was a consistent outlier for all genes, data from all pools were included in our analysis. Differences between parental pools increase our confidence intervals for the interspecific expression difference and can reduce our power for detecting *trans*-regulatory differences. To deal with this issue, interspecific expression differences for genes with unusually large variance between the four original pools were measured in an additional three parental pools.

The normalization procedure used to correct cDNA measurements for experimental bias (Supplementary Information) prohibits a standard nested analysis of variance, but a *t*-test provided a simple and robust test of our null hypotheses. Two-tailed *t*-tests were used to identify *cis*-regulatory divergence ($H_0$: $Mel_{F1}/Sim_{F1} = 1$), interspecific expression differences ($H_0$: Mel/Sim = 1) and parent-of-origin effects ($H_0$: $Mel_{F1}/Sim_{F1} = Mel_{F1}/Sim_{F1}$ in reciprocal crosses). To identify *trans*-regulatory divergence, two-sided *t*-tests (with the Cochran correction for unequal variances) and nonparametric Mann–Whitney *U*-tests were used to compare relative expression between hybrid and parental pools. The decision to accept or reject the null hypothesis ($H_0$: $Mel_{F1}/Sim_{F1}$ = Mel/Sim) was the same for both tests for all except three genes, and *t*-test significance was ultimately used to infer *trans*-regulatory divergence. All statistical analyses were preformed with SAS software v. 8.2 (SAS Institute, Cary, North Carolina) and are shown in Supplementary Table 2.

1. Davidson, E. H. *Genomic Regulatory Systems: Development and Evolution* (Academic, San Diego, 2001).
2. Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Blackwell Science, Oxford, 2001).
3. Wray, G. A. *et al.* The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20**, 1377–1419 (2003).
4. Cowles, C. R., Hirschhorn, J. N., Altshuler, D. & Lander, E. S. Detection of regulatory variation in mouse genes. *Nature Genet.* **32**, 432–437 (2002).
5. Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B. & Kinzler, K. W. Allelic variation in human gene expression. *Science* **297**, 1143 (2002).
6. Powell, J. R. *Progress and Prospects in Evolutionary Biology: The Drosophila Model* (Oxford Univ. Press, New York, 1997).
7. Pastinen, T. *et al.* A survey of genetic and epigenetic variation affecting human gene expression. *Physiol. Genomics* **16**, 184–193 (2004).
8. Ahmadian, A. *et al.* Single-nucleotide polymorphism analysis by pyrosequencing. *Anal. Biochem.* **280**, 103–110 (2000).
9. Neve, B. *et al.* Rapid SNP allele frequency determination in genomic DNA pools by pyrosequencing. *Biotechniques* **32**, 1138–1142 (2002).
10. Rifkin, S. A., Kim, J. & White, K. P. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nature Genet.* **33**, 138–144 (2003).
11. Ranz, J. M., Castillo-Davis, C. I., Meiklejohn, C. D. & Hartl, D. L. Sex-dependent gene expression and evolution of the *Drosophila transcriptome*. *Science* **300**, 1742–1745 (2003).
12. Davidson, E. H., McClay, D. R. & Hood, L. Regulatory gene networks and the properties of the developmental process. *Proc. Natl Acad. Sci. USA* **100**, 1475–1480.
13. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29 (2000).
14. The FlyBase Consortium, The FlyBase database of the *Drosophila* genome projects and community literature. *Nucleic Acids Res.* **31**, 172–175 (2003).
15. Yvert, G. *et al.* *Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nature Genet.* **35**, 57–64 (2003).
16. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755 (2002).
17. Schadt, E. E. *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297–302 (2003).
18. Montooth, K. L., Marden, J. H. & Clark, A. G. Mapping determinants of variation in energy metabolism, respiration and flight in *Drosophila*. *Genetics* **165**, 623–635 (2003).
19. Davis, A. W. *et al.* Rescue of hybrid sterility in crosses between *D. melanogaster* and *D. simulans*. *Nature* **380**, 157–159 (1996).
20. Sawamura, K., Yamamoto, M. T. & Watanabe, T. K. Hybrid lethal systems in the *Drosophila melanogaster* species complex. II. The *Zygotic hybrid rescue* (*Zhr*) gene of *D. melanogaster*. *Genetics* **133**, 307–313 (1993).
21. Hollocher, H., Agopian, K., Waterbury, J., O'Neill, R. W. & Davis, A. W. Characterization of defects in adult germline development and oogenesis of sterile and rescued female hybrids in crosses between *Drosophila simulans* and *Drosophila melanogaster*. *J. Exp. Zool.* **288**, 205–218 (2000).
22. Barbash, D. A., Roote, J. & Ashburner, M. The *Drosophila melanogaster* hybrid male rescue gene causes inviability in male and female species hybrids. *Genetics* **154**, 1747–1771 (2000).
23. Otto, P., Kephart, D., Bitner, R., Huber, S. & Volkerding, K. Separate isolation of genomic DNA and total RNA from single samples using the SV Total RNA Isolation System. *Promega Notes* **69**, 19–24 (1998).

---

# Evidence for dynamically organized modularity in the yeast protein–protein interaction network

**Jing-Dong J. Han**[1], **Nicolas Bertin**[1], **Tong Hao**[1], **Debra S. Goldberg**[2], **Gabriel F. Berriz**[2], **Lan V. Zhang**[2], **Denis Dupuy**[1], **Albertha J. M. Walhout**[1]*, **Michael E. Cusick**[1], **Frederick P. Roth**[2] & **Marc Vidal**[1]

[1]*Center for Cancer Systems Biology and Department of Cancer Biology, Dana-Farber Cancer Institute, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA*
[2]*Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA*

* Present address: Program in Gene Function and Expression, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA

**In apparently scale-free protein–protein interaction networks, or 'interactome' networks**[1,2]**, most proteins interact with few partners, whereas a small but significant proportion of proteins, the 'hubs', interact with many partners. Both biological and non-biological scale-free networks are particularly resistant to random node removal but are extremely sensitive to the targeted removal of hubs**[1]**. A link between the potential scale-free topology of interactome networks and genetic robustness**[3,4]** seems to exist, because knockouts of yeast genes**[5,6]** encoding hubs are approximately threefold more likely to confer lethality than those of non-hubs**[1]**. Here we investigate how hubs might contribute to robustness and other cellular properties for protein–protein interactions dynamically regulated both in time and in space. We uncovered two types of hub: 'party' hubs, which interact with most of their partners simultaneously, and 'date' hubs, which bind their different partners at different times or locations. Both *in silico* studies of network connectivity and genetic interactions described *in vivo* support a model of organized modularity in which date hubs organize the proteome, connecting biological processes—or modules**[7]**—to each other, whereas party hubs function inside modules.**

The biological role of topological hubs, so far considered in static representations of interactome networks without information on the functional states of these networks—that is, dynamic or steady state[8]—might vary depending on the timing and location of the interactions they mediate (Fig. 1a). Because accurate temporal parameters are not yet available for many protein–protein interactions, we estimated temporal characteristics of hubs and their partners by using compilations of yeast messenger RNA expression profiling data[9].

Hubs connected by false-positive interactions[10] would be uncorrelated in mRNA expression with their interaction partners[9,11], and would resemble date hubs. To minimize false positives, we first generated a high-quality yeast interaction data set by intersecting data generated by several different interaction detection methods (see Methods). The resulting 'filtered yeast interactome' (FYI) data set contains 2,493 high-confidence interactions, each observed by at least two different methods (Supplementary Fig. 1). FYI is a high-quality network enriched for genuine positives (Supplementary Information and Supplementary Fig. 2). The FYI network contains 1,379 proteins with an average degree of 3.6 interactions per protein and a large connected component of 778 proteins. Its degree distribution follows the power law that characterizes scale-free networks (Supplementary Fig. 3). FYI hubs were characterized with an expression-profiling compendium of 315 data points for most yeast genes across five different experimental conditions (referred to below as the 'yeast expression compendium'[9]). For each hub we calculated the average of Pearson

correlation coefficients between the hub and each of its respective partners for mRNA expression (see Methods). Strikingly, the average PCCs of hubs, defined as nodes (proteins) with degree $k$ greater than 5, follow a bimodal distribution in the whole compendium (Fig. 1b, red curve). In contrast, the average PCCs of non-hubs, defined as nodes with degree $k$ of 5 or less, show a normal distribution centred on 0.1 (Fig. 1b, cyan curve, and Supplementary Fig. 4). In randomized interactome networks of the same topology (Supplementary Methods), the average PCCs of hubs also show a normal distribution centred on 0 (Fig. 1b, black curve). This bimodal distribution suggests that hubs can be split into two distinct populations: one with relatively high average PCCs (party hubs) and the other with relatively low average PCCs (date hubs) (Supplementary Information). Futhermore, the bimodal distribution suggests a natural boundary for separating or partitioning date hubs from party hubs.

Party and date hubs were analysed for each individual condition of the yeast expression compendium. Average PCCs of FYI hubs show a clear bimodal distribution for two conditions: 'stress response' and 'cell cycle' (containing 174 and 77 data points, respectively) (Fig. 1b). The three remaining conditions ('phero-mone treatment', 'sporulation' and 'unfolded protein response') contain fewer data points (45, 10 and 9, respectively), which may explain the absence of a clear bimodal distribution (Fig. 1b). For our subsequent analyses, party hubs are those with an average PCC higher than the threshold indicated by the arrow, in at least one of the five conditions in Fig. 1b (exact cutoffs used in Methods and Supplementary Information). All other hubs were defined as date hubs. Using these criteria, we found 91 date hubs and 108 party hubs in FYI (Supplementary Table 1) after excluding ribosomal hub proteins (Supplementary Information).

Highlighting the biological significance of our date/party hub partitioning is the fact that average PCC values correctly predict the expected date versus party behaviour for several well-charac-terized protein hubs (Supplementary Table 1 and Supplementary Information). The dynamics of interactome networks should be considered not only by expression timing but also spatial distribution—that is, subcellular localization. We estimated the localization diversity of partners of hubs by using a proteome-wide cellular localization data set[12]. Partners of date hubs are significantly more diverse in spatial distribution than partners of party hubs (protein localization diversity evaluated by entropy
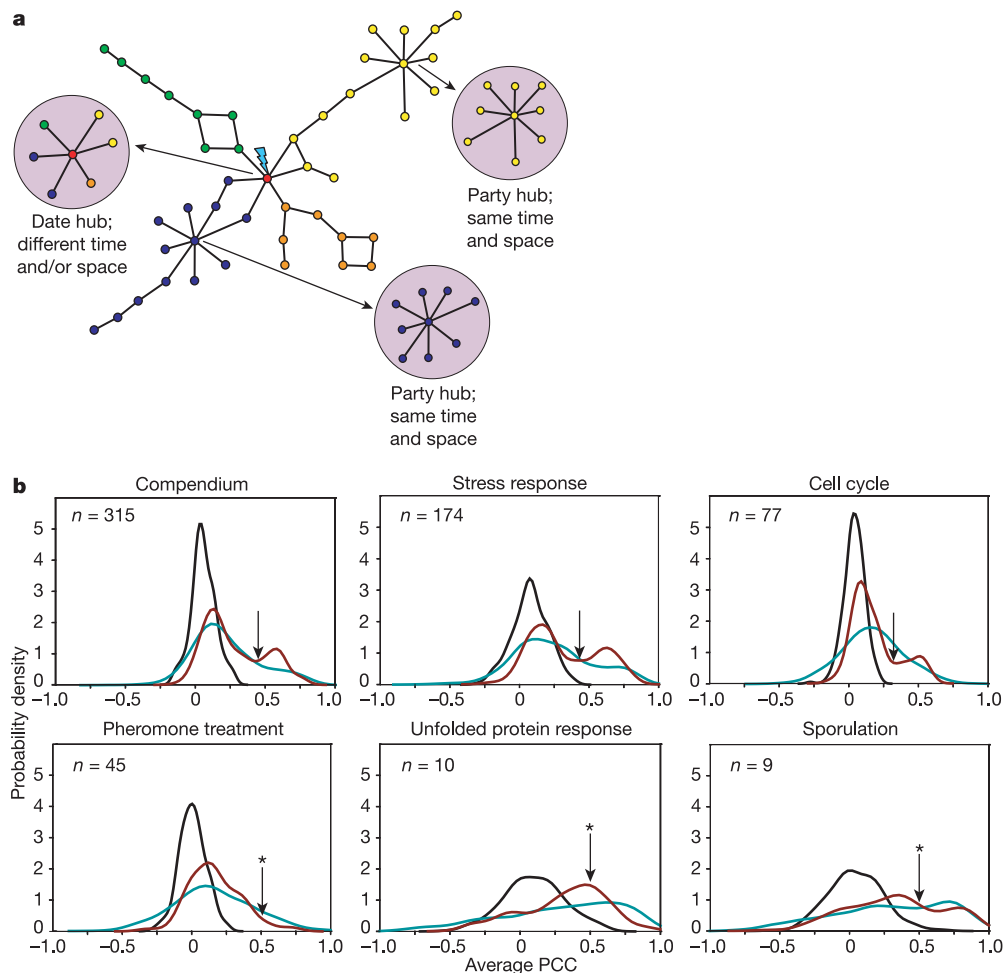


**Figure 1** Date and party hubs. **a**, In this schematic protein interaction network, proteins are coloured according to mutual similarity in their mRNA expression patterns. 'Party' hubs are highly correlated in expression with their partners, and presumably interact with them at similar times. The partners of 'date' hubs exhibit more limited co-expression, and presumably the corresponding physical interactions occur at different times and/or different locations. **b**, Probability densities (Supplementary Methods) of the average PCCs were calculated from a global expression profiling compendium[9] (top left panel). Average PCCs were also independently calculated for each condition constituting the compendium. The number $n$ in each panel refers to the number of data points for each gene for each condition. Average PCCs for hubs in the FYI (red curve) show a clear bimodal distribution that is used to separate date and party hubs (located by the arrow) for the conditions shown in the top panels. For the conditions in the bottom panels that do not show a clear bimodal distribution, an arbitrary average PCC cutoff of 0.5 was used (details in Methods and Supplementary Information). No bimodal distribution is observed with the average PCCs of non-hub proteins (cyan curve) or for hubs in randomized networks (black curve).

calculation; Student's *t*-test $P < 0.05$). Hence, the distinction between date and party hubs obtained from gene expression is recapitulated by protein localization data.

When removed from the interactome network, party and date hubs have distinct effects on the overall topology. We used an *in silico* strategy[13] that simulates the effect of specifically removing (attacking) hubs in the FYI network on the characteristic path length of the main component of the network. The characteristic path length, defined as the average distance (shortest path length) between node pairs, reflects the overall network connectivity[13]. As expected, successive attacks against FYI hubs, starting from the most con-

nected hubs, without distinguishing between party and date hubs, have a significantly more deleterious effect on the network integrity than the removal of random proteins (failure)[13] (Fig. 2a, b). However, this *in silico* experiment revealed an unexpected and striking difference between party and date hubs. Removal of party hubs does not affect connectivity and thus resembles failures (Fig. 2a, b), whereas attacks directed against date hubs account for a vast majority of the effect observed when attacking all hubs (Fig. 2a, b).

To rule out the possibility that small variations of local topology cause the differences observed above, we performed additional
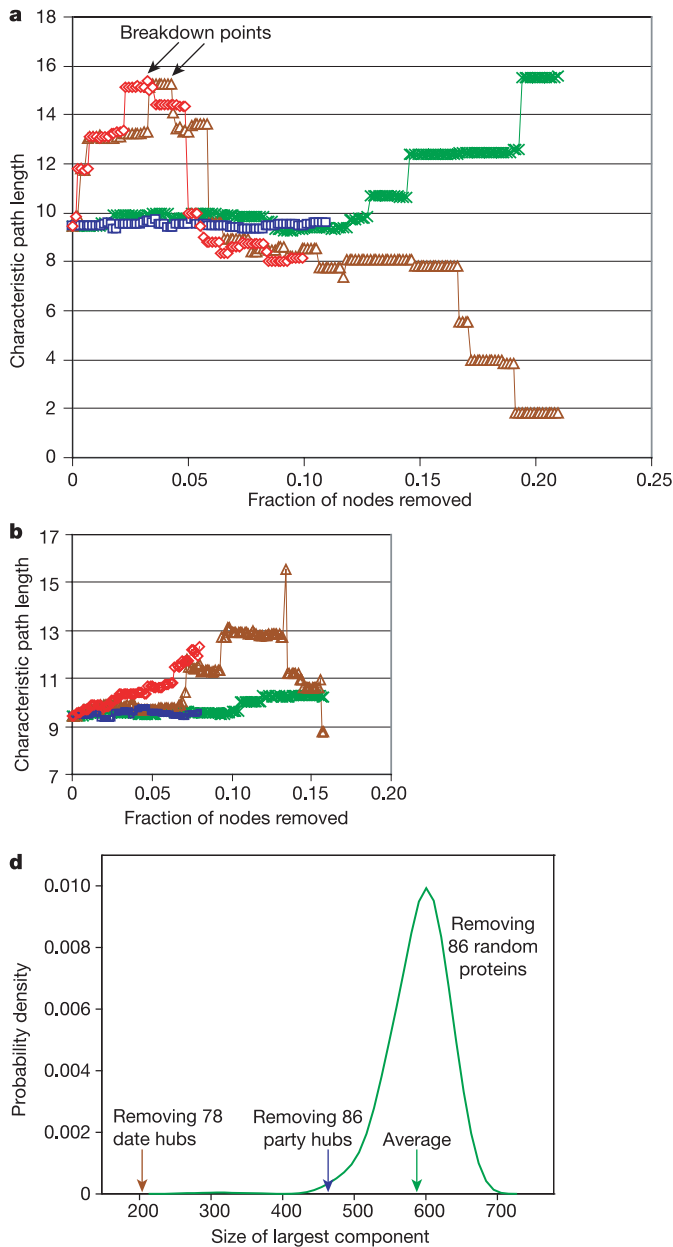


**Figure 2** Date hubs are central to network topology. **a**, The effects on the characteristic path length of the network on gradual node removal. Random removal of nodes ('failures') is represented by the green line, attacks against all hubs by the brown line, attacks against party hubs by the blue line, and attacks against date hubs by the red line. The 'breakdown point' is the threshold after which the main component of the network starts disintegrating. **b**, Subsets of date and party hubs with comparable degree ($k$) and clustering coefficients ($C_v$)[18] were selected for attack (lines are coloured as in **a**). **c**, The main component of the FYI network (top panel) splits into small subnetworks (middle panel) after the removal of date hubs, whereas it stays almost intact after the removal of party hubs (bottom panel). **d**, The sizes of the largest remaining component after removing all 78 date hubs (red arrow), all 86 party hubs (blue arrow) or 86 randomly selected proteins (green curve). This last experiment has been repeated 1,000 times to determine the empirical *P* values of recovering sizes similar to those obtained upon removal of date and party hubs (empirical *P* values are equal to $6 \times 10^{-3}$ and less than $10^{-3}$, respectively).

simulations by attacking FYI with subsets of date and party hubs that show comparable values of $C_v$ (clustering coefficient, a measure of neighbourhood density) and $k$ (number of interaction partners) (see Methods). The differences between date and party hubs were similar to those noted above (Fig. 2b and Supplementary Information). Thus, date and party hubs have markedly different global properties in the interactome network.

The main component that remains after the removal of party hubs is significantly larger than that remaining after the removal of date hubs (Fig. 2c, d, and Supplementary Fig. 5). Conversely, the subnetworks released by date hub removals tend to be larger in size and number than those obtained by party hub removals (Fig. 2c). To test whether FYI subnetworks obtained after the complete removal of date hubs corresponds to small interaction maps of specific biological processes, or modules[7], we estimated their functional homogeneity by using annotations from the Munich Information Center for Protein Sequences (MIPS) database[14]. In comparison with control networks of the same size distribution, most FYI subnetworks were more homogeneous in function (Fig. 3a, Supplementary Information). We could assign a 'most likely' function for each subnetwork by determining the most enriched function category among all nodes over the entire FYI data set (Supplementary Table 2). Thus, subnetworks derived from the FYI by using a definition based on non-biased functional states, and not biased by topology alone, often correspond to known biological modules.

Subnetworks represent not only stable molecular machines or complexes (for example the ribosomal RNA synthesis complex) but also more loosely connected regulatory pathways (for example osmosensing). Indeed, subnetworks have a broad range of average values of PCCs between all protein pairs involved (Fig. 3b). Protein pairs inside subnetworks corresponding to protein complexes tend to show high PCC values, whereas less densely connected regulatory pathway modules tend to show lower PCC values (Fig. 3b).

These results support a model of organized modularity for the yeast proteome, as illustrated when date hubs are reconnected to the modular subnetworks (module compositions at http://vidal.dfci.harvard.edu/fyi/moduleNet.pl; Fig. 4a). In this model, date hubs represent global, or 'higher level'[15], connectors between modules, and party hubs function inside modules, at a 'lower level'[15] of the organization of the proteome. For example, the date hub calmodulin (Cmd1) connects four different biological modules, 'homeostasis of cations', 'protein folding and stabilization', 'budding, cell polarity and filament formation' and 'endoplasmic reticulum', whereas the party hubs Sec17, Sec 22 and Vti1 all function within the 'endoplasmic reticulum' module (Fig. 4a, inset).

The organized modularity model predicts that experimental perturbations of date hubs *in vivo* should confer different effects from perturbations of party hubs. In single-gene knockout experiments[5,6], similar proportions of party and date hubs score as essential (Fig. 4b). Although party hubs tend to mediate their role locally within modules, they can still mediate unique functions in essential modules and thus score as essential genes, explaining the similar essentiality rate between date and party hubs.

In contrast, genetic perturbations of date hubs tend to sensitize the proteome to other perturbations, more so than perturbations of party hubs (Fig. 4c). Among all genetic interactions published by individual laboratories and curated in MIPS[14], genetic interactions involving date hubs are twice as prevalent as those involving party hubs or only non-hub proteins ($P < 10^{-5}$) (Fig. 4c and Supplementary Methods). Assuming that date hubs are not more likely to be studied than party hubs, the higher rate of observed genetic interactions for date hubs suggests that they have a central role in organizing the modularity of the yeast proteome. Conversely, the lower rate of observed genetic interactions for party hubs reflects their localized role within isolated, encapsulated regions of the interactome.
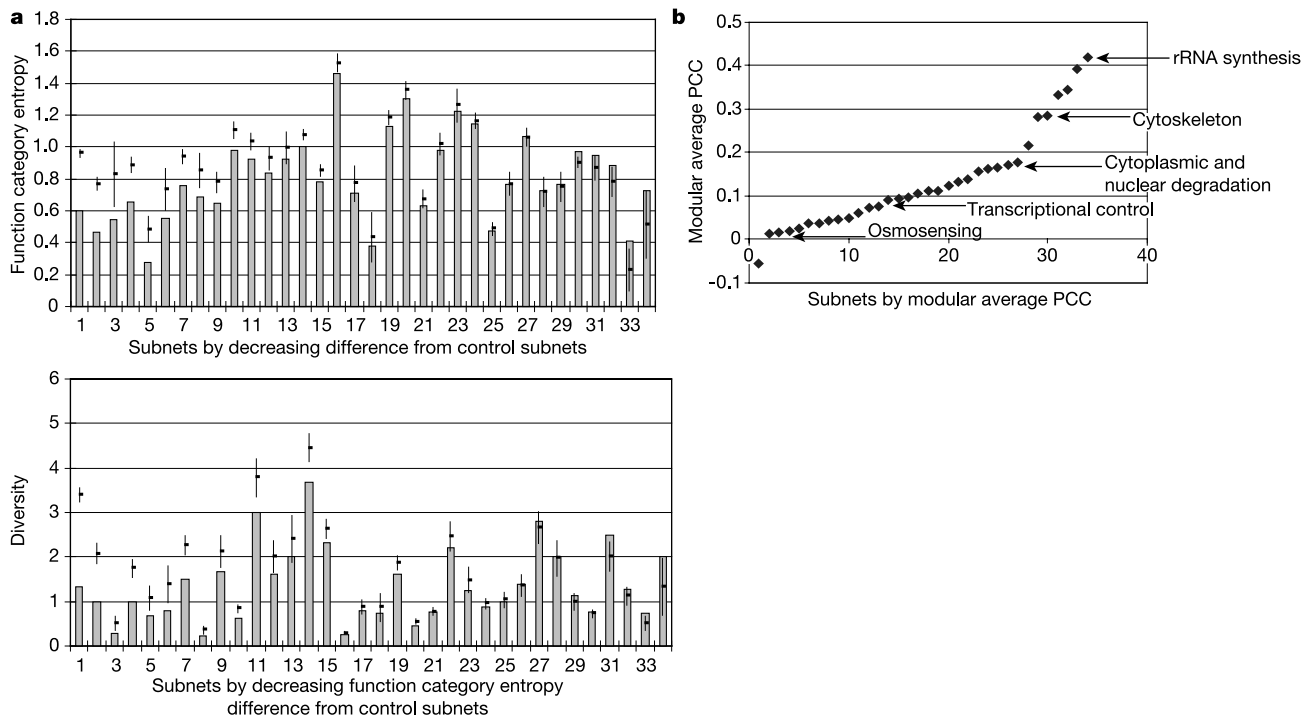


**Figure 3** Properties of subnetworks. **a**, Most subnetworks generated by removing date hubs represent functionally homogeneous modules. Grey bars represent the function category entropy (upper panel) or diversity (lower panel) of the subnets revealed by removing date hubs from the original largest component. Lower entropy means greater homogeneity. Dots are the average function category entropy or diversity values of 200 control subnetworks; lines mark one standard deviation on each side of the average. **b**, Subnetworks are probably both complexes and more loosely connected modules. PCC values between all genes within each module are plotted against rank on increasing PCCs. The arrows indicate several examples.
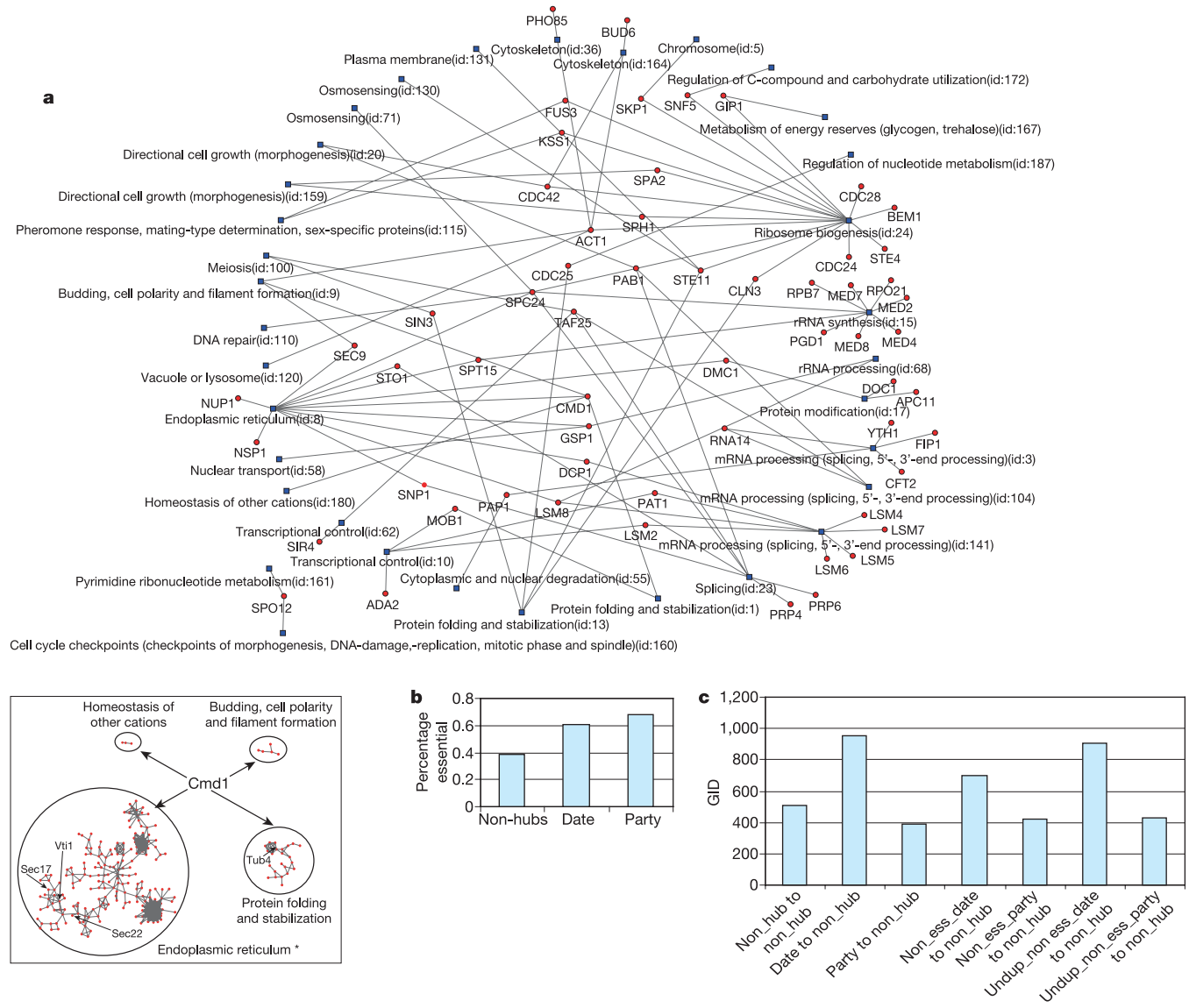
**Figure 4** Organized modularity model. **a**, Date-hub/module network representation of the FYI. Date hubs are represented as red circles and modules are represented as blue squares. The inset (below left) illustrates modular organization in detail; the date hub Cmd1 connects four modules at 'higher level', whereas the nearby party hub Sec22 connects to eight proteins within an 'endoplasmic reticulum' module. **b**, Date and party hubs are both more likely to be essential than non-hubs, but their single knockout affects cellular viability to the same extent. **c**, Date hubs participate in more genetic interactions than party hubs or non-hubs, as measured by genetic interaction density (GID) based on genetic interactions gathered at MIPS[14].

Thus, hubs in the yeast interactome network can be classified into date and party hubs on the basis of their partners' expression profiles. This distinction suggests a model of organized modularity for the yeast proteome, with modules connected through regulators, mediators or adaptors, the date hubs. Party hubs represent integral elements within distinct modules and, although important for the functions mediated by these modules (and therefore likely to be essential proteins), tend to function at a lower level of the organization of the proteome. We propose that date hubs participate in a wide range of integrated connections required for a global organization of biological modules in the whole proteome network (although some date hubs could simply be 'shared' between, and mediate local functions inside, overlapping modules). Emergent properties of the interactome network, such as genetic robustness and plasticity towards a wide range of external conditions, might be better understood by using such an organized modularity model as a framework.

Presuming that a modular network organization has selective advantages for reasons of stability and flexibility, similar partitioning might uncover modularity in metazoan interactome networks[2,16]. Similar temporal or spatial dynamic analysis might also be applied to non-biological networks, such as the World Wide Web, epidemiological networks and social networks[17]. Finally, it is possible that discriminating between date and party hubs might also help to define new therapeutic drug targets.  □

## Methods

### Protein interaction data sets

The following protein interaction lists were used to create the FYI: first, high-throughput yeast two-hybrid (HT-Y2H) projects[18–21] (5,249 potential interactions obtained from the union of the available data sets (including single hits)); second, systematic affinity purification of tagged proteins followed by mass spectrometric identification of associated proteins[22,23] (6,630 potential interactions obtained using a 'spoke' representation[24] of the union of 'Gavin' and 'Ho' data sets); third, *in silico* computational predictions of interactions[10] (7,446 potential interactions from the 'von Mering' data set obtained from the union of gene co-occurrence[25], gene neighbourhood[26] and gene fusion[25] predictions); fourth, all 'MIPS protein complexes' published singly in the literature[14] (9,597 potential

interactions obtained by using a 'matrix' representation; that is, all pairwise interactions between all components of a complex), and last, the MIPS physical interactions list (excluding genome-scale experiments: 1,285 interactions).

## Transcriptome profiling data set and average PCC calculation

The 'conditions' expression profile compendium was obtained from ref. 9. For average PCC calculation over all profiles, each of the five condition data sets in this compendium was normalized with $Z$-score normalization[10]; that is, the expression measurement for each gene was adjusted to have a mean of 0 and a standard deviation of 1 across all conditions. For the calculation of average PCC within each condition, the original $\log_2$ fold change values were used, with or without filtering for genes that displayed at least 1.5-fold changes.

## Average PCC cutoff to divide date and party hubs

For those conditions that showed a bimodal distribution, we selected the average PCC cutoff at the valley between the two peaks. For those conditions that did not show a clear division, we used an arbitrary cutoff of 0.5, a value slightly higher than those corresponding to bimodal distributions. In addition, we tested an alternative method, which yielded 104 party hubs (Supplementary Information). The data based on the second partition are very similar to those presented in the main text (data not shown). Thus, our partitioning strategy is tolerant to a small fraction of spillover between the date and party hubs.

## Attacking hubs with comparable clustering coefficients and degrees

The difference between local clustering density or degree distribution of the date and party hubs could explain their different behaviour in our *in silico* network attacks. To refute this hypothesis, we calculated $C_v$ for each hub as described[27] and subsequently selected a subset of 62 date and 62 party hubs in the largest component of comparable $k$ and $C_v$ by first removing the 8 date hubs with the highest $k$ and the 16 party hubs with the lowest $k$, and subsequently removing the 8 date hubs with the lowest $C_v$ and the 8 party hubs with the highest $C_v$. This resulted in sets of date and party hubs with nearly identical average values of $k$ or $C_v$ (Wilcoxon rank sum test $P > 0.5$ for both; $\chi^2$ test $P = 0.72$ and $0.37$ for $C_v$ and $k$, respectively, based on counts at either side of their mean values).

## Calculations

*Localization entropy of partners of a hub.* This was calculated as $-\Sigma L_i \log L_i$ (ref. 28), where $L_i$ is the frequency of appearance of a subcellular localization $i$. $L_i = T_i/\Sigma_i^n T_i$, where $T_i$ is the number of times that the subcellular localization $i$ associated with all partners of a hub and $n$ is the number of distinct subcellular localizations associated with all partners of the hub. A large-scale localization data set[12], excluding the broad high-level categories 'cytoplasm' and 'nucleus', was used.

*Function category entropy of a subnet.* This was calculated as $-\Sigma F_i \log F_i$ (ref. 28), where $F_i$ is the frequency of appearance of a function category $i$. $F_i = T_i/\Sigma_i^n T_i$, where $T_i$ is the number of times that the function category $i$ appears in the subnetwork and $n$ is the number of distinct function categories present in the subnetwork. MIPS function categories, excluding the broad high-level categories 'cytoplasm', 'mitochondrion' and 'nucleus', were used.

*Function diversity of a subnet.* This was calculated as the number of unique function categories of a subnetwork divided by the total number of members in the subnetwork. It measures the number of unique function categories per member and hence serves as an intuitive way of determining functional diversity.

*Genetic interaction density.* This was calculated as the ratio of the actual number of interactions found divided by the potential number of interactions[11]. A detailed description can be found in Supplementary Methods.

1. Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411,** 41–42 (2001).
2. Li, S. *et al.* A map of the interactome network of the metazoan *C. elegans. Science* **303,** 540–543 (2004).
3. Wagner, A. Robustness against mutations in genetic networks of yeast. *Nature Genet.* **24,** 355–361 (2000).
4. Gu, Z. *et al.* Role of duplicate genes in genetic robustness against null mutations. *Nature* **421,** 63–66 (2003).
5. Winzeler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285,** 901–906 (1999).
6. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418,** 387–391 (2002).
7. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402,** C47–C52 (1999).
8. Papin, J. A., Price, N. D., Wiback, S. J., Fell, D. A. & Palsson, B. O. Metabolic pathways in the post-genome era. *Trends Biochem. Sci.* **28,** 250–258 (2003).
9. Kemmeren, P. *et al.* Protein interaction verification and functional annotation by integrated analysis of genome-scale data. *Mol. Cell* **9,** 1133–1143 (2002).
10. von Mering, C. *et al.* Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* **417,** 399–403 (2002).
11. Ge, H., Liu, Z., Church, G. M. & Vidal, M. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae. Nature Genet.* **29,** 482–486 (2001).
12. Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425,** 686–691 (2003).
13. Albert, R., Jeong, H. & Barabasi, A. L. Error and attack tolerance of complex networks. *Nature* **406,** 378–382 (2000).
14. Mewes, H. W. *et al.* MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **30,** 31–34 (2002).
15. Ideker, T. & Lauffenburger, D. Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends Biotechnol.* **21,** 255–262 (2003).
16. Giot, L. *et al.* A protein interaction map of *Drosophila melanogaster. Science* **302,** 1727–1736 (2003).
17. Le Pont, F. *et al.* A new scale for measuring dynamic patterns of sexual partnership and concurrency: application to three French Caribbean regions. *Sex. Transm. Dis.* **30,** 6–9 (2003).
18. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae. Nature* **403,** 623–627 (2000).
19. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98,** 4569–4574 (2001).
20. Fromont-Racine, M., Rain, J. C. & Legrain, P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet.* **16,** 277–282 (1997).
21. Fromont-Racine, M. *et al.* Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins. *Yeast* **17,** 95–110 (2000).
22. Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415,** 180–183 (2002).
23. Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415,** 141–147 (2002).
24. Bader, G. D., Betel, D. & Hogue, C. W. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.* **31,** 248–250 (2003).
25. Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O. & Eisenberg, D. A combined algorithm for genome-wide prediction of protein function. *Nature* **402,** 83–86 (1999).
26. Dandekar, T., Snel, B., Huynen, M. & Bork, P. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* **23,** 324–328 (1998).
27. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393,** 440–442 (1998).
28. Snel, B., Bork, P. & Huynen, M. A. The identification of functional modules from the genomic association of genes. *Proc. Natl Acad. Sci. USA* **99,** 5890–5895 (2002).

••••••••••••••••••••••••••••••••••••••••••••••••

# PTK7/CCK-4 is a novel regulator of planar cell polarity in vertebrates

Xiaowei Lu[1], Annette G. M. Borchers[2], Christine Jolicoeur[1], Helen Rayburn[1], Julie C. Baker[2] & Marc Tessier-Lavigne[1]*

[1]*Department of Biological Sciences, Howard Hughes Medical Institute,*
[2]*Department of Genetics, Stanford University, Stanford, California 94305, USA*

\* Present address: Genentech., 1 DNA Way, South San Francisco, California 94080, USA

.................................................................

**In addition to the apical–basal polarity pathway operating in epithelial cells, a planar cell polarity (PCP) pathway establishes polarity within the plane of epithelial tissues and is conserved from *Drosophila* to mammals. In *Drosophila*, a 'core' group of PCP genes including *frizzled (fz), flamingo/starry night, dishevelled (dsh), Van Gogh/strabismus* and *prickle*, function to regulate wing hair, bristle and ommatidial polarity[1–3]. In vertebrates, the PCP pathway regulates convergent extension movements and neural tube closure[3–5], as well as the orientation of stereociliary bundles of sensory hair cells in the inner ear[6]. Here we show that a mutation in the mouse *protein tyrosine kinase 7* (*PTK7*) gene, which encodes an evolutionarily conserved transmembrane protein with tyrosine kinase homology, disrupts neural tube closure and stereociliary bundle orientation, and shows genetic interactions with a mutation in the mouse *Van Gogh* homologue**