

Proteome survey reveals modularity of the yeast cell machinery

Anne-Claude Gavin^{1*†}, Patrick Aloy^{2*}, Paola Grandi¹, Roland Krause^{1,3}, Markus Boesche¹, Martina Marzioch¹, Christina Rau¹, Lars Juhl Jensen², Sonja Bastuck¹, Birgit Dimpelfeld¹, Angela Edelmann¹, Marie-Anne Heurtier¹, Verena Hoffman¹, Christian Hoefert¹, Karin Klein¹, Manuela Hudak¹, Anne-Marie Michon¹, Malgorzata Schelder¹, Markus Schirle¹, Marita Remor¹, Tatjana Rudi¹, Sean Hooper², Andreas Bauer¹, Tewis Bouwmeester¹, Georg Casari¹, Gerard Drewes¹, Gitte Neubauer¹, Jens M. Rick¹, Bernhard Kuster¹, Peer Bork², Robert B. Russell² & Giulio Superti-Furga^{1,4}

Protein complexes are key molecular entities that integrate multiple gene products to perform cellular functions. Here we report the first genome-wide screen for complexes in an organism, budding yeast, using affinity purification and mass spectrometry. Through systematic tagging of open reading frames (ORFs), the majority of complexes were purified several times, suggesting screen saturation. The richness of the data set enabled a *de novo* characterization of the composition and organization of the cellular machinery. The ensemble of cellular proteins partitions into 491 complexes, of which 257 are novel, that differentially combine with additional attachment proteins or protein modules to enable a diversification of potential functions. Support for this modular organization of the proteome comes from integration with available data on expression, localization, function, evolutionary conservation, protein structure and binary interactions. This study provides the largest collection of physically determined eukaryotic cellular machines so far and a platform for biological data integration and modelling.

Genomes are remarkable in that they encode most of the functions necessary for their interpretation and propagation¹. However, many principles as to how individual gene products form the structures required for biological activity are still unknown. Biological processes, such as the cell cycle and replication, require precise organization of molecules in time and space. Complexes are among the fundamental units of macromolecular organization². They are thought to assemble in a particular order, and often require energy-driven conformational changes, specific post-translational modifications or chaperone assistance for proper formation³. Their composition is also known to vary according to cellular requirements.

Affinity purification methods are well suited for studying complexes under near-physiological conditions^{4,5}. They allow macromolecules physically associated with a tagged bait to be retrieved and identified by mass spectrometry^{6,7}. These methods have been applied as large-scale screens in prokaryotic and eukaryotic cells, and have led to a growing collection of cellular machines^{8–11} that, in combination with large-scale yeast two-hybrid studies^{12,13}, are powerful integrators of additional biological data^{14–16}. However, in the absence of a genome-wide screen, where many complexes are retrieved repeatedly through a ‘reverse purification’ process, assignment of a component to a particular complex relied heavily on experimental stringency and arbitrary thresholds. Here we report the first genome-wide screen for complexes to investigate the underlying organizational principles of the eukaryotic cellular machinery.

Genome-wide characterization of complexes

We applied the tandem-affinity-purification method coupled to

mass spectrometry (TAP-MS)^{6–8} to all 6,466 ORFs of *Saccharomyces cerevisiae* as annotated in 2002 (refs 17, 18; Fig. 1 and Supplementary Information). We employed standardized protocols and successfully purified 1,993 unique TAP-fusion proteins, of which 88% retrieved at least one partner (Fig. 1; Supplementary Table S1). From all purifications, we processed 52,000 samples for mass spectrometry and identified 36,000 proteins, of which 2,760 were distinct (Fig. 1; Supplementary Figs S2–S5). These represent about 60% of the estimated proteome for exponentially growing yeast^{19–21}, and cover all functional classes and subcellular localizations. The absolute abundances of the identified proteins show a wide range, from 32 to 500,000 copies per cell¹⁹, although coverage varied considerably, being highest for the most abundant proteins (>16,000 copies per cell: 80% coverage), and lowest for the rarest proteins (<500 copies: 40% coverage) (Supplementary Fig. S1). We measured reproducibility by performing 139 purifications in duplicate (99 soluble; 40 membrane), and found that, on average, 69% of recovered proteins were common to both, giving an approximation of false-positive/negative rates within the raw data. However, as complexes are retrieved in several purifications, interactions observed repeatedly are more likely to be correct (see below).

The purification data contains 73% of known complexes from the Munich Information Center for Protein Sequences (MIPS) database²² (217 complexes) and our own literature mining (62 complexes). We found no evidence for 74 known complexes, possibly because they may not assemble under our growth conditions or because the tag interferes with complex assembly⁸. This is the case for the partially recovered CCT (chaperonin-containing tailless complex

¹Cellzome AG, Meyerhofstrasse 1, 69117 Heidelberg, Germany. ²EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany. ³MPI-MG, MPI-IB, Charité Campus Mitte, Schumannstrasse 21/22, 10117 Berlin, Germany. ⁴Center for Molecular Medicine of the Austrian Academy of Sciences, Lazarettgasse 19, 1090 Vienna, Austria. [†]Present address: EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany.

*These authors contributed equally to this work.

polypeptide 1) complex—the carboxy termini of the eight subunits in the ring-like core of the complex lie on interaction interfaces²³. However, these situations could often be rescued: 30% of TAP-tagged proteins that we could not purify were detected in purifications using other complex components.

We used a modified purification procedure for membrane proteins and successfully purified 340 of the 628 that were tagged. For example, we retrieved the Q/t-SNARE complex, including both integral membrane components of the trimeric receptor (Use1, Sec20 and Ufe1) and the peripheral membrane machinery (Dsl1, Sec39, Tip20) required for stability²⁴. We also detected novel links such as that between the Akr1 palmitoyl transferase (a six-transmembrane-segment protein) and Ste4 (the G β subunit of the pheromone receptor-coupled G protein), which is consistent with genetic evidence²⁵ and supports a role for protein acylation in the pheromone response.

De novo definition of protein complexes

The proportion of new proteins identified per purification dropped asymptotically during the progression of the screen, suggesting that the procedure was to near saturation (Supplementary Fig. S6a). We also observed that 64% of known complexes²² were retrieved several times resulting in a high coverage of known components (Supplementary Fig. S6b). We exploited this redundancy to define complexes computationally. Current approaches for defining complexes from binary interactions²⁶ were not deemed appropriate as these are not directly inferable from purifications. We also explicitly avoided the incorporation of prior knowledge to circumvent any bias towards well-studied proteins.

We first derived a ‘socio-affinity’ index (see the Methods) that quantifies the propensity of proteins to form partnerships. It measures the log-odds of the number of times two proteins are observed together, relative to what would be expected from their frequency in the data set, and encompasses both the ‘spoke’ and the ‘matrix’ models for assigning binary interactions within purifications. The index accounts for the frequency of proteins within

the data set and thus naturally discriminates true from spurious interactions involving very promiscuous partners. For instance, Vma2, which was seen in 552 purifications and would have been ignored under previous high-frequency filtering strategies^{8,9}, showed high indices only with proteins it is known to associate with (Vma5, Vma6, Vma10 and Rav1). Generally, pairs with socio-affinity indices below 5 should be considered with caution (reproducibility <70%), though those above 5 are more reliable (89%). These indices capture some biochemical properties of protein–protein interactions: there is a tentative correlation with the few dissociation constants available in the literature ($P < 0.08$) and protein pairs with high socio-affinity indices are more likely to be in direct contact as measured either by three-dimensional structures or the yeast two-hybrid system (Supplementary Fig. S7). To our knowledge, this is the first attempt to re-create numbers approximating physical measurements purely from proteomics data.

If each protein only belonged to a single complex, we could generate a definitive set by a single clustering step using socio-affinity indices. However, it is well established that proteins can be present in multiple complexes; a property we reasoned could be captured by an iterative procedure. Briefly, we first used the socio-affinity indices to form a matrix for all pairs of proteins studied, and then applied cluster analysis to generate an initial list of complexes. We then subtracted a penalty from the initial matrix values and repeated clustering. Tight associations are not drastically affected by the penalty, while looser ones are gradually eroded, and can be replaced by others not present initially. We varied the clustering parameters (number of iterations, clustering type, penalty values, and so on) over a sensible range to produce 1,784 different complex sets, and compared each to a manually curated group of known complexes used for structural analysis¹⁴. We computed both coverage (that is, the fraction of proteins in known complexes that we retrieved) and accuracy (that is, the fraction of the retrieved complexes components that match those already known; Fig. 1). The best conditions generated a collection of 491 complexes with 83% coverage and 78% accuracy. However, inspection revealed that known complex

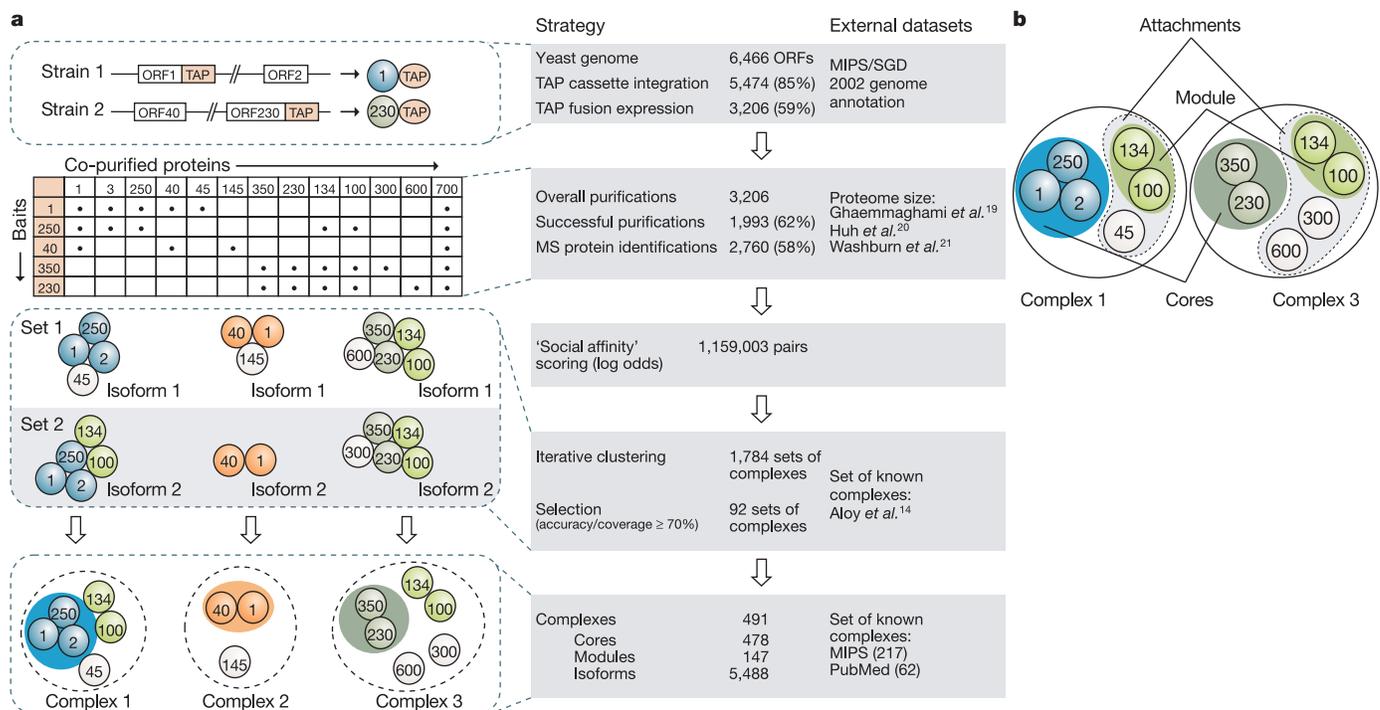


Figure 1 | Synopsis of the genome-wide screen for complexes and data analysis. a, Summary of the overall experimental strategy. MIPS/SGD, Munich Information Center for Protein Sequences/*Saccharomyces* Genome

Database. **b**, Definition and terminology used to define protein-complex architecture.

components could be found under clustering conditions with slightly poorer accuracy or coverage. Therefore, we grouped similar complexes from conditions with coverage and accuracy above 70%. The resulting 5,488 different protein-complex variations were termed ‘complex isoforms’ (Fig. 1). This procedure increased the overall coverage to 90%. The inclusion of parameters resulting in accuracy/coverage below 70% did not increase the coverage, but significantly decreased accuracy (data not shown).

Comparison with the complete collection of known complexes (279 from MIPS and the literature) showed that 257 of 491 complexes were entirely novel, and just 20 of those previously known lacked novel components (Supplementary Table S2). Of the known

complexes not recovered by the procedure above, 36 were partially found in single purifications (Supplementary Table S4) but produced a signal too weak to be recovered automatically.

Modular organization of the cell machinery

The above procedure partitions proteins in complexes into two types: core components that are present in most isoforms, and attachments present in only some of them (Fig. 1). This is reminiscent of an organization structure proposed previously that was based on a small-scale analysis²⁷. Complex cores ranged from 1–23 proteins in size (average 3.1 ± 2.5). Among the attachments, we noticed several instances where two or more proteins were always together and present in multiple complexes, which we call ‘modules’ (Supplementary Table S3; on average, associated with 3.3 ± 1.6 cores).

We tested whether this organization was a reflection of biological phenomena by first looking at transcriptional control of the complex components. A quality controlled set of 975 differentially expressed genes derived from microarray analyses¹⁵ showed that a large percentage of pairs of proteins within cores were coexpressed at the

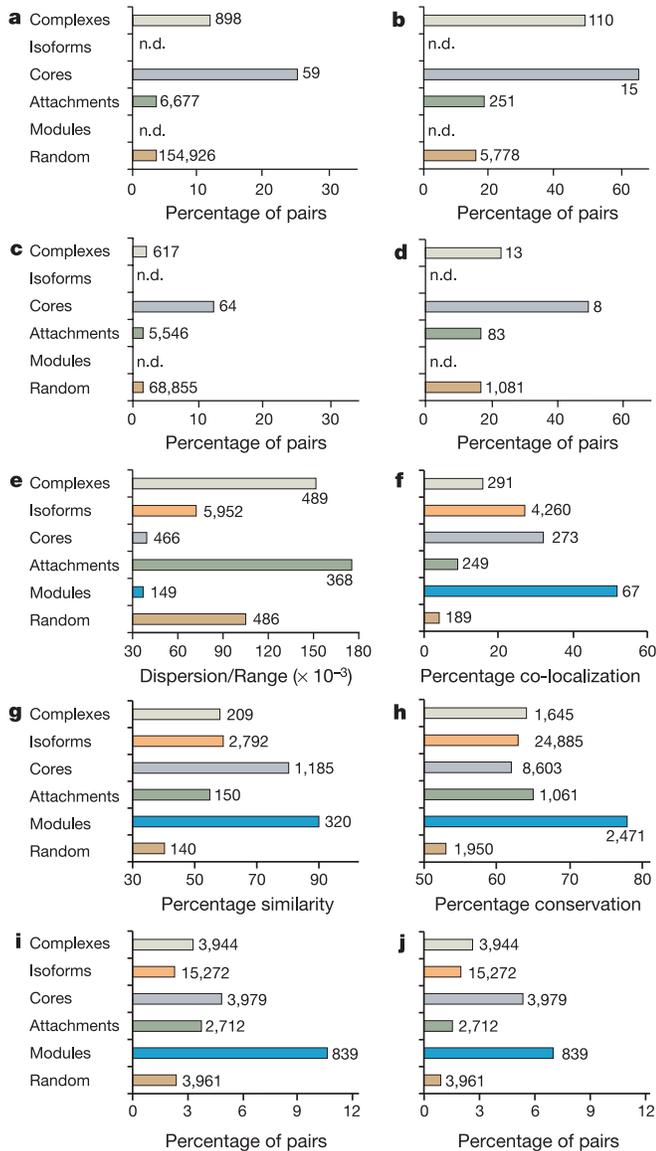


Figure 2 | Evidence supporting complex organization. Proteins in each organization level (cores, and so on) are referred to as groups. **a**, Percentage of cell cycle co-regulated genes found in the same group. **b**, Percentage of co-regulated proteins in the same group expressed at the same time during the cell cycle. **c**, **d**, are as for **a**, **b**, but for sporulation genes. **e**, Average dispersion ranges for protein abundance within each group. **f–h**, Percentage of groups having exactly the same subcellular localizations, cellular functions or phylogenetic conservation, respectively. **i**, **j**, Percentage of pairs for which a direct interaction is known from three-dimensional structures or yeast two-hybrid experiments, respectively. Values on each bar show the total number of counts; n.d., not determined. See Supplementary Information for further details.

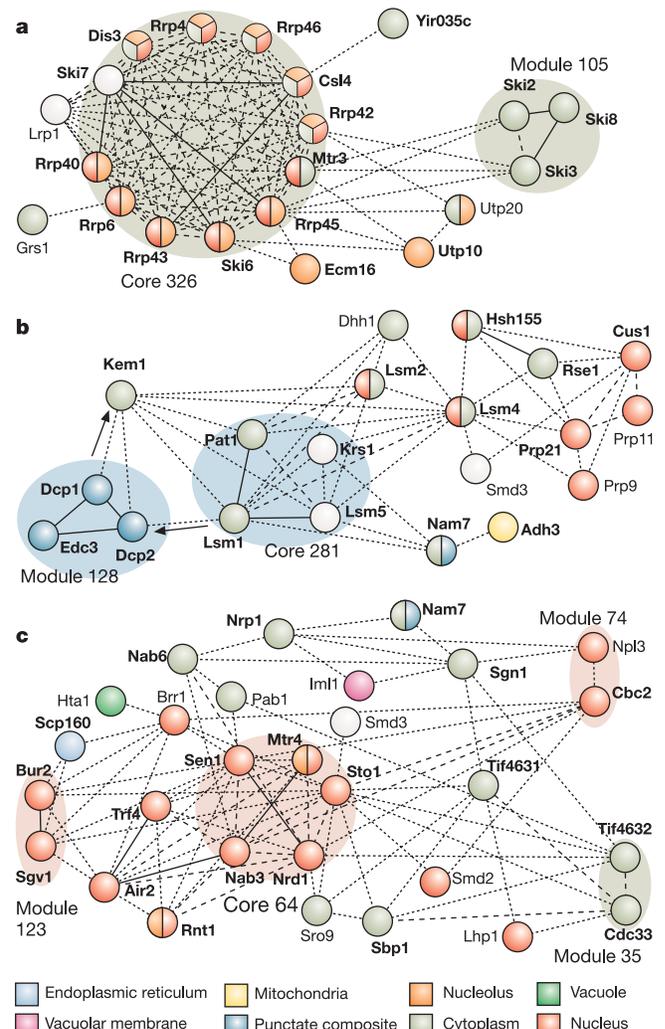


Figure 3 | Architecture and modularity of complexes. Proteins are coloured according to their localization²⁰. The line attribute corresponds to socio-affinity indices: dotted lines, 5–10; dashed lines, 10–15; plain lines, >15. Bait proteins are shown in bold and shaded circles around groups of proteins indicate cores and modules. **a**, The exosome and the Ski module. **b**, Stages in de-adenylation-dependent mRNA degradation; arrows show the order of events. **c**, Two distinct families of cap-binding proteins: the nuclear CBC (cap-binding complex) and the cytoplasmic eIF4F.

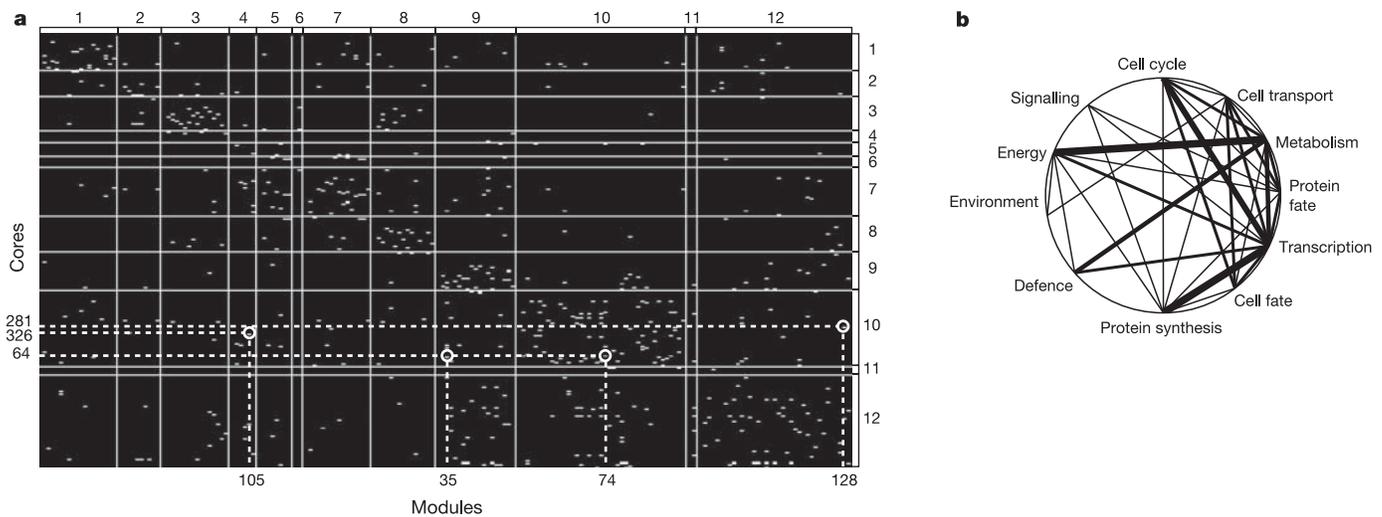


Figure 4 | Modularity of the yeast cellular machinery. **a**, Modularity matrix across cellular function. The *x* and *y* axes show modules and cores, respectively, clustered according to functional categories (1–12): cell cycle, cell fate, cell transport, defence, energy, environment, metabolism, protein fate, protein synthesis, transcription, signalling and unknown. Whenever a

module combines with a core the intersection is highlighted. Dotted lines show the modularity of the complexes in Fig. 3. **b**, Frequency of cross-talk between different cellular processes. The thickness of the lines between the functional classes are proportional to the frequency of core–module interactions between them.

same time during the cell cycle and sporulation (Fig. 2a–d), consistent with the view that core components represent functional units. Comparison with genome-wide protein abundance and localization studies^{19,20} revealed that cores and modules were also more likely to be expressed at a similar copy number (Fig. 2e) and to be co-localized in the cell (Fig. 2f). Notably, attachments showed a greater heterogeneity in expression levels than expected from random, supporting the notion that they might represent non-stoichiometric components. Cores and modules showed the greatest degree of similarity in terms of annotated function (Fig. 2g). When considering orthologous proteins in other species, cores and modules were least likely to be present partially: that is, if one component was present (or absent), the others usually were also (Fig. 2h). Finally, proteins within cores and modules were most likely to be in direct physical contact, as assessed both by three-dimensional structures (Fig. 2i) and the yeast two-hybrid system (Fig. 2j). Overall, the greatest degree of functional similarity and physical association was found between proteins within cores or modules, thus strongly supporting the model.

Examples of protein-complex architecture

The analysis was able to capture architectural details of known complexes. Attachments often specify a particular function for a complex. The exosome contains the complete Ski complex among its attachments (Fig. 3a), supporting previous reports that this association is required for cytoplasmic messenger RNA 3′-to-5′ decay²⁸. The modular architecture can also capture sequential events associated with pathways, providing a dynamic view of cellular processes. Complex 281 captured three discrete functional stages in de-adenylation-dependent RNA degradation (Fig. 3b). The core of the complex binds to de-adenylated mRNAs, a module (Edc3–Dcp1–Dcp2; known as the mRNA de-capping complex) removes the 5′ cap, and the attachment protein Kem1 (a 5′-3′ exonuclease) digests the RNA²⁹.

We identified 87 mutually exclusive modules in 48 complexes. Of these, 31 appeared to be related to differences in subcellular locations and might thus specify subtle differences in function. Among them, two mutually exclusive cap-binding modules were in different isoforms of complex 64 (Fig. 3c). The first, Tif4632–Cdc33 (or eIF4F), is cytoplasmic and essential for cap-dependent translation, while the second is nuclear and plays a direct role in pre-mRNA processing and export^{30,31}.

Other architectures hinted at novel regulatory mechanisms. Complex 437, formed around the yeast 14-3-3 protein Bmh2,

contained three metabolic enzymes involved in the heat stress response³²: Nth1, a neutral trehalase and the serine palmitoyl-transferase complex Lcb1–Lcb2. Nth1 contained three predicted 14-3-3-binding motifs and formed a core with Bmh2. The presence of Lcb1–Lcb2 as a module suggested the assembly of alternative complexes around Bmh2. A common control mechanism for Nth1 and Lcb1–Lcb2 might ensure the coordinated production of two metabolites central to the heat shock response—trehalose and sphingolipids. Similar coordinated control of metabolic enzymes through phosphorylation and subsequent binding to 14-3-3 is established in plants³³ and has recently been proposed for human cells³⁴.

A modularity matrix across functions

We derived a matrix representing a global view of the connections between cores and modules (Fig. 4a). There was a strong tendency for modules to combine with cores in the same functional category, suggesting coherence in our assignment of core and module composition. Using the ‘guilt-by-association’ principle, it is possible to suggest functions for modules. For example, the novel module 78 (Kre33 and Ygr145w) combined with several cores involved in ribosome biogenesis, suggesting a role in this process. Module 115 (Sgn1 and Ygr250c) associated with the translation initiation complex eIF4G, supporting previous genetic evidence for a role in RNA metabolism³⁵.

The degree of core–module cross-talk between functional categories (Fig. 4b) highlights many known connections, such as that between protein synthesis, transcription and the cell cycle, in addition to others less well established. For instance, the many links between metabolism and transcription are supported by recent findings of roles for metabolic enzymes in transcriptional regulation³⁶. Similarly, strong links between cell metabolism and defence argue for a re-evaluation of yeast metabolic pathways as targets for anti-fungal drug discovery.

Complexes as a scaffold for genetic data

Interaction networks have been used previously to study the effect of gene knockouts, for example showing that proteins central in networks tend to be lethal when deleted³⁷. More recently, studies have systematically monitored the effects of loss of function under a series of different conditions^{38,39} leading to phenotypic profiles, which are ideal for probing protein-complex architecture (Fig. 5). We found 20

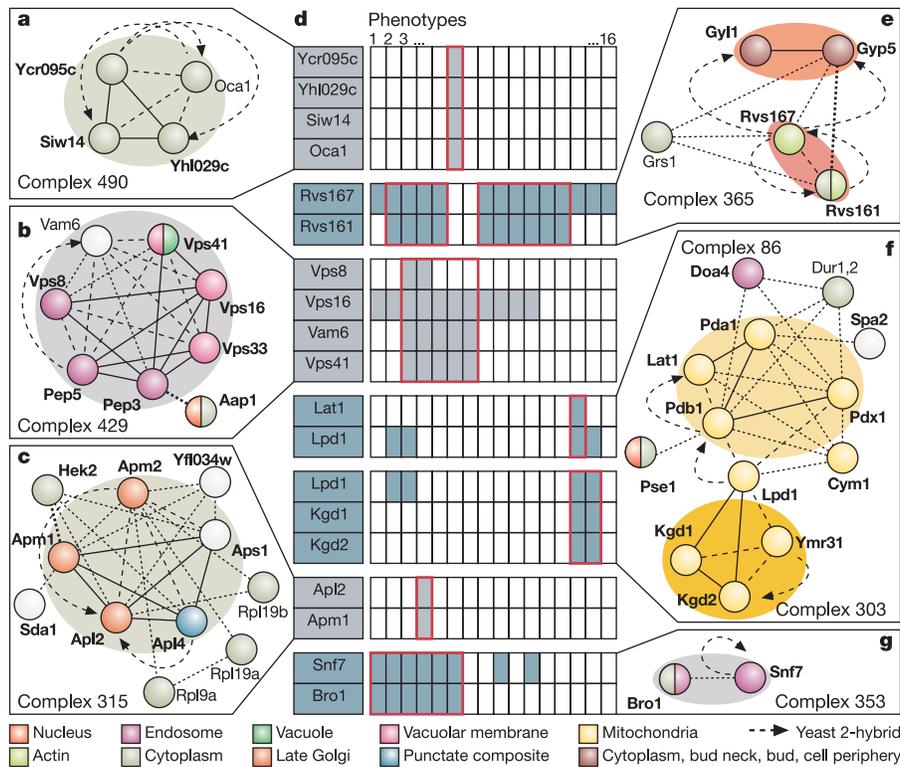


Figure 5 | Phenotypic data mapped to complexes. **a**, Novel complex 490; **b**, HOPS (homotypic fusion and vacuole protein sorting) complex⁴¹; **c**, API adaptor complex; **e**, Rvs161–Rvs167 amphiphysin-like complex and the module Gyl1–Gyp5⁴²; **f**, Pyruvate and α -ketoglutarate dehydrogenase complexes⁴³; **g**, Bro1–Snf7 complex. Details are as for Fig. 3. **d**, Phenotypic effect of deletion of complex components³⁸. Shaded cells indicate a growth defect (slow growth or no growth relative to the control); those boxed in red

represent the phenotypic signature of the complex. Similarities (mean number of phenotypes shared by components/total number of phenotypes) were calculated for 20 complexes. Sensitivity phenotypes (1–16): paraquat, ethanol, CdCl₂, hygromycin-B, CaCl₂, caffeine, rapamycin, cycloheximide, hydroxyurea, galactose, high salt, raffinose, glycerol, lactate, benomyl and low phosphate.

complexes with at least two proteins present in a data set of yeast phenotypes³⁸, of which 16 showed similar phenotypic patterns (Fig. 5d; random behaviour would predict only five). In one case, profile similarity supported the authenticity of a novel complex (Fig. 5a). In others, there is evidence that shared proteins play wider roles than the individual complexes they are part of. For example, the pyruvate and α -ketoglutarate dehydrogenase complexes show similar phenotypes, but the lipoamide dehydrogenase subunit (Lpd1) shared between them has other phenotypes, suggesting that it could have additional functions (Fig. 5f). These examples highlight the promise for the molecular machinery described here to provide a molecular rationale for gene-to-phenotype relationships.

Discussion

This analysis represents only a snapshot of the proteome averaged over all phases of the cell cycle. Nevertheless, this is the first screen for complexes run to saturation and, as such, it serves as a guide for the future exploration of protein interactions under other physiological states. For example, we do not expect protein-complex cores to vary extensively under different conditions, whereas we expect significant changes to occur in attachment proteins. Extrapolation based on the fraction of known complexes recovered suggests that there may be an additional 300 core machines, leading to a total of 800 in yeast. In a rough approximation, based on the ratio of gene numbers between species, we estimate some 3,000 core human complexes.

The number of protein-complex cores is small compared to the many cellular processes mediated by them, and shuffling functional modules provides an efficient means to multiply functionality and simplify temporal and spatial regulation. The modularity is highly reminiscent of that seen elsewhere in nature, for example the combinatorial use of amino acids to build polypeptides, or domains

to create proteins with complex biochemical properties. Modularity might very well represent a general attribute of living matter, with *de novo* invention being rare and reuse the norm.

Genome sequencing and functional genomics have provided a parts-list and partial knowledge of how these parts are arranged in space and time. The next challenge is to integrate these data into rational models of entire systems. Our analysis makes some first steps in this direction, providing a collection of individual integrative subsystems—the machines—but also a view on how they might coordinate cellular functions through sharing functional modules. As such, it may be a very useful platform for systems biology and indeed new applications in nano- and synthetic-biology that seek to re-engineer the cellular machinery towards new processes.

METHODS

Experimental procedures. We created a library of strains with TAP-tag cassettes at the 3' end of each ORF by homologous recombination. We prepared protein extracts from exponentially growing haploid yeast strains grown in 21 of complete medium. Tandem-affinity purification (TAP)–mass spectrometry (MS) characterization of complexes was performed as previously described⁸. For membrane proteins, we used a special protocol provided as Supplementary Information.

Socio-affinity and iterative clustering to generate protein-complex sets. We defined a socio-affinity index ($A(i,j)$) that quantifies the tendency for proteins to identify each other when tagged (the spoke model, S) and to co-purify when other proteins are tagged (the matrix model, M)⁴⁰:

$$A(i,j) = S_{i,j|i=bait} + S_{i,j|j=bait} + M_{i,j}; \quad S_{i,j|i=bait} = \log \left(\frac{n_{i,j|i=bait}}{f_i^{bait} n_{bait} f_j^{prey} n_{i=bait}^{prey}} \right)$$

$$M_{i,j} = \log \left(\frac{n_{i,j}^{prey}}{f_i^{prey} f_j^{prey} \sum_{\text{all baits}} n_{prey} (n_{prey} - 1) / 2} \right)$$

For the spoke model terms (S), $n_{i,j|i=\text{bait}}$ is the number of times that protein i retrieves j when i is tagged; f_i^{bait} is the fraction of purifications where protein i was bait; f_j^{prey} is the fraction of all retrieved preys that were protein j ; n_{bait} is the total number of purifications (that is, baits); and $n_{i=\text{bait}}^{\text{prey}}$ is the number of preys retrieved with protein i as bait. For the matrix model term (M), $n_{i,j}^{\text{prey}}$ is the number of times that proteins i and j are seen in purifications with baits other than i or j ; f_i^{prey} and f_j^{prey} are as above; and n_{prey} is the number of preys observed with a particular bait (excluding itself).

We used socio-affinity indices to populate the upper-diagonal of a pair-wise matrix (that is, one value for each pair of proteins in the data set). We assigned a value of zero to all pairs of proteins that had never been seen together. We generated a first set of clusters using the OC program (G. Barton, University of Dundee) and then subtracted a penalty from each pair-wise value associated with the set. We then repeated the cluster generation a number of times, each time adding any new clusters to a growing list. To generate different sets of complexes using this procedure, we varied the number of iterations (2–10), the socio-affinity threshold to define clusters (1–10), the penalty value (0.5, 1 or 2), and the type of clustering (UPGMA, single or complete linkage).

Received 17 October; accepted 15 December 2005.

Published online 22 January 2006.

- Hood, L., Heath, J. R., Phelps, M. E. & Lin, B. Systems biology and new technologies enable predictive and preventative medicine. *Science* **306**, 640–643 (2004).
- Alberts, B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291–294 (1998).
- Goh, C. S., Milburn, D. & Gerstein, M. Conformational changes associated with protein–protein interactions. *Curr. Opin. Struct. Biol.* **14**, 104–109 (2004).
- Kemmeren, P. *et al.* Protein interaction verification and functional annotation by integrated analysis of genome-scale data. *Mol. Cell* **9**, 1133–1143 (2002).
- Edwards, A. M. *et al.* Bridging structural biology and genomics: assessing protein interaction data with known complexes. *Trends Genet.* **18**, 529–536 (2002).
- Puig, O. *et al.* The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**, 218–229 (2001).
- Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**, 1030–1032 (1999).
- Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
- Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
- Bouwmeester, T. *et al.* A physical and functional map of the human TNF- α /NF- κ B signal transduction pathway. *Nature Cell Biol.* **6**, 97–105 (2004).
- Butland, G. *et al.* Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433**, 531–537 (2005).
- Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
- Rual, J. F. *et al.* Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **437**, 1173–1178 (2005).
- Aloy, P. *et al.* Structure-based assembly of protein complexes in yeast. *Science* **303**, 2026–2029 (2004).
- de Lichtenberg, U., Jensen, L. J., Brunak, S. & Bork, P. Dynamic complex formation during the yeast cell cycle. *Science* **307**, 724–727 (2005).
- Kelley, R. & Ideker, T. Systematic interpretation of genetic interactions using protein networks. *Nature Biotechnol.* **23**, 561–566 (2005).
- Mewes, H. W. *et al.* MIPS: A database for genomes and protein sequences. *Nucleic Acids Res.* **30**, 31–34 (2002).
- Kumar, A. *et al.* An integrated approach for finding overlooked genes in yeast. *Nature Biotechnol.* **20**, 58–63 (2002).
- Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).
- Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
- Washburn, M. P., Wolters, D. & Yates, J. R. III Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.* **19**, 242–247 (2001).
- Mewes, H. W. *et al.* MIPS: Analysis and annotation of proteins from whole genomes. *Nucleic Acids Res.* **32**, D41–D44 (2004).
- Liou, A. K. & Willison, K. R. Elucidation of the subunit orientation in CCT (chaperonin containing TCP1) from the subunit composition of CCT micro-complexes. *EMBO J.* **16**, 4311–4316 (1997).
- Kraynack, B. A. *et al.* Dsl1p, Tip20p, and the novel Dsl3(Sec39) protein are required for the stability of the Q/t-SNARE complex at the endoplasmic reticulum in yeast. *Mol. Biol. Cell* **16**, 3963–3977 (2005).
- Kao, L. R., Peterson, J., Ji, R., Bender, L. & Bender, A. Interactions between the ankyrin repeat-containing protein Akrlp and the pheromone response pathway in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **16**, 168–178 (1996).
- Bader, G. D. & Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4**, 2 (2003).
- Dezso, Z., Oltvai, Z. N. & Barabasi, A. L. Bioinformatics analysis of experimentally determined protein complexes in the yeast *Saccharomyces cerevisiae*. *Genome Res.* **13**, 2450–2454 (2003).
- Orban, T. I. & Izaurralde, E. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA* **11**, 459–469 (2005).
- Sheth, U. & Parker, R. Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* **300**, 805–808 (2003).
- Fortes, P. *et al.* The yeast nuclear cap binding complex can interact with translation factor eIF4G and mediate translation initiation. *Mol. Cell* **6**, 191–196 (2000).
- McKendrick, L., Thompson, E., Ferreira, J., Morley, S. J. & Lewis, J. D. Interaction of eukaryotic translation initiation factor 4G with the nuclear cap-binding complex provides a link between nuclear and cytoplasmic functions of the m⁷ guanosine cap. *Mol. Cell Biol.* **21**, 3632–3641 (2001).
- Jenkins, G. M. & Hannun, Y. A. Role for *de novo* sphingoid base biosynthesis in the heat-induced transient cell cycle arrest of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **276**, 8574–8581 (2001).
- Sehnke, P. C. & Ferl, R. J. Plant 14-3-3s: Omnipotent metabolic phosphopartners? *Sci. STKE* **2000**, PE1 (2000) (doi:10.1126/stke.2000.56.pe1).
- Pozuelo Rubio, M. *et al.* 14-3-3-affinity purification of over 200 human phosphoproteins reveals new links to regulation of cellular metabolism, proliferation and trafficking. *Biochem. J.* **379**, 395–408 (2004).
- Winstall, E., Sadowski, M., Kuhn, U., Wahle, E. & Sachs, A. B. The *Saccharomyces cerevisiae* RNA-binding protein Rbp29 functions in cytoplasmic mRNA metabolism. *J. Biol. Chem.* **275**, 21817–21826 (2000).
- Shi, Y. & Shi, Y. Metabolic enzymes and coenzymes in transcription—a direct link between metabolism and transcription? *Trends Genet.* **20**, 445–452 (2004).
- Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
- Dudley, A. M., Janse, D. M., Tanay, A., Shamir, R. & Church, G. M. A global view of pleiotropy and phenotypically derived gene function in yeast. *Mol. Syst. Biol.* published online 29 March 2005 (doi:10.1038/msb4100004).
- Said, M. R., Begley, T. J., Oppenheim, A. V., Lauffenburger, D. A. & Samson, L. D. Global network analysis of phenotypic effects: protein networks and toxicity modulation in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **101**, 18006–18011 (2004).
- Bader, G. D. & Hogue, C. W. Analyzing yeast protein–protein interaction data obtained from different sources. *Nature Biotechnol.* **20**, 991–997 (2002).
- Subramanian, S., Woolford, C. A. & Jones, E. W. The Sec1/Munc18 protein, Vps33p, functions at the endosome and the vacuole of *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **15**, 2593–2605 (2004).
- 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* **170**, 555–568 (2005).
- Ross, J., Reid, G. A. & Dawes, I. W. The nucleotide sequence of the *LPD1* gene encoding lipamide dehydrogenase in *Saccharomyces cerevisiae*: comparison between eukaryotic and prokaryotic sequences for related enzymes and identification of potential upstream control sites. *J. Gen. Microbiol.* **134**, 1131–1139 (1988).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Cohen, S. Artavanis-Tsakonas, B. Seraphin and L. Serrano for support and suggestions throughout the work, and F. Weisbrodt for assistance with the graphics.

Author Information Purification and complex data have been deposited at the IntAct database (<http://www.ebi.ac.uk/intact/>) with accession numbers EBI-768904 (purifications) and EBI-765905 (author inferred complexes). The data, including the MS protein identifications, are accessible at <http://yeast-complexes.embl.de>, and the yeast strains are available from Euroscarf (http://web.uni-frankfurt.de/fb15/mikro/euroscarf/col_index.html). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.S.-F. (gsupert@cemm.oeaw.ac.at), R.B.R. (russell@embl-heidelberg.de).