value for each centrality score.

Analyses of non-CCo enrichment in the CCo network and protein complex enrichment in the *SEC63* GIs

To determine non-CCos with a significantly high number of interactions in the CCo GI network, we calculated a hypergeometric distribution:

$$P(K \geq k) = \sum_{K=k}^{n} h(K;N,m,n) = \sum_{K=k}^{n} \frac{\binom{m}{K}\binom{N-m}{n-K}}{\binom{N}{n}}$$

where k = number of interactions with CCo in the network; n = number of CCos in the network; m = number of interactions in the network; N = total number of genes in the network.

To obtain CYC2014 complexes[18] that are significantly enriched with *SEC63* interactions from the negative GI, positive GI and GI profile correlation similarity (GIPC; Table S2, ESI†) datasets, we used the same formula as described above with the following variables: k = number of interactions with a CYC2014 complex in the *SEC63* GI dataset; n = number of members in a CYC2014 complex; m = total number of *SEC63* interactions in a particular dataset; N = total number of genes in the particular GI dataset.

Physicochemical properties, protein abundance and protein domain enrichment analyses

Six properties were gathered for all proteins (CCo and non-CCo) in the GIPC, negative GI, and positive GI datasets: (1) molecular weight, (2) isoelectric point (pI), (3) hydrophobicity score,[19] (4) instability index[20] and (5) aliphatic index,[21] and (6) protein abundance in the cell.[22] Protein domain enrichments were performed by gathering all domains of CCo and non-CCo proteins from the Pfam database[23] followed by selection of the top most enriched ($P \leq 0.05$) domains. Hierarchical clustering of enriched domains and spectral count scores from mass spectrometry (MS) data was performed using the uncentered Pearson correlation metric and the centroid linkage method from the Cluster 3.0 program.[24] Generation of the heat maps for enriched domains was performed in the Java Treeview program.[25] CCo gene distribution for each dataset, significant non-CCo connectivity in the CCo network, interactor physicochemical property distribution plots and bar-plots of protein complex analyses were performed in the program JMP®, Version 9 (SAS Institute Inc., Cary, NC).

To determine the number of interactions between CCo and complexes in the CYC2014 catalog,[18] we calculated the Jaccard index for CCo_i and CYC_j. We then summed the Jaccard indices over all complexes for Fig. 6 or over all CCos for Fig. 9:

Sum of Jaccard indices

$$= \sum_{\text{all } i, \text{all } j} \left(\frac{\text{Int}(i,j)}{\text{Total Int CCo}_i + \text{Mem CYC}_j - \text{Int}(i,j)} \right)$$

where $\text{Int}(i,j)$ = number of common interactions between CCo_i and CYC_j; total IntCCo_i = total interactions of CCo_i; Mem CYC_j = total number of subunit members of CYC_j.

Pulldowns and mass spectrometry

To carry out the pulldown experiments described in Fig. 7B, the following strains were generated:

YK199 (S288C background, *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 CBF5-FLAG::KANMX*), YK200 (S288C background, *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 NOP10-FLAG::KANMX*), YK201 (S288C background, *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 GAR1-FLAG::KANMX*), YK202 (S288C background, *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 NHP2-FLAG::KANMX*). The following strain was obtained from Open Biosystems BY4741 (S288C background, *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*).

Yeast cells containing specific ORFs having a C-terminal 3×FLAG-tag were grown in YPD medium to log phase (OD600 ≈ 0.6) at 30 °C. Cell pellets were frozen and then lysed using a coffee grinder cooled with dry ice. Protein complexes were purified using FLAG antibody resin (A2220; Sigma-Aldrich) according to established protocols.[26] The isolated proteins were digested with trypsin and then subjected to an Orbitrap mass spectrometer. The resulting MS/MS spectra were searched against the protein coding sequences of the derivative strain using the SEQUEST search engine. High-confidence matches were then evaluated using spectral counts and probability scores generated by the STATQUEST algorithm. In order to quantify the MS data, we used the normalized spectral abundance factor (NSAF) method as previously described,[27] which utilizes spectral counts normalized to protein length and the total spectral count in each experiment.

Membrane yeast two-hybrid (MYTH) for Sec63

MYTH mapping of Sec63 interactors was carried out as described before.[28] The yeast MYTH reporter strain THY.AP4 expressing an endogenously tagged Sec63 encoding the C-terminal fragment of ubiquitin (Cub) fused to an artificial LexA-VP16 transcription factor was used. Reporter cells expressing the tagged bait were screened using a high-efficiency transformation protocol against the cDNA and the genomic prey libraries, which are expressing proteins that are fused to the N-terminal fragment of ubiquitin (NubG). Subsequently, cells that are expressing both the bait and prey proteins were identified by growth selection for the MYTH reporter system on nutrient-deficient media. Selected plasmid DNA construct expressing prey interactors were amplified and identified by sequencing. These candidate preys were then retested in MYTH for interaction with Sec63 along with the vacuolar ABC transporter Adp1, which is an unrelated control bait protein. Sec63-specific interacting prey proteins were selected as those that again interacted with Sec63 but did not interact with Adp1. The list of specific prey proteins interacting with Sec63 are given in Table S1 (ESI†).

Results

Topological architecture of the CCos in the whole yeast genome GI network

To find distinct features of the CCos in the GI network,[4] we compared the topology of CCos against all non-CCo yeast genes (Fig. 1). We used four well-established centrality measures: degree, eigenvector, betweenness, and closeness (see Methods for details). The degree centrality of a node is typically calculated on the basis of number of links a node has in the network. The eigenvector centrality is an extension of degree centrality but differs by not awarding the same weight to all connections. It is based on the notion that a node is important if it is linked to other important nodes. The betweenness centrality is calculated by counting the number of times a node acts as a bridge along the shortest paths between two other nodes in the network. Lastly, a node's closeness centrality is defined as the average shortest distance between a given node and all other nodes in the network.

As shown in Fig. 1, CCo genes have significantly higher centrality scores for all four parameters ($P < 0.01$). In addition, as a control, we bootstrapped (1000 times) using a randomly selected sample sets of non-CCos having the same size as the CCo set. The different centrality scores were calculated for each randomly sampled dataset with respect to CCo (see Methods), resulting in P-value based distributions (Fig. S1, ESI†). The median P-values obtained were between 0.029 and 0.074 (Fig. S1, ESI†), suggesting that the centrality scores for the
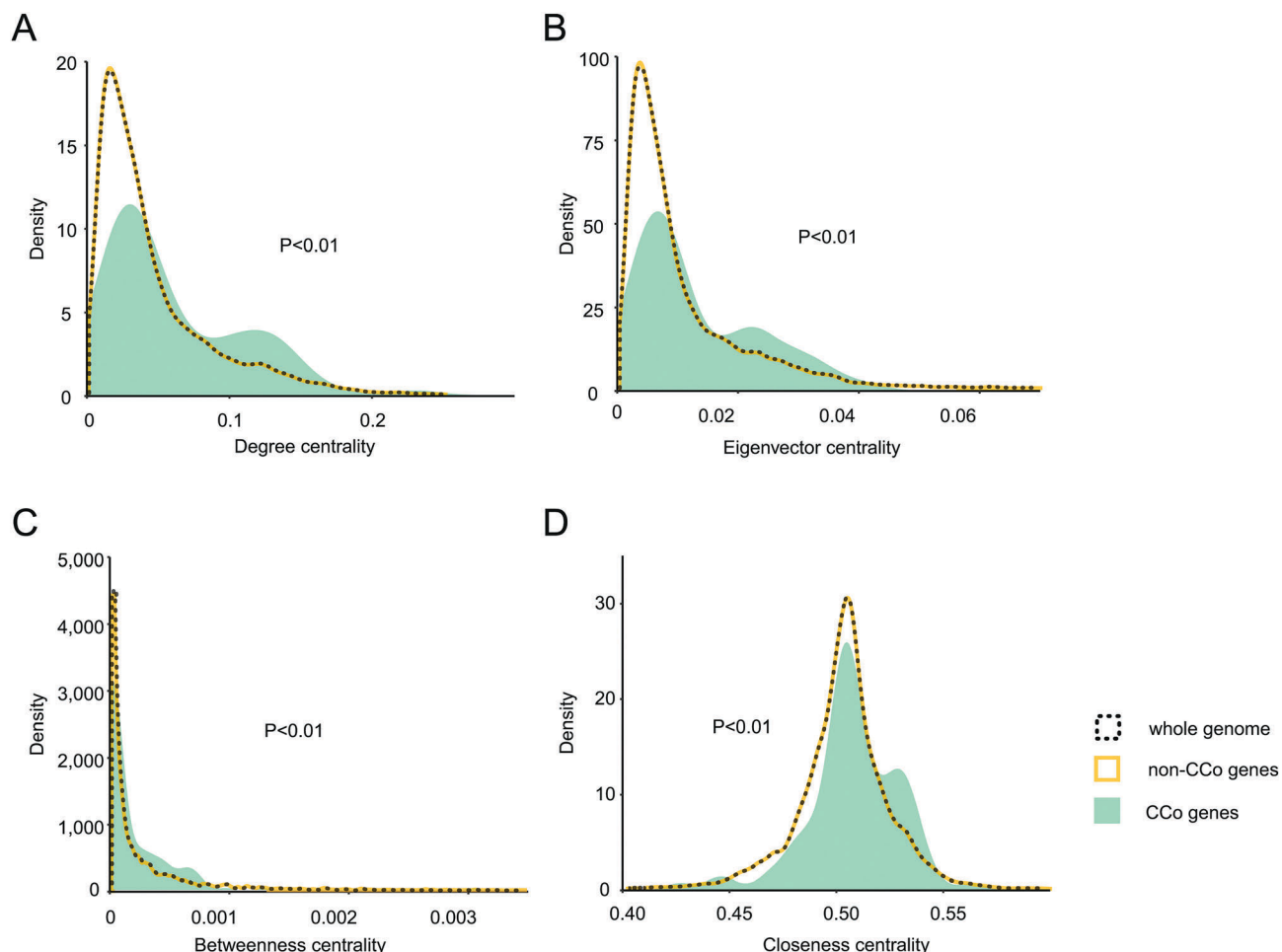
Fig. 1 Topology of the CCos in the whole yeast genome GI network. Distribution of the CCo topology parameters compared to all genes and non-CCo genes in the whole yeast genome GI network for: (A) degree centrality, (B) eigenvector centrality, (C) betweenness centrality, and (D) closeness centrality parameters. CCos have higher centrality compared to the whole genome network (Wilcoxon–Mann–Whitney test to calculate *P*-value).

CCo dataset are not random. Indeed, 10% of CCos ranked highest in all centrality scores parameters with the top three being *HSP10*, the Hsp90 cochaperone *CDC37*, and the Hsp70 *LHS1*. This suggests that overall, CCos have a greater influence in the network by playing a more central role in mediating interactions among numerous, less connected genes, reflecting their critical functions in the cell.

Overview of CCo centrality in the GI network

Biological networks are intrinsically inhomogeneous with certain genes having greater connectivity and are indispensable compared to others with less connections.[29] Many CCos have been found to be highly interactive among each other or with other proteins throughout the network.[1,3] To characterize the topological features driven by individual CCo gene in the GI network, we computed the degree, betweenness, closeness, and the normalized centrality measures for each individual CCo (Fig. 2). We find that the essential *HSP10* cochaperonin of the mitochondrial matrix chaperone Hsp60 (orthologue of bacterial GroEL) has the highest degree and betweenness centrality in both the positive and negative GI datasets, and the highest

closeness centrality in the positive GI dataset (Fig. 2). This makes the essential *HSP10* one of the most connected CCo hubs and, based on its high betweenness centrality, it serves as a bridge by bringing the denser subnetworks together. Interestingly, despite the fact that Hsp10 acts as the cap for Hsp60, *HSP60* has less than half of the normalized centrality value of *HSP10*. This could suggest that Hsp10 has functions independent of Hsp60.

In addition, we find the Hsp90 cochaperones *CDC37* and *CPR7*, the Hsp70 (Hsp110 subfamily) *SSE1* and *LHS1*, the Hsp70 *SSZ1*, and the prefoldin *GIM3* to also have high betweenness and closeness centralities compared to other CCos in the different datasets. This suggests that, they also connect subnetworks and are part of closely connected clusters.

The GI profile of a given gene is composed of a set of positive and negative GIs with other genes in the genome. Genes whose GI profiles correlate tend to be part of the same complex or function in similar pathways.[4] Consistent with this, we find the Hsp70 *SSZ1* with the highest degree centrality value as well as among the highest betweenness centrality values and consequently to be the most central CCo in the GI profile correlation
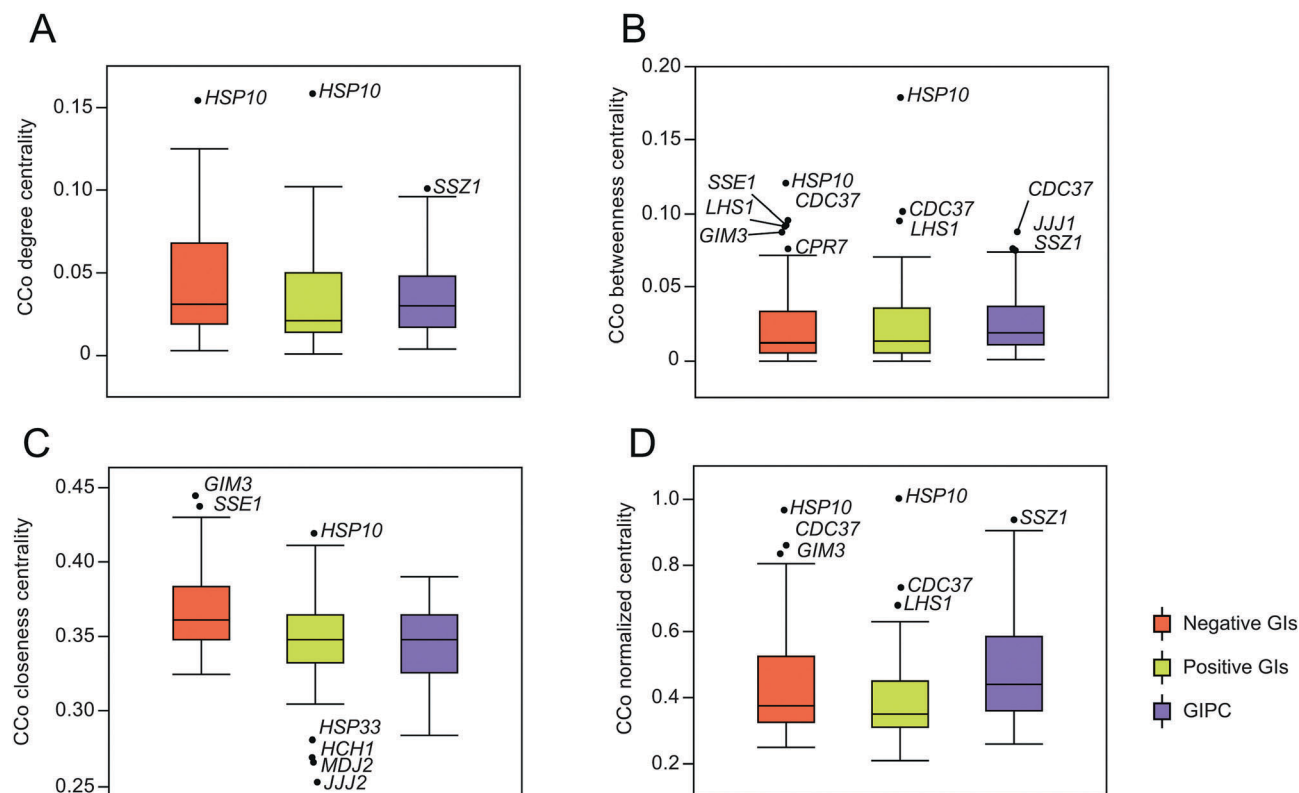
Fig. 2 Topology of the CCos in the CCo network. Boxplot distributions of: (A) degree centrality, (B) betweenness centrality, (C) closeness centrality, and (D) normalized centrality of the CCo gene's negative GI, positive GI and GIPC in the CCo network. Outlier genes are labeled.
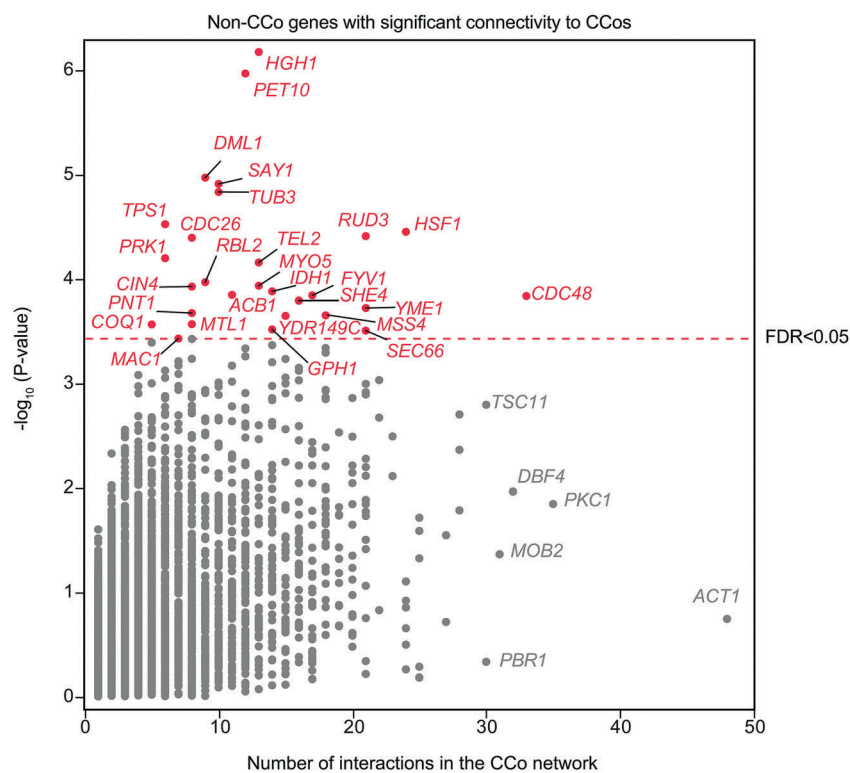

Fig. 3 Top non-CCo genes associated with CCos. Scatter-plot of non-CCo interactors enriched in the CCo network. Non-CCo genes that have a significant (FDR < 0.05) degree of connectivity are highlighted in red and are labelled. Genes with an FDR > 0.05 are colored in grey and six of those genes with the highest number of interactions in the CCo network are labelled.

similarity (GIPC) network (Fig. 2A, B and D). This chaperone is largely involved with the ribosome-associated complex (RAC), but has also been shown to be involved in various other cellular processes such as transcriptional regulation, stress response, lipid metabolism, and ER-associated degradation (ERAD).[30] In summary, *HSP10, SSZ1, CDC37, CPR7, GIM3, SSE1*, and *LHS1* are pivotal players giving shape to the CCo GI network's topology.

GI network reveals the highly associated CCo clients in yeast

To systematically identify highly-dependent CCo clients, we quantified the association of non-CCo genes with CCos by looking at their degree connectivity by GIs. Because a highly connected non-CCo gene can have many connections with both CCos as well as non-CCos, we looked for interactors that have significantly (FDR ≈ 5%) higher connectivity to the CCo network *versus* the rest of the yeast genome. This framework allowed us to identify 28 non-CCo genes (from a total of 4503 ORFs) with strong association to chaperones (Fig. 3). These include genes involved in: cytoskeleton (*TUB3, MYO5, SHE4, RBL2, MSS4, PRK1*), metabolism (*MTL1, ACB1, GPH1, TPS1, COQ1*), mitochondria (*DML1, IDH1, PNT1, YME1*), ER processes (*SEC66, SAY1, CDC48, ERV46*), poorly characterized functions (*HGH1, PET10, FYV1, YDR149C*), chromatin remodeling (*TEL2, CIN4*),
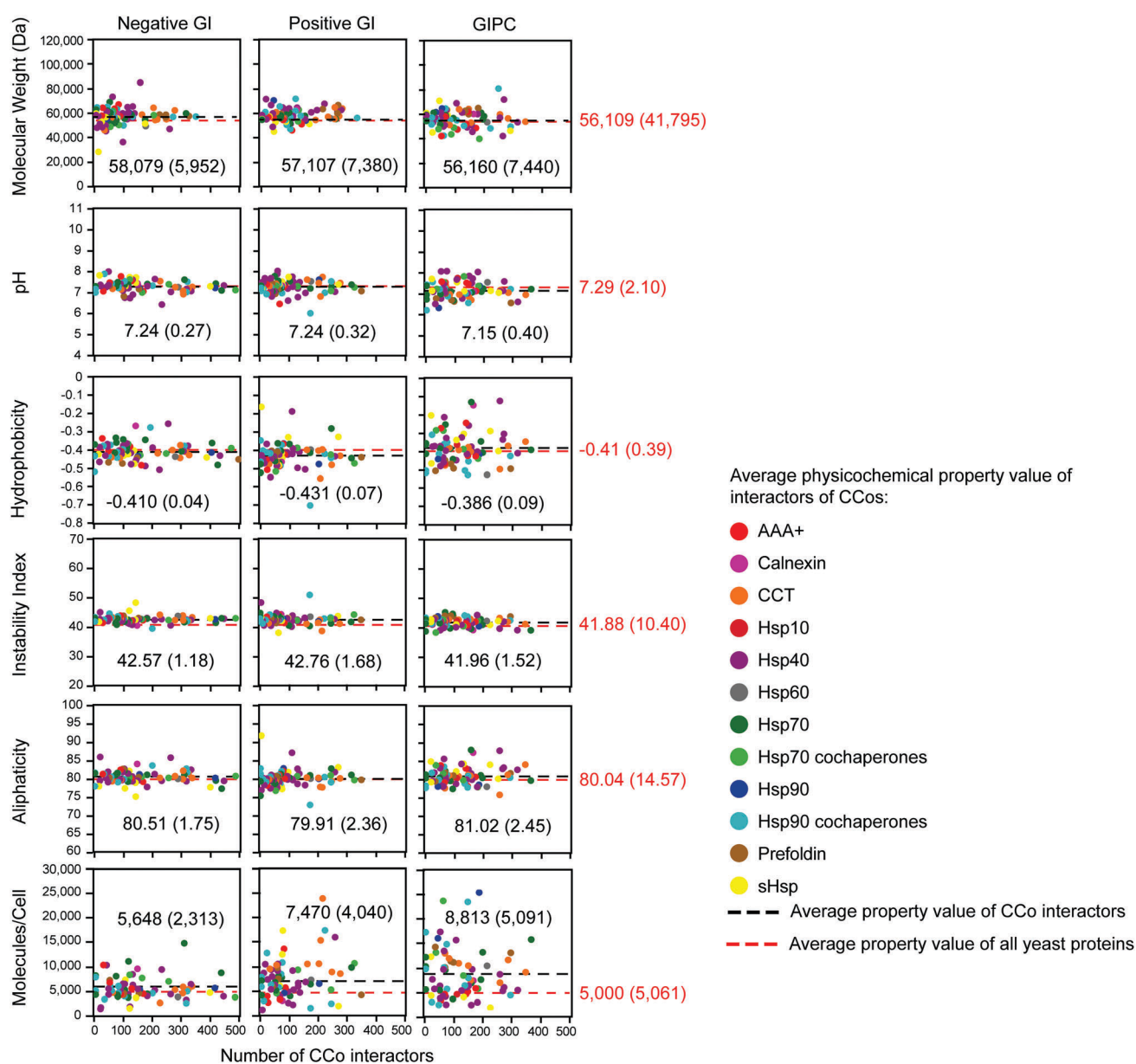

Fig. 4 Physicochemical properties of CCo interactors. Scatter plots of molecular weight, isoelectric point (pI), hydrophobicity as GRAVY score,[19] instability index,[20] aliphatic index,[21] and CCo protein interactor abundance in the cell.[22] Average property values for interactors of individual CCos from the negative GI, positive GI, and GIPC datasets are shown. CCos are color coded based on their corresponding family as indicated. Mean values are indicated and marked with a dashed line and the standard deviations are in parentheses. The average property value for all yeast proteins is also shown.

cell cycle (*CDC26*), Golgi-related (*RUD3*), and the heat-shock transcription factor (*HSF1*) processes.

Interestingly, the two most CCo-associated genes, *HGH1* and *PET10*, have putative annotated functions related to ribosomal and lipid processes, respectively. As well, *HGH1* and *PET10* have significant (PCC $\geq$ 0.1) GIPC with the Hsp90 CCo *CPR7* and the Hsp70 *SSE1*, respectively. Another interesting observation is that *CDC48*, which itself could be considered a chaperone but was not annotated as such in our analysis, is very strongly connected to many other chaperones (Fig. 3). We find *CDC48* to have significant (PCC $\geq$ 0.1) GIPC to various subunits of the proteasomal 19S regulatory particle, the lid, and 20S subunits. Indeed, the AAA+ Cdc48 is involved in a variety of degradation pathways in the cell including ERAD, inner-nuclear-membrane-associated degradation (INMAD), mitotic spindle disassembly, and ribosome-associated degradation, among others. These results suggest that Cdc48 is associated to CCo functions related to protein homeostasis.

Global view of the physicochemical properties of CCo clients

Since the negative GI, positive GI, and GIPC similarity were identified for CCo interactors, we next analyzed the physicochemical

properties of these interactors to understand the preferred properties CCos may have towards client proteins. Specifically, 6 physicochemical properties were tested, namely (Fig. 4): (1) molecular weight, (2) isoelectric point (pI), (3) hydrophobicity score,[19] (4) instability index,[20] (5) aliphatic index,[21] and (6) protein abundance in the cell.[22] In each of these properties, the average value and standard deviation for each CCo gene in all three GI datasets were computed. In addition, the average physico-chemical property values of the yeast proteome are indicated (Fig. 4). We find that CCos generally interact with genes whose encoded proteins have similar properties to the rest of the proteome. This likely shows the breadth and promiscuity of the CCos.

Overview of protein domain enrichment in CCo interactors

We performed enrichment analyses ($P \leq 0.05$) for protein domains using Pfam database[23] of CCo interactors in the three GI datasets. The domain enrichments were calculated for each individual CCo by obtaining the top domains present in the set of interactors for that CCo. Subsequently, hierarchical clustering was done to determine domains enriched for multiple CCo interactions. Only such domains that interact with at least three CCos are shown (Fig. 5).
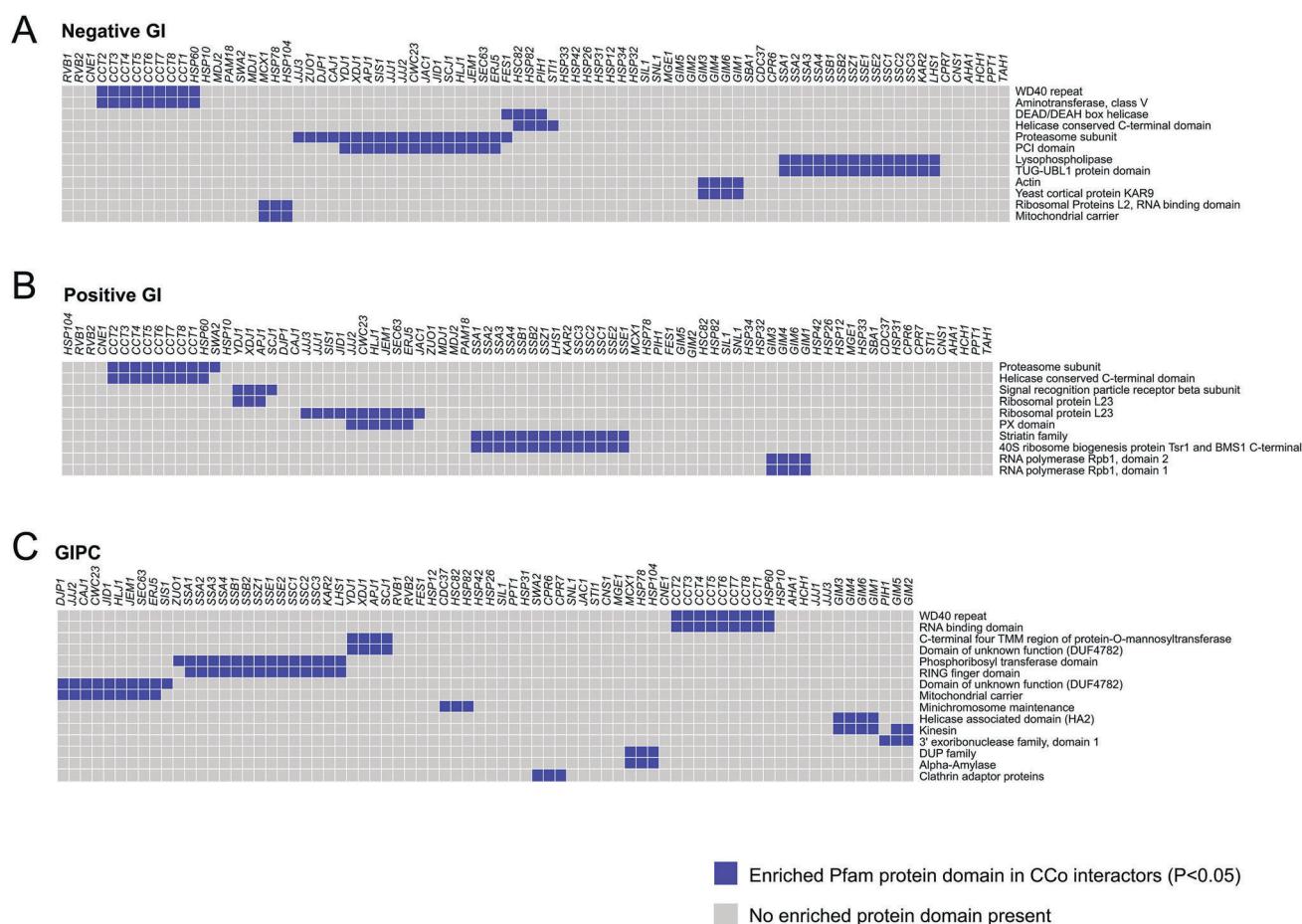

Fig. 5 Protein domains enriched in CCo interactors. Hierarchically clustered map of enriched ($P < 0.05$) Pfam protein domains of CCo interactors obtained from (A) negative GI, (B) positive GI, and (C) GIPC are shown for each CCo. Only Pfam domain clusters interacting with at least three CCos are shown.

Interestingly, CCo interactors in both the negative GI and the GIPC datasets are enriched for WD40 repeat domains that are involved in variety of functions ranging from signal transduction and transcription regulation to cell cycle control and apoptosis (Fig. 5A and C). WD40 domains have been previously observed to be enriched in physical interactions with the CCT chaperone complex.[31,32] Domains enriched ($P \leq 0.05$) in the positive GI dataset mainly consist of ribosomal, RNA polymerase-related, and proteasome subunit domains (Fig. 5B). Indeed, a strong positive GI bias has been previously observed in genes whose proteins are involved in proteostasis, including chaperones and the proteasome.[4] Furthermore, this was largely due to genetic suppression, where a fitness defect associated with a hypomorphic temperature sensitive allele of an essential gene was suppressed by a second mutation in a proteasome-encoding gene.
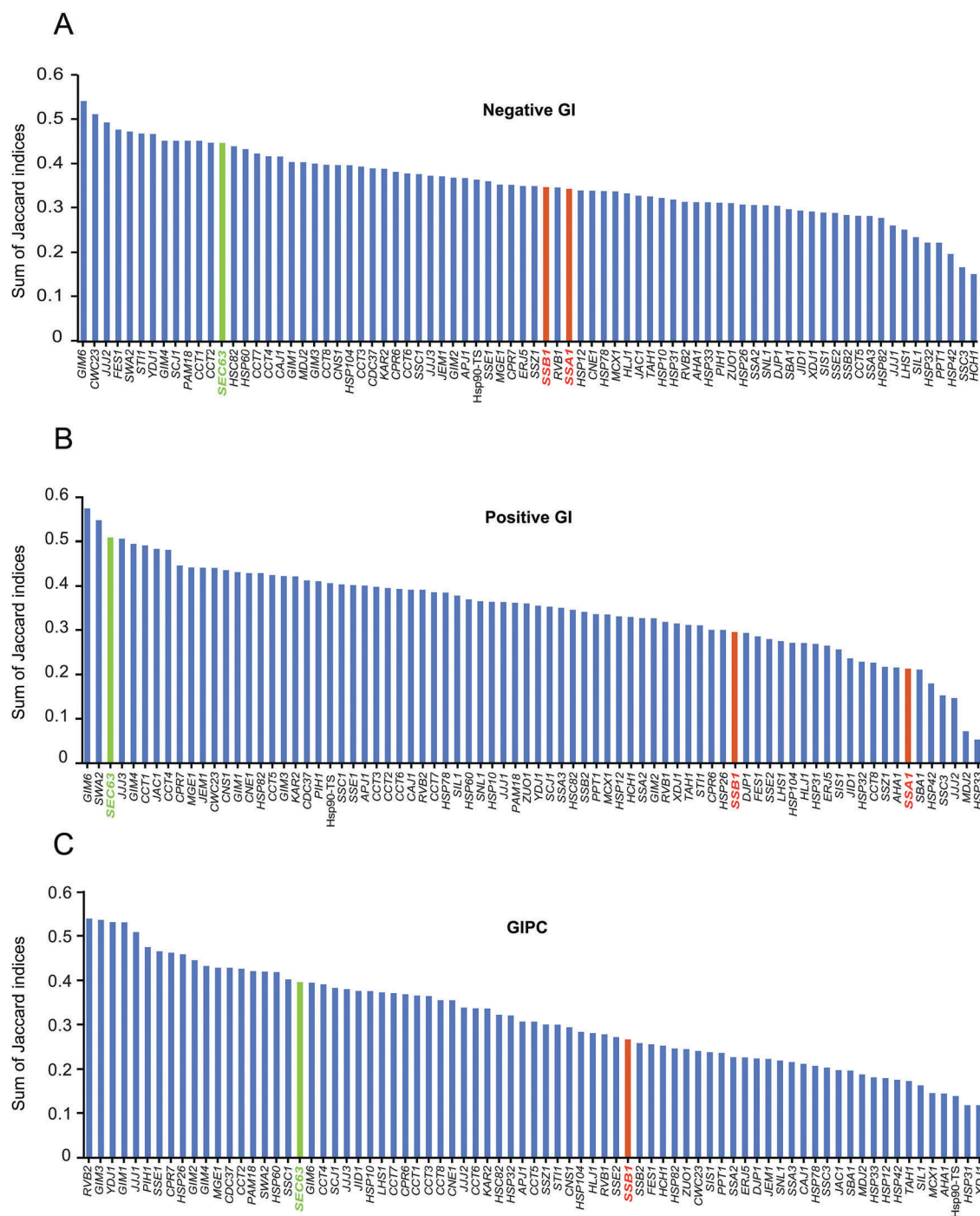

Fig. 6 Interaction of CCos with protein complexes. The sum of Jaccard indices for each CCo interaction with manually curated CYC2014 protein complexes[18] were calculated for (A) negative GI, (B) positive GI, and (C) GIPC datasets as described in the methods. *SEC63*, *SSB1* and *SSA1* are highlighted.

Role of CCos in protein complex assembly

In addition to folding proteins and promoting protein stability, several CCos have been found to facilitate protein complex assembly. To obtain a global view of what CCos might be involved in protein complex assembly, the interaction of CCos with subunits of the manually curated protein complexes given in the CYC2014 catalogue[18] were assessed as described in the methods. This was done for the negative GI, positive GI, and GIPC datasets.

As shown in Fig. 6, the different datasets resulted in different ordering of the CCos; however, CCT, the prefoldin complex and various Hsp40s typically have high numbers of interactions with the subunits of established complexes. Interestingly, for the GIPC dataset, which might more closely reflect physical interactions,[33] the AAA+ Rvb2, the prefoldins, Hsp40s Ydj1 and Jjj1, and the Hsp90 cofactor Pih1 have the highest interactors with complexes (Fig. 6C).

Rvb2 and Pih1 are members of the R2TP chaperone complex that also contains Rvb1 and Tah1 proteins and indeed has been proposed to be involved in complex assembly.[34,35]

The results shown in Fig. 6 were further verified for two specific cases. In the first illustrative example, we performed MYTH[28] using Sec63, which is an essential membrane subunit of the ER translocon containing a J domain, as a bait and also has a high number of interactions with complexes as shown in Fig. 6. A total of 49 prey proteins were obtained in MYTH (Table S1, ESI†) with 12 also found to have a significant GI with *SEC63*. These 12 proteins are members of eight different protein complexes from the CYC2014 catalog that are significantly ($P \leq 0.03$) interacting with *SEC63* (Fig. 7A and Table S2, ESI†) where one, the Ssh1 translocon complex, is present in all three datasets and the rest are distributed between the negative GI and GIPC (Fig. 7A). Among these complexes, we find Sec62/Sec63, Ssh1, and signal peptidase complexes as known interactors of Sec63, validating the protein complex interactions obtained in GIPC and negative GIs.

In the second example, we note that *SSA1* shows no complex interactions in the GIPC dataset (Fig. 6A), while *SSB1* has a relatively high number of interacting complexes. Also, *SSB1* has more complex interactors than *SSA1* in both the negative GI and positive GI datasets (Fig. 6A and B). Ssa1 is a cytoplasmic Hsp70 that is constitutively expressed and is involved in protein folding and translocation, while the Hsp70 Ssb1 is primarily a ribosome associated chaperone largely involved in nascent chain folding and assembly. Hence, generally, Ssb1 seems to be more involved in complex assembly than Ssa1. To further verify
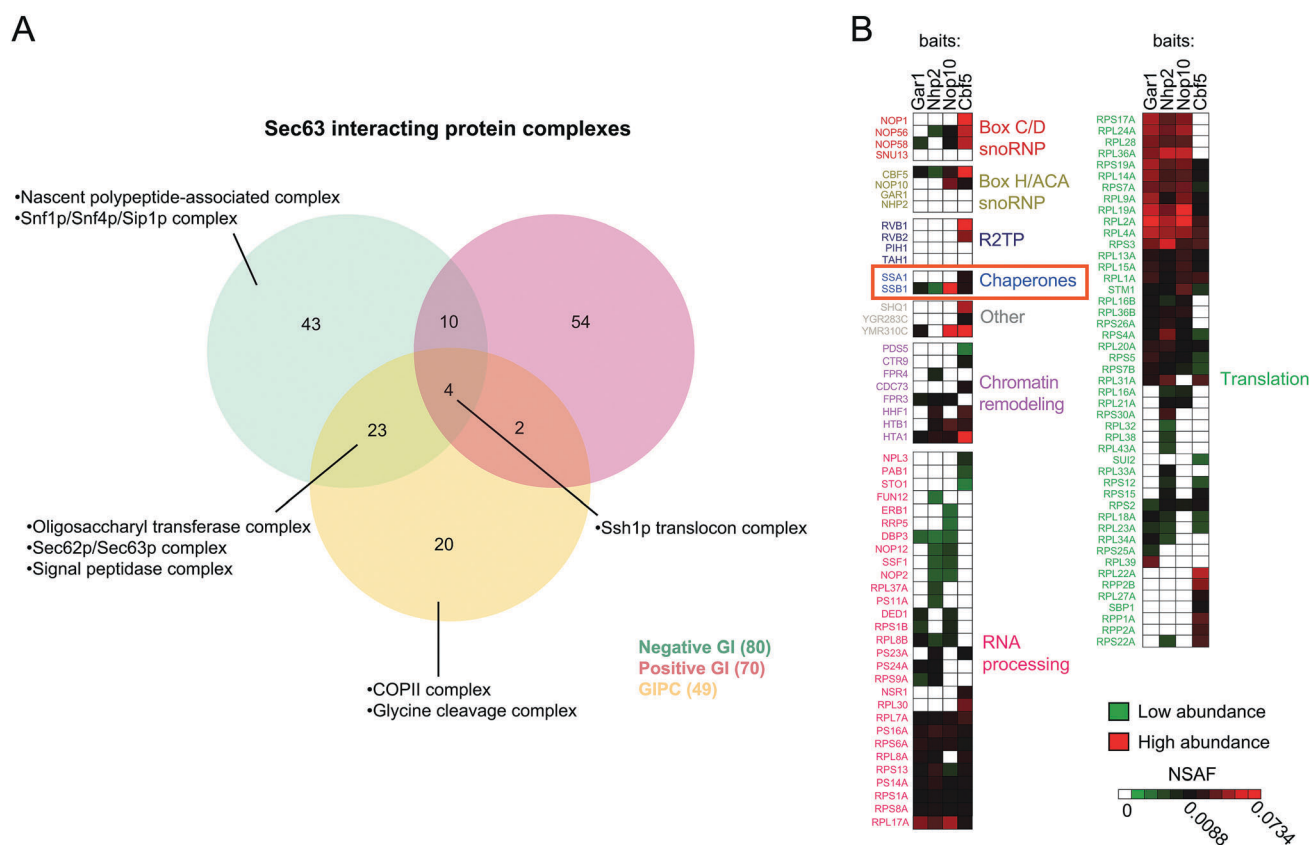

Fig. 7 Experimental validation of Sec63, Ssb1 and Ssa1 interactions with protein complexes. (A) Venn diagram showing the distribution of significant *SEC63* interacting protein complexes ($P < 0.03$) from CYC2014[18] present in the negative GIs, positive GIs and GIPC datasets. Protein complexes that contain prey proteins identified in the Sec63 MYTH are listed. (B) Proteomic discovery of interactors of endogenous C-terminally FLAG-tagged box H/ACA proteins: Gar1, Nhp2, Nop10, and Cbf5. Bound proteins were identified by mass spectrometry and the relative amounts of each protein were calculated using NSAF values.[27]

this observation, we pulled down protein components of the box H/ACA snoRNP complex, which is involved in pseudo-uridylation of pre-rRNA and is composed of a distinct H/ACA RNA and a set of four highly conserved proteins: Cbf5, Nop10, Gar1, and Nhp2.[36] The pull-down assays were carried out using strains expressing endogenously FLAG-tagged box H/ACA proteins (Fig. 7B). The results show that the majority of prey proteins are involved in RNA processing and translation pathways. Importantly, we find Ssb1 interacting with the four box H/ACA proteins, while Ssa1 interacts with only one (Cbf5; Fig. 7B). This is consistent with the result of Fig. 6.

Finally, to gain further insights into the complexes inter-acting with the R2TP complex, we plotted the complexes that are in common between *PIH1* and the other member subunits of R2TP from the GIPC dataset (Fig. 8). Pih1 is considered as the main defining subunit of R2TP.[34] We find that 52% of *PIH1* complexes interact with other members of R2TP, among these complexes are ribosomal/translation (large and small subunits, Npa2p, mRNA CFI, Noc1/2, tRNA synthase complexes), cyto-skeleton (CBF3 and RSC complexes), transcription factors (TFIII, TFIID, TFIIF complexes), chromatin remodeling (Ino80, nuclear condensin, nucleotide excision repair, capping complexes), spliceosomal (commitment complex), RNA pol II complex, signaling (TORC1 complex), mitochondrial (PAM and sorting and assembly machinery complex). Taken together, this highlights the proposed function of R2TP in the assembly of many different types of complexes.

In addition to finding that some CCos tend to preferentially interact with certain protein complexes, we looked at the complexes that are most associated with CCos. To do this, we selected the top 30 complexes highly interactive with CCos (based on Jaccard indices) from the negative GI, positive GI, and the GIPC datasets, and then ranked them from highest to lowest based on GIPC Jaccard index values (Fig. 9). Indeed, we find that the top ranked CCo-associated complexes are involved in protein synthesis (*e.g.* large and small ribosomal subunits, Kornbergs mediator SRB complex, and Rpdl3) and protein degradation (*e.g.* protea-somal 19/22S regulator complex, and 20S proteasome among others). This highlights the central role of CCos in maintaining protein homeostasis in the cell. Interestingly, despite the differences seen between negative and positive GIs in the whole yeast genome, we find CCos to interact with similar protein complexes throughout the three datasets (Fig. 9).

Discussion

Here, we performed a comparative analysis of the topological properties of CCos in the GI network and found CCos to form a unique group of genes with significant centrality in yeast (Fig. 1).
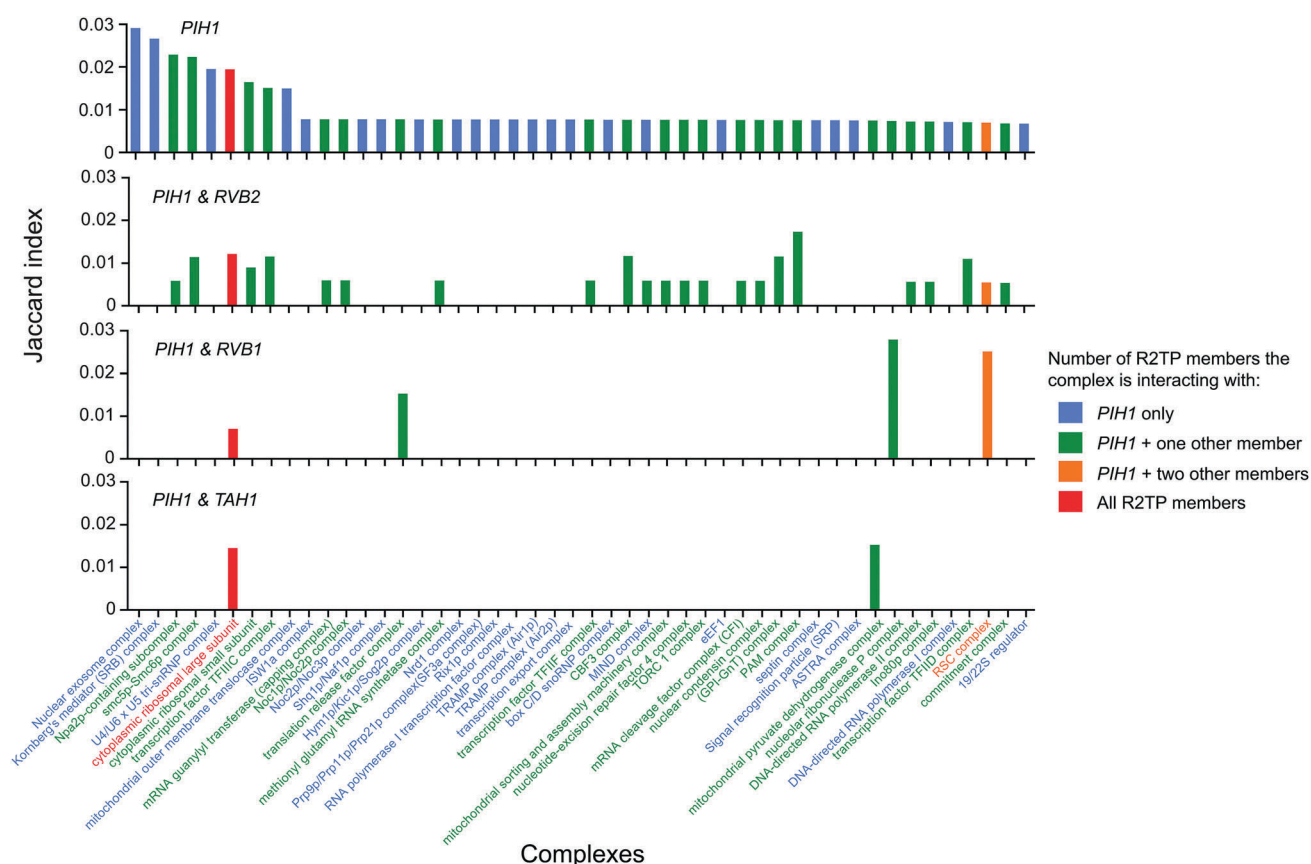

Fig. 8 Complexes interacting with R2TP proteins. Complexes interacting with *PIH1*, *PIH1 & RVB2*, *PIH1 & RVB1*, and *PIH1 & TAH1* in the GIPC dataset are listed based on Jaccard indices.
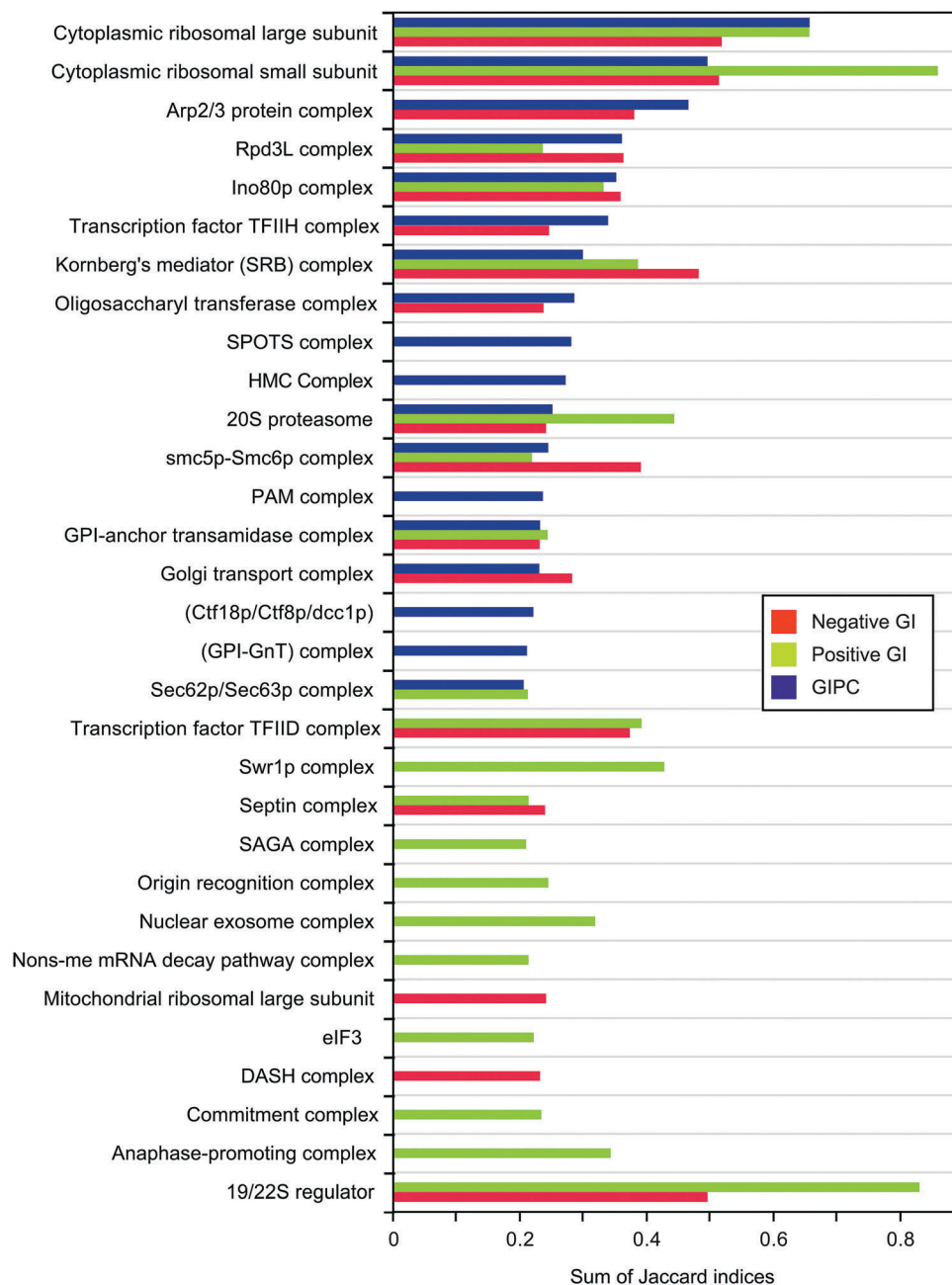
Fig. 9 CCo-associated complexes. The sum of Jaccard indices for each CYC2014 protein complex[18] with CCos were calculated for negative GI, positive GI, and GIPC datasets as described in the methods. Complexes are ranked from highest to lowest based on sum of Jaccard indices from the GIPC dataset.

Consistent with other biological networks,[9] the topological properties of the CCo network is likewise a scale-free network, which is a connectivity distribution following a power-law and is characterized by hub genes playing an important role in keeping the whole network together.[8,9] The analysis revealed that *HSP10, SSZ1, CDC37, CPR7, GIM3, SSE1*, and *LHS1* are key genes found to have the highest normalized centrality measures among all CCos in the negative and positive GIs (Fig. 2). This suggests that they are responsible for linking subnetworks or complexes within the chaperone network.

To get detailed information of the client proteins found to interact with CCos in the network, we looked at the non-CCo

genes that had a significant (FDR ≈ 5%) degree of association with CCo genes *versus* the rest of the genes (Fig. 3). 28 interacting genes were identified whose proteins have a myriad of biological functions involving cytoskeleton, metabolism, cell cycle, and mitochondrial processes. Interestingly, most of the top CCo interacting genes were found to encode for proteins with an undetermined function in particular *HGH1* and *PET10*, which have high GIPC scores with *CPR7* and *SSE1* CCos, respectively.

Our comprehensive analysis of the physicochemical properties of CCo interactors (Fig. 4) highlights the promiscuity of CCos, showing no particular bias towards any property. Furthermore,

interactor protein domain enrichment analysis (Fig. 5) highlights distinct CCo clusters, showing biased preference for WD40, actin-related and proteasomal domains. We then looked at CCo–protein-complex interactions in our GI datasets and found that, although many CCos have a GI with a known yeast complex, the GIPC dataset was a better predictor of such interactions (Fig. 6) as verified experimentally (Fig. 7). The R2TP complex was found to be one of the highest interactor of protein complexes, highlighting its putative function as a scaffolding chaperone.[35] CCos are also found to be highly associated with a variety of complexes in an unbiased manner in terms of the three GI datasets used. Taken together, by analyzing our recently generated yeast GI chaperone network,[3] this work provides a step towards understanding the overall chaperone cellular network architecture.

Conflicts of interest

There are no competing or conflicting interests for any of the authors.

Acknowledgements

References

1 Y. Gong, Y. Kakihara, N. Krogan, J. Greenblatt, A. Emili, Z. Zhang and W. A. Houry, *Mol. Syst. Biol.*, 2009, **5**, 275.

2 W. Voth and U. Jakob, *Trends Biochem. Sci.*, 2017, **42**, 899–913.

3 K. Rizzolo, J. Huen, A. Kumar, S. Phanse, J. Vlasblom, Y. Kakihara, H. A. Zeineddine, Z. Minic, J. Snider, W. Wang, C. Pons, T. V. Seraphim, E. E. Boczek, S. Alberti, M. Costanzo, C. L. Myers, I. Stagljar, C. Boone, M. Babu and W. A. Houry, *Cell Rep.*, 2017, **20**, 2735–2748.

4 M. Costanzo, B. VanderSluis, E. N. Koch, A. Baryshnikova, C. Pons, G. Tan, W. Wang, M. Usaj, J. Hanchard, S. D. Lee, V. Pelechano, E. B. Styles, M. Billmann, J. van Leeuwen, N. van Dyk, Z. Y. Lin, E. Kuzmin, J. Nelson, J. S. Piotrowski, T. Srikumar, S. Bahr, Y. Chen, R. Deshpande, C. F. Kurat, S. C. Li, Z. Li, M. M. Usaj, H. Okada, N. Pascoe, B. J. San Luis, S. Sharifpoor, E. Shuteriqi, S. W. Simpkins, J. Snider, H. G. Suresh, Y. Tan, H. Zhu, N. Malod-Dognin, V. Janjic, N. Przulj, O. G. Troyanskaya, I. Stagljar, T. Xia, Y. Ohya, A. C. Gingras, B. Raught, M. Boutros, L. M. Steinmetz, C. L. Moore, A. P. Rosebrock, A. A. Caudy, C. L. Myers, B. Andrews and C. Boone, *Science*, 2016, **353**, aaf1420–aaf1420.

5 A. C. Gavin, M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon, C. M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer and G. Superti-Furga, *Nature*, 2002, **415**, 141–147.

6 N. J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, T. Punna, J. M. Peregrin-Alvarez, M. Shales, X. Zhang, M. Davey, M. D. Robinson, A. Paccanaro, J. E. Bray, A. Sheung, B. Beattie, D. P. Richards, V. Canadien, A. Lalev, F. Mena, P. Wong, A. Starostine, M. M. Canete, J. Vlasblom, S. Wu, C. Orsi, S. R. Collins, S. Chandran, R. Haw, J. J. Rilstone, K. Gandi, N. J. Thompson, G. Musso, P. St Onge, S. Ghanny, M. H. Lam, G. Butland, A. M. Altaf-Ul, S. Kanaya, A. Shilatifard, E. O'Shea, J. S. Weissman, C. J. Ingles, T. R. Hughes, J. Parkinson, M. Gerstein, S. J. Wodak, A. Emili and J. F. Greenblatt, *Nature*, 2006, **440**, 637–643.

7 M. Babu, J. Vlasblom, S. Pu, X. Guo, C. Graham, B. D. Bean, H. E. Burston, F. J. Vizeacoumar, J. Snider, S. Phanse, V. Fong, Y. Y. Tam, M. Davey, O. Hnatshak, N. Bajaj, S. Chandran, T. Punna, C. Christopolous, V. Wong, A. Yu, G. Zhong, J. Li, I. Stagljar, E. Conibear, S. J. Wodak, A. Emili and J. F. Greenblatt, *Nature*, 2012, **489**, 585–589.

8 E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai and A. L. Barabasi, *Science*, 2002, **297**, 1551–1555.

9 A. L. Barabasi and Z. N. Oltvai, *Nat. Rev. Genet.*, 2004, **5**, 101–113.

10 J. Dong and S. Horvath, *BMC Syst. Biol.*, 2007, **1**, 24.

11 J. Yoon, A. Blumer and K. Lee, *Bioinformatics*, 2006, **22**, 3106–3108.

12 A. L. Barabasi, N. Gulbahce and J. Loscalzo, *Nat. Rev. Genet.*, 2011, **12**, 56–68.

13 C. Soti and P. Csermely, *Exp. Gerontol.*, 2007, **42**, 113–119.

14 P. C. Echeverria, A. Bernthaler, P. Dupuis, B. Mayer and D. Picard, *PLoS One*, 2011, **6**, e26044.

15 P. C. Echeverria and D. Picard, in *The Molecular Chaperones Interaction Networks in Protein Folding and Degradation*, ed. W. A. Houry, Springer New York, New York, NY, 2014, pp. 133–149, DOI: 10.1007/978-1-4939-1130-1_6.

16 A. F. Jarnuczak, C. E. Eyers, J. M. Schwartz, C. M. Grant and S. J. Hubbard, *Proteomics*, 2015, **15**, 3126–3139.

17 A. Chatr-Aryamontri, R. Oughtred, L. Boucher, J. Rust, C. Chang, N. K. Kolas, L. O'Donnell, S. Oster, C. Theesfeld, A. Sellam, C. Stark, B. J. Breitkreutz, K. Dolinski and M. Tyers, *Nucleic Acids Res.*, 2017, **45**, D369–D379.

18 S. Pu, J. Wong, B. Turner, E. Cho and S. J. Wodak, *Nucleic Acids Res.*, 2009, **37**, 825–831.

19 J. Kyte and R. F. Doolittle, *J. Mol. Biol.*, 1982, **157**, 105–132.

20 K. Guruprasad, B. V. Reddy and M. W. Pandit, *Protein Eng.*, 1990, **4**, 155–161.

21 A. Ikai, *J. Biochem.*, 1980, **88**, 1895–1898.

22 S. Ghaemmaghami, W. K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea and J. S. Weissman, *Nature*, 2003, **425**, 737–741.

23 R. D. Finn, P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry, A. L. Mitchell, S. C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G. A. Salazar, J. Tate and A. Bateman, *Nucleic Acids Res.*, 2016, **44**, D279–D285.

24 M. B. Eisen, P. T. Spellman, P. O. Brown and D. Botstein, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 14863–14868.

25 A. J. Saldanha, *Bioinformatics*, 2004, **20**, 3246–3248.

26 X. Shen, *Methods Enzymol.*, 2004, **377**, 401–412.

27 B. Zybailov, A. L. Mosley, M. E. Sardiu, M. K. Coleman, L. Florens and M. P. Washburn, *J. Proteome Res.*, 2006, **5**, 2339–2347.

28 J. Snider and I. Stagljar, *Cold Spring Harb. Protoc.*, 2016, **2016**(1), DOI: 10.1101/pdb.top077560.

29 H. Jeong, S. P. Mason, A. L. Barabasi and Z. N. Oltvai, *Nature*, 2001, **411**, 41–42.

30 J. Verghese, J. Abrams, Y. Wang and K. A. Morano, *Microbiol. Mol. Biol. Rev.*, 2012, **76**, 115–158.

31 K. Siegers, B. Bolter, J. P. Schwarz, U. M. Bottcher, S. Guha and F. U. Hartl, *EMBO J.*, 2003, **22**, 5230–5240.

32 C. Dekker, P. C. Stirling, E. A. McCormack, H. Filmore, A. Paul, R. L. Brost, M. Costanzo, C. Boone, M. R. Leroux and K. R. Willison, *EMBO J.*, 2008, **27**, 1827–1839.

33 A. H. Tong, G. Lesage, G. D. Bader, H. Ding, H. Xu, X. Xin, J. Young, G. F. Berriz, R. L. Brost, M. Chang, Y. Chen, X. Cheng, G. Chua, H. Friesen, D. S. Goldberg, J. Haynes, C. Humphries, G. He, S. Hussein, L. Ke, N. Krogan, Z. Li, J. N. Levinson, H. Lu, P. Menard, C. Munyana, A. B. Parsons, O. Ryan, R. Tonikian, T. Roberts, A. M. Sdicu, J. Shapiro, B. Sheikh, B. Suter, S. L. Wong, L. V. Zhang, H. Zhu, C. G. Burd, S. Munro, C. Sander, J. Rine, J. Greenblatt, M. Peter, A. Bretscher, G. Bell, F. P. Roth, G. W. Brown, B. Andrews, H. Bussey and C. Boone, *Science*, 2004, **303**, 808–813.

34 Y. Kakihara and W. A. Houry, *Biochim. Biophys. Acta*, 2012, **1823**, 101–107.

35 W. A. Houry, E. Bertrand and B. Coulombe, *Trends Biochem. Sci.*, 2018, **43**, 4–9.

36 J. Venema and D. Tollervey, *Annu. Rev. Genet.*, 1999, **33**, 261–311.