Visualisation and analysis of the complexome network of *Saccharomyces cerevisiae*

Simone S. Li[1], Kai Xu[2], Marc R. Wilkins[1,§]

[1] Systems Biology Initiative, School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, Australia.

[2] National ICT Australia Ltd, Australian Technology Park, Eveleigh, NSW, Australia and Interaction Design Centre, School of Engineering and Information Sciences, Middlesex University, London, UK.

[§] Corresponding author: Prof. Marc Wilkins, m.wilkins@unsw.edu.au

Abstract:

Most processes in the cell are delivered by protein complexes, rather than individual proteins. While the association of proteins has been studied extensively in protein-protein interaction networks (the interactome), an intuitive and effective representation of complex-complex connections (the complexome) is not yet available. Here, we describe a new representation of the complexome of *Saccharomyces cerevisiae*. Using the core-module-attachment data of Gavin et al. (Nature 2006, 440, 631-6), protein complexes in the network are represented as nodes; these are connected by edges that represent shared core and/or module protein subunits. To validate this network, we examined the network topology and its distribution of biological processes. The complexome network showed scale-free characteristics, with a power law-like node degree distribution and clustering coefficient independent of node degree. Connected complexes in the network showed similarities in biological process that were non-random. Further to this,

clusters of interacting complexes reflected a higher-level organisation of many cellular functions. The strong functional relationships seen in these clusters, along with literature evidence, allowed 45 uncharacterised complexes to be assigned putative functions using guilt-by-association. We demonstrate our network model using the GEOMI visualisation platform, on which we have developed capabilities to integrate and visualise complexome data.

Keywords:

Complexome, complex-complex network, *S. cerevisiae*, interaction network, GEOMI, network visualisation, interactome

Introduction:

Protein-protein interactions have been studied extensively in many species, primarily using two-hybrid techniques [1, 2, 3, 4, 5]. The resulting data have been used to construct protein-protein interaction networks. These have been the subject of detailed analysis, leading to important insights into the type and nature of protein interactions, as well as the molecular basis of biological phenomena such as disease [6] and pleiotropy [3].

In yeast, in addition to studies of protein-protein interactions by two-hybrid and other protein complementation assays, there have also been intensive studies of the 'complexome' – the entirety of protein and molecular complexes in the cell. This offers a higher order view of the proteome. Protein complexes have typically been characterised using a combination of affinity purification and tandem mass

spectrometry [7, 8]. Studies have been conducted on a large scale in *S. cerevisiae*, whereby almost all proteins in the yeast proteome have been analysed in this fashion. This has been repeated a number of times [9, 10, 11, 12]. All together, as documented in the MIPS database, the yeast complexome consists of about 875 heteromeric protein complexes [13]. This is consistent with estimates from Gavin et al. [9]. However, techniques used to detect and analyse complexes so far have favoured the discovery of stable, rather than transient interactions.

In two proteome-scale studies, Gavin et al. observed that repeated affinity purification of the same tagged protein resulted in complexes that contained different subunits [9, 10]. Deemed to be isoforms of complexes, they are likely to represent protein complexes that the cell uses under different intra- or extra-cellular conditions [9, 14]. To reflect this observation, Gavin et al. proposed a core-module-attachment architecture for protein complexes [9]. Core proteins were those present in two-thirds of the isoforms and proposed to be responsible for the basic function of the complex. The core can be supplemented by attachment proteins (found only in some isoforms), some of which may alter the function of the complex. Modules are groups of two or more attachment proteins that were found in more than one complex. Core and module proteins within the same complex typically share cellular function; have similar abundances and half-lives; and are more likely to be co-expressed and co-localised inside the cell [9, 15]. More importantly, the protein complex dataset generated by Gavin et al. [9] was the first whereby proteins could be members of multiple complexes. This is in contrast to other efforts, for example, the study by Krogan et al. [12] where proteins could only belong to one complex, due to the clustering procedure used to derive the dataset. The sharing of subunits between

protein complexes provides a potential mechanism by which complexes can be related to one another.

Studies into the graphical representation of the complexome are not as advanced as those for the interactome. Previously, we showed that protein-protein interaction networks could highlight known protein complexes through the use of alphanumeric labelling and the use of a special layout to coalesce protein subunits of a complex [16]. Other efforts to model and visualise the complexome as a network have represented protein complexes and the manner in which their subunits can be shared [12, 17, 18]. These showed all subunits within each complex, which resulted in visualisations that were dense and complicated, which was exacerbated by the extensive sharing of subunits. As a consequence, these approaches do not scale well. Hart et al. [19] showed that a complexome network can be built, whereby each node in the network is a complex. This provided some of the first high-level views of the complexome. However, there was no strategy to comprehensively understand the complexes and the significance of their connections.

Here we present an approach for the construction and analysis of the complexome network. Using the core-module-attachment model as a means to define nodes and edges for the graph, we show that a biologically relevant network can be built. The network shows properties of scale-free networks and, in this manner, is similar to protein-protein interaction networks. The biological relationships of many complexes in the eukaryotic cell are also highlighted in our network model, allowing guilt-by-association predictions for complexes of unknown function. Software, user guide and data files are available for download at http://www.systemsbiology.org.au.

Materials and methods

Data sources and processing

The dataset by Gavin et al. [9] was used in this study. It describes 491 heteromeric complexes (153 of which are of known identity) and their most likely subcellular localisation, as well as the standard gene names, complex membership and core-module-attachment classification information for 1,487 proteins. The standard names of some proteins were updated to reflect changes in the *Saccharomyces* Genome Database (SGD) [20]. Gene Ontology (GO) information was also obtained from the SGD. Protein names were from UniProtKB release 15.5 [21]. The information described above was processed and stored in a PostgreSQL relational database.

GEOMI plugins for the visualisation of the complexome network

Gavin et al. described each protein complex as being composed of core, module and attachment proteins [9]. These proteins can be members of one or more complexes. We utilised this feature to devise a visualisation strategy of the complexome as a network, whereby nodes are assigned as complexes and edges represent the proteins shared between 2 complexes. A plugin for the GEOMI platform [22] was developed to generate the network visualisations.

Assignment of biological process to complexes

There is currently no known centralised source of functional information specific to the level of the protein complex. To functionally annotate our complexome network, the biological process of each protein complex was derived from their constituent proteins. The mapping of the broader GO Slim groupings on a protein-protein interaction network has previously been demonstrated to be an effective method to

display and identify trends between connected proteins [16]. The GO Slim biological process annotations were retrieved for each protein subunit in a complex and tallied. A majority vote method was used, whereby the biological process with the highest frequency was assigned to each complex. To achieve a comprehensible visualisation, each protein complex was assigned only one biological process, for example, translation. In situations where a complex had two or more processes of highest frequency, the GO process most commonly found in its immediate neighbours in the network was chosen. If the biological process could still not be determined from this, precedence was given to the more commonly occurring process in the complexome network. The results from each of the steps described above can be examined using the 'Show Details' option in GEOMI, where users can also view the biological processes of each protein subunit (Figure 1).

To visualise the GO process of protein complexes, we developed an additional GEOMI plugin. This mapped the 10 most frequently occurring GO biological process categories onto the network using colour. The remaining categories were combined into a single 'other' group. Limiting the number of categories to be visualised was required to facilitate interpretation of the network; a smaller set of colours is more distinguishable to the human eye, and thus allows relationships between the complexes to be more easily identified.

Network analysis of the complexome

We built a new GEOMI plugin to measure topological features of a complexome network. It utilises the Java Universal Network/Graph (JUNG) library version 1.7.6 [23] to calculate network measures that are commonly used to characterise the interactome [24]. These include the (i) diameter and mean path length, (ii) degree

distribution, and (iii) clustering coefficient. Using these measures, we could determine if the complexome network was random, scale-free or hierarchical.

We also measured the extent to which interacting complexes were of the same biological process. For the CM network, we counted the number of edges that connected complexes of identical process. This was done on a pairwise basis, ignoring neighbourhood effects (above). Subsequently, we generated 6×10^6 random networks by shuffling the node colours, preserving the topology of the original network as well as node colour frequency. Edges of all randomised networks were counted, as above, allowing us to assess the significance of trends exhibited in our CM network.

Other improvements to GEOMI

A new feature in GEOMI allows networks to be generated from tab-delimited text files. Previously, networks were drawn using information in relational databases [25] or an XML-based format [16]. Search functionality is now also included in the updated version. This permits one or more proteins to be found in interaction networks by gene, locus or protein name, or other user-specified properties. In the complexome network, users can search for a particular complex by name or other identifiers. GEOMI can also highlight complexes that contain protein/s of interest using the parameters outlined above.

Edges in the networks can now also carry user-defined data, which can be accessed via the ‘Show Details’ menu. This allows information that is specific to the relationship shared by the two nodes to be easily shown. For example, in protein interaction networks, this can include evidence for a particular interaction and relevant literature references. In the complexome network, edges can show the

proteins that are shared between the two complexes and the type of protein they are (that is, core, module or attachment) (Figure 1). Moreover, labels on the edges allow additional information to be displayed on the network. Further details of these features can be found in the User Guide, which can be downloaded at http://www.systemsbiology.org.au.

Results:

Strategy for a complexome network

We devised a new strategy for the representation of the complexome as a network. To build the complexome network, protein complexes are represented as nodes. When two complexes have one or more protein subunit in common, they are connected by an edge (Figures 2A, 2B). An edge can thus represent one or many proteins. A common protein subunit between three complexes is represented by a triangle conformation, with three nodes interconnected by three edges. However, an identical representation is also used in cases where the three complexes share different protein subunits (Figures 2C, 2D).

It is important to note that our complexome network is fundamentally different from a protein-protein interaction network. A complexome network provides a higher-level view of the proteome as each node in the network is a multiprotein complex for which there is experimental evidence. Connections in a complexome network between nodes do not represent a single protein-protein interaction but the sharing of one or more proteins between complexes (Figure 3). By contrast, an interactome network, representing pairwise protein-protein interactions, does not provide strong evidence for the existence of simultaneous and cooperative interactions that are required for the formation of multiprotein complexes. The complexome network, as compared to an interactome, thus provides a complex-centric view of protein-protein interactions.

The complexome as a network of protein complexes

There were 491 protein complexes defined by Gavin et al. [9]. Of these complexes, almost all shared one or more core, module and/or attachment protein. Only 15 complexes did not share any protein subunits (3%). Using this information, we first built a network where complexes (nodes) were connected to other complexes if they shared any proteins, be they core, module or attachments. The resulting network, however, was highly interconnected and did not display any discernable features. This is contrary to what one might expect. For this entire network, each complex was connected to an average of 39.9 other complexes via a total of 9,788 edges. The network diameter was 5 and the mean shortest path was 2.3. Together, these measures reinforce the extensive interconnectivity of the network and its homogenous nature. The network of the nuclear complexome, built using this method, illustrates this (Figure 4A).

To improve the visual clarity of the complexome network, and its biological coherence, it was necessary to remove some edges from the network. We sought to understand which network connections were of high confidence and relevance and should be retained, and to identify connections of low confidence that could be excluded. Gavin et al. [9] found that attachment proteins, in comparison to core and modules, had the lowest conservation of annotated function with the complexes they were associated with. They also investigated the degree to which proteins in complexes were present in known 3-D structures, and if their interactions were also seen in two-hybrid analyses. They concluded that the evidence for core and module proteins being in physical contact was stronger than for attachment proteins. Similarly, Pang et al. showed that interactions within and between core and module

proteins had the highest proportion of domain-domain interactions, and thus have a strong structural basis [15]. Interactions involving attachment proteins showed the lowest proportion of domain-domain interactions. Together, these observations suggest that the interactions of attachment proteins are likely to be of lower confidence than those of core or module proteins, and hence not strong candidates for use as edges in complexome networks.

We built an alternative complexome network, whereby interactions between complexes were shown when complexes shared core and/or module proteins. We term this the Core-Module (CM) network (Figure 4B). Edge colours represent shared core proteins (green), shared module proteins (blue) or both (red). In contrast to the core-module-attachment network, the CM network showed a very clear structure. Each complex was connected to an average of 4.9 other complexes via a total of 992 edges. The biggest connected component included a number of areas of high and low connectivity. A number of smaller connected components were also present. A moderate proportion of complexes (19%) did not share any core or module proteins with other complexes, which is to be expected. For the sake of completeness, we generated two other types of complexome networks where nodes were connected by only shared core proteins or only shared module proteins respectively. However, these networks contained an extremely large proportion of singletons (41% and 55%, respectively) and were very sparse (data not shown). Since strong structural and functional relationships have been observed in and between core and module proteins (above), and the CM network appeared to be informative and comprehensible, we chose the CM network as the strongest model with which to represent the complexome.

The Core-Module network shows scale-free properties

We investigated the topology of the CM network. Its biggest connected component had a diameter of 15, and mean shortest path length of 5.9. Strikingly, it showed properties of scale-free networks [24, 26]. The node degree distribution followed a power law distribution, which is a key property of scale-free networks (Figure 5A). There was also a strong negative correlation between node degree *k* and P(*k*) (R^2=0.86). The CM network had a degree exponent γ = 1.87 (its biggest connected component having γ = 1.69); this is consistent with scale -free networks, which typically have degree exponents of 2 to 3 but can also exist with exponents less than 2 [24, 27]. Networks with exponents less than 2 are known to have small world properties; this was supported by the clustering coefficient of the CM network which, although showing a weak negative association between node degree *k* and clustering coefficient C(*k*) (R^2=0.24), was predominantly independent of node degree (Figure 5B). Finally, it was seen that some complexes were highly connected in the network, similar to hub and bottleneck proteins in protein-protein interaction networks; these are also characteristic of scale-free networks.

Functional relationships of connected complexes

A complexome network is expected to show connections between complexes that are of related function. An examination of complexes in the CM network, solely by name, revealed direct connections between complexes known to be related, such as RNA polymerases I, II and III (Figure 6A). The MIND and CTF19 complexes were

also connected (Figure 6B); these are both components of the kinetochore [28]. There were, however, also cases where known families of complexes were not connected in the CM network. For example, the adaptin complexes, AP-1, AP-2 and AP-3, are known to be structurally and functionally homologous [29]. However, the 3 complexes were not connected in the CM network, as they did not share any core or module proteins (data not shown). On close examination, this accurately reflected their function, as they associate with different subcellular organelles. It also accurately represented the composition and nature of their subunits, being paralogous but no longer interchangeable. They represent ‘parallel complexes’, as defined by Periera-Leal and Teichmann [30], demonstrating that the CM network can represent these intricate relationships with accuracy.

To further investigate functional relationships between connected complexes in the CM network, we mapped Gene Ontology biological process (of constituent core and module proteins) onto nodes in the network, using the method described (Figure 7). Forty-one unique categories were represented in the network; the 10 most frequent were each assigned a colour, with the remaining classifications combined into a single ‘other’ group. The most common biological processes in the network were RNA metabolic process (117 complexes), transport (97), translation (89), and ribosome biogenesis (73). This revealed that many connected complexes were involved in the same physiological process. For example, the two large, highly interconnected green clusters in the network primarily consisted of ribosomal protein complexes (Figure 7). Notably, all of the green protein complexes in the network were concentrated in these two clusters. A smaller, isolated cluster contained complexes involved in the regulation of transcription. To investigate if the functional relationships observed within the network were not due to chance, we counted the

number of interactions that occurred between complexes of identical biological process. This was 364 in our CM network. We then compared this to more than 6 x 10^6 randomised networks (see materials and methods). These networks had an average of 237 interactions between complexes of identical GO process, with a standard deviation of 17. Statistically, the CM network showed a significantly higher number of connections between nodes of identical biological process than a random network (p =1.6 x 10^{-7}) (Supplementary Figure 1). This allows us to conclude that the functional relationships in the CM network are non-random and that many complexes with common protein subunits are involved in the same biological process.

The CM network, as investigated above, clearly showed regions of functional homogeneity. However, the CM network also contained heterogeneous regions where complexes of different GO biological process classification were directly connected (Figure 7). The reasons for this are obscure but could reveal insights into the sequence of events that occur during related but not identical biological processes, reflect the reuse of protein subunits at different places and /or times within the cell, or the crosstalk between cellular processes as a means of regulating function.

Connected complexes reflect higher organisation of the cell

To further validate the CM network, we examined the detailed functional relationships that existed in regions of high interconnectivity. Interestingly, many complexes involved in related (but not identical) processes localised to common areas of the network and showed complex-complex interactions. Many of these links are supported by the literature (Table 1). In effect, these regions reflected a higher

organisation of the cell associated with the sharing of complex subunits.

One example is a highly interconnected cluster of complexes involved in RNA metabolic process within the CM network (Figure 8A). All of the known complexes within this cluster are involved in pre-mRNA splicing, but different aspects of this process. We observed high interconnectivity between the Lsm2-8, U1 snRNP, U2 snRNP and U4/U6.U5 tri-snRNP complexes, as well as the Prp19-associated complex (also known as NTC). Together, these complexes make up the complete spliceosome in the nucleus [31]. Our CM network also connected the spliceosome with the cytoplasmic mRNA processing bodies (also known as P-bodies), and in turn connected this to the mRNA decapping complex. Interestingly, these connections confirm a recently proposed model in which the P-bodies form by a combination of the decapping and LSM complexes [32]. The connectivity of U1 snRNP complex was also notable, as it shared different protein subunits with complexes within and outside of the cluster, serving as an interface. It is depictive of bottleneck hub nodes that are often found in interaction networks [33].

A second cluster of interconnected complexes of related function was also seen (Figure 8B). This included the TFIID, SAGA and mediator complexes as well as complexes associated with the modification of histones; all are basal initiation machinery of transcription. The TFIID and SAGA complexes have similar structure and function [34] and are both capable of delivering the TATA-binding protein Spt15p to the promoter. They share 14 protein subunits in the CM network. In addition, the SAGA complex is known to contain histone acetyltransferase (HAT) Tra1p subunit [35], which it shares with neighbouring NuA4 HAT complex. The mediator is also a known component of the preinitiation complex, which includes TFIID and SAGA [36]; these relationships are clear in the CM network. The connection between the NuA4 HAT

and Sin3 histone deacetylase (HDAC) complexes in the CM network is interesting as they have opposing functions. However, commonalities have been found in the binding properties and regulation of the two complexes [37, 38, 39], which could explain this observation. Interestingly, SAGA and NuA4 are HAT complexes that are known to be recruited to the promoters of active protein-coding genes, whereas HDACs such as Sin3 are targeted to specific sets of genes associated with distinct cellular functions. The functional relationships of these complexes are clear from their interconnectivity in the same cluster of the CM network.

A small group of related complexes in the CM network further supported the notion of subunit-associated high-level organisation of the cell (Figure 8C). It consists of six complexes, three of which are known to be involved in the cell cycle and cytoskeleton organisation. The actin capping protein complex and motorproteins 2 complex share Aim21p, as well as the F-actin-capping proteins Cap1p and Cap2p. The actin-associated motorproteins 1 and 2 complexes are connected through common myosin heavy chain proteins. Collectively, this would suggest that these protein complexes are involved in cytokinesis. This is further supported by the presence of the Myo5p and Cmd1p module proteins, as Cmd1p has been found to activate Myo5p-induced actin polymerisation [40]. Interestingly, the She2p/She3p/Myo4p complex, which transports mRNA transcripts to the bud tip of growing yeast [41], can also be found as a component inside complex 487.

Predicting function of uncharacterised proteins and complexes using guilt-by-association

There were many cases in the CM network where, as in the examples above, uncharacterised complexes were connected to one or more other complexes of

known function (Supplementary Table 1). In protein-protein interaction networks, 'guilt-by-association' has been effectively used to predict the function of uncharacterised proteins (for example, see [16]). The manner in which the CM network brings together complexes of related biological process, as outlined above, suggests that guilt-by-association could also be applied to predict the function of complexes in highly interconnected regions of the complexome networks.

We observed the presence of uncharacterised complexes in areas of the CM network that contained the spliceosome, basal transcriptional machinery and actin-associated protein complexes. Analysis of the constituent proteins of these complexes strongly supported the notion that they are indeed functionally related to known complexes in the same cluster. For example, in the cluster of complexes containing the spliceosome (Figure 8A), it is likely that complexes 54, 175, 418 and 436 are also involved in the splicing of pre-mRNA. The function of their constituent proteins supports this (data not shown). Similarly, uncharacterised complexes 191, 351 and 487, which are connected to actin-associated protein complexes (Figure 8C, Table 2), all contain myosin heavy and light chain proteins. This suggests that they are also involved in cytokinesis.

We applied the guilt-by-association principle to regions of high interconnectivity in the CM network. We first identified clusters of highly interconnected protein complexes (of both known and unknown function) in the CM network that were of identical biological process. Literature investigation was used to check that the known complexes within each cluster were of similar function, or involved in the same process or pathway. This resulted in the definition of 14 clusters containing a total of 49 complexes of known function (Table 1). Finally, we used guilt-by-

association, along with relevant literature, to assign potential functions to unknown complexes in clusters. A total of 45 uncharacterised protein complexes, outlined in Table 1, were assigned putative functions by this approach.

Discussion:

In this study, we have generated a new view of the yeast complexome. Represented as a network of protein complexes connected by shared protein subunits, our model differs from the established view of the interactome as a network of individual proteins joined by their interactions. Since the majority of functions in the cell are delivered by protein complexes, rather than individual proteins, we believe our network brings a unique biological context to each complex and a valuable higher-level view of the proteome.

The complexome exhibited a highly connected network topology, alluding to a level of component reuse by protein complexes. Whilst much of this connectivity was mediated by attachment proteins [9], we found that a network built through the use of shared core or module proteins resulted in a structure which was topologically consistent with other biological networks. The CM network contained regions of high and low connectivity, similar to interactome networks. Network measures also indicated that the CM network followed a power-law node degree distribution, with clustering coefficient independent of node degree. These features are consistent with, and thus reinforce, the notion that biological networks are typically scale-free [24]. This is further supported by the presence of 'hub complexes' that have a large number of connections, a feature similar to the hub proteins commonly found in protein-protein interaction networks [42].

The CM network revealed insights into the nature of inter-complex interactions. Many connections that were observed between known complexes are described in the literature, confirming the relevance of our methodology and our resulting network.

For example, a highly interconnected region in the network was found to contain all of the protein complexes of the spliceosome (Figure 8A). Specifically, the Lsm2-8 complex is known to bind to the U6 snRNP, and in fact forms its core [43]. In addition, the NTC is known to stabilise the interaction of U6 with U2 [44] and is also involved in the dissociation of the Lsm2-8 complex from U6 [45]. Notably, however, there is no known link between the nuclear Lsm2-8 complex and cytoplasmic P-bodies. We observed that the Lsm8 protein is not actually a member of the Lsm2-8 complex described by Gavin et al. [9]. Indeed, it is not present in the dataset at all. This could in part be due to the low abundance of the protein, measured at 1440 molecules/cell in log phase SD medium [46]. Conversely, Lsm1 (at 3490 molecules/cell) is described as a member of the Lsm2-8 complex (among others), even though the cytoplasmic Lsm1-7 complex is not present in the dataset. Moreover, it has been found that the Lsm2-Lsm7 subunits can be exchanged between the two complexes under certain conditions [47]. Hence, it is possible that both Lsm1-7 and Lsm2-8 complexes were experimentally purified, however, as the two complexes differ only by a single protein, they were combined into a single entity due to the nature of the clustering algorithm that was used to process the raw data. Similar clustering effects are also present in other protein complex datasets [12, 19, 39] where, for example, RNA polymerases I, II and III were combined into a single entity due to presence of several common subunits between the 3 complexes. Therefore, the Lsm2-8 complex described in the dataset is likely to also represent the Lsm1-7 complex, which is a known component of cytoplasmic P-bodies [48], thus accounting for the connection observed between the Lsm2-8 complex with the P-bodies.

The complexome network revealed strong functional relationships between

interacting complexes. Mapping biological process to the network revealed homogenous regions of interconnected complexes that were involved in the same processes inside the cell. This suggests a hierarchical organisation of the yeast complexome, a feature that has been observed elsewhere [19], and also at the level of the interactome [39, 49]. Regions of functional homogeneity allowed us to suggest putative functions of uncharacterised complexes through guilt-by-association. Using this technique, we proposed the biological process and function of 51 unknown complexes (Table 1). These can be further validated by incorporating subcellular localisation data onto the network, in a similar fashion to the technique used by Wang et al. [39]. Wet-lab experimental analyses will be able to then prove or disprove the suggested functions.

The quality of our complexome network is dependent on the accuracy of the underlying data. Current protein complex datasets are by no means complete. Limitations of current high-throughput techniques, such as the inability to detect transient or weak protein interactions, will lead to inaccuracies and missing data in resulting visualisations. For example, the dataset used in this study, which covers about a third of the yeast proteome, contained a high proportion of complexes (338 out of 491) that have not been described in the literature [9]. Many of these complexes are small, with 220 containing 10 or less proteins. This suggests that they may be functional subunits of larger known complexes or, more simply, experimental artefacts. Unfortunately, these cannot be immediately resolved through use of multiple datasets; differences in clustering procedures used to establish membership of complexes have resulted in very little overlap between the 4 available protein complex datasets [9, 10, 11, 12]. Nevertheless, integrating these datasets with data from

small-scale experiments can increase coverage of the yeast complexome in the networks. There have been a number of different approaches taken to achieve this [19, 39, 49, 50], however, a recurring challenge has been how to build connections between protein complexes. A feature that distinguished the data from Gavin et al. [9] was the presence of proteins that belonged to multiple complexes. This formed the basis of complex-complex connections in our CM network, as well as that in another study [19]. For datasets where protein membership is mutually exclusive, it would be possible to establish links between protein complexes using binary interaction data (from, for example, protein complementation assays). Two complexes could then be connected if they contain a pair of proteins that have been found to interact, hence enabling a complementary complexome network to be built. However, the overlap between available protein-protein interaction and protein complex datasets remains poor and insufficient to build a cohesive complex-centric network [3, 51]. This is due to physicochemical biases of the different experimental techniques. Another means to connect complexes would be to include connections mediated by attachment proteins. Correlations with complementary data from, for example, yeast two-hybrid or co-expression studies, could be used to determine the likelihood of an interaction between an attachment protein and a complex.

The network generated in this study represents a biologically accurate and coherent model of the yeast complexome. Visualisation of the complexome as a network has reinforced our current understanding of protein complexes, and the manner in which sharing of their subunits is central to many related processes in the cell. This is highlighted by the functional associations observed between connected complexes in the CM network. Our work has enabled the assignment of putative function to

unknown complexes, and revealed potentially novel links between different cellular processes and pathways. This generates new hypotheses that are not apparent in protein-protein interaction networks. Studies have found that essential proteins are often concentrated in protein complexes, which are the actual dictators of the phenotype of the cell upon deletion [14, 19, 39, 52]. Hence, a move to visualise and analyse a higher organisation of the cell is crucial in gaining a more complete understanding of the integration and coordination of cellular processes and functions.

Acknowledgements:

MRW and SSL acknowledge financial support from the New South Wales State Government Science Leveraging Fund, the EIF Super Science scheme and the University of New South Wales. We thank Mark Tanaka for statistical advice and Seok-Hee Hong for her willingness to share the GEOMI code base with us.

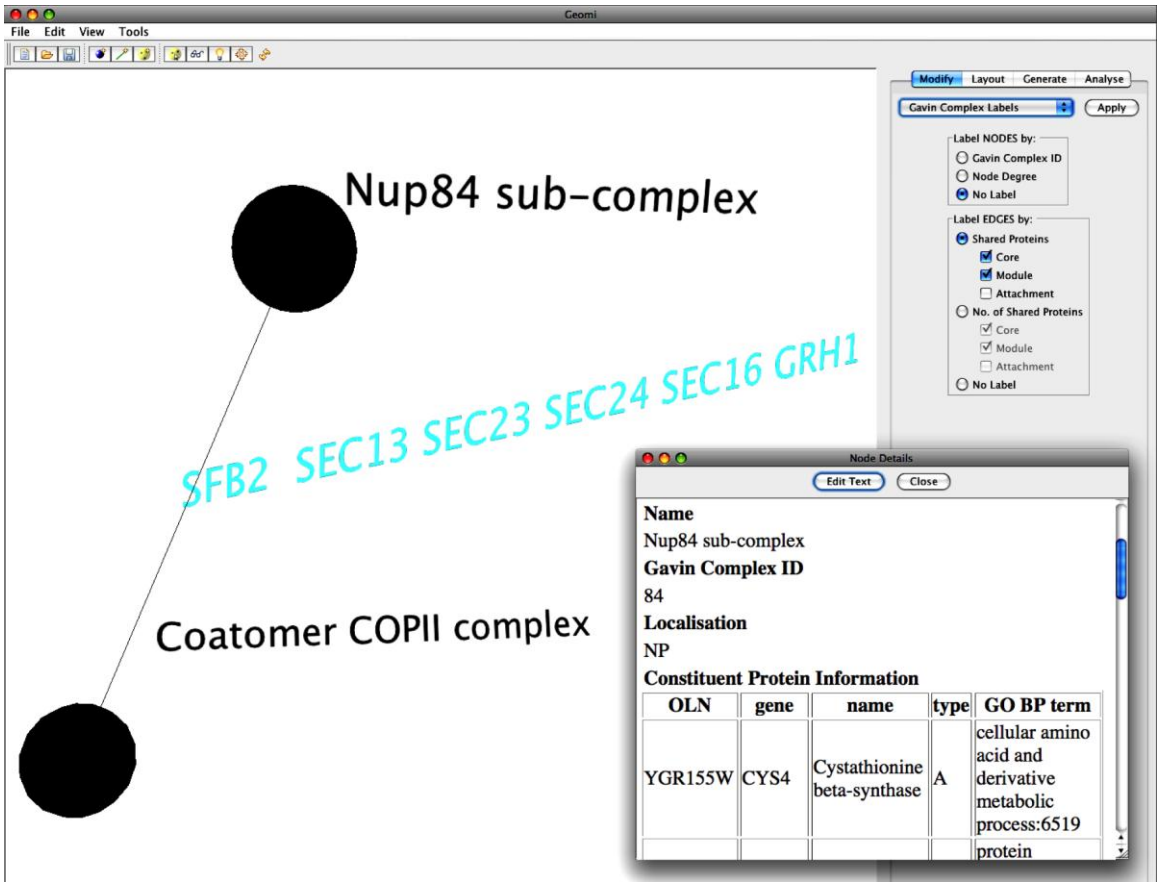


Figure 1: GEOMI software for visualising the complexome.

Information for each complex can be retrieved and displayed. In this case, the Nup84 sub-complex has been selected and its protein subunits, their type (core, attachment, or module) and Gene Ontology biological process can be seen via the 'Show Details' option (inset).

An edge label allows common protein subunits to be co-visualised on the network. In this case, the Nup84 subcomplex and coatomer COPII complexes share proteins SFB2, SEC13, SEC23, SEC24, SEC16 and GRH1. The control panel on the top right has many options for labelling the nodes and edges.

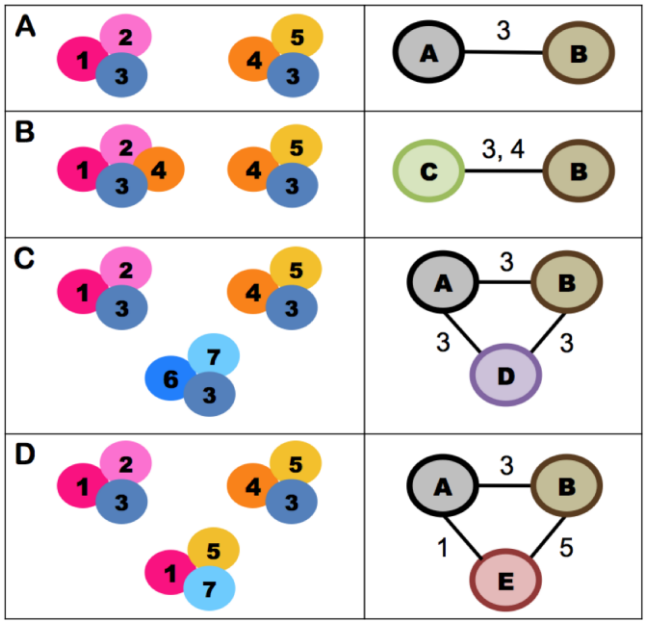


Figure 2: Representations of a complexome network.

As represented in (A) and (B), complexes with one or multiple common protein subunits are connected by a single edge, which can be labelled with the names of the shared proteins.
(C) and (D) highlight different ways to interpret identical topology in the CM network. In (C), complexes A, B and D all contain the protein 3 and are thus connected to each other. However, in (D), we have the same topology that has been conferred by different common subunits between complexes A, B and E. Hence, complexes that are completely interconnected (that is, form a clique) do not necessarily share identical protein subunits.

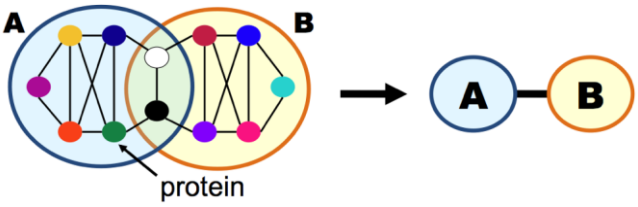


Figure 3: From interactome to complexome.

In this example, protein complexes A and B share two proteins, which are coloured black and white. In the interactome, this is visualised as a highly interconnected cluster, and may be misinterpreted to be a single complex. In the complexome, however, A and B are collapsed into single nodes, each representing a unique protein complex. Their common protein subunits become the edge connecting A and B.

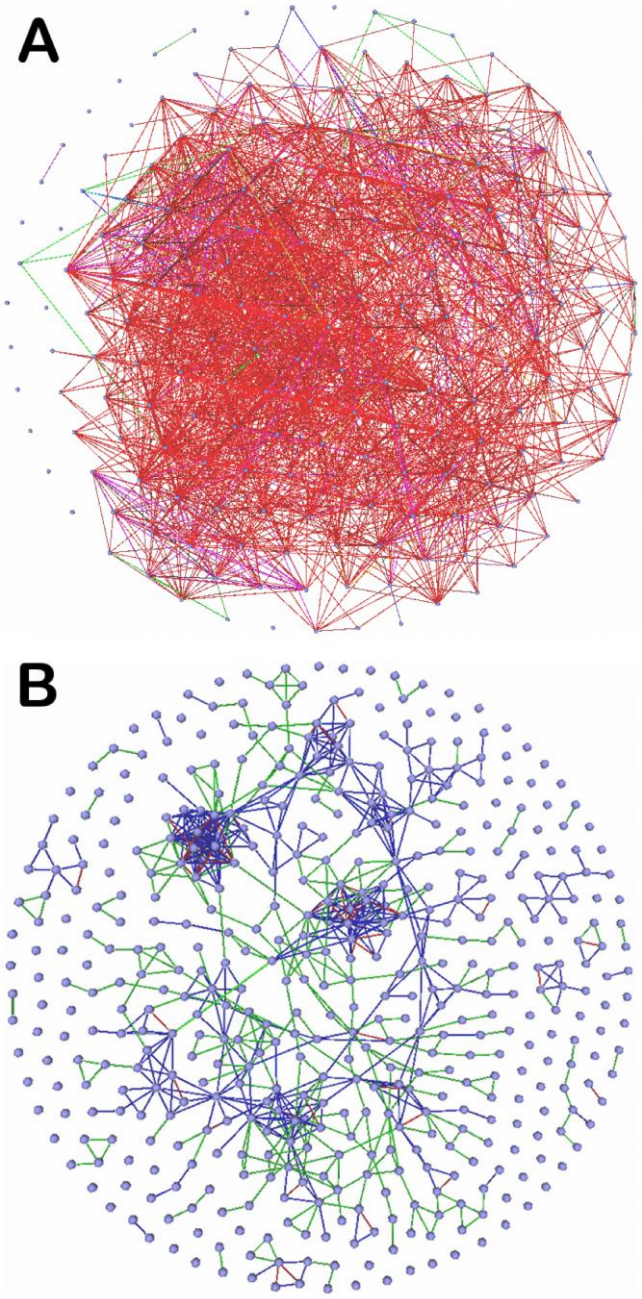


Figure 4: The complexome network.

(A) The core-module-attachment network of the yeast nuclear complexome. Edge colours represent the different combinations of core, module and attachment proteins shared between two complexes.
Key: Red (attachment proteins only), blue (module proteins only), green (core proteins only), cyan (core and module proteins), pink (module and attachments), yellow (core and attachments), black (core, module and attachments).

The network is clearly dominated by red edges. This indicates that attachment proteins are widely shared between complexes, and suggests that much of the complexity in the interactome is mediated by attachment proteins.
[GEOMI visualisation parameters: Force Directed Layout, spring = 50, origin = 80, repulsion = 12, planar = 100]

(B) The CM (core-module) network of the entire complexome. It shows regions of both high and low interconnectivity between protein complexes. Complexes with no partners, and hence do not share protein subunits with others, are not shown.

Key: Green (shared core proteins), blue (shared module proteins), red (core and modules).

[GEOMI visualisation parameters: Force Directed Layout, spring = 50, origin = 80, repulsion = 12, planar = 100]

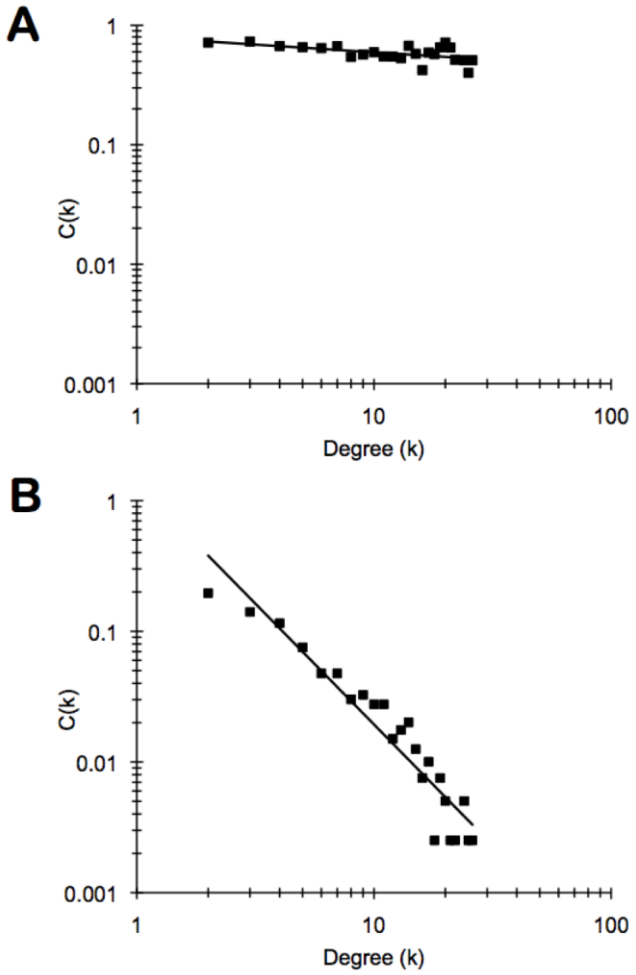


Figure 5: Network measures of the CM network.

The node degree distribution (A) and clustering coefficient (B) of the CM network. The linear relationships of these two network measures suggest that the CM network is scale-free.

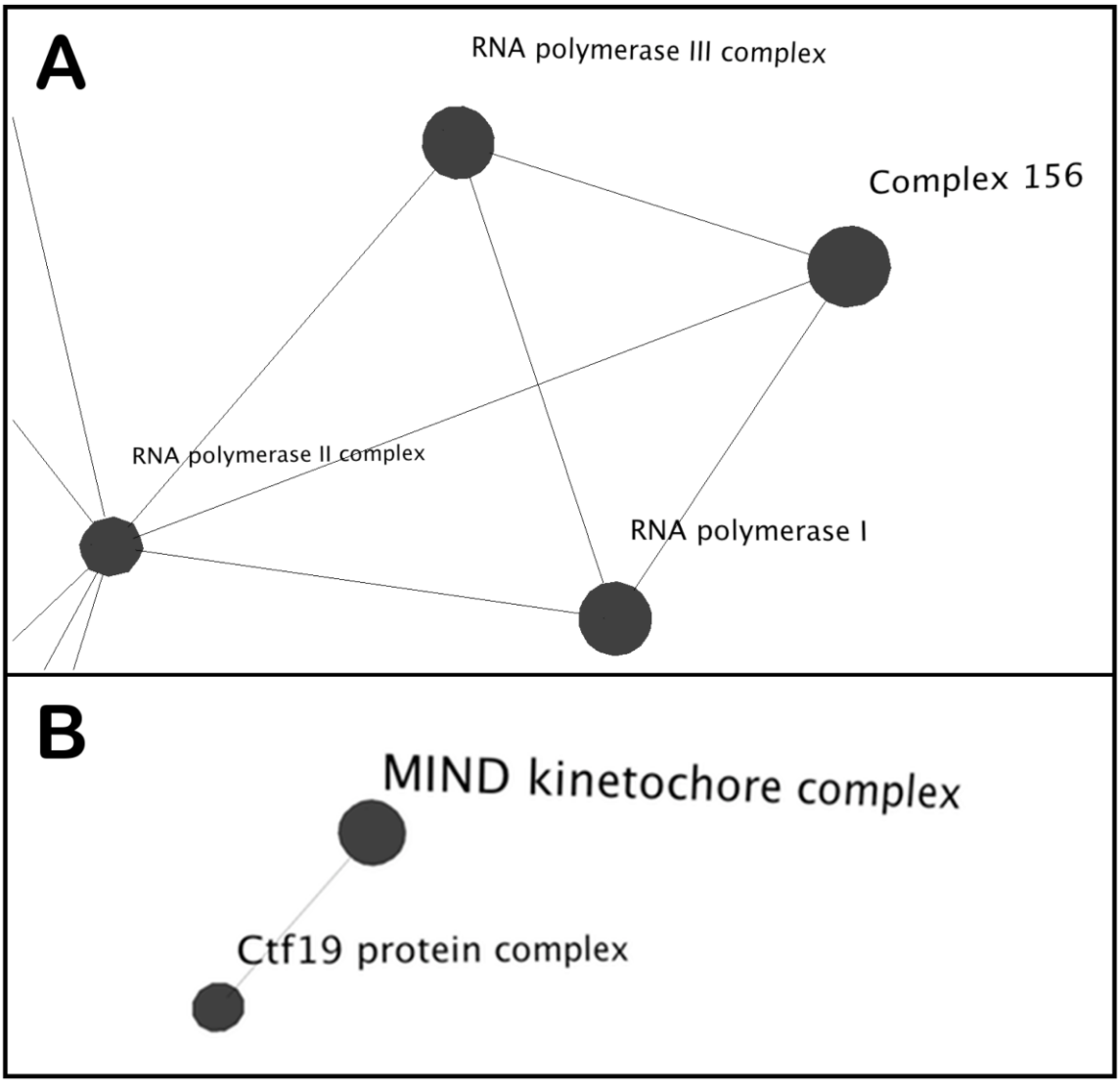


Figure 6: Details of the CM network.

The RNA polymerase complexes I to III show interconnection (A), as do the MIND kinetochore and CTF19 complexes (B).

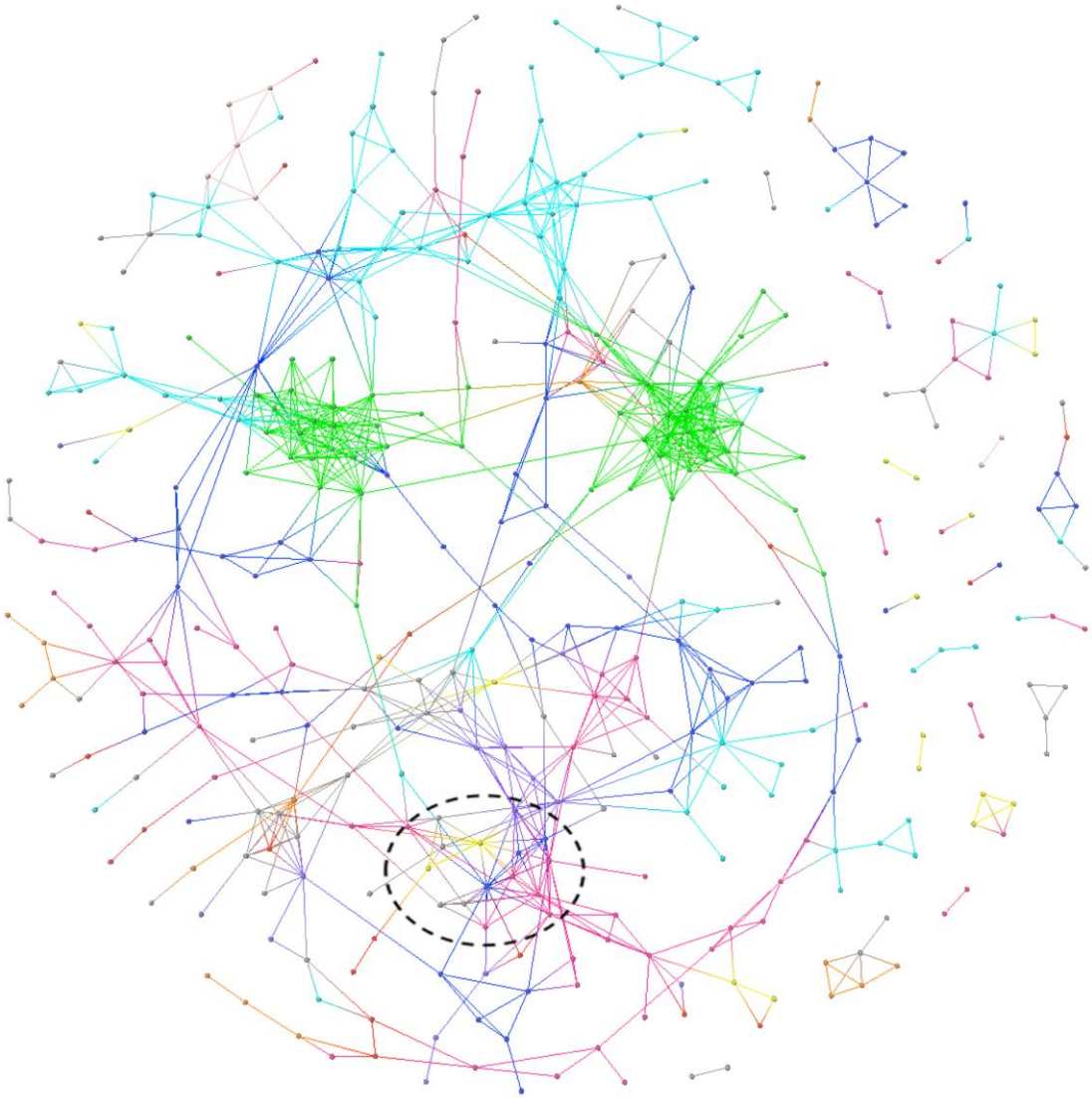


Figure 7: The CM network, coloured by biological process.

The 10 most frequently occurring biological processes in the network are given a unique colour, and remaining groups are combined into an ‘other’ category.

The circle highlights regions of heterogeneity of biological process between protein complexes.

Key: RNA metabolic process (cyan), transport (magenta), translation (blue), ribosome biogenesis (green), response to stress (yellow), cell cycle (orange), cellular

amino acid and derivative metabolic process (purple), vesicle-mediated transport (brown), transcription (light blue), DNA metabolic process (pink), and others (grey).

Complexes with no connections, and hence do not share protein subunits with others, are not shown for clarity.

[GEOMI visualisation parameters: Force Directed Layout, spring = 35, origin = 30, repulsion = 30, planar = 100]

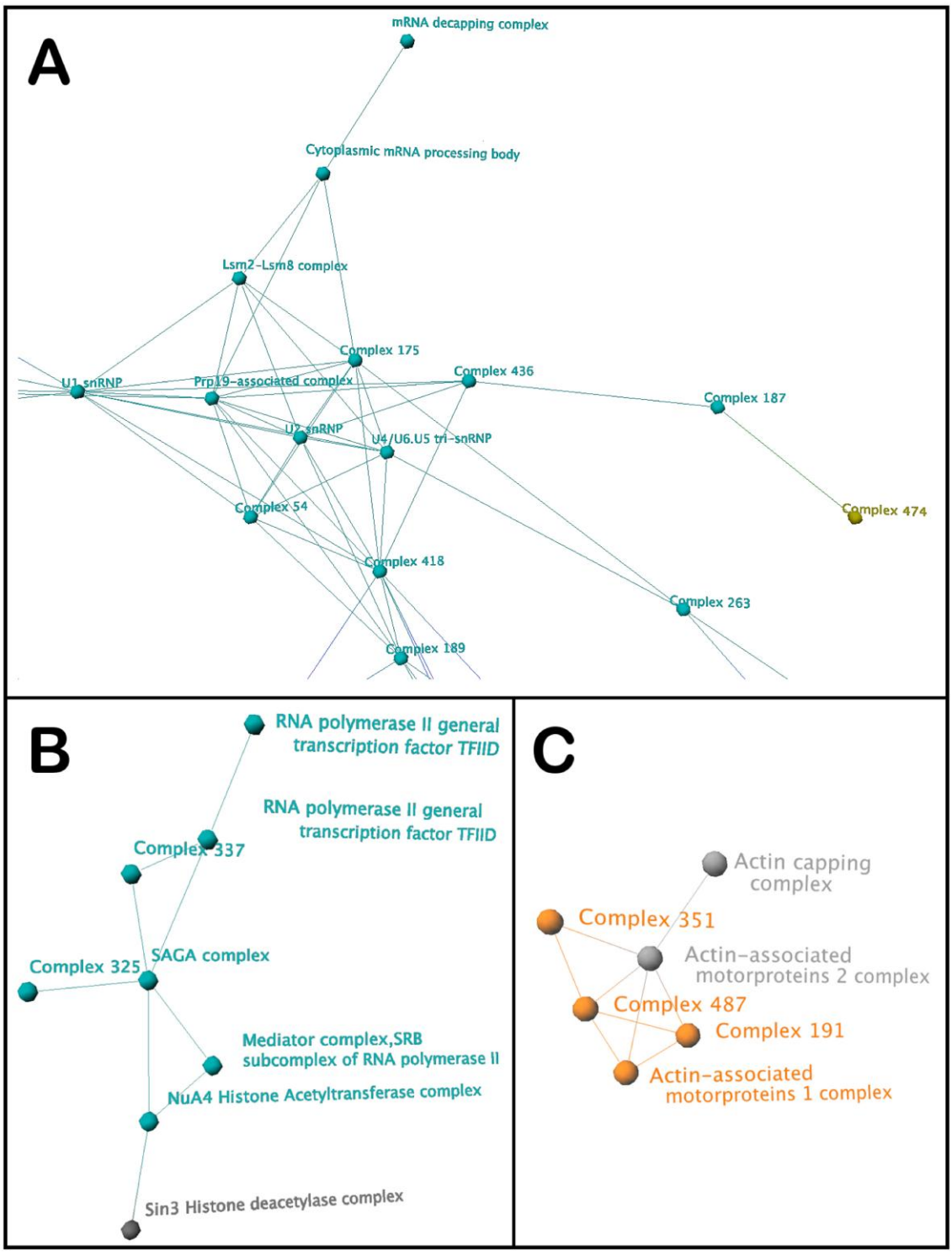


Figure 8: Connected complexes in the CM network reflect higher order of the cell.

(A) A highly interconnected cluster of complexes, almost all of which are involved in

the RNA metabolic process, as indicated by the cyan colour. (All complexes in this cluster, except Complex 474, are cyan).

(B) The SAGA, mediator and histone acetylases and deacetylases, which are involved in different stages of transcription, show interconnection.

(C) An isolated cluster of complexes primarily involved in cytokinesis.

Key: RNA metabolic process (cyan), response to stress (yellow), cell cycle (orange), others (grey).

Table 1: 'Higher order' functional groupings in the CM network.

Interacting complexes of related function	Supporting literature	Associated complexes of unknown function	Putative function of uncharacterised complexes
Actin capping complex Actin-associated motorproteins 2 complex Actin-associated motorproteins 1 complex	[53]	Complex 191 Complex 351 Complex 487	Actin-associated
V0 vacuolar ATPase complex V1 vacuolar ATPase complex RAVE/Skp1 complex	[54]	Complex 104 Complex 359 Complex 484	ATPase-associated
Mediator complex, SRB subcomplex of RNA polymerase II RNA polymerase II general transcription factor TFIID RNA polymerase II general transcription factor TFIID SAGA complex	[36]	Complex 204 Complex 325 Complex 337 Complex 395	Basal transcription machinery
Sit4/Sap190 complex Sit4/Sap185 protein phosphatase complex	[55]	Complex 151 Complex 483	TOR signalling cascade
Pab1/Pan2/Pan3 complex Pab1/eIF4G/eIF4E complex PAN complex	[56]	Complex 119 Complex 299 Complex 377 Complex 425	mRNA processing
Ku complex SWI-SNF global transcription complex RSC,abundant chromatin remodeling complex ISW1/IOC3 complex Ccr4/Not protein complex DNA polymerase alpha-primase complex DNA polymerase epsilon complex Ctf18/Rfc2/Rfc3/Rfc4/Rfc5 complex	[57, 58, 59]	Complex 12 Complex 99 Complex 162 Complex 240 Complex 278	DNA replication
U1 snRNP U2 snRNP U4/U6.U5 tri-snRNP Prp19-associated complex Lsm2-Lsm8 complex Cytoplasmic mRNA processing body mRNA decapping complex	[31, 32]	Complex 54 Complex 175 Complex 187 Complex 189 Complex 207 Complex 263 Complex 418 Complex 436 Complex 474	Spliceosomal processes
2-oxoglutarate dehydrogenase complex Pyruvate dehydrogenase	[60]	Complex 58 Complex 216 Complex 234	2-oxoglutarate metabolic processes or Acetyl-CoA biosynthetic process

19/22S regulator 20S core particle of the proteasome	[61]	Complex 216	Protein degradation
Pre-mRNA cleavage and polyadenylation factor 1A Polyadenylation factor I Protein phosphatase 1 complex	[62]	Complex 252 Complex 355	RNA processing and polyadenylation
Gyp5/Gyl1 complex Rvs161/Rvs167 complex	[63]	Complex 5 Complex 286 Complex 302	Cytoskeleton and vesicle transport
RNA polymerase I RNA polymerase II complex RNA polymerase III complex	[64]	Complex 156	RNA synthesis
Golgi transport complex Kel1/Lte1 complex Protein phosphatase 2A complex	[65, 66]	Complex 87 Complex 348	Golgi-associated transport or Mitosis
SKI complex Exosome 3'-5' exoribonuclease complex Rai1/Rat1 complex Ribosome-associated complex (RAC)*	[67]	Complex 260 Complex 385 Complex 396	RNA processing and degradation

* The ribosome-associated complex was directly connected to the SKI, exoribonuclease and Rai1/Rat1 complexes in the CM network. Literature linking it to these complexes, however, could not be found.

Table 2. Protein complexes and their core and module subunits inside a cluster.

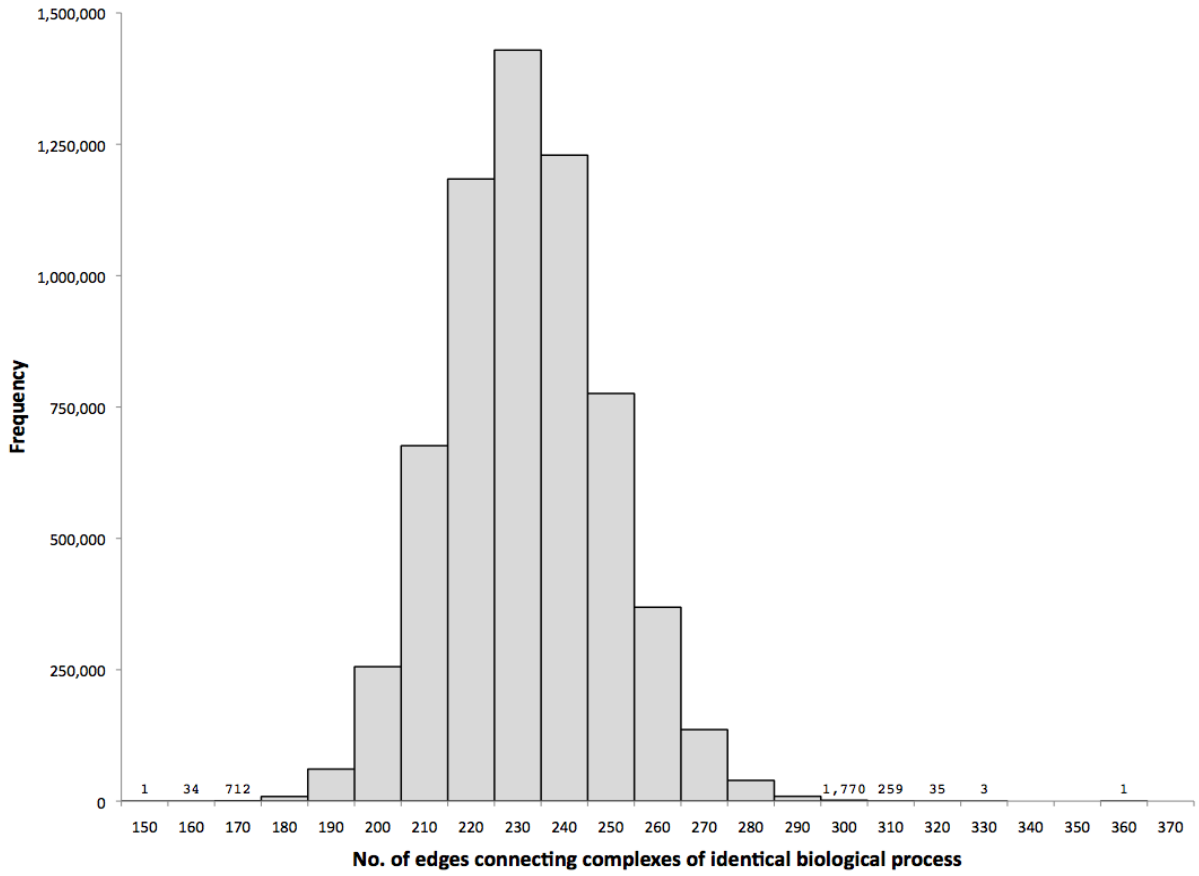
Complex	Core and modules protein subunits [9]
Actin capping complex	Cap1p, Cap2p, YER071C, YIR003W/Aim21p
Actin-associated motorproteins 1 complex	Myo2p, She4p Cmd1p, Mlc1, Myo1p, Myo5p
Actin-associated motorproteins 2 complex	Mlc1p, Myo4p, She2p, She3p Cap1p, Cap2p, Cdc47p, Cmd1p, Erg26p, Myo1p, Myo2p, Myo5p, She4p, YIR003W
Complex 191	Myo5p Cmd1p
Complex 351	Cdc47p, Erg26p
Complex 487	Cdc47p, Myo5p Cmd1p, Mlc1p, Myo1p, Myo4p, She2p, She3p

Supporting Information Available:

- Figure S1: Measuring the prevalence of interconnectivity between complexes of identical biological process in the CM network.

- Table S2: List of uncharacterised protein complexes in the CM network, and known protein complexes they are connected to in the CM network. It also includes putative biological processes these complexes are involved in, based on their constituent core and module proteins.

- GEOMI network visualisation platform, data files and user manual can be downloaded from http://www.systemsbiology.org.au.



Supplementary Figure S1: Measuring the prevalence of interconnectivity between complexes of identical biological process in the CM network.

The CM network was scored, and compared with those of 6,176,768 randomised networks. Its result of 364 edges (that connect two complexes of the same biological process) was the highest of all the networks, and it is a clear outlier as observed in this frequency distribution.

Supplementary Table S2: Interactions of complexes of unknown function with those of known function.

A list of likely biological processes the complexes are involved in, based on their constituent core and module proteins, is included.

Complex of unknown function	Likely biological process	Connected complexes of known function
Complex 5	ribosome biogenesis	Gyp5/Gyl1 complex
Complex 6	cellular lipid metabolic process, fungal-type cell wall organisation, transcription, cellular amino acid and derivative metabolic process, chromosome organisation, RNA metabolic process, protein modification process	Histone acetyltransferase B complex
Complex 10	chromosome organisation, RNA metabolic process, protein modification process, transport, ribosome biogenesis, translation	Gcn1/Gcn2/Gcn20 complex INO80 protein complex Anthranilate synthase complex
Complex 11	RNA metabolic process, ribosome biogenesis	2-oxoglutarate dehydrogenase complex
Complex 12	transcription, chromosome organisation, RNA metabolic process	ISW1/IOC3 complex RSC,abundant chromatin remodeling complex
Complex 16	response to chemical stimulus, response to stress	RNA polymerase II complex
Complex 18	ribosome biogenesis	Small subunit processome Small subunit processome
Complex 21	translation	Ribosome-associated complex (RAC) Translation initiation factor eIF3 complex
Complex 23	ribosome biogenesis	Small subunit processome Small subunit processome 2-oxoglutarate dehydrogenase complex
Complex 27	ribosome biogenesis	Small subunit processome Small subunit processome
Complex 28	translation	Cytoskeleton-regulatory Lsb3/Las17 complex Mitochondrial ribosomal large subunit Mitochondrial ribosomal large subunit COPI complex
Complex 36	transport	COPI complex
Complex 37	ribosome biogenesis	Small subunit processome Small subunit processome
Complex 51	transport	Kap104/Hrp1 complex Serine palmitoyltransferase complex
Complex 54	RNA metabolic process	U1 snRNP U2 snRNP U4/U6.U5 tri-snRNP

		Prp19-associated complex
Complex 58	ribosome biogenesis	Small subunit processome Karyopherins complex 2-oxoglutarate dehydrogenase complex
Complex 62	RNA metabolic process	RNA polymerase II complex
Complex 64	RNA metabolic process	U1 snRNP Pab1/eIF4G/eIF4E complex Prp19-associated complex Svg1/cyclin Bur2 protein complex mRNA cap-binding -Bur2/Sgv1 complexes
Complex 66	translation	Gcn1/Gcn2/Gcn20 complex Anthranilate synthase complex Phosphoribosyl diphosphate synthase complex
Complex 72	cytoskeleton organisation, response to stress, transport, vesicle-mediated transport, cellular membrane organisation	Cytoskeleton-regulatory Lsb3/Las17 complex
Complex 76	cofactor metabolic process, cellular amino acid and derivative metabolic process	Ribosome-associated complex (RAC) Anthranilate synthase complex Serine palmitoyltransferase complex
Complex 87	cell budding, cell cycle, signaling process, pseudohyphal growth, cytoskeleton organisation, protein modification process, translation	Protein phosphatase 2A complex Golgi transport complex
Complex 88	cellular lipid metabolic process, cofactor metabolic process, generation of precursor metabolites and energy, protein modification process, cellular carbohydrate metabolic process	GPI transamidase complex
Complex 90	generation of precursor metabolites and energy, mitochondrion organisation, cellular respiration, translation	Mitochondrial ribosomal small subunit
Complex 99	RNA metabolic process, transcription	ISW1/IOC3 complex DNA polymerase epsilon complex RSC,abundant chromatin remodeling complex
Complex 102	transport, RNA metabolic process, pseudohyphal growth, translation, transcription	U1 snRNP
Complex 103	RNA metabolic process, translation	Pab1/eIF4G/eIF4E complex mRNA cap-binding -Bur2/Sgv1 complexes

Complex 104	cellular homeostasis	V1 vacuolar ATPase complex
Complex 107	RNA metabolic process	mRNA guanylyl transferase complex
Complex 119	RNA metabolic process, translation	Pab1/Pan2/Pan3 complex
Complex 128	RNA metabolic process	RNA polymerase II complex
Complex 129	DNA metabolic process, transport, RNA metabolic process, transcription	THO complex
Complex 130	cell cycle, transport, signaling process, response to stress, cytoskeleton organisation, heterocycle metabolic process	F0/F1 ATP synthase
Complex 131	RNA metabolic process, ribosome biogenesis, protein complex biogenesis, cellular protein catabolic process	Small subunit processome
Complex 133	RNA metabolic process, translation	Kap104/Hrp1 complex
Complex 135	cellular amino acid and derivative metabolic process	Clathrin Arginine-specific carbamoyl-phosphate synthase complex Histone acetyltransferase B complex
Complex 136	transport	Gcn1/Gcn2/Gcn20 complex
Complex 139	transport, ribosome biogenesis, translation	Translation initiation factor eIF3 complex
Complex 141	DNA metabolic process, cell cycle, RNA metabolic process, signaling process, ribosome biogenesis, response to stress, protein modification process, cytokinesis, transcription	Translation elongation factor EF1
Complex 149	RNA metabolic process, ribosome biogenesis	Small subunit processome Small subunit processome
Complex 151	transport	U1 snRNP
Complex 156	other	RNA polymerase I RNA polymerase II complex RNA polymerase III complex
Complex 157	ribosome biogenesis	Small subunit processome Small subunit processome
Complex 158	DNA metabolic process, chromosome organisation, RNA metabolic process, response to stress, transcription	RNA polymerase II complex
Complex 159	RNA metabolic process	Small subunit processome Small subunit processome
Complex 162	DNA metabolic process, cell cycle, response to stress	DNA polymerase alpha-primase complex Ctf18/Rfc2/Rfc3/Rfc4 /Rfc5 complex
Complex 165	transport	Gcn1/Gcn2/Gcn20 complex Anthranilate synthase complex

		Serine palmitoyltransferase complex Heteromeric p24 complex 1
Complex 166	cell cycle, transport, mitochondrion organisation, translation	Kap104/Hrp1 complex Serine palmitoyltransferase complex
Complex 169	transport, RNA metabolic process	Gcn1/Gcn2/Gcn20 complex
Complex 171	cofactor metabolic process, cellular amino acid and derivative metabolic process, response to stress	Gcn1/Gcn2/Gcn20 complex Anthranilate synthase complex
Complex 173	mitochondrion organisation, translation	Mitochondrial ribosomal small subunit
Complex 174	transport, ribosome biogenesis, translation	Ribosome-associated complex (RAC)
Complex 175	RNA metabolic process	U1 snRNP U2 snRNP U4/U6.U5 tri-snRNP Lsm2-Lsm8 complex Prp19-associated complex Cytoplasmic mRNA processing body
Complex 178	cell budding, transport, cell cycle, vesicle-mediated transport, vesicle organisation, cytokinesis, cellular membrane organisation	Mitochondrial ribosomal large subunit
Complex 186	cellular amino acid and derivative metabolic process, cytoskeleton organisation	Trehalose-6-phosphate synthase/phosphatase complex
Complex 189	RNA metabolic process, ribosome biogenesis	U2 snRNP Prp19-associated complex Karyopherins complex
Complex 191	cell budding, transport, cell cycle, vesicle-mediated transport, cellular membrane organisation	Actin-associated motorproteins 1 complex Actin-associated motorproteins 2 complex
Complex 199	ribosome biogenesis	Small subunit processome
Complex 203	DNA metabolic process, transport	THO complex
Complex 206	mitochondrion organisation, translation	Mitochondrial ribosomal small subunit
Complex 210	cell budding, conjugation, cell cycle, RNA metabolic process, signaling process, cellular component morphogenesis, ribosome biogenesis, pseudohyphal growth, response to chemical stimulus, cytoskeleton organisation, cytokinesis	Pab1/eIF4G/eIF4E complex mRNA cap-binding -Bur2/Sgv1 complexes
Complex 211	RNA metabolic process,	Small subunit processome

	ribosome biogenesis	Small subunit processome
Complex 215	mitochondrion organisation, heterocycle metabolic process, translation	Mitochondrial ribosomal small subunit
Complex 216	cell cycle, cellular protein catabolic process	20S core particle of the proteasome 19/22S regulator Karyopherins complex Pyruvate dehydrogenase
Complex 220	DNA metabolic process, cell cycle	Clathrin Cohesin complex
Complex 222	cell budding, cell cycle, transport, vesicle-mediated transport, cytoskeleton organisation, cytokinesis, cellular membrane organisation	Cytoskeleton-regulatory Lsb3/Las17 complex
Complex 223	RNA metabolic process, mitochondrion organisation, translation	Translational release factor complex
Complex 225	RNA metabolic process, ribosome biogenesis	Pab1/eIF4G/eIF4E complex mRNA cap-binding -Bur2/Sgv1 complexes
Complex 234	RNA metabolic process, ribosome biogenesis	Small subunit processome 2-oxoglutarate dehydrogenase complex Ric1/Rgp1 complex
Complex 237	RNA metabolic process	U1 snRNP Pab1/eIF4G/eIF4E complex
Complex 240	DNA metabolic process, response to stress	Pab1/eIF4G/eIF4E complex ISW1/IOC3 complex DNA polymerase epsilon complex Ctf18/Rfc2/Rfc3/Rfc4 /Rfc5 complex
Complex 246	transport	Gcn1/Gcn2/Gcn20 complex
Complex 247	cellular amino acid and derivative metabolic process, cellular aromatic compound metabolic process	Anthranilate synthase complex Serine palmitoyltransferase complex
Complex 250	RNA metabolic process, ribosome biogenesis	Small subunit processome Small subunit processome
Complex 252	translation	Ribosome-associated complex (RAC) Translation initiation factor eIF3 complex Pre-mRNA cleavage and polyadenylation factor 1A Polyadenylation factor I

Complex 254	RNA metabolic process, translation	Gcn1/Gcn2/Gcn20 complex INO80 protein complex Anthranilate synthase complex
Complex 255	RNA metabolic process, ribosome biogenesis	Small subunit processome
Complex 257	transport	Kap104/Hrp1 complex Serine palmitoyltransferase complex
Complex 260	RNA metabolic process	Ribosome-associated complex (RAC)
Complex 261	RNA metabolic process, transcription	RNA polymerase II complex
Complex 263	RNA metabolic process	U4/U6.U5 tri-snRNP
Complex 264	mitochondrion organisation, heterocycle metabolic process, translation	Mitochondrial ribosomal small subunit
Complex 265	translation	Ribosome-associated complex (RAC) Translation initiation factor eIF3 complex
Complex 266	ribosome biogenesis	Small subunit processome Small subunit processome
Complex 267	transport, cellular amino acid and derivative metabolic process, heterocycle metabolic process	alpha-Aminoadipate-semialdehyde dehydrogenase complex
Complex 272	response to chemical stimulus, protein modification process, translation	Karyopherins complex
Complex 282	cellular amino acid and derivative metabolic process, cellular aromatic compound metabolic process, generation of precursor metabolites and energy, cellular carbohydrate metabolic process	Spindle pole body Anthranilate synthase complex Serine palmitoyltransferase complex
Complex 284	RNA metabolic process, ribosome biogenesis	Small subunit processome
Complex 285	translation	Karyopherins complex
Complex 286	RNA metabolic process, ribosome biogenesis	Gyp5/Gyl1 complex
Complex 290	cellular homeostasis, DNA metabolic process, cell cycle, cofactor metabolic process, cellular amino acid and derivative metabolic process, cellular aromatic compound metabolic process, signaling	Clathrin

	process, protein folding, response to chemical stimulus, response to stress, heterocycle metabolic process, translation, vitamin metabolic process	
Complex 299	RNA metabolic process, translation	Pab1/eIF4G/eIF4E complex Pab1/Pan2/Pan3 complex
Complex 302	transport, vesicle-mediated transport, translation, cellular membrane organisation	Rvs161/Rvs167 complex
Complex 311	unknown	Karyopherins complex
Complex 314	generation of precursor metabolites and energy	Spindle pole body
Complex 317	response to chemical stimulus, response to stress, cellular protein catabolic process	Golgi associated retrograde protein complex
Complex 322	other	Gcn1/Gcn2/Gcn20 complex INO80 protein complex
Complex 323	chromosome organisation, RNA metabolic process, protein modification process, translation, transcription	Histone acetyltransferase B complex
Complex 325	RNA metabolic process, transcription	SAGA complex
Complex 327	cellular lipid metabolic process, RNA metabolic process	Translational release factor complex
Complex 331	ribosome biogenesis	Small subunit processome Small subunit processome
Complex 337	RNA metabolic process, protein complex biogenesis, transcription	SAGA complex RNA polymerase II general transcription factor TFIID
Complex 339	DNA metabolic process, response to stress	Act1/Sac6 complex
Complex 344	protein folding	Clathrin
Complex 345	cellular amino acid and derivative metabolic process, cellular aromatic compound metabolic process	Anthranilate synthase complex
Complex 348	cellular lipid metabolic process, fungal-type cell wall organisation, cellular amino acid and derivative metabolic process, other	Golgi transport complex
Complex 351	DNA metabolic process, cell cycle, RNA metabolic process, response to stress, transcription	Actin-associated motorproteins 2 complex
Complex 352	cellular amino acid and derivative metabolic process, cellular aromatic compound	Anthranilate synthase complex Serine palmitoyltransferase complex

	metabolic process	
Complex 354	ribosome biogenesis	Small subunit processome Small subunit processome
Complex 355	translation	Pre-mRNA cleavage and polyadenylation factor 1A Polyadenylation factor I
Complex 357	transport	Kap104/Hrp1 complex Serine palmitoyltransferase complex
Complex 358	translation	Gcn1/Gcn2/Gcn20 complex Anthranilate synthase complex
Complex 359	transport, vesicle-mediated transport, signaling process, protein modification process, heterocycle metabolic process	mRNA guanylyl transferase complex RAVE/Skp1 complex
Complex 360	cellular homeostasis, DNA metabolic process, cell cycle, cofactor metabolic process, signaling process, protein folding, response to chemical stimulus, response to stress, translation, cellular carbohydrate metabolic process	Clathrin
Complex 362	transport	Gcn1/Gcn2/Gcn20 complex
Complex 362	transport	INO80 protein complex
Complex 377	cell budding, RNA metabolic process, signaling process, translation, cytokinesis	Pab1/Pan2/Pan3 complex
Complex 389	transport, vesicle-mediated transport, protein folding, translation, cellular membrane organisation	Cytoskeleton-regulatory Lsb3/Las17 complex Clathrin
Complex 396	translation	Ribosome-associated complex (RAC) Translation initiation factor eIF3 complex
Complex 399	transport, vesicle-mediated transport, response to chemical stimulus, cytoskeleton organisation, response to stress, cellular membrane organisation	Cytoskeleton-regulatory Lsb3/Las17 complex
Complex 401	transport, vesicle-mediated transport	Karyopherins complex
Complex 406	translation	Trehalose-6-phosphate synthase/phosphatase complex Mitochondrial ribosomal large subunit Pab1/eIF4G/eIF4E complex mRNA cap-binding -Bur2/Sgv1 complexes Mitochondrial ribosomal large subunit
Complex 412	transport	Gcn1/Gcn2/Gcn20 complex

Complex 414	RNA metabolic process	Mitochondrial ribosomal small subunit
Complex 418	RNA metabolic process	U1 snRNP U2 snRNP U4/U6.U5 tri-snRNP Prp19-associated complex Karyopherins complex
Complex 419	heterocycle metabolic process	Phosphoribosyl diphosphate synthase complex
Complex 425	translation	Pab1/eIF4G/eIF4E complex mRNA cap-binding -Bur2/Sgv1 complexes
Complex 427	transport, vesicle-mediated transport	SNARE complex
Complex 431	ribosome biogenesis, translation	Translation initiation factor eIF3 complex
Complex 433	transport, cell cycle, cellular amino acid and derivative metabolic process, protein folding, mitochondrion organisation	Cyclin-dependent protein kinase complex
Complex 436	RNA metabolic process	U1 snRNP U2 snRNP Prp19-associated complex
Complex 437	cellular lipid metabolic process, response to stress, cellular carbohydrate metabolic process	Serine palmitoyltransferase complex
Complex 443	protein folding	Chaperonine containing T-complex TRiC
Complex 444	transport, mitochondrion organisation	Cytoskeleton-regulatory Lsb3/Las17 complex
Complex 446	transport, cellular amino acid and derivative metabolic process, RNA metabolic process, cellular aromatic compound metabolic process, generation of precursor metabolites and energy, ribosome biogenesis, translation, cellular carbohydrate metabolic process	Spindle pole body
Complex 447	ribosome biogenesis	Small subunit processome Small subunit processome
Complex 449	transport, chromosome organisation	Cytoskeleton-regulatory Lsb3/Las17 complex Clathrin INO80 protein complex

Complex 452	unknown	Scs2/Opi1 complex
Complex 461	protein folding, mitochondrion organisation, translation	Translational release factor complex
Complex 467	RNA metabolic process	Pab1/eIF4G/eIF4E complex ISW1/IOC3 complex Bre5/Ubp3 complex mRNA cap-binding -Bur2/Sgv1 complexes
Complex 468	ribosome biogenesis	Small subunit processome Small subunit processome
Complex 475	ribosome biogenesis	Small subunit processome
Complex 477	generation of precursor metabolites and energy, cellular carbohydrate metabolic process	Spindle pole body
Complex 480	protein folding, response to stress	Clathrin
Complex 483	cell cycle, RNA metabolic process, signaling process	Sit4/Sap190 complex
Complex 484	transport, vesicle-mediated transport	mRNA guanylyl transferase complex RAVE/Skp1 complex
Complex 485	response to stress	Scs2/Opi1 complex
Complex 487	cell cycle	Actin-associated motorproteins 1 complex Actin-associated motorproteins 2 complex

REFERENCES

1. Li, S.; Armstrong, C. M.; Bertin, N.; Ge, H.; Milstein, S.; Boxem, M.; Vidalain, P. O.; Han, J. D.; Chesneau, A.; Hao, T.; Goldberg, D. S.; Li, N.; Martinez, M.; Rual, J. F.; Lamesch, P.; Xu, L.; Tewari, M.; Wong, S. L.; Zhang, L. V.; Berriz, G. F.; Jacotot, L.; Vaglio, P.; Reboul, J.; Hirozane-Kishikawa, T.; Li, Q.; Gabel, H. W.; Elewa, A.; Baumgartner, B.; Rose, D. J.; Yu, H.; Bosak, S.; Sequerra, R.; Fraser, A.; Mango, S. E.; Saxton, W. M.; Strome, S.; Van Den Heuvel, S.; Piano, F.; Vandenhaute, J.; Sardet, C.; Gerstein, M.; Doucette-Stamm, L.; Gunsalus, K. C.; Harper, J. W.; Cusick, M. E.; Roth, F. P.; Hill, D. E.; Vidal, M., A map of the interactome network of the metazoan C. elegans. *Science* **2004,** *303* (5657), 540-3.
2. Giot, L.; Bader, J. S.; Brouwer, C.; Chaudhuri, A.; Kuang, B.; Li, Y.; Hao, Y. L.; Ooi, C. E.; Godwin, B.; Vitols, E.; Vijayadamodar, G.; Pochart, P.; Machineni, H.; Welsh, M.; Kong, Y.; Zerhusen, B.; Malcolm, R.; Varrone, Z.; Collis, A.; Minto, M.; Burgess, S.; McDaniel, L.; Stimpson, E.; Spriggs, F.; Williams, J.; Neurath, K.; Ioime, N.; Agee, M.; Voss, E.; Furtak, K.; Renzulli, R.; Aanensen, N.; Carrolla, S.; Bickelhaupt, E.; Lazovatsky, Y.; DaSilva, A.; Zhong, J.; Stanyon, C. A.; Finley, R. L., Jr.; White, K. P.; Braverman, M.; Jarvie, T.; Gold, S.; Leach, M.; Knight, J.; Shimkets, R. A.; McKenna, M. P.; Chant, J.; Rothberg, J. M., A protein interaction map of Drosophila melanogaster. *Science* **2003,** *302* (5651), 1727-36.
3. Yu, H.; Braun, P.; Yildirim, M. A.; Lemmens, I.; Venkatesan, K.; Sahalie, J.; Hirozane-Kishikawa, T.; Gebreab, F.; Li, N.; Simonis, N.; Hao, T.; Rual, J. F.; Dricot, A.; Vazquez, A.; Murray, R. R.; Simon, C.; Tardivo, L.; Tam, S.; Svrzikapa, N.; Fan, C.; de Smet, A. S.; Motyl, A.; Hudson, M. E.; Park, J.; Xin, X.; Cusick, M. E.; Moore, T.; Boone, C.; Snyder, M.; Roth, F. P.; Barabasi, A. L.; Tavernier, J.; Hill, D. E.; Vidal, M., High-quality binary protein interaction map of the yeast interactome network. *Science* **2008,** *322* (5898), 104-10.
4. Rual, J. F.; Venkatesan, K.; Hao, T.; Hirozane-Kishikawa, T.; Dricot, A.; Li, N.; Berriz, G. F.; Gibbons, F. D.; Dreze, M.; Ayivi-Guedehoussou, N.; Klitgord, N.; Simon, C.; Boxem, M.; Milstein, S.; Rosenberg, J.; Goldberg, D. S.; Zhang, L. V.; Wong, S. L.; Franklin, G.; Li, S.; Albala, J. S.; Lim, J.; Fraughton, C.; Llamosas, E.; Cevik, S.; Bex, C.; Lamesch, P.; Sikorski, R. S.; Vandenhaute, J.; Zoghbi, H. Y.; Smolyar, A.; Bosak, S.; Sequerra, R.; Doucette-Stamm, L.; Cusick, M. E.; Hill, D. E.; Roth, F. P.; Vidal, M., Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **2005,** *437* (7062), 1173-8.
5. Dortay, H.; Gruhn, N.; Pfeifer, A.; Schwerdtner, M.; Schmulling, T.; Heyl, A., Toward an interaction map of the two-component signaling pathway of Arabidopsis thaliana. *J Proteome Res* **2008,** *7* (9), 3649-60.
6. den Hollander, A. I.; Koenekoop, R. K.; Mohamed, M. D.; Arts, H. H.; Boldt, K.; Towns, K. V.; Sedmak, T.; Beer, M.; Nagel-Wolfrum, K.; McKibbin, M.; Dharmaraj, S.; Lopez, I.; Ivings, L.; Williams, G. A.; Springell, K.; Woods, C. G.; Jafri, H.; Rashid, Y.; Strom, T. M.; van der Zwaag, B.; Gosens, I.; Kersten, F. F.; van Wijk, E.; Veltman, J. A.; Zonneveld, M. N.; van Beersum, S. E.; Maumenee, I. H.; Wolfrum, U.; Cheetham, M. E.; Ueffing, M.; Cremers, F. P.; Inglehearn, C. F.; Roepman, R., Mutations in LCA5, encoding the ciliary protein lebercilin, cause Leber congenital amaurosis. *Nat Genet* **2007,** *39* (7), 889-95.
7. Neubauer, G.; Gottschalk, A.; Fabrizio, P.; Seraphin, B.; Luhrmann, R.; Mann, M., Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc Natl Acad Sci U S A* **1997,** *94* (2), 385-90.
8. Mann, M.; Hendrickson, R. C.; Pandey, A., Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem* **2001,** *70*, 437-73.
9. Gavin, A. C.; Aloy, P.; Grandi, P.; Krause, R.; Boesche, M.; Marzioch, M.; Rau, C.; Jensen, L. J.; Bastuck, S.; Dumpelfeld, B.; Edelmann, A.; Heurtier, M. A.; Hoffman, V.; Hoefert, C.; Klein, K.; Hudak, M.; Michon, A. M.; Schelder, M.; Schirle, M.; Remor, M.; Rudi, T.; Hooper, S.; Bauer, A.; Bouwmeester, T.; Casari, G.; Drewes, G.; Neubauer, G.; Rick, J.

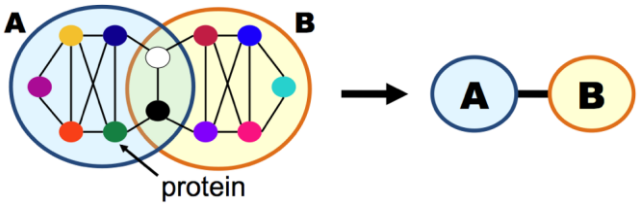
M.; Kuster, B.; Bork, P.; Russell, R. B.; Superti-Furga, G., Proteome survey reveals modularity of the yeast cell machinery. *Nature* **2006,** *440* (7084), 631-6.
10. Gavin, A. C.; Bosche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J. M.; Michon, A. M.; Cruciat, C. M.; Remor, M.; Hofert, C.; Schelder, M.; Brajenovic, M.; Ruffner, H.; Merino, A.; Klein, K.; Hudak, M.; Dickson, D.; Rudi, T.; Gnau, V.; Bauch, A.; Bastuck, S.; Huhse, B.; Leutwein, C.; Heurtier, M. A.; Copley, R. R.; Edelmann, A.; Querfurth, E.; Rybin, V.; Drewes, G.; Raida, M.; Bouwmeester, T.; Bork, P.; Seraphin, B.; Kuster, B.; Neubauer, G.; Superti-Furga, G., Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **2002,** *415* (6868), 141-7.
11. Ho, Y.; Gruhler, A.; Heilbut, A.; Bader, G. D.; Moore, L.; Adams, S. L.; Millar, A.; Taylor, P.; Bennett, K.; Boutilier, K.; Yang, L.; Wolting, C.; Donaldson, I.; Schandorff, S.; Shewnarane, J.; Vo, M.; Taggart, J.; Goudreault, M.; Muskat, B.; Alfarano, C.; Dewar, D.; Lin, Z.; Michalickova, K.; Willems, A. R.; Sassi, H.; Nielsen, P. A.; Rasmussen, K. J.; Andersen, J. R.; Johansen, L. E.; Hansen, L. H.; Jespersen, H.; Podtelejnikov, A.; Nielsen, E.; Crawford, J.; Poulsen, V.; Sorensen, B. D.; Matthiesen, J.; Hendrickson, R. C.; Gleeson, F.; Pawson, T.; Moran, M. F.; Durocher, D.; Mann, M.; Hogue, C. W.; Figeys, D.; Tyers, M., Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **2002,** *415* (6868), 180-3.
12. Krogan, N. J.; Cagney, G.; Yu, H.; Zhong, G.; Guo, X.; Ignatchenko, A.; Li, J.; Pu, S.; Datta, N.; Tikuisis, A. P.; Punna, T.; Peregrin-Alvarez, J. M.; Shales, M.; Zhang, X.; Davey, M.; Robinson, M. D.; Paccanaro, A.; Bray, J. E.; Sheung, A.; Beattie, B.; Richards, D. P.; Canadien, V.; Lalev, A.; Mena, F.; Wong, P.; Starostine, A.; Canete, M. M.; Vlasblom, J.; Wu, S.; Orsi, C.; Collins, S. R.; Chandran, S.; Haw, R.; Rilstone, J. J.; Gandi, K.; Thompson, N. J.; Musso, G.; St Onge, P.; Ghanny, S.; Lam, M. H.; Butland, G.; Altaf-Ul, A. M.; Kanaya, S.; Shilatifard, A.; O'Shea, E.; Weissman, J. S.; Ingles, C. J.; Hughes, T. R.; Parkinson, J.; Gerstein, M.; Wodak, S. J.; Emili, A.; Greenblatt, J. F., Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* **2006,** *440* (7084), 637-43.
13. Guldener, U.; Munsterkotter, M.; Kastenmuller, G.; Strack, N.; van Helden, J.; Lemer, C.; Richelles, J.; Wodak, S. J.; Garcia-Martinez, J.; Perez-Ortin, J. E.; Michael, H.; Kaps, A.; Talla, E.; Dujon, B.; Andre, B.; Souciet, J. L.; De Montigny, J.; Bon, E.; Gaillardin, C.; Mewes, H. W., CYGD: the Comprehensive Yeast Genome Database. *Nucleic Acids Res* **2005,** *33* (Database issue), D364-8.
14. Pache, R. A.; Babu, M. M.; Aloy, P., Exploiting gene deletion fitness effects in yeast to understand the modular architecture of protein complexes under different growth conditions. *BMC Syst Biol* **2009,** *3*, 74.
15. Pang, C. N.; Krycer, J. R.; Lek, A.; Wilkins, M. R., Are protein complexes made of cores, modules and attachments? *Proteomics* **2008,** *8* (3), 425-34.
16. Ho, E.; Webber, R.; Wilkins, M. R., Interactive three-dimensional visualization and contextual analysis of protein interaction networks. *J Proteome Res* **2008,** *7* (1), 104-12.
17. Krause, R.; von Mering, C.; Bork, P.; Dandekar, T., Shared components of protein complexes--versatile building blocks or biochemical artefacts? *Bioessays* **2004,** *26* (12), 1333-43.
18. Hu, Z.; Mellor, J.; Wu, J.; Kanehisa, M.; Stuart, J. M.; DeLisi, C., Towards zoomable multidimensional maps of the cell. *Nat Biotechnol* **2007,** *25* (5), 547-54.
19. Hart, G. T.; Lee, I.; Marcotte, E. R., A high-accuracy consensus map of yeast protein complexes reveals modular nature of gene essentiality. *BMC Bioinformatics* **2007,** *8*, 236.
20. SGD, p., Saccharomyces Genome Database.
21. Apweiler, R.; Bairoch, A.; Wu, C. H.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Natale, D. A.; O'Donovan, C.; Redaschi, N.; Yeh, L. S., UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* **2004,** *32* (Database issue), D115-9.
22. Ahmed, A. D., T.; Forster, M.; Fu, X.; Ho, J.; Hong, S.-H.; Koschutzki, D.; Murray, C.; Nikolov, N. S.; Taib, R.; Tarassov, A.; Xu, K In *GEOMI: GEOmetry for Maximum Insight*,

Proceeding of 13th International Symposium on Graph Drawing, Limerick, Ireland, September 2005, Limerick, Ireland, Limerick, Ireland, 2005; pp 468– 479.
23. Fisher, D.; O'Madadhain, J.; Smyth, P.; White, S.; Boey, Y.-B., Analysis and Visualization of Network Data using JUNG. *Journal of Statistical Software*.
24. Barabasi, A. L.; Oltvai, Z. N., Network biology: understanding the cell's functional organization. *Nat Rev Genet* **2004,** *5* (2), 101-13.
25. Goel, A.; Li, S. S.; Wilkins, M. R.; Goel A, L. S., Wilkins MR (2010). ., submitted., Four-dimensional visualisation and analysis of protein-protein interaction networks. *Proteomics* **2011,** *submitted*.
26. Albert, R., Scale-free networks in cell biology. *J Cell Sci* **2005,** *118* (Pt 21), 4947-57.
27. Seyed-Allaei, H.; Bianconi, G.; Marsili, M., Scale-free networks with an exponent less than two. *Phys Rev E Stat Nonlin Soft Matter Phys* **2006,** *73* (4 Pt 2), 046113.
28. De Wulf, P.; McAinsh, A. D.; Sorger, P. K., Hierarchical assembly of the budding yeast kinetochore from multiple subcomplexes. *Genes Dev* **2003,** *17* (23), 2902-21.
29. Dell'Angelica, E. C.; Ooi, C. E.; Bonifacino, J. S., Beta3A-adaptin, a subunit of the adaptor-like complex AP-3. *J Biol Chem* **1997,** *272* (24), 15078-84.
30. Pereira-Leal, J. B.; Teichmann, S. A., Novel specificities emerge by stepwise duplication of functional modules. *Genome Res* **2005,** *15* (4), 552-9.
31. Brow, D. A., Allosteric cascade of spliceosome activation. *Annu Rev Genet* **2002,** *36*, 333-60.
32. Franks, T. M.; Lykke-Andersen, J., The control of mRNA decapping and P-body formation. *Mol Cell* **2008,** *32* (5), 605-15.
33. Yu, H.; Kim, P. M.; Sprecher, E.; Trifonov, V.; Gerstein, M., The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol* **2007,** *3* (4), e59.
34. Timmers, H. T.; Tora, L., SAGA unveiled. *Trends Biochem Sci* **2005,** *30* (1), 7-10.
35. Grant, P. A.; Schieltz, D.; Pray-Grant, M. G.; Steger, D. J.; Reese, J. C.; Yates, J. R., 3rd; Workman, J. L., A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation. *Cell* **1998,** *94* (1), 45-53.
36. Sikorski, T. W.; Buratowski, S., The basal initiation machinery: beyond the general transcription factors. *Curr Opin Cell Biol* **2009,** *21* (3), 344-51.
37. Lottersberger, F.; Panza, A.; Lucchini, G.; Longhese, M. P., Functional and physical interactions between yeast 14-3-3 proteins, acetyltransferases, and deacetylases in response to DNA replication perturbations. *Mol Cell Biol* **2007,** *27* (9), 3266-81.
38. Shi, X.; Kachirskaia, I.; Walter, K. L.; Kuo, J. H.; Lake, A.; Davrazou, F.; Chan, S. M.; Martin, D. G.; Fingerman, I. M.; Briggs, S. D.; Howe, L.; Utz, P. J.; Kutateladze, T. G.; Lugovskoy, A. A.; Bedford, M. T.; Gozani, O., Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. *J Biol Chem* **2007,** *282* (4), 2450-5.
39. Wang, H.; Kakaradov, B.; Collins, S. R.; Karotki, L.; Fiedler, D.; Shales, M.; Shokat, K. M.; Walther, T. C.; Krogan, N. J.; Koller, D., A complex-based reconstruction of the Saccharomyces cerevisiae interactome. *Mol Cell Proteomics* **2009,** *8* (6), 1361-81.
40. Grotsch, H.; Giblin, J. P.; Idrissi, F. Z.; Fernandez-Golbano, I. M.; Collette, J. R.; Newpher, T. M.; Robles, V.; Lemmon, S. K.; Geli, M. I., Calmodulin dissociation regulates Myo5 recruitment and function at endocytic sites. *EMBO J* **2010,** *29* (17), 2899-914.
41. Shepard, K. A.; Gerber, A. P.; Jambhekar, A.; Takizawa, P. A.; Brown, P. O.; Herschlag, D.; DeRisi, J. L.; Vale, R. D., Widespread cytoplasmic mRNA transport in yeast: identification of 22 bud-localized transcripts using DNA microarray analysis. *Proc Natl Acad Sci U S A* **2003,** *100* (20), 11429-34.
42. Han, J. D.; Bertin, N.; Hao, T.; Goldberg, D. S.; Berriz, G. F.; Zhang, L. V.; Dupuy, D.; Walhout, A. J.; Cusick, M. E.; Roth, F. P.; Vidal, M., Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **2004,** *430* (6995), 88-93.
43. Tharun, S., Roles of eukaryotic Lsm proteins in the regulation of mRNA function. *Int Rev Cell Mol Biol* **2009,** *272*, 149-89.

44. Hogg, R.; McGrail, J. C.; O'Keefe, R. T., The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochem Soc Trans* **2010,** *38* (4), 1110-5.
45. Chan, S. P.; Kao, D. I.; Tsai, W. Y.; Cheng, S. C., The Prp19p-associated complex in spliceosome activation. *Science* **2003,** *302* (5643), 279-82.
46. Ghaemmaghami, S.; Huh, W. K.; Bower, K.; Howson, R. W.; Belle, A.; Dephoure, N.; O'Shea, E. K.; Weissman, J. S., Global analysis of protein expression in yeast. *Nature* **2003,** *425* (6959), 737-41.
47. Spiller, M. P.; Reijns, M. A.; Beggs, J. D., Requirements for nuclear localization of the Lsm2-8p complex and competition between nuclear and cytoplasmic Lsm complexes. *J Cell Sci* **2007,** *120* (Pt 24), 4310-20.
48. Sheth, U.; Parker, R., Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* **2003,** *300* (5620), 805-8.
49. Collins, S. R.; Kemmeren, P.; Zhao, X. C.; Greenblatt, J. F.; Spencer, F.; Holstege, F. C.; Weissman, J. S.; Krogan, N. J., Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae. *Mol Cell Proteomics* **2007,** *6* (3), 439-50.
50. Pu, S.; Vlasblom, J.; Emili, A.; Greenblatt, J.; Wodak, S. J., Identifying functional modules in the physical interactome of Saccharomyces cerevisiae. *Proteomics* **2007,** *7* (6), 944-60.
51. Fasolo, J.; Sboner, A.; Sun, M. G.; Yu, H.; Chen, R.; Sharon, D.; Kim, P. M.; Gerstein, M.; Snyder, M., Diverse protein kinase interactions identified by protein microarrays reveal novel connections between cellular processes. *Genes Dev* **2011,** *25* (7), 767-78.
52. Michaut, M.; Baryshnikova, A.; Costanzo, M.; Myers, C. L.; Andrews, B. J.; Boone, C.; Bader, G. D., Protein complexes are central in the yeast genetic landscape. *PLoS Comput Biol* **2011,** *7* (2), e1001092.
53. Cooper, J. A., The role of actin polymerization in cell motility. *Annu Rev Physiol* **1991,** *53*, 585-605.
54. Seol, J. H.; Shevchenko, A.; Deshaies, R. J., Skp1 forms multiple protein complexes, including RAVE, a regulator of V-ATPase assembly. *Nat Cell Biol* **2001,** *3* (4), 384-91.
55. Jablonowski, D.; Taubert, J. E.; Bar, C.; Stark, M. J.; Schaffrath, R., Distinct subsets of Sit4 holophosphatases are required for inhibition of Saccharomyces cerevisiae growth by rapamycin and zymocin. *Eukaryot Cell* **2009,** *8* (11), 1637-47.
56. Dunn, E. F.; Hammell, C. M.; Hodge, C. A.; Cole, C. N., Yeast poly(A)-binding protein, Pab1, and PAN, a poly(A) nuclease complex recruited by Pab1, connect mRNA biogenesis to export. *Genes Dev* **2005,** *19* (1), 90-103.
57. van Attikum, H.; Gasser, S. M., ATP-dependent chromatin remodeling and DNA double-strand break repair. *Cell Cycle* **2005,** *4* (8), 1011-4.
58. Wang, Z.; Wu, X.; Friedberg, E. C., DNA repair synthesis during base excision repair in vitro is catalyzed by DNA polymerase epsilon and is influenced by DNA polymerases alpha and delta in Saccharomyces cerevisiae. *Mol Cell Biol* **1993,** *13* (2), 1051-8.
59. Shiomi, Y.; Masutani, C.; Hanaoka, F.; Kimura, H.; Tsurimoto, T., A second proliferating cell nuclear antigen loader complex, Ctf18-replication factor C, stimulates DNA polymerase eta activity. *J Biol Chem* **2007,** *282* (29), 20906-14.
60. Schnarrenberger, C.; Martin, W., Evolution of the enzymes of the citric acid cycle and the glyoxylate cycle of higher plants. A case study of endosymbiotic gene transfer. *Eur J Biochem* **2002,** *269* (3), 868-83.
61. Coux, O.; Tanaka, K.; Goldberg, A. L., Structure and functions of the 20S and 26S proteasomes. *Annu Rev Biochem* **1996,** *65*, 801-47.
62. Nedea, E.; Nalbant, D.; Xia, D.; Theoharis, N. T.; Suter, B.; Richardson, C. J.; Tatchell, K.; Kislinger, T.; Greenblatt, J. F.; Nagy, P. L., The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Mol Cell* **2008,** *29* (5), 577-87.

63. Friesen, H.; Colwill, K.; Robertson, K.; Schub, O.; Andrews, B., Interaction of the Saccharomyces cerevisiae cortical actin patch protein Rvs167p with proteins involved in ER to Golgi vesicle trafficking. *Genetics* **2005,** *170* (2), 555-68.
64. Archambault, J.; Friesen, J. D., Genetics of eukaryotic RNA polymerases I, II, and III. *Microbiol Rev* **1993,** *57* (3), 703-24.
65. Yoshida, S.; Ichihashi, R.; Toh-e, A., Ras recruits mitotic exit regulator Lte1 to the bud cortex in budding yeast. *J Cell Biol* **2003,** *161* (5), 889-97.
66. Jiang, Y., Regulation of the cell cycle by protein phosphatase 2A in Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* **2006,** *70* (2), 440-9.
67. Butler, J. S., The yin and yang of the exosome. *Trends Cell Biol* **2002,** *12* (2), 90-6.

Synopsis:



We describe a novel representation of the yeast complexome, as a network of protein complexes connected by common protein subunits. Extensive validation shows this network is of high biological relevance. Connected complexes in the network were found to be of identical or related biological process, revealing a higher-level organisation of cellular function and allowing putative function to be assigned to 45 uncharacterised complexes.